JURNAL

ILMU PENGETAHUAN & TEKNOLOGI KOMPUTER

Vol. 6. No. 2 Februari 2021

ISSN: 2685-8223 (Printed)

ISSN: 2527-4864 (Online)

Publisher:
Pusat Penelitian Pengabdian Masyarakat STMIK Nusa Mandiri
Jl. Damai No. 8 (Margasatwa) Pasar Minggu
http://ejournal.nusamandiri.ac.id/index.php/jitk/index

Vol. 6 No. 2 Februari 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v6i2

Vol. 6 No. 2 Februari 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v6i2

EDITORIAL BOARD

Editorial Address : Kampus STMIK Nusa Mandiri
Jl. Kramat Raya No, 18, Kwitang Senen Jakarta Pusat
Indonesia

Website : http://ejournal.nusamandiri.ac.id/index.php/jitk

Editorial Email : redaksi.jitk@nusamandiri.ac.id or jurnal.jitk@gmail.com

Vol. 6 No. 2 Februari 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v6i2

PREFACE

Editor of JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), said praise and gratitude to the presence of Allah S.W.T, creator of the universe who mastered knowledge as wide as heaven and earth, for the abundance of grace and gifts that have been given to JITK editors to publish JITK Vol. 5, No. 2 February 2020, which is used by lecturers, researching, and professionals as a medium or media to publish publications on the findings of research conducted in each semester.

JITK is published 1 (one) year for 2 (two) times at the end of each semester, JITK editors receive scientific articles from the results of research, reports / case studies, information technology studies, and information systems, which are oriented to the latest in science and information technology in order to be a source of scientific information that is able to contribute to the increasingly complex development of information technology.

The editor invited fellow researchers, scientists from various tertiary institutions to make scientific contributions, both in the form of research results and scientific studies in the fields of management, education, and information technology. The editors really expect input from readers, information technology professionals, or those related to publishing, for the sake of increasing the quality of journals as we all hope.

The editor hopes that the scientific articles contained in the JITK scientific journal will be useful for academics and professionals working in the world of management, education, and information technology

Chief Editor

Vol. 6 No. 2 Februari 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v6i2

Vol. 6 No. 2 Februari 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v6i2

TABLE OF CONTENTS

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1939

COMPARISON OF REGIONAL CLUSTER ANALYSIS ACCORDING TO INCLUSIVE DEVELOPMENT INDICATORS IN JAVA ISLAND 2018 BETWEEN HIERARCHICAL AND PARTITIONING CLUSTERING STRATEGIES

Akhmad Fatikhurrizqi[1]; Arie Wahyu Wijayanto[2]

Statistika[1.2]
Politeknik Statistika STIS
www.stis.ac.id
211709518@stis.ac.id [1]; ariewahyu@stis.ac.id [2]

Abstract— Gross Domestic Product (GDP) is one of the most common indicators to reflect a nation's development. Indonesia's GDP has an average growth rate of 5 percent over the 2015-2019 period with the highest growth rate occurred in 2018. Furthermore, the provinces in Java Island contributed the most out of any province to Indonesia's GDP in that year. However, the development in Java Island still has several issues, such as high poverty, unequal income distribution, and high unemployment. This problem indicates that the economic growth in Java Island has not been inclusive concerning development. This study aims to group regencies/municipalities in Java Island based on indicators of inclusive growth. These indicators refer to McKinley (2010) in a journal published by the Asian Development Bank (ADB). The cluster methods used to represent each hierarchical and partitioning are the Agglomerative Nesting (AGNES) and K-Means methods. The results of this study show that there are 3 clusters based on the AGNES method and 4 clusters based on the K-Means method. Clusters with good inclusive growth characteristics are dominated by municipality areas based on the K-Means method. Meanwhile, the clusters with low inclusive growth characteristics are dominated by regencies/municipalities on Madura Island based on the K-Means and AGNES methods. The comparison of the appropriate methods in this study based on the silhouette value is the AGNES method.

Keywords: Inclusive Growth, Cluster Analysis, AGNES, K-Means.

Abstrak— *Produk Domestik Bruto (PDB) adalah salah satu indikator yang paling umum untuk mencerminkan pembangunan suatu bangsa. PDB Indonesia memiliki tingkat pertumbuhan rata-rata 5 persen selama periode 2015-2019 dengan laju pertumbuhan tertinggi terjadi pada tahun 2018. Selain itu, provinsi-provinsi di Pulau Jawa memberikan kontribusi terbesar dibanding provinsi lainnya terhadap PDB Indonesia pada tahun tersebut. Namun demikian, pembangunan di Pulau Jawa masih memiliki beberapa permasalahan, seperti kemiskinan yang tinggi, distribusi pendapatan yang tidak merata, dan pengangguran yang tinggi. Masalah ini menunjukkan bahwa pertumbuhan ekonomi di Pulau Jawa belum inklusif dalam pembangunan. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Pulau Jawa berdasarkan indikator pertumbuhan inklusif. Indikator tersebut mengacu pada McKinley (2010) dalam jurnal yang diterbitkan oleh Asian Development Bank (ADB). Metode klaster yang digunakan untuk merepresentasikan setiap hierarki dan partisi adalah metode Agglomerative Nesting (AGNES) dan K-Means. Hasil penelitian menunjukkan bahwa terdapat 3 klaster berdasarkan metode AGNES dan 4 cluster berdasarkan metode K-Means. Klaster dengan karakteristik pertumbuhan inklusif yang baik didominasi oleh kawasan perkotaan berdasarkan metode K-Means. Sedangkan klaster dengan karakteristik pertumbuhan inklusif rendah didominasi oleh kabupaten/kota di Pulau Madura berdasarkan metode K-Means dan AGNES. Perbandingan metode yang sesuai dalam penelitian ini berdasarkan nilai silhouette adalah dengan metode AGNES.*

***Kata Kunci**:* Pertumbuhan Inklusif, Analisis Klaster, AGNES, *K-Means.*

INTRODUCTION

The concept of inclusive growth is interpreted narrowly and broadly [1]. In a narrow sense, inclusive growth focuses more on economic growth, where human capabilities are seen as an instrument to increase economic output. Whereas in a broad sense, inclusive growth more focuses on comprehensive and sustainable growth, which will create and expand economic opportunities, and ensure wider access to these opportunities, so that every member of society can participate and feel

the benefit from this growth. This shows that economic growth, which is commonly used as a measure of a country's development, only includes a narrow concept, where the measure of economic growth is emphasized on an increase in output or Gross Domestic Product [2]. *Badan Pusat Statistik* (BPS) noted that Indonesia's economic growth rate in 2015-2019 experienced relatively good and quite stable with an average growth of around 5 percent, where the highest GDP growth rate occurred in 2018 [3]. Meanwhile, based on the regional contribution to Indonesia's GDP, Java Island contributed more than half of it. The contribution of Java Island to Indonesia's GDP in 2018 was 58.48 percent [3]. This is also supported by the large number of industrial estates in Java, which are the largest in number compared to other islands. Java Island is considered more ready to become a location for industrial development in terms of human resources, infrastructure, and natural resources [4]. Therefore, it can be said that Indonesia's economic growth is largely supported by the island of Java, even though it only has six provinces. The amount of Java Island's GRDP contribution to national GDP still leaves several problems, such as high poverty, unequal income distribution, and high unemployment. In Table 1 it can be seen that all provinces in Java Island have relatively high economic growth, which is above the national economic growth value of only 5.17 percent in 2018 [3]. However, not all provinces have achieved high economic growth make a good contribution to public welfare. The emphasis on increasing GDP leads to a subordination of problems such as poverty, discrimination, unemployment, and income distribution [2]. Also, GDP per capita as a measure of economic growth is marred by issues of inequality and severe poverty. Therefore, inclusive growth requires the growth and improvement of individual welfare, especially in the income and non-income aspects of individual and societal well-being [5]. This condition goes along with the condition of several problems in Java Island, such as the percentage of poverty in Central Java, DI Yogyakarta, and East Java having values above 10 percent and exceeding the national average; the open unemployment rate in DKI Jakarta, West Java, and Banten also scored above the national average [6]; as well as the distribution of income inequality in DKI Jakarta, West Java, and DI Yogyakarta also scored above the national average.

Table 1. The value of economic growth, the percentage of poverty, the open unemployment rate, and the Gini ratio of provinces in Java Island in 2018

Province	Economic Growth	Percentage of Poverty	Unemployment Rate	Gini Ratio
DKI Jakarta	6.17	3.55	6.24	0.39
Jawa Barat	5.66	7.25	8.17	0.41
Jawa Tengah	5.31	11.19	4.51	0.36
DI Yogyakarta	6.20	11.81	3.35	0.42
Jawa Timur	5.50	10.85	3.99	0.37
Banten	5.82	5.25	8.52	0.37
Nasional	**5.17**	**9.66**	**5.34**	**0.38**

Source: BPS [7]

Based on Tabel 1, Conditions of poverty, unemployment, and high inequality of income distribution indicate that economic growth in Java is not yet inclusive. Inclusive growth is growth followed by a reduction in poverty, a reduction in inequality in income distribution, and an increase in labor absorption [8]. Inclusive growth reflects more on regional growth in a multidimensional manner, so that good linkages between indicators can support inclusive growth. For example, income inequality affects inclusive education [9]. Early childhood education is an effective target for fiscal spending to create employment and gender-inclusive growth [10]. Also, case studies in Africa show the health sector to play a major role in inclusive growth [11].

Even though it isn't a new issue, inclusive growth is still a topic that attracts attention and is considered important because it can inclusively explain the state of development, so that it is useful for promoting sustainable development. Therefore, this study discusses the grouping of regencies/municipalities in Java based on their level of growth inclusiveness. The objectives of this study are as follows:

1) Grouping regencies/municipalities in Java Island based on indicators of inclusive growth based on the cluster analysis method [12].
2) Knowing the characteristics of regencies/municipalities in Java Island based on the clusters formed.
3) Comparing between Agglomerative Nesting (AGNES) and K-Means clustering methods as representatives of each hierarchical and partitioning strategy.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1939

MATERIALS AND METHODS

Data Source

The data source of this study is the Central Bureau of Statistics (BPS). Among other things, the publication of each province in Java, such as "*Provinsi Dalam Angka 2019", "Statistik Kesejahteraan Rakyat Provinsi 2018", dan "Keadaan Angkatan Kerja Provinsi 2018"*, as well as the publication of tables and indicators from the website of the BPS. The units of analysis used in this study were 119 regencies/municipalities in Java Island in 2018.

Research Variable

The relatively new concept of inclusive growth makes it difficult to determine a theory that supports the argument for the direction of the influence of the variables used in describing inclusive growth [13]. This study uses the concept of the Asian Development Bank which explains the dimensions of inclusive growth [1]. The variables used in this study which refer to these dimensions can be seen in Table *2* below.

Table 2. Research Variable

Dimension	Variable	Definition	Unit of Measurement
Economic Growth	X1	Grow rate of GRDP per capita	Percent
Productive Workforce	X2	Percentage of Labor in the Industrial Sector	Percent
	X3	Unemployment rate (TPT)	Percent
Economic Infrastructure	X4	Percentage of households with access to electricity	Percent
	X5	Percentage of households that own a computer/laptop	Percent
Poverty	X6	Percentage of the poor population	Percent
Inequality	X7	Gini Ratio	Ratio
Gender Equality	X8	Gender Development Index (IPG)	Ratio
	X9	Gender Empowerment Index (IDG)	Ratio
Health	X10	Life expectancy at birth (UHH)	Year
Education	X11	The expectation of old school years	Percent
Water and Sanitation	X12	Percentage of households that have access to proper sanitation	Year
	X13	Percentage of households that have access to decent drinking water	Percent
Social Protection	X14	Percentage of households buying/receiving "Raskin"	Percent
	X15	Percentage of the population having social protection card (KPS)	Percent

Source: McKinley [1], modified

Method of Analysis

1) AGNES (Agglomerative Hierarchical Clustering)

An agglomerative hierarchical clustering method uses a bottom-up strategy. It typically starts by letting each object form its cluster and iteratively merges clusters into larger and larger clusters, until all the objects are in a single cluster or certain termination conditions are satisfied. The single cluster becomes the hierarchy's root. For the merging step, it finds the two clusters that are closest to each other (according to some similarity measure) and combines the two to form one cluster. Because two clusters are merged per iteration, where each cluster contains at least one object, an agglomerative method requires at most *n* iterations [14]. How AGNES works can be seen in Figure 1 below.

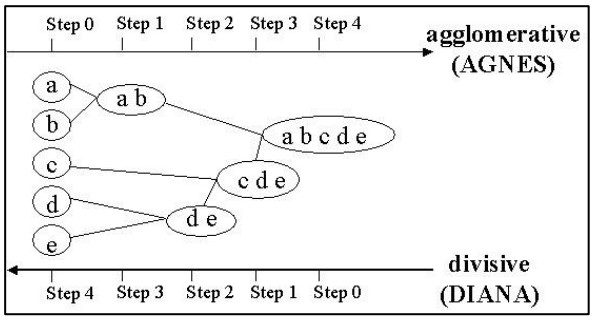


Figure 1. Concept of Hierarchical Clustering

2) K-means

K-Means Clustering is an unsupervised clustering algorithm. This method groups several objects into a cluster. Membership of objects is seen based on the object's distance to the center point (centroid) of a cluster. In the k-means approach, the initial step is to determine k as the

number of clusters to be formed. The k-means algorithm can be written as follows [15]:

a) For n objects, initiate k the cluster center point
b) Enter each object in the closest cluster, calculating the distance.

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad \text{..... (1)}$$

c) Update the cluster center point for each incoming object, looking at the average member.
d) Repeat steps 2 and 3 until there is no change in the center point of the cluster.

The systematics and workings of the k-means algorithm can be seen in Figure *2* below.

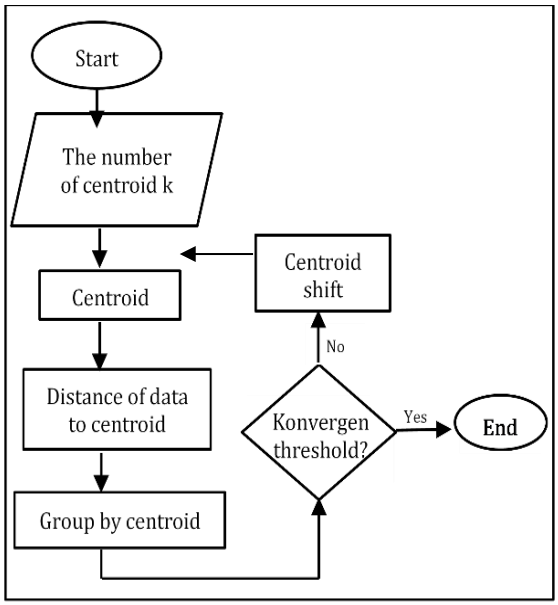


Figure 2. Flowchart of K-Means Algorithm

RESULTS AND DISCUSSION

Pre-processing Data

Missing Value

Based on the available data, there is a missing value in the Gini ratio variable in the regencies/municipalities of DKI Jakarta Province. Therefore, steps that can be taken to overcome this problem are to estimate the Gini ratio using the linear regression method. Modeling was carried out on other regencies/municipalities in Java Island. The independent variables used are per capita expenditure (pkp) and the level of poverty depth (p1). The model equation that is formed is as follows:

$$Gini\ ratio_i = 0{,}276 + (6{,}551\ x\ 10^{-6})pkp_i - 0.011\ p1_i \quad \text{..... (2)}$$

Standardization / Normalization

Normalization is a step for adjusting the data values into specific ranges such as between 0 to 1 or -1 to 1 [16]. The purpose of normalization is to equalize the units in all variables. Generally, the methods used are minimum-maximum, z-score, and decimal scaling. In this study, the z-score normalization was used.

Feature Selection

Feature selection is the stage of selecting the variables to be used. This stage can reduce the number of variables, get rid of unrelated, repetitive, or noise variables, thereby speeding up the data mining algorithm, increasing accuracy, and producing the best model [17]. In this study, the variables were selected based on the assumption of non-multicollinearity. A low correlation value between variables indicates the absence of data redundancy. Therefore, a variable that has a high correlation value is selected, which is above 0.7. Figure 3 shows the pairs of variables that correlate 0.7 are X5-X6, X6-X14, X6-X15, and X14-X15. The decision was taken, namely to issue X14 and X15 because it was already represented by X6, and to issue X5 because other variables represented the dimensions of economic infrastructure.

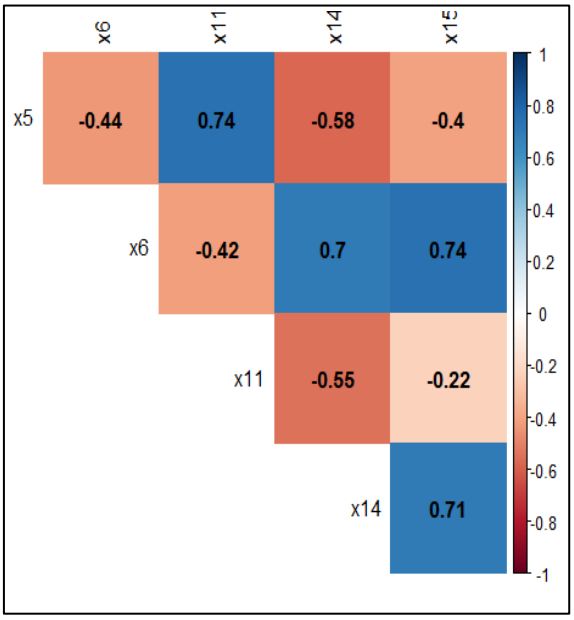


Figure 3. Correlation of X5, X6, X11, X14, X15

Result of the AGNES (Agglomerative Nesting) Clustering Analysis

The AGNES method is a hierarchical cluster method that moves from bottom to top. As for the hierarchical method, there are several methods such as Single Linkage, Average Linkage, Ward, and Complete. In determining the most appropriate method, the Agglomerative Coefficient can be used. The coefficient value close to 1 indicates the grouping is getting stronger or better. The

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1939

following is the Agglomerative Coefficient value for each method:

Table 3. Value of Agglomerative Coefficient

Method	Agglomerative Coefficient
Average	0,7541
Single Linkage	0,6534
Complete Linkage	0,8088
Ward	0,8838

Based on Table 3, the best method of hierarchical grouping in this study is the ward method, with an agglomerative coefficient value of 0.88. In the AGNES hierarchy method, it can be seen that the cluster formation is based on the formed dendrogram. Figure 4 below is the dendrogram that is formed.

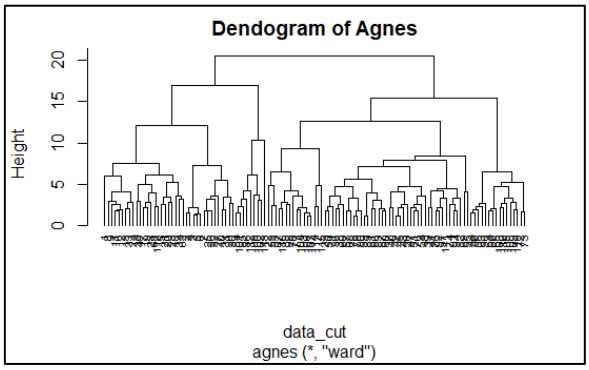


Figure 4. Dendrogram of Grouping with AGNES Method

Furthermore, in determining the maximum number of clusters to be formed based on the AGNES method with the ward, the NbClust () package is used in the R application. The results obtained show that the best number of clusters is three clusters. The number of members in each cluster formed is 37, 72, and 9 regencies/municipalities. As for the distribution of members in each cluster, it can be seen in Figure 5 below.

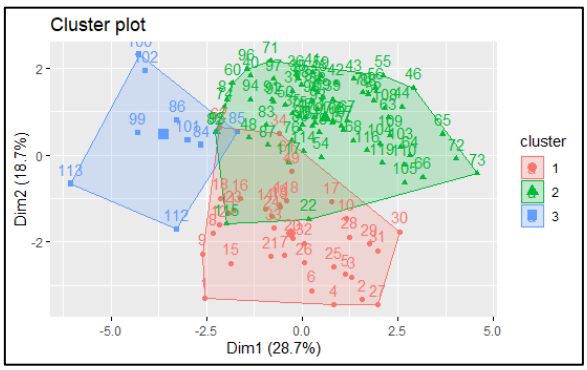


Figure 5. Clustering Plot based AGNES Method

Based on the average value of the variables in each cluster (Table 4), the characteristics of each variable in each cluster are obtained, which are as follows:

1. Cluster 1 has unfavorable characteristics in several variables, namely the high level of open unemployment and the low percentage of households that have access to proper drinking water.
2. Cluster 2 has good characteristics in all variables, except poverty. This is indicated by the average percentage of poor people in cluster 2 of 9.76 percent, which is still above the national average (9.66 percent).
3. Cluster 3 has poor characteristics in several variables, namely the per capita GRDP growth rate and the lowest percentage of industrial workers among other clusters, the high average percentage of poor people, and the low percentage of households with access to proper sanitation.

In general, all the clusters that were formed were still experiencing their problems, so that no cluster had good inclusive growth. Therefore, if sorted according to the level of inclusive growth from the highest are cluster 2, cluster 1, and cluster 3. The categories for each cluster are based on their inclusive growth, namely cluster 2 is quite good, cluster 1 is medium, and cluster 3 is not good.

Table 4. Average of Variable Value Based on AGNES Clustering

No	Variable	Cluster 1	Cluster 2	Cluster 3
1	X1	8.69	7.81	7.77
2	X2	18.14	25.16	17.62
3	X3	7.42	4.42	4.26
4	X4	99.96	99.9	99.67
5	X6	7.94	9.76	15.37
6	X7	0.36	0.33	0.28
7	X8	89.58	92.14	84.83
8	X9	66.78	69.32	58.58
9	X10	71.86	73.27	67.63
10	X11	12.70	13.23	12.6
11	X12	64.10	84.53	33.45
12	X13	36.13	76.84	62.94

The results of grouping regencies/municipalities based on inclusive growth clusters using the AGNES method can be seen in Table 5. Cluster 1 is dominated by regencies/municipalities in DKI Jakarta and West Java. Cluster 2 is dominated by regencies/municipalities in Central Java and DI Yogyakarta. The whole cluster 3 is a regency area and is dominated by areas on Madura Island. The results of grouping using the AGNES method show the tendency of regencies/municipalities within a

province horde in a cluster. This indicates there is a spatial influence between regions in explaining the inclusiveness of a region's growth [18].

Table 5. Results of Grouping of Regencies/Municipalities with AGNES Method

Cluster	Regencies/Municipalities
Cluster 1	Kep. Seribu, Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar, Cilacap, Blora, Tegal, Brebes, Kota Serang.
Cluster 2	Bekasi, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal, Kulon Progo, Bantul, Gunung Kidul, Sleman, Kota Yogyakarta, Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu, Tangerang, Serang, Kota Tangerang, Kota Cilegon, Kota Tangerang Selatan.
Cluster 3	Bondowoso, Situbondo, Probolinggo, Bangkalan, Sampang, Pamekasan, Sumenep, Pandeglang, Lebak

Results of the K-means Cluster Analysis

The initial stage of k-means is to determine the optimum number of clusters. The method used in determining the number of clusters in this study is the Elbow Method. The smaller the variance in the cluster (within), the better the grouping is done. Figure 6 shows an elbow plot where the horizontal axis is the number of clusters and the vertical axis is the total number of squares in the cluster (total within the sum of squares). Determining the optimum number of clusters is by determining the point that has decreased sharply from the total value within the sum of the square, where this value can determine the variance value in the cluster.

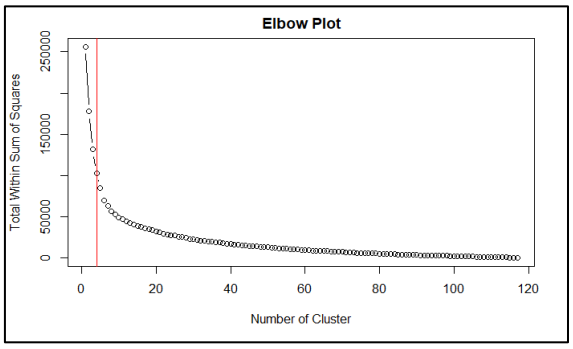


Figure 6. Elbow Plot

Based on these results, it illustrates that the number of clusters is 4 can significantly reduce the within variance, while the number of clusters more than 4 can only reduce the within variance which is less, so it is less efficient. Also, Figure 7 shows that the cluster formed can separate objects well and there is no overlapping.

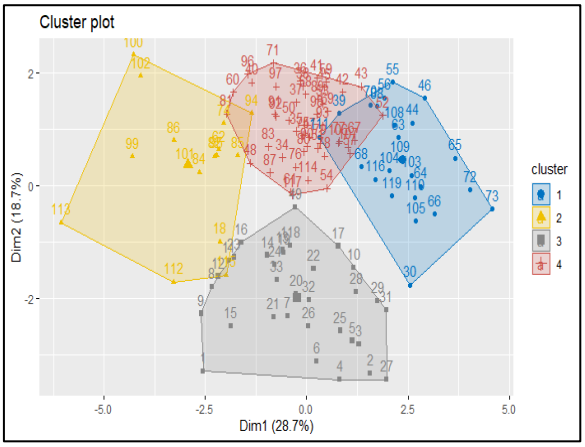


Figure 7. Clustering Plot based K-Means Method

The results of grouping using the K-means method indicate that there are 4 clusters formed, each of which has members of 24, 16, 33, and 46 regencies/municipalities. The variance between clusters formed was 516.8 and the variance within each cluster was 143.8; 199.2; 267.8; 288.4. This shows that the cluster formed already has good cluster characteristics, namely variance between high clusters (heterogeneous) and variance in low clusters (homogeneous). Based on the average value of the variables in each cluster (Table *6*), the characteristics of each variable in each cluster are obtained, which are as follows:

1. Cluster 1 has good characteristics in all variables, meaning that it can be said that the regencies/municipalities in Cluster 1 already have inclusive development.
2. Cluster 2 has unfavorable characteristics in several variables, namely the lowest percentage of workers in the industrial sector among other clusters, high poverty, low levels

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1939

of gender equality, and a low percentage of households with access to proper sanitation.

3. Cluster 3 has unfavorable characteristics in several variables, namely the high level of open unemployment and the low percentage of households that have access to proper drinking water.
4. Cluster 4 has fairly good characteristics in all dimensions, except poverty. This is indicated by the average percentage of poor people of 11.15 percent which has a value above the national average (9.66 percent).

Therefore, if all clusters sorted by the level of inclusive growth from the highest, there are cluster 1, cluster 4, cluster 3, and cluster 2. The categories for each cluster are based on inclusive growth, namely cluster 1 is good, cluster 4 is good enough, cluster 3 is moderate, and cluster 4 is not good.

Table 6. Average of Variable Value Based on K-Means Clustering

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X1	7.72	8.02	8.84	7.75
X2	22.77	18.24	19.20	25.97
X3	4.53	5.02	7.45	4.35
X4	99.92	99.66	99.97	99.93
X6	6.68	13.99	7.51	11.15
X7	0.35	0.29	0.37	0.32
X8	94.90	86.25	89.76	90.98
X9	76.12	61.40	66.45	66.44
X10	74.56	68.50	71.93	72.98
X11	14.36	12.56	12.73	12.69
X12	93.10	47.31	61.66	82.98
X13	77.95	63.36	32.10	77.58

The results of grouping regencies/municipalities based on inclusive growth clusters using the K-means method can be seen in Table 7. Cluster 1 is dominated by municipality areas and several regencies that are quite advanced, such as Semarang and Sleman. In Cluster 2, all of its members are regencies, even all regencies on Madura Island are included in this cluster. Cluster 3 has a fairly even distribution between regency and municipality areas. This cluster also occurs in the grouping of all areas in DKI Jakarta. Finally, cluster 4 is dominated by regencies in Central Java and East Java.

Table 7. Results of Grouping of Regencies/Municipalities with K-means Method

Cluster	Regencies / Municipalities
Cluster 1	Kota Depok, Purworejo, Sukoharjo, Karanganyar, Semarang, Temanggung, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Tegal, Bantul, Sleman, Kota Yogyakarta, Sidoarjo, Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu, Kota Tangerang, Kota Tangerang Selatan
Cluster 2	Indramayu, Brebes, Pacitan, Jember, Bondowoso, Situbondo, Probolinggo, Ngawi, Bojonegoro, Bangkalan, Sampang, Pamekasan, Sumenep, Pandeglang, Lebak, Serang.
Cluster 3	Kep. Seribu, Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Cimahi, Kota Tasikmalaya, Kota Banjar, Blora, Kota Serang.
Cluster 4	Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Boyolali, Klaten, Wonogiri, Sragen, Grobogan, Rembang, Pati, Kudus, Jepara, Demak, Kendal, Batang, Pekalongan, Pemalang, Tegal, Kota Pekalongan, Kulon Progo, Gunung Kidul, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Banyuwangi, Pasuruan, Mojokerto, Jombang Nganjuk Madiun, Magetan, Tuban, Lamongan, Gresik, Kota Probolinggo, Kota Pasuruan, Tangerang, Kota Cilegon.

Comparison of The Result Clustering Analysis between AGNES and K-Means method

To compare the cluster method, this study used silhouette values. The results of the silhouette coefficient calculation can vary from -1 to 1. The coefficient of an object is 1 indicating that the object is in the right cluster. If it is 0, it indicates that the object is not certain which cluster it belongs to. Meanwhile, if the value is -1, it indicates that the cluster structure that is formed is overlapping.

Table 8. Silhouette Value

Method	Silhouette Value
AGNES	0,17
K-Means	0,15

Based on Table 8, the comparison of silhouette values between the AGNES and K-Means methods is not too different, namely 0.17 and 0.15. Apart from having a higher silhouette value, the AGNES method also has a smaller number of clusters. This shows that the AGNES method is more efficient in grouping objects. Therefore, it can be concluded that the AGNES method is more appropriate than the K-Means method in this study.

CONCLUSION

The grouping of regencies/municipalities in Java Island based on their level of inclusive growth resulted in several different clusters per method.

The AGNES method produces three clusters and the K-Means method produces four clusters. Clusters that have good inclusive growth characteristics are dominated by municipality areas based on the K-Means method. Meanwhile, regional clusters that have low inclusive growth based on the K-Means and AGNES methods are regencies/municipalities on Madura Island. The appropriate comparison method in this study between AGNES and K-means is based on silhouette values and efficiency in grouping regencies/municipalities in the AGNES method.

REFERENCE

[1] T. McKinley, "Inclusive growth criteria and indicators: an inclusive growth index for diagnosis of country progress," *Asian Dev. Bank Work. Pap.*, no. 14, pp. 1–34, 2010, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Inclusive+Growth+Criteria+and+Indicators+:+An+Inclusive+Growth+Index+for+Diagnosis+of+Country+Progress#0.

[2] M. P. Todaro and S. C. Smith, *Economic Development Twelfth Edition*, 12th Editi. New York: Pearson, 2014.

[3] Badan Pusat Statistik, "Data Online Badan Pusat Statistik: Tabel Dinamis Subjek Produk Domestik Bruto," 2019. https://www.bps.go.id/subject/11/produk-domestik-bruto--lapangan-usaha-.html#subjekViewTab3.

[4] Kementerian Perindustrian RI, "Rencana Induk Pembangunan Industri Nasional 2015 - 2035," *Rencana Induk Pembang. Ind. Nas. 2015-2035*, pp. 1–98, 2015, [Online]. Available: http://www.depkop.go.id.

[5] I. C. Oluseye and A. A. Gabriel, "Determinants of Inclusive Growth in Nigeria : An ARDL Approach," vol. 7, no. 3, pp. 97–109, 2017, DOI: 10.5923/j.economics.20170703.01.

[6] T. Sukwika, "Peran Pembangunan Infrastruktur terhadap Ketimpangan Ekonomi Antarwilayah di Indonesia," *J. Wil. dan Lingkung.*, vol. 6, no. 2, p. 115, 2018, doi: 10.14710/jwl.6.2.115-130.

[7] Badan Pusat Statistik, "Data Online Badan Pusat Statistik: Tabel Dinamis Subjek Produk Domestik Bruto, Kemiskinan dan Ketimpangan, serta Tenaga Kerja.," 2020, [Online]. Available: https://www.bps.go.id/subject/11/produk-domestik-bruto--lapangan-usaha-.html#subjekViewTab3.

[8] T. Ji Long and E. Pasaribu, "Analisis Spasial Determinan Pertumbuhan Inklusif Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2017," *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, pp. 416–423, 2020, doi: 10.34123/semnasoffstat.v2019i1.11.

[9] S. A. Asongu, J. Nnanna, and P. N. Acha-Anyi, "Finance, inequality and inclusive education in Sub-Saharan Africa," *Econ. Anal. Policy*, vol. 67, pp. 162–177, 2020, DOI: 10.1016/j.eap.2020.07.006.

[10] K. Kim, İ. İlkkaracan, and T. Kaya, "Public investment in care services in Turkey: Promoting employment & gender-inclusive growth," *J. Policy Model.*, vol. 41, no. 6, pp. 1210–1229, 2019, DOI: 10.1016/j.jpolmod.2019.05.002.

[11] S. Tella and O. Alimi, "Determinants of Inclusive Growth in Africa: Role of Health and Demographic Changes," *African J. Econ. Rev.*, vol. 4, no. 2, pp. 138–146, 2016.

[12] S. Wahyuni and Y. A. Jatmiko, "Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering," *J. Apl. Stat. Komputasi Stat.*, vol. 10, no. 1, pp. 1–8, 2019, doi: https://doi.org/10.34123/jurnalasks.v10i1.197.

[13] N. Ngepah, "A review of theories and evidence of inclusive growth: an economic perspective for Africa," *Curr. Opin. Environ. Sustain.*, vol. 24, pp. 52–57, 2017, DOI: 10.1016/j.cosust.2017.01.008.

[14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Third Edit. Amsterdam: Morgan Kauffman, 2012.

[15] S. K. Chandel, "Intrusion Detection System using K-Means Data Mining and Outlier Detection Approach," Masaryk University, 2017.

[16] S. A. Alasadi and W. S. Bhaya, "Alasadi, S. A. (2017). Review of data preprocessing techniques in data mining 12(16), 4102-4107.," *J. Eng. Appl. Sci.*, vol. 12, no. 16, pp. 4102–4107, 2017, DOI: http://dx.doi.org/10.36478/jeasci.2017.4102.4107.

[17] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature selection: An ever-evolving frontier in data mining," *Featur. Sel. data Min.*, vol. 10, pp. 4–13, 2010, [Online]. Available: http://proceedings.mlr.press/v10/liu10b/liu10b.pdf.

[18] C. Park and R. Claveria, "DOES REGIONAL INTEGRATION MATTER FOR INCLUSIVE GROWTH ? EVIDENCE FROM THE MULTIDIMENSIONAL REGIONAL ADB

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1939

ECONOMICS Does Regional Integration Matter for Inclusive Growth ? Evidence from the Multidimensional Regional Integration Index," *ADB Econ. Work. Pap. Ser.*, vol. Oktober, no. 559, 2018, [Online]. Available: https://www.adb.org/sites/default/files/publication/460681/ewp-559-regional-integration-inclusive-growth.pdf.

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1909

APRIORI ALGORITHM IMPLEMENTATION TO DETERMINE PURCHASE PATTERNS OF RAW MATERIALS AT PT PENJALINDO NUSANTARA

Dedi Rozaq Prastyo[1]; Sri Dianing Asri[2*]

Informatics Study Program[1,2]
Universitas Mercu Buana
www.mercubuana.ac.id
41516320005@mercubuana.ac.id [1]; dianing.asri@mercubuana.ac.id [2*]
(*) Corresponding Author

Abstract— PT. Penjalindo Nusantara is a manufacturing company in the packaging field where production depends on customer demand or what is commonly known as job orders so that timely production work and availability of sufficient materials are mandatory for the company. There was a problem in the implementation of the raw material supply strategy by PT. Penjalindo Nusantara caused delays in the supply of raw material stocks. The solution to this problem is to apply the Apriori algorithm to find out what raw materials are being purchased simultaneously so that it can be the basis for implementing a purchasing strategy in supporting the effectiveness of procurement of raw material stocks and also saving time in sending raw materials by suppliers. This research uses a Web-based data mining application to find the raw material purchase pattern. The result of this research is obtained 11 patterns of purchasing raw materials using a minimum value of 90% support and a minimum of 100% confidence with a lift ratio of 1 as a reference for determining which raw materials will be purchased at the same time.

Keywords: Data Mining, Apriori Algorithm, Raw Materials, Web-Based.

Abstrak— *PT. Penjalindo Nusantara adalah suatu perusahaan yang bergerak di bidang manufaktur kemasan yang di mana produksi bergantung pada permintaan pelanggan atau yang biasa disebut job order, sehingga pengerjaan produksi yang tepat waktu dan ketersedian material yang mencukupi menjadi hal yang wajib dilakukan oleh perusahaan. Terjadi permasalahan dalam penerapan strategi persediaan bahan baku oleh PT. Penjalindo Nusantara yang tidak efektif menyebabkan terjadi keterlambatan dalam penyedian stok material produksi. Solusi dari masalah ini adalah dengan melakukan penerapan algoritma Apriori untuk mengetahui material produksi apa saja yang dibeli secara bersamaan sehingga dapat menjadi dasar untuk menerapkan strategi pembelian dalam menunjang efektivitas pengadaan stok material produksi dan juga menghemat waktu pengiriman material produksi oleh supplier. Penelitian ini menggunakan aplikasi data mining berbasis Web-based untuk menemukan pola pembelian bahan baku. Hasil dari penelitian ini mendapatkan 11 pola pembelian material produksi dengan menggunakan nilai minimum support 90% dan minimum confidence 100% dengan lift ratio 1 sebagai acuan untuk menentukan material produksi mana saja yang akan dibeli pada waktu bersamaan.*

Kata Kunci*: Data Mining, Algoritma Apriori, Bahan Baku, Web-Based.*

INTRODUCTION

To make the raw material inventory management and inventory cost more efficient, controlling the raw material purchases according to the production plan so that there are no shortages and excesses in the raw material inventory[1]. The Supply Chain department oversees the raw materials purchase planning and controlling the company materials including providing raw materials, controlling the amount of inventory, as well as a system for storing and managing goods in and out of the warehouse. Controlling the amount of inventory is one of the most important factors in optimizing inventory. Inventory Control is an activity in managing inventory to suit the needs and maintain the raw material stability. PT. Penjalindo Nusantara is a manufacturing company in the packaging field where production depends on customer demand[2], there is a problem with the effectiveness of raw material stock control[3] and delays in ordering raw materials because the company does not have the right raw material order scheduling[4][5] caused the problem for the production activity. So far, the company has

overcome the problem if there is no stock for the raw material that is needed in production by using existing materials with the same specifications according to product needs, but it cost a lot of production cost[5].

To fix problems in the material procurement planning, the solution is to determine which raw materials will be purchased at the same time, the basis of determining the raw material purchased at the same time is to analyze what raw material is often purchased at the same time for 11 months using data mining. One of the algorithms that belong to an association in data mining that matches is the Apriori algorithm. Apriori algorithm aims to find frequent itemsets in a set of data. Apriori algorithm defined a process to find Apriori rules that meet minimum requirements for support and minimum requirements for confidence.[6]

This research seeks to analyze the purchase of raw materials in 11 months by using the Apriori algorithm and obtained the association rules. So, the results can be identified by seeing which raw materials are often purchased simultaneously for over 11 months.

Literature Research

There is some literature research related to Apriori Algorithm. Munarwan et al [7]. This research discusses the Market Basket Analysis method works by using the Apriori Algorithm to determine the purchase patterns in one transaction. The results obtained the average minimum support by 0,1563 and minimum confidence by 0,3486, and the highest support with value by 0,3125 and the highest confidence by 0,7143 contains 3 itemsets that if buy AM and Buy SN Then Buy RT. Another research in 2018 D. S. Wijaya et al[8] aims to analyze the data to determine the correlation between fish catch that may affect another fish, this research using the Association Rules method and obtained the value of highest minimum support by 50% and the highest minimum confidence by 80%, the result is if getting Yellowfin Tuna is also getting the Bigeye Tuna. And then in 2020, Y. Nawawi et al[9] presents a study to find the relation of the book borrowed at the same time by using the Association Rule. The results obtained 50 association rules with the highest rule is lending of communication science and psychology book with support and confidence of 8.17%. And then the research related to maintaining the item stock is done by Junaidi[10] in 2019, The goal of this research is to determine the inventory of goods by analyzing which items are sold simultaneously using the Ariori algorithm, the process is getting the results by setting the minimum support 60% and the minimum confidence by 90%, the results are obtained 24 associative rules with the minimum support 60% and the minimum confidence 90% that contains the items which items are sold simultaneously and can be used as a reference for providing stock of items. And then the research in 2018 by D. A. O. Turang et al [11], this research analysed the inventory stock system and implementing the system using the Differential Evolution Algorithm, this research analyzed the inventory data and sales data, sales and inventory data are calculated to determine the estimated safety stock and reorder point so that the status of the raw material inventory position can be determined. Lot sizing method is successfully used to calculate the number of orders and the period of ordering in order to achieve optimal costs and achieve efficiency of the company.

The GAP analysis from the above research with this research is using the Apriori algorithm to determine the raw material purchasing pattern which can be used as a reference for determining the supply of raw material stocks to be more effective and reduce production costs because the required raw materials are always on stock and there is no need to use other raw materials which can burden production costs.

MATERIALS AND METHODS

The analysis is an attempt to observe in detail a thing or object by describing its constituent or constituent components for further research[12]. In collecting data, it is necessary to carry out certain methods to obtain data that suits your needs. The method used in this research is to make a direct observation of the work process by conducting interviews with several users to obtain the required data and then recording the data systematically based on the facts in PT. Penjalindo Nusantara, study the literature to collect references from books, articles, papers, journals about the apriori algorithm[13]. Association analysis or association rule mining is a data mining technique to find associative rules between a combination of items[14]. Association rule is a method that aims to find patterns that often appear among many transactions, where each transaction consists of several items. One of the stages in this method is called high-frequency pattern analysis. The importance of an associative rule can be determined by two parameters, namely support and confidence[15]. Support is the percentage of the combination of these items in the database, while confidence (certainty value) is the strong relationship between items in the association rule.[16]

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1909

The basic methodology of association analysis is divided into two stages:

1. Analyze the high-frequency pattern.

This stage looks for a combination of items that meet the minimum requirements of the support value in the database. The value of an item's support is obtained by the formula:[17]

$$Support\ (A) = \frac{\sum Transactions\ contains\ A}{\sum Transactions} \times 100\% \quad \text{.....(1)}$$

While the support value of 2 items is obtained from the following formula:

$$Support\ (A,B) = \frac{\sum Transactions\ contains\ A,B}{\sum Transactions} \times 100\% \quad \text{.....(2)}$$

And then to find the combination of 3 items, the support value of 3 items is obtained from the following formula:

$$Support\ (A,B,C) = \frac{\sum Transactions\ contains\ A,B,C}{\sum Transactions} \times 100\% \quad \text{.....(3)}$$

2. The determination of association rules

After all high frequent itemset patterns are found, then the associative rules that meet the minimum requirements for the confidence by calculating the confidence of associative rules A → B can be obtained from the following formula:

$$Confidence(B|A) = \frac{\sum Transactions\ Contains\ A,B}{\sum Transactions\ A} \times 100\% \quad \text{.....(4)}$$

Research Stages

The author using the systematic research steps that contain an outline of this research to help the research flow to be well directed. The steps that the author implemented can be seen below:[18]

1. Preliminary Study

The preliminary study includes a literature study related to the research topic. This step is to search the problem and study the problem that exists in the PT Penjalindo Nusantara, and then decide the background of the problem, scope of the problem, and then search some literature that related to the problem and implementing the solutions that suitable for the problems.

2. Collecting the Data

Collect the data by conducting some interviews with users related to the process that has a problem. This step is to aim the data that contains information of raw material purchase history data that purchased in January 2020 to November 2020, the data collected will be cleaned first to streamline the next process.

3. Data Mining Implementation

After collecting the data, the next process is identifying the type of raw material that want to be processed with the data mining, in this research, the raw material type that will be used is raw material ink, and then the next step is to determine the data mining method, in this research the data mining method is using the apriori algorithm to find the purchase pattern of raw material.

4. Processing the Data

After the data is ready to process with the apriori algorithm. The next process to processing the data by using a web-based application that has a function to analyze the data with the apriori algorithm, this application is running with PHP programming languages and using MySQL for the database. This application will import the data that will be processed with the .csv extension.

5. Analyze the Results

Analysis of the results by determining the final association rules done by calculating the data with the application.

6. Conclusions

The results of the purchase patterns calculation using the apriori algorithm is creating the percentages of raw material ink purchase patterns and showing the final results for the support value percentages and the confidence value percentages.

RESULTS AND DISCUSSION

In the process of analysis, several criteria are needed to be used in the implementation of data mining. In this research, the author collects the data with the appropriate criteria and attributes from the purchase transaction. The attributes used are the total purchase transactions and the type of material purchased and then processing the data with the apriori algorithm.

1. List of Raw Materials Types

The data of raw material ink with the item id-data that obtained for the data mining process using the apriori algorithm method can be seen in Table 1 below:

Table 1. List of Materials Type of PT Penjalindo Nusantara

No	Item ID	Description
1	RMI0000178	PU_SRKF 0001 MEDIUM UF T3_CT
2	RMI0000187	PU_VYS WHITE UN @15 KG_HII
3	RMI0000498	PU_LAMIC F 739 BLUE @180 KG_HII
4	RMI0000478	PU_LAMIC F 716 RED @180 KG_HII
5	RMI0000481	PU_LAMIC F 950 ORANGE @180 KG_HII
6	RMI0000487	PP_LNA 723 YELLOW @180 KG_HII
7	RMI0000488	PP_LNA 739 BLUE @180 KG_HII
8	RMI0000525	PP_LNA 795 BLACK @15 KG (PJ)_HII

No	Item ID	Description
9	RMI0000484	PU_LAMIC F MEDIUM @180 KG_HII
10	RMI0000476	PU_LAMIC F 702 RED @180 KG_HII
11	RMI0000485	PU_LAMIC F 722 TR YELLOW @180 KG_HII
12	RMI0000491	PP_LNA 702 RED @180 KG_HII
13	RMI0000086	PU_LAMIC F 779 GREEN_HII
14	RMI0000526	PP_LNA 817 RED @180 KG_HII

Source: [19]

2. Raw Material Ink Purchase Pattern

The data of raw material ink from the January 2020 – November 2020 period can be seen in Table 2 below:

Table 2. Raw Material Ink Purchase Pattern

Month	Description
January	RMI0000178, RMI0000187, RMI0000498, RMI0000478, RMI0000481, RMI0000487, RMI0000488, RMI0000525, RMI0000484, RMI0000476, RMI0000485, RMI0000491
February	RMI0000484, RMI0000498, RMI0000478, RMI0000481, RMI0000485, RMI0000086, RMI0000488, RMI0000525, RMI0000187, RMI0000476, RMI0000526, RMI0000491, RMI0000487, RMI0000178
March	RMI0000498, RMI0000476, RMI0000478, RMI0000481, RMI0000488, RMI0000487, RMI0000525, RMI0000187, RMI0000484, RMI0000485, RMI0000086, RMI0000526, RMI0000491, RMI0000178
April	RMI0000484, RMI0000498, RMI0000485 RMI0000478, RMI0000525, RMI0000476 RMI0000481, RMI0000488, RMI0000487 RMI0000491, RMI0000086
May	RMI0000484, RMI0000498, RMI0000478, RMI0000485, RMI0000481, RMI0000086, RMI0000487, RMI0000525, RMI0000476, RMI0000488, RMI0000491, RMI0000526
June	RMI0000498, RMI0000476, RMI0000478, RMI0000485, RMI0000484, RMI0000487, RMI0000481, RMI0000488, RMI0000491, RMI0000525
July	RMI0000484, RMI0000478, RMI0000485, RMI0000481, RMI0000526, RMI0000487, RMI0000525, RMI0000491, RMI0000488, RMI0000476, RMI0000498
August	RMI0000526, RMI0000491, RMI0000525, RMI0000484, RMI0000476, RMI0000478, RMI0000488, RMI0000487, RMI0000485, RMI0000481
September	RMI0000476, RMI0000478, RMI0000487, RMI0000498, RMI0000525, RMI0000526
October	RMI0000498, RMI0000491, RMI0000525, RMI0000488, RMI0000526, RMI0000487, RMI0000484, RMI0000478
November	RMI0000481, RMI0000476

Source: [19]

3. High-Frequency Patterns Calculations

Calculation of 1 itemset

To calculate 1 itemset or C1 with a minimum support value of 90% can be obtained by the following formula:

$$Support\ (A) = \frac{\sum Transactions\ contains\ A}{\sum Transactions} \times 100\% \text{ .....(5)}$$

The following Table 3 contains the candidate of 1 item sets:

Table 3. List of 1 itemset candidate

Item Set	QTY	Support
RMI0000086	4	36%
RMI0000178	3	27%
RMI0000187	3	27%
RMI0000476	10	91%
RMI0000478	10	91%
RMI0000481	9	82%
RMI0000484	9	82%
RMI0000485	8	73%
RMI0000487	10	91%
RMI0000488	9	82%
RMI0000491	9	82%
RMI0000498	9	82%
RMI0000525	10	91%
RMI0000526	7	64%

Source:[19]

With the minimum support value of 90%, the data with the support value below 90% will be removed, looks like the following Table 4:

Table 4. List of 1 itemset with a minimum support value of 90%

Item Set	QTY	Support
RMI0000476	10	91%
RMI0000478	10	91%
RMI0000487	10	91%
RMI0000525	10	91%

Source:[19]

Calculation of 2 itemset

To calculate 2 itemsets or C2 with a minimum support value of 90% can be obtained by the following formula:

$$Support\ (A,B) = \frac{\sum Transactions\ contains\ A,B}{\sum Transactions} \times 100\% \text{ .....(6)}$$

The following Table 5 contains the candidate of 1 item sets:

Table 5. List of 2 itemset candidate

Item Set	QTY	Support
RMI0000476, RMI0000478	9	82%
RMI0000476, RMI0000487	9	82%
RMI0000476, RMI0000525	9	82%
RMI0000476, RMI0000481	9	82%
RMI0000478, RMI0000487	10	91%
RMI0000478, RMI0000525	10	91%
RMI0000478, RMI0000484	9	82%
RMI0000478, RMI0000488	9	82%
RMI0000478, RMI0000491	9	82%
RMI0000478, RMI0000498	9	82%
RMI0000487, RMI0000525	10	91%
RMI0000487, RMI0000484	9	82%
RMI0000487, RMI0000488	9	82%
RMI0000487, RMI0000491	9	82%
RMI0000487, RMI0000498	9	82%
RMI0000525, RMI0000484	9	82%
RMI0000525, RMI0000488	9	82%
RMI0000525, RMI0000491	9	82%
RMI0000525, RMI0000498	9	82%

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1909

RMI0000484, RMI0000488	9	82%
RMI0000484, RMI0000491	9	82%
RMI0000488, RMI0000491	9	82%

Source:[19]

With the minimum support value of 90%, the data with the support value below 90% will be removed, looks like the following Table 6:

Table 6. List of 2 itemsets with a minimum support value of 90%

Item Set	QTY	Support
RMI0000478, RMI0000487	10	91%
RMI0000478, RMI0000525	10	91%
RMI0000487, RMI0000525	10	91%

Source: [19]

Calculation of 3 itemset

To calculate 3 itemsets or C3 with a minimum support value of 90% can be obtained by the following formula:

$$Support\ (A,B,C) = \frac{\sum Transactions\ contains\ A,B,C}{\sum Transactions} \times 100\% \quad \text{.....(7)}$$

The following Table 7 contains the candidate of 3 item sets:

Table 7. List of 3 itemset candidate

Item Set	QTY	Support
RMI0000476, RMI0000478, RMI0000487	9	82%
RMI0000476, RMI0000478, RMI0000525	9	82%
RMI0000476, RMI0000487, RMI0000525	9	82%
RMI0000478, RMI0000487, RMI0000525	10	91%
RMI0000478, RMI0000487, RMI0000484	9	82%
RMI0000478, RMI0000487, RMI0000488	9	82%
RMI0000478, RMI0000487, RMI0000491	9	82%
RMI0000478, RMI0000487, RMI0000498	9	82%
RMI0000478, RMI0000525, RMI0000484	9	82%
RMI0000478, RMI0000525, RMI0000488	9	82%
RMI0000478, RMI0000525, RMI0000491	9	82%
RMI0000478, RMI0000525, RMI0000498	9	82%
RMI0000478, RMI0000484, RMI0000488	9	82%
RMI0000478, RMI0000484, RMI0000491	9	82%
RMI0000478, RMI0000488, RMI0000491	9	82%
RMI0000487, RMI0000525, RMI0000484	9	82%
RMI0000487, RMI0000525, RMI0000488	9	82%
RMI0000487, RMI0000525, RMI0000491	9	82%
RMI0000487, RMI0000525, RMI0000498	9	82%
RMI0000487, RMI0000484, RMI0000488	9	82%
RMI0000487, RMI0000484, RMI0000491	9	82%
RMI0000487, RMI0000488, RMI0000491	9	82%

Source:[19]

With the minimum support value of 90%, the data with the support value below 90% will be removed, looks like the following Table 8:

Table 8. List of 3 itemsets with a minimum support value of 90%

Item Set	QTY	Support
RMI0000478, RMI0000487, RMI0000525	10	91%

Source: [19]

Calculation of Association Rules

After calculating all the high frequent itemset patterns and get the data with the minimum requirement[20], then the next step is finding the associative rules that meet the minimum requirements for confidence by 100% by calculating 2 itemsets with the following formula:

$$Confidence(B|A) = \frac{\sum Transactions\ Contains\ A,B}{\sum Transactions\ A} \times 100\% \quad \text{.....(8)}$$

The following Table 9 contains the association candidate:

Table 9. Association rules candidate

Item Set	Support	QTY	Confidence
if buy RMI0000478 then buy RMI0000498	82%	9/9	100%
if buy RMI0000478 then buy RMI0000491	82%	9/9	100%
if buy RMI0000491 then buy RMI0000484	82%	9/9	100%
if buy RMI0000525 then buy RMI0000491	82%	9/9	100%
if buy RMI0000525 then buy RMI0000484	82%	9/9	100%
if buy RMI0000488 then buy RMI0000491	82%	9/9	100%
if buy RMI0000491 then buy RMI0000488	82%	9/9	100%
if buy RMI0000488 then buy RMI0000484	82%	9/9	100%
if buy RMI0000484 then buy RMI0000488	82%	9/9	100%
if buy RMI0000525 then buy RMI0000488	82%	9/9	100%
if buy RMI0000487 then buy RMI0000491	82%	9/9	100%
if buy RMI0000487 then buy RMI0000484	82%	9/9	100%
if buy RMI0000487 then	91%	10/10	100%

Item Set	Support	QTY	Confidence
buy RMI0000525			
if buy RMI0000487 then buy RMI0000488	82%	9/9	100%
if buy RMI0000476 then buy RMI0000481	82%	9/9	100%
if buy RMI0000525 then buy RMI0000487	91%	10/10	100%
if buy RMI0000484 then buy RMI0000491	82%	9/9	100%
if buy RMI0000478 then buy RMI0000525	91%	10/10	100%
if buy RMI0000487 then buy RMI0000498	82%	9/9	100%
if buy RMI0000525 then buy RMI0000498	82%	9/9	100%
if buy RMI0000487 then buy RMI0000478	91%	10/10	100%
if buy RMI0000478 then buy RMI0000487	91%	10/10	100%
if buy RMI0000525 then buy RMI0000478	91%	10/10	100%
if buy RMI0000478 then buy RMI0000488	82%	9/9	100%
if buy RMI0000478 then buy RMI0000484	82%	9/9	100%

Source: [19]

A final association rule selected by the author is based on a combination of two item sets with the minimum support value of 90% with the minimum confidence of 100%, the other association rules that below the minimum requirement will be eliminated, and then the final association rules that meet the minimum requirement can be seen by the following Table 10.

Table 10. List of final association rules with a minimum support value of 90% and a minimum confidence value of 100%

Item Set	Support	QTY	Confidence
if buy RMI0000487 then buy RMI0000478	91%	10/10	100%
if buy RMI0000478 then buy RMI0000487	91%	10/10	100%
if buy RMI0000525 then buy RMI0000478	91%	10/10	100%
if buy RMI0000478 then buy RMI0000525	91%	10/10	100%
if buy RMI0000525 then buy RMI0000487	91%	10/10	100%
if buy RMI0000487 then buy RMI0000525	91%	10/10	100%

Source: [19]

From Table 10 above, the data that can be analyzed which combination of raw materials ink that purchased by PT Penjalindo Nusantara simultaneously can be seen in Figure 1 below:
Source: [19]

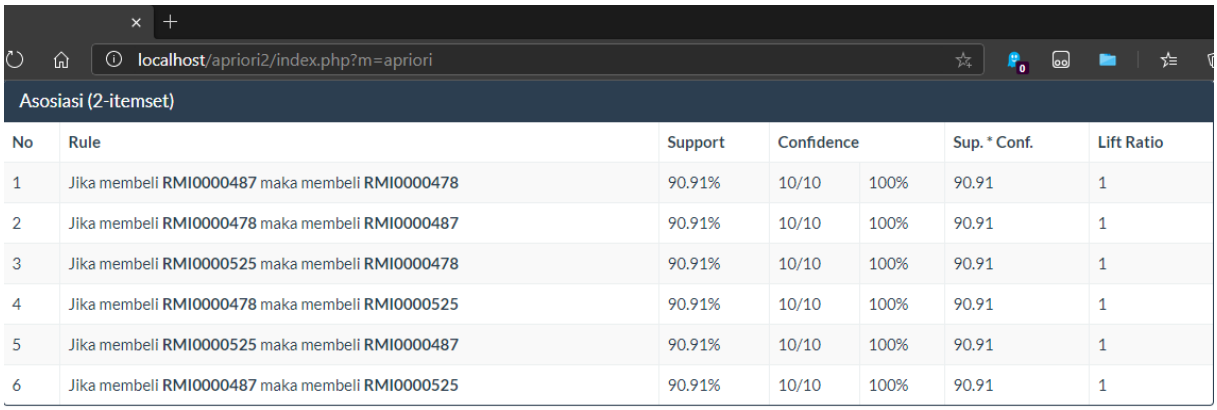


Figure 1. The Results of Final Association Rules Processed by Web Application

The results of the research based on the final association rules obtained from January 2020 to November 2020 there are 3 types of raw materials that are often purchased at the same time so that these results can be used as a reference for production material procurement officers or material planners in procuring raw material stock items and increase the effectiveness of raw material planning strategy.

CONCLUSION

The results of research by processed with the Apriori algorithm using the web-based application can solve the problems of raw material planning with the raw material purchase patterns that have a high percentage. The percentage of purchasing with the raw material PP_LNA 723 YELLOW @180 KG_HII and the raw material PU_LAMIC F 716 RED @180 KG_HII is 90% of support value and the percentage of confidence is 100%. The percentage of purchasing with the raw material PU_LAMIC F 716 RED @180 KG_HII and the raw material PP_LNA 723 YELLOW @180 KG_HII is 90% of support value and the percentage of confidence is

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1909

100%. The percentage of purchasing with the raw material PP_LNA 795 BLACK @15 KG (PJ)_HII and the raw material PU_LAMIC F 716 RED @180 KG_HII is 90% of support value and the percentage of confidence is 100%. The percentage of purchasing with the raw material PU_LAMIC F 716 RED @180 KG_HII and the raw material PP_LNA 795 BLACK @15 KG (PJ)_HII is 90% of support value and the percentage of confidence is 100%. The percentage of purchasing with the raw material PP_LNA 795 BLACK @15 KG (PJ)_HII and the raw material PP_LNA 723 YELLOW @180 KG_HII is 90% of support value and the percentage of confidence is 100%. The percentage of purchasing with the raw material PP_LNA 723 YELLOW @180 KG_HII and the raw material PP_LNA 795 BLACK @15 KG (PJ)_HII is 90% of support value and the percentage of confidence is 100%.

REFERENCE

[1] K. Hidayat, J. Efendi, and R. Faridz, "Analisis Pengendalian Persediaan Bahan Baku Kerupuk Mentah Potato Dan Kentang Keriting Menggunakan Metode Economic Order Quantity (EOQ)," *Performa Media Ilm. Tek. Ind.*, vol. 18, no. 2, pp. 125–134, 2020.

[2] N. Wahyu and E. P. Ningrum, "Penerapan Pengendalian Internal Melalui Pengelolaan Persediaan Bahan Baku (Studi Kasus Pada CV. Cipta Gemilang Karya Mandiri)," *CAKRAWALA*, vol. 1, no. 1, 2018.

[3] F. Sulaiman and Nanda, "PENGENDALIAN PERSEDIAAN BAHAN BAKU DENGAN MENGGUNAKAN METODE EOQ PADA UD. ADI MABEL," *TEKNOVASI*, vol. 2, no. 1, pp. 1–11, 2015.

[4] T. Pujadi, "Model Pemesanan Bahan Baku menggunakan Peramalan Time Series dengan CB Predictor," *ComTech Comput. Math. Eng. Appl.*, vol. 5, no. 2, pp. 954–962, 2014.

[5] N. Apriyani and A. Muhsin, "Analisis Pengendalian Persediaan Bahan Baku Dengan Metode Economic Order Quantity Dan Kanban Pada Pt Adyawinsa Stamping Industries," *Opsi*, vol. 10, no. 2, p. 128, 2017.

[6] P. B. I. S. Putra, N. P. S. M. Suryani, and S. Aryani, "Analysis of Apriori Algorithm on Sales Transactions to Arrange Placement of Goods on Minimarket," *IJEET Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 13–17, 2018.

[7] Munarwan, A. Sinaga, and U. Nugraha, "Implementation of Apriori Algorithm for Determining Purchase Patterns in One Transaction," *Int. J. Eng. Technol.*, vol. 7, no. 4.33, pp. 204–207, 2018.

[8] D. S. Wijaya and D. Fitrianah, "Data Mining Analysis with Association Rules Method to Determine the Result of Fish Catch using FP-Growth Algorithm," *Int. J. Comput. Appl.*, vol. 181, pp. 7–15, 2018.

[9] Y. Nawawi and I. Suharjo, "Implementasi Data Mining Menentukan Rekomendasi Penempatan Buku Berdasarkan Pola Peminjaman Dengan Menggunakan Association Rule," *J. Multimed. Artif. Intell.*, vol. 4, no. 1, 2020.

[10] A. Junaidi, "Implementasi Algoritma Apriori dan FP-Growth Untuk Menentukan Persediaan Barang," *J. SISFOKOM*, vol. 8, no. 1, 2019.

[11] D. A. O. Turang and S. R. Astari, "SISTEM MANAJEMEN INVENTORY MENGGUNAKAN ALGORITMA DIFFERENTIAL EVOLUTION," *Semin. Nas. Inform. 2018 (semnasIF 2018)*, 2018.

[12] P. H. Simbolon, "Implementasi Data Mining Pada Sistem Persediaan Barang Menggunakan Algoritma Apriori (Studi Kasus : Srikandi Cash Credit Elektronic dan Furniture)," *J. Ris. Komput.*, vol. 6, no. 4, pp. 401–406, 2019.

[13] G. A. Syaripudin and E. Faizal, "Implementasi Algoritma Apriori Dalam Menentukan Persediaan Obat," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, 2017.

[14] S. Wahyuni, Suherman, and L. P. Harahap, "Implementasi Data Mining dalam Memprediksi Stok Barang Menggunakan Algoritma Apriori," *J. Tek. DAN Inform.*, vol. 5, no. ISSN 2089-5490, 2018.

[15] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data," *Knowl. Eng. Data Sci.*, vol. 1, no. 1, pp. 20–25, 2018.

[16] M. T. Permatasari and H. D. Purnomo, "Penerapan Algoritma Apriori Untuk Mengetahui Pola Penempatan Buku dan Rekomendasi Persediaan Buku di Perpustakaan Daerah Kota Salatiga," 2019.

[17] R. Wahyuning, A. M. Kom, L. S. M. Kom, H. Wijaya, and T. Informatika, "Penerapan Association Rule Pada Data Persediaan Bahan Baku Di Pro Ab Chicken Jambi," *Fortech*, vol. 1, no. 1, pp. 43–48, 2017.

[18] L. Kurniawati, A. E. Kusuma, and B. Dewansyah, "Implementasi Algoritma Apriori Untuk Menentukan Persediaan Spare Part Compressor," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 6, 2019.

[19] D. R. Prastyo, "Implementasi Algoritma

Apriori Untuk Menentukan Pola Pembelian Bahan Baku di PT Penjalindo Nusantara," Jakarta, 2020.

[20] S. Anggraeni, M. A. Iha, W. Erawati, and S. Khairunnas, "Analysis of Sales by Using Apriori and FP-Growth at PT. Panca Putra Solusindo," *Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 3, no. 2, pp. 41–46, Mar. 2019.

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1377

NEW STUDENT ADMISSIONS INFORMATION SYSTEM WITH CLIENT SERVER BASED SMS GATEWAY

Zulhipni Reno Saputra Elsi[1*]; Gita Rohana[2]; Vera Nuranjani[3]

Teknologi Informasi[1,2,3]
Universitas Muhammadiyah Palembang
www.um-palembang.ac.id
zulhipni_renosaputra@um-palembang.ac.id, gitarohana306@gmail.com, veranuranjani2016@gmail.com
(*) Corresponding Author

Abstract—Palembang State Junior High School 58 every year accepts new students, every year SMP N 58 always uses the form in registration and collects diplomas, SKHUN, and other files in hardcopy format so that they often experience file loss for that we need a new student admission information system based client-server and SMS gateway. This information system was designed using the Software Development Life Cycle (SDLC) method, and an analysis and design were carried out using a Data Flow Diagram. This information system has the function of saving, delete, update, report automatically, and can send information in the form of SMS. The new student admission information system is user friendly so that it can be easily used, so the admission process is more effective and there are no more missing files.

Keywords: New Student Admission; Information System, SMS Gateway; Software Development Life Cycle.

Intisari—*Sekolah Menengah Pertama Negeri 58 Palembang setiap tahun melakukan penerimaan siswa baru, setiap tahun SMP N 58 selalu menggunakan formulir dalam pendaftaran dan mengumpulkan ijazah, SKHUN dan berkas lainya dalam bentuk hardcopy sehingga sering mengalami kehilangan berkas untuk itu diperlukan sebuah sistem informasi penerimaan siswa baru yang berbasis client server dan sms gateway. Sistem informasi ini dirancang dengan menggunakan metode Software Development life Cycle (SDLC), serta dilakukan analisa dan perancangan dengan Data Flow Diagram. Sistem informasi ini mempunyai fungsi simpan, hapus, update, laporan automatis, serta dapat mengirim informasi berupa SMS. Sistem informasi penerimaan siswa baru bersifat user friendly sehingga dapat dengan mudah digunakan, sehingga proses penerimaan siswa baru lebih efektif dan tidak adanya lagi berkas berkas yang hilang..*

***Kata Kunci**: Sistem Informasi; Penerimaan Siswa Baru SMS Gateway;Software Development Life Cycle.*

INTRODUCTION

With the rapid development of technology, the new student registration process should be done anywhere and anytime, for example at home, outside the city, and no longer need to come to school to queue up for forms[1]and the payment process can be done online[2].

In the research, Umi Kholifah said that the implementation of the New Student Admissions Information System provides time efficiency in the processing of new student admissions[3]. Currently, computer-based data processing has developed rapidly, the use of computer programs is very helpful for its users to solve difficult problems so that data processing can be done well besides that it also has high accuracy so that it will make it easier and not take time. long time searching for the required data.

Table 1. Previous Research

NO	NAME	PROBLEM	METHOD	RESULT
1	Wahyu Hidayat, Ramadhian Agus Triono, Sukadi Pembangunan Sistem Informasi Pendaftaran Siswa Baru SMP Negeri 2 Sudimoro Pacitan	Data collection for new student registration at SMPN 2 Sudimoro Pacitan is done conventionally.	Designed using context diagrams	With this new student registration information system, the PSB committee can accelerate the process of recording student data, reducing mistakes in recording and losing data[4].

NO	NAME	PROBLEM	METHOD	RESULT
2	Santoso Sistem Informasi Pendaftaran Siswa Baru Pada SMK Diponegoro Tulakan	The system in processing new student data at SMK Diponegoro has not been computerized	Make observations of the object being observed then perform problem-solving analysis	With the new student registration information system that uses the Microsoft Visual Basic 6.0 programming language, data processing of new student candidates becomes easier and more efficient[5].
3	Fajar Sidik, Mari Rahmawati Perancangan Sistem Informasi Pendaftaran Siswa Baru Berbasis Web Pada SMK Bina Putra Jakarta	The registration system is still offline, uses paper, which results in queues when registering, and risks losing data	Waterfall method	The existence of an online information system reduces paper usage and does not occur data loss so that queues do not occur when registering[6].

Based on table 1, it can be concluded that the problems that exist in schools at the time of admission of new students are the systems that still use paper in the registration process and new student data collection and this problem can be overcome by developing a new student admission information system even though using different methods.

Junior High School (SMP) Negeri 58 Palembang was established on June 23, 2014. SMP Negeri 58 Palembang every year will accept new students (PSB), in this PSB process PSB officers are still managing registration data and admission of prospective new students still using the media paper for registration.

In the first process, prospective students are required to fill out a registration form and submit the required documents, such as diploma, SKHUN, and others to the PSB committee. After the registration form and required documents are received by the PSB committee, then the forms and documents are verified. If the required forms and documents are declared complete, the PSB committee will record the registration of prospective students in the registration book as well as provide details of education payments. In the next stage, prospective students are required to make educational payments to the PSB committee after receiving details of education payments [7].

When students return the registration files to the PSB committee and the registration files for new prospective students are still kept in an archive, so they are vulnerable to file loss and damage because too many students register[8]. Then the PSB committee recaps into the computer in the process of data recording the committee using Microsoft Excel[9].

Currently, SMPN 58 Palembang requires proper data processing to create efficiency and accuracy of data to support operational processes, management, and decision-making processes properly. With the above problems, it is necessary to improve the system that is running, for that we need a client-based application. Server and SMS Gateway to simplify the PSB process.

MATERIALS AND METHODS

This research refers to the research framework described in Figure 1.

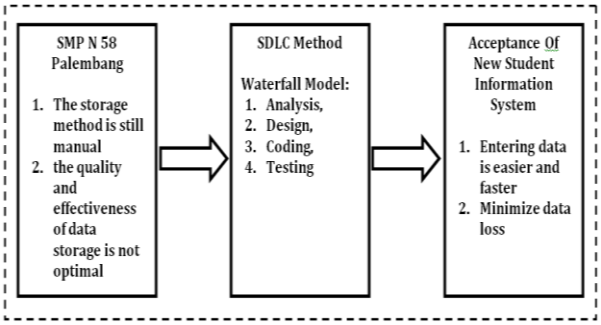


Figure 1. Research Framework

In Figure 1, the stages carried out in research where SMP Negeri 58 Palembang have 2 problems. The problem is that the system developer uses the SDLC method with the waterfall model and finally produces a new information system.

In the development of this information system, research data was collected, the research data were obtained by two methods, namely:

1) Interview

Researchers conducted direct interviews with the principal of SMPN 58 Palembang and the head of administration at SMP N 58 Palembang, where the results of the interview revealed the procedures for new student admissions, input, and data reports of new students that had been running at SMPN 58 Palembang.

2) Observation

Observation is data collection carried out by direct observation of the object of research[10]. Researchers made observations at SMPN 58 Palembang, by directly surveying how new student data input and new student data reports. The

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1377

results of observations made by the author at SMPN 58 Palembang are examples of registration forms that have been filled in by new student candidates submitted to the new student admissions committee or administrative staff at SMPN 58 Palembang then selected according to the requirements for new student admission then announced on the school walls After that, the student data received is recapitulated to a computer using the Microsft Excel application, wherein Microsoft Excel, not all new student data is entered, this is due to time efficiency in the processing process.

In this study, the system was developed using the Software Development Life Cycle (SDLC) method. The stages in this study refer to the Waterfall model stages, namely analysis, design, coding, testing, and maintenance [11] [1][12][13].

A. Needs Analysis

Data input is a process of entering the relevant data into one place (in this case software) with the aim that the data is not lost and can be reopened. For example, at SMPN 58 Palembang, this government agency engaged in education uses Microsoft Excel software. In entering data for new students at SMPN 58 Palembang, the administrative staff who served as the entry committee were found to be obstacles, the extent to which the data was lost, and also the ineffective time used to enter the data.

This new student admission application is designed to help overcome problems faced by the committee or new admissions staff when inputting and reporting new student data at SMPN 58 Palembang. The general function of this application design is to input new student data, in this new student admission application is equipped with edit, cancel, search, save, delete, and print facilities.

B. Design

The design is part of the Waterfall model stage, this stage is carried out after the analysis stage. At this stage, it provides a detailed description. In system design, the system design that will be built is described before coding into a programming language [4].

Figure 2 illustrates the data flow used during data processing, the data flow is drawn with a Data Flow Diagram. Data Flow Diagrams are a way to document the design system process to show the flow of information and information transformation applied as data flowing from input and output[3][14].

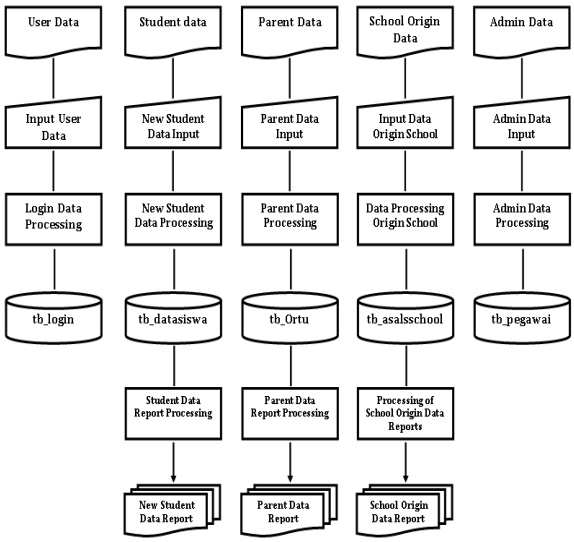


Figure 2. Flow Chart

Use case new student information system is depicted in the diagram in Figure 3.

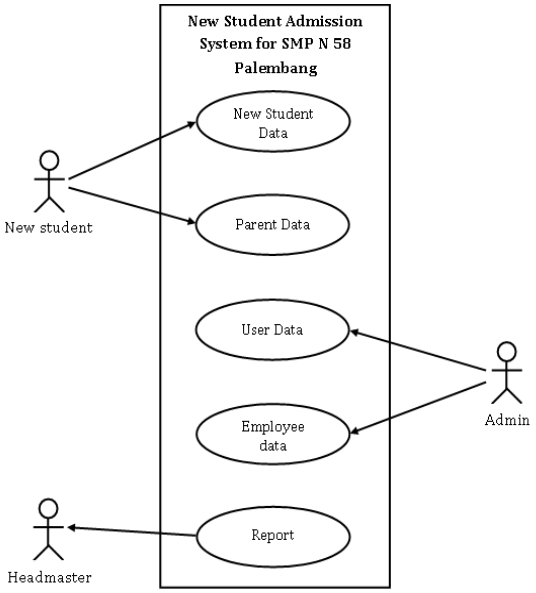


Figure 3. Uses case

Figure 3 explains those who play a role in using the information system, namely the admin (committee), the principal, and new students.

C. Encoding

At this stage, the preparation or writing of the Delphi programming language is carried out according to the system design that has been made so that it becomes the required information system. The Client-Server Information System topology is depicted in Figure 4.

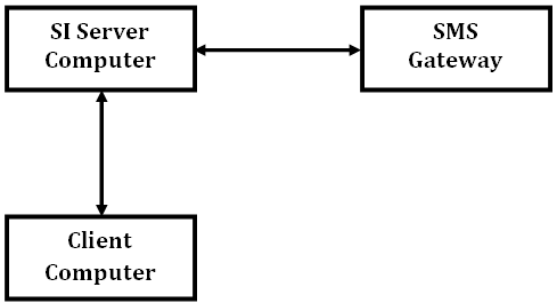


Figure 4. Client-Server Information System Topology

Figure 4 shows the client-server topology, the information system used consists of a server computer connected to the SMS gateway and a server computer connected to the client computer.

D. Testing

At this stage, testing is carried out on a system that has been completed using a network with two computers, with one computer being used as a server and the other as a client. This testing system uses the User Accepted Test (UAT) method with the UAT type Black Box Testing. Black Box Testing is functionality testing, where the end-user will test the function of the software regardless of the internal code structure.

RESULTS AND DISCUSSION

The results of the design of a new admissions information system with data flow diagrams can be implemented in the system. Implementation is the steps or procedures carried out in completing an approved system design, to test, install, and start a new system or a repaired system to replace the old system.

The dashboard for the new student admission information system only has the login and CANCEL menu, on this menu application users who want to open the information system must first log in to be able to the main information system dashboard.

Figure 5. Information System Dashboard

Figure 5 shows the information system dashboard, there are only log in and Exit buttons. After the user logs in, the main dashboard will appear as depicted in Figure 6.

Figure 6. Main Dashboard of Information Systems

In Figure 6 6 buttons can be used by users, these 5 buttons can call the new student-parent data input menu function, new student data input, user data, employee data input, report list, and 1 button to exit.

The new student data form is used to input data related to new students such as NIS, NISN, student name, address, date of birth, and some others. In this form there is a menu for the process of saving, editing, exiting, deleting, refreshing, adding, printing, and also searching for data. The image of the new student data input form can be seen in Figure 7.

Figure 7. New Student Data Form

The process of filling in the new student data input form by filling in each textbox on the form, after all the textboxes are filled, new student data can be saved after saving by clicking the SAVE button.

After all the data is stored, the data will enter into the dBGrid and immediately provide an SMS notification to the new students that the data attached is complete. The SMS notification can be seen in Figure 8.

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1377

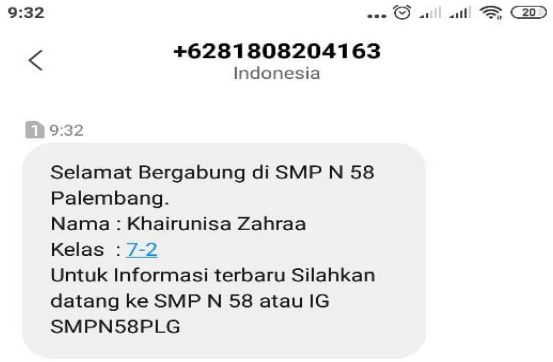


Figure 8. SMS notification

The parent data form is used to input data related to the parents of new students such as the name of the father, mother, occupation, address, and NISN. In this form there is a menu for the process of saving, editing, exiting, deleting, refreshing, adding, printing, and also searching for data. The image of the new student parents' data input form can be seen in Figure 9.

Figure 9. Parent Data Form

The employee data form is used to input data related to employees such as id, name, gender, place of birth, date of birth, and position. In this form there is a menu for the save, edit, out, delete, refresh, and add processes. The image of the employee data input form can be seen in Figure 10.

Figure 10. Employee Data Form

A user data form is for inputting user data such as admin and user, as for inputted data such as name, password, position, and status. The image of the user data form can be seen in Figure 11.

Figure 11. User Data Form

In Figure 11 there is a combo box status, this combo box functions for users of the new student admission information system.

In this report menu, the admin and user can select a list of reports such as new student data reports, new student parents data reports, new student school data reports, student data reports based on the year of entry, and student data reports by gender. The menu display can be seen in Figure 12.

Figure 12. SI report

Figure 13. New Student Report

Figure 13 shows the new student admission report form based on new student data.

Figure 14. Report of New Students with School of Origin

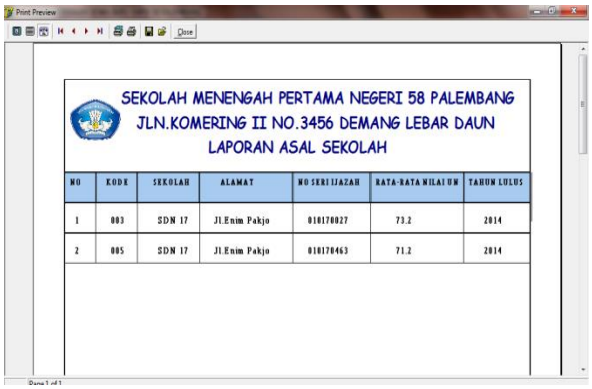


Figure 15. Report of New Student School Origin Data

Figure 14 shows the new student admission report form based on the student's school of origin.

Whereas Figure 15 shows the school data report based on the origin of the new student's school.

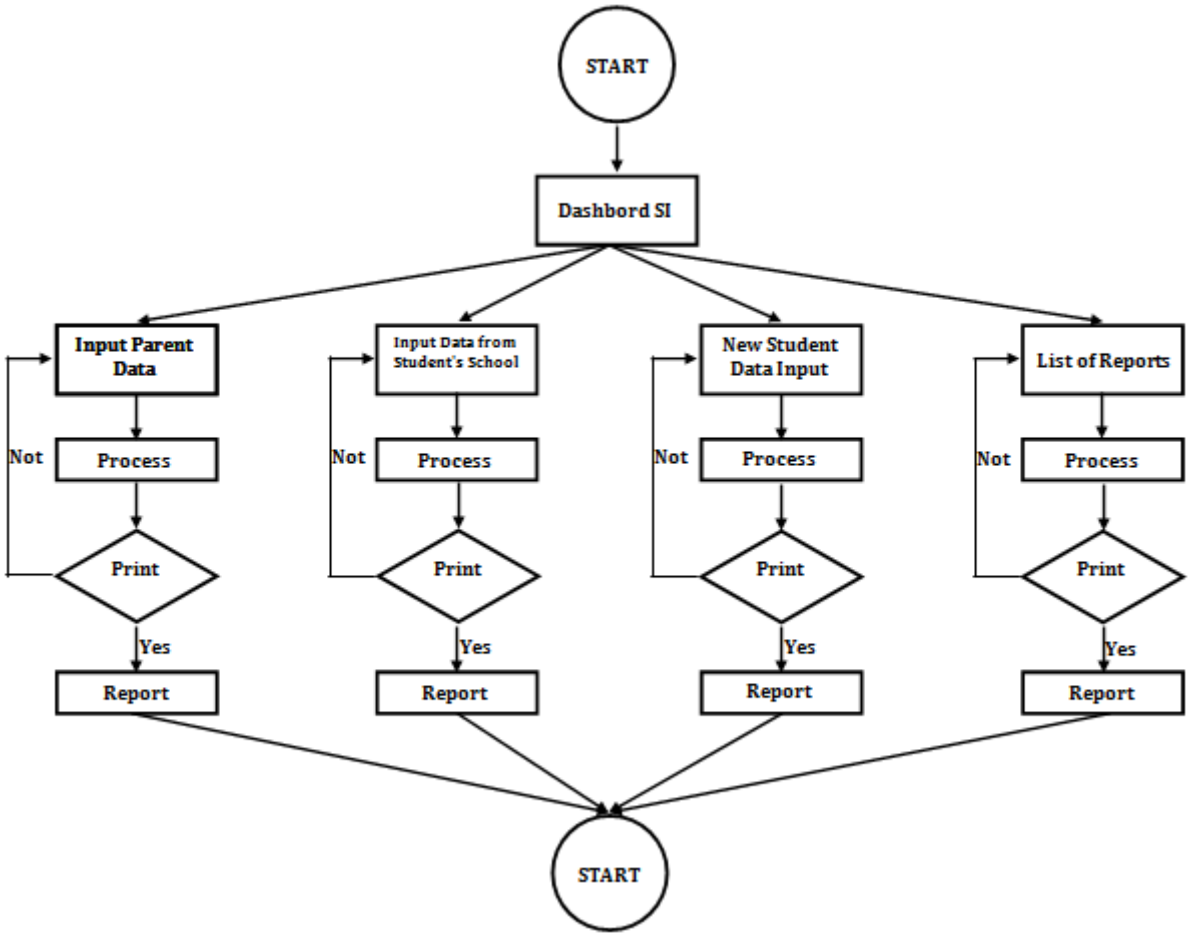


Figure 16. Flowchart of new student admission information system

Figure 16 illustrates the flowchart of the use of the new student admissions information system starting with a star then processed, in this first process 6 processes can be used. Of these 6 processes, there are 4 processes, while the other 2 processes have no reports.

To test the information system that has been created, information system testing is carried out using a server computer, an SMS gateway, and a server computer. Based on the testing that has been done, the results are:

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1377

a. Data processing can run better, all data inputted can be stored in the database.
b. The new admissions information system is user friendly.
c. The process of creating reports relating to new students can be done quickly.

From the results of testing the information system for new student admissions, there are advantages and disadvantages, including:

1) Advantage

In the data entry process, it is younger, because it uses a database so that the process of searching for data and storing data is younger and safer. The new student admission information system can create new student reports, new student parent data reports, and new student school data reports.

2) Loss

In this new student admission application, it is still limited to internal, where only the school can input and see, external or external parties cannot because it is not online yet.

The test results with Black Box Testing can be seen in Table 1 below.

Table 1. Black box testing success percentage

Menu	Successful (%)	Failed (%)	Timed Out (%)	Not Played (%)
New Student Data	98.76	0.44	0.62	0.18
Parent Data	99.01	0.38	0.42	0.19
User Data	99.63	0.24	0.13	0
Employee data	98.96	0.57	0.36	0.11
Report	99.84	0.08	0.08	0
Total	99.24	0.342	0.322	0.096

CONCLUSION

Successfully creating a new student admission system based on client-server and SMS gateway at SMP Negeri 58 Palembang, this information system can input new student data, new student parents data, user data, employee data, and reports. The report is related to data entry so that the new student data processing information system can help new student data entry at SMPN 58 Palembang more effectively and new students can receive information via SMS. Based on the results of the UAT test, the acceptance rate of this new student admission system is on average 99.24 percent, the failure of this system is on average 0.342 percent, Timed Out is on average 0.322 and Not Played is on average 0.096.

REFERENCE

[1] Sunarti, "Implementasi Sistem Informasi Penerimaan Siswa Baru," *J. Evolusi - Vol. 3 No 2 - 2015 - lppm3.bsi.ac.id/jurnal*, 2015.

[2] Jimmie, M. I. Ma'ruf, and D. Mahendra, "Sistem Informasi Penerimaan Siswa Baru Berbasis Web Di Sekolah Menengah Pertama Negeri 43 Palembang," *J. Digit. Teknol. Inf.*, vol. 1, no. 2, pp. 98–105, 2018.

[3] I. U. W. Umi Kholifah, "Sistem Informasi Pendaftaran Peserta Didik Baru Pada Sekolah Menengah Kejuruan Negeri 1 Sudimoro," *IJNS – Indones. J. Netw. Secur.*, 2014, doi: 10.1123/IJNS.V3I3.1254.

[4] W. Hidayat, R. A. Triono, and Sukadi, "Pembangunan Sistem Informasi Pendaftaran Siswa Baru Smp Negeri 2 Sudimoro Pacitan," *IJNS – Indones. J. Netw. Secur.*, vol. 4, no. 3, pp. 1–6, 2013.

[5] Santoso, "Sistem Informasi Pendaftaran Siswa Baru Pada SMK Diponegoro Tulakan," *J. Speed – Sentra Penelit. Eng. dan Edukasi*, vol. 7, no. 1, pp. 46–52, 2015.

[6] F. Sidik and M. Rahmawati, "Perancangan Sistem Informasi Pendaftaran Siswa Baru Berbasis Web Pada SMK Bina Putra Jakarta," *Paradigma*, vol. 20, no. 1, pp. 119–128, 2018.

[7] Ruhul Amin, "Rancang Bangun Sistem Informasi Penerimaan Siswa Baru Pada SMK Budhi Warman 1 Jakarta," *J. ILMU Pengetah. DAN Teknol. Komput.*, vol. 2, no. 2, pp. 113–121, 2017.

[8] S. Suherni, N. Rubiati, and H. Khumaini, "Aplikasi Pendaftaran Dan Penerimaan Siswa Baru Di Smk Negeri 1 Rupat Berbasis Web Dan SMS Gateway," *J. Inform. Manaj. dan Komput.*, vol. 9, no. 2, pp. 15–24.

[9] D. A. Wasesha, "Implementasi Metodologi Waterfall Dalam Sistem Pendukung Keputusan Penyedia Informasi Akademik Universitas," *INTI Nusa Mandiri*, vol. 14, no. 1, pp. 41–48, 2019.

[10] D. Haryanto and Apriansyah, "Kesiapan Industri Kelapa Sawit Dalam Implementasi Enterprise Resource Planning," *J. Teknol. Inf. Mura*, vol. 12, no. 1, pp. 29–43, 2020.

[11] R. Febryan, R. I. Euclerr, Y. Wardaningsih, and F. Masya, "Web-Based Event Seminar Registration Information System Application," *J. Ris. Inform.*, vol. 2, no. 3, pp. 159–168, Jun. 2020, DOI: 10.34288/jri.v2i3.144.

[12] S. Shofia and D. A. Anggoro, "Sistem Informasi Manajemen Administrasi Dan Keuangan Pada Tk-It Permata Hati Sumberrejo-Bojonegoro," *J. ILMU Pengetah. DAN Teknol. Komput.*, vol. 5, no. 1, pp. 221–230, 2020.

[13] Karnadi and T. Ismail, "Sistem Informasi Penerimaan Siswa Baru Berbasis Web Di SMA Muhammadiyah 1 Muara Padang," *J. Digit. Teknol. Inf.*, vol. 1, no. 2, pp. 56–63, 2018.

[14] Ermatita, "Analisis Dan Perancangan Sistem Informasi Perpustakaan," *J. Sist. Inf.*, vol. 8, no. 1, pp. 966–977, 2016.

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1968

COMPARATIVE ANALYSIS OF SOFTWARE EFFORT ESTIMATION USING DATA MINING TECHNIQUE AND FEATURE SELECTION

Abdul Latif [1*], Lady Agustin Fitriana [2], Muhammad Rifqi Firdaus [3]

Computer Science[1,2,3]
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri[1,2,3]
www.nusamandiri.ac.id
14002369@nusamandiri.ac.id [1], 14002386@nusamandiri.ac.id [2], 14002384@nusamandiri.ac.id [3]
(*) Author Corresponding

Abstract—Software development involves several interrelated factors that influence development efforts and productivity. Improving the estimation techniques available to project managers will facilitate more effective time and budget control in software development. Software Effort Estimation or software cost/effort estimation can help a software development company to overcome difficulties experienced in estimating software development efforts. This study aims to compare the Machine Learning method of Linear Regression (LR), Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Decision Tree Random Forest (DTRF) to calculate estimated cost/effort software. Then these five approaches will be tested on a dataset of software development projects as many as 10 dataset projects. So that it can produce new knowledge about what machine learning and non-machine learning methods are the most accurate for estimating software business. As well as knowing between the selection between using Particle Swarm Optimization (PSO) for attributes selection and without PSO, which one can increase the accuracy for software business estimation. The data mining algorithm used to calculate the most optimal software effort estimate is the Linear Regression algorithm with an average RMSE value of 1603,024 for the 10 datasets tested. Then using the PSO feature selection can increase the accuracy or reduce the RMSE average value to 1552,999. The result indicates that, compared with the original regression linear model, the accuracy or error rate of software effort estimation has increased by 3.12% by applying PSO feature selection.

Keywords: Software Effort Estimation, Machine Learning, Feature Selection, PSO, RMSE.

Abstrak— *Pengembangan perangkat lunak melibatkan beberapa faktor yang saling terkait yang mempengaruhi upaya pengembangan dan produktivitas. Meningkatkan teknik estimasi yang tersedia untuk manajer proyek akan memfasilitasi kontrol waktu dan anggaran yang lebih efektif dalam pengembangan perangkat lunak. Software Effort Estimation atau estimasi biaya / usaha perangkat lunak dapat membantu perusahaan pengembang perangkat lunak untuk mengatasi kesulitan yang dialami dalam memperkirakan upaya pengembangan perangkat lunak. Penelitian ini bertujuan untuk membandingkan metode Machine Learning Regresi Linier (LR), Multilayer Perceptron (MLP), Radial Basis Function (RBF), dan Decision Tree Random Forest (DTRF) untuk menghitung estimasi biaya / upaya software. Kemudian kelima pendekatan ini akan diujikan pada dataset proyek pengembangan perangkat lunak sebanyak 10 dataset proyek. Sehingga dapat menghasilkan pengetahuan baru tentang pembelajaran mesin dan metode pembelajaran non mesin apa yang paling akurat untuk memperkirakan bisnis software. Serta mengetahui antara pemilihan antara menggunakan Particle Swarm Optimization (PSO) untuk pemilihan atribut dan tanpa PSO, yang mana dapat meningkatkan akurasi software effort estimation. Algoritma data mining yang digunakan untuk menghitung estimasi upaya perangkat lunak yang paling optimal adalah algoritma Regresi Linier dengan nilai RMSE rata-rata 1603.024 untuk 10 dataset yang diuji. Kemudian menggunakan pemilihan fitur PSO dapat meningkatkan akurasi atau menurunkan nilai rata-rata RMSE menjadi 1552.999. Hasil penelitian menunjukkan bahwa, dibandingkan dengan model regresi linier asli, akurasi atau tingkat kesalahan estimasi upaya perangkat lunak telah meningkat sebesar 3,12% dengan menerapkan pemilihan fitur PSO.*

***Kata Kunci**: Software Effort Estimation, Machine Learning, Feature Selection, PSO, RMSE.*

INTRODUCTION

Software Effort Estimation (SEE) is needed because software development is limited by predetermined costs and schedules. Estimation is the activity of estimating how many resources are needed to complete a project plan. Developer often faces a variety of difficult situations that make it fail,

such as software being delivered late, unreliable, using costs several times higher than originally estimated, and often exhibiting poor performance characteristics so that project managers have difficulty estimating projects which it runs. The project failure was caused by a management approach to developing software[1].

The success of a development project is influenced by many factors, including executive support, user involvement in the project, project management experience, clear business objectives, software infrastructure, and the use of formal development methodologies. Other factors are factors related to the timing and scope of the project, including the minimal scope and reliable estimation. A software project is considered a failure if the project exceeds 50% of the planning cost and passes the predetermined schedule. The accuracy of estimating and measuring a software project is very important in facilitating the resource manpower and estimation effort on an IT project [2]. Ideally, in estimating software effort/software effort estimation, machine learning techniques can be used to predict, control, or significantly reduce the effort associated with building software [3].

In this study, researchers will apply machine learning methods to predict software business estimates using the Linear Regression (LR)[4], Multilayer Perceptron (MLP)[5], Radial Basis Function (RBF)[6], and Decision Tree Random Forest (DTRF)[7]. It is hoped that this will generate new knowledge on what machine learning methods are most accurate for estimating software effort. As well as knowing how effective the use of Particle Swarm Optimization (PSO) is in increasing the accuracy of software effort estimation [7].

Several effort estimation techniques exist and they can be classified under three main categories [8]. These categories are:

1. Expert judgment: In this category, a project estimator tends to use his or her expertise which is based on historical data and similar projects to estimate software. This method is very subjective and it lacks standardizations and thus, cannot be reusable. Another drawback of this method is the lack of analytical argumentation because of the frequent use of phrases such as "I believe that . . ." or "I feel that . . .".
2. Algorithmic models: This is still the most popular category in the literature [9]. These models include COCOMO [10], SLIM [11], and SEER-SEM [12]. The main cost driver of these models is the software size, usually the Source Lines of Code (SLOC). Algorithmic models either use a linear regression equation, like the one used by Kok et al. (1990), or non-linear regression equations, those which are used by Boehm (1981).
3. Machine learning: Recently, machine learning techniques are being used in conjunction or as alternatives to algorithmic models. These techniques include neural networks, fuzzy logic, neuro-fuzzy, Genetic Algorithm, and regression trees. Machine learning models can incorporate historical data and can be trained to better predict software effort.

This study aims to provide a framework that enables managers to make reasonable estimates of resources, costs, and schedules. Then, these estimates are made within a limited time frame at the start of the project and must be updated regularly as the project progresses. So that we get the best scenario and the worst scenario and the project results can be limited. The benefits of this research are expected to minimize software project failures by providing a framework that enables project managers to make reasonable estimates of resources, costs, and schedules [4]. This paper is a continuation of previous research, which compares Linear Regression (LR), Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Decision Tree Random Forest (DTRF) to get the best machine-learning algorithm to predict software effort estimation and improve the accuracy with feature selection Particle Swarm Optimizer (PSO).

Several previous studies by Nassif have discussed software effort estimation using a log-linear regression model based on the use case point (UCP) model to calculate software effort as well as fuzzy logic and multilayer perceptron neural network (MLP) models [8]. Then, Nassif continued his research using Regression Fuzzy Models [13]. In other research, BaniMustafa has predicted software effort estimation using three machine learning methods include Naïve Bayes, Logistic Regression, and DTRF [7]. Other research using kNN [14], neural networks [15], Artificial Neural Network (ANN) [16], Cascade-Correlation Neural Network (CNN) [17], Radial Basis and Generalized Regression [18]. In other cases, researchers used the particle swarm optimizer (PSO) feature selection to improve the accuracy of software effort estimation using the Artificial Neural Network algorithm [19]. Then a comparative study using tree/rule-based models (M5 and CART), linear models (ordinary least squares regression with and without various transformations, ridge regression (RR), and robust regression (RoR)), nonlinear models (MARS, least squares support vector machines, multilayered perceptron neural networks (NN), and radial basis function (RBF) [5]. Previous literature in Software Effort Estimation can be seen in Table 1.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1968

Table 1. Previous literature of Software Effort Estimation

Author	Title	Algorithm	Dataset	Result
[8]	*Towards an early software estimation using log-linear regression and a multilayer perceptron model*	-log-LR -MLP -UCP -Schneider	Western CompuTop ISBSG	MMER, RMSE, MAE, SD
[13]	*Software Development Effort Estimation Using Regression Fuzzy Models*	-*Regression fuzzy logic* -ANN	ISBSG	MAE, MBRE, MIBRE, SA, Scott-Knott
[7]	*Predicting Software Effort Estimation Using Machine Learning Techniques*	-NB -LR -DTRF	COCOMO	MAE AUC
[14]	Kombinasi *Median Weighted Information Gain* Dengan *K-Nearest Neighbor* Pada *Dataset Label Months Software Effort Estimation*	-KNN+ Median WIG	China Desharnais Kitchenham	RMSE
[18]	*Software Effort Estimation using Radial Basis and Generalized Regression Neural Networks*	-Radial Basis NN - GRNN	COCOMO81	MMRE, MARE, VARE, Mean BRE
[15]	*Empirical Validation of Neural Network Models for Agile Software Effort Estimation based on Story Points*	- GRNN - PNN - GMDH - CNN	agile projects	MSE. MMRE
[16]	*Proposing an Enhanced Artificial Neural Network Prediction Model to Improve the Accuracy in Software Effort Estimation*	-ANN -COCOMO II	Cocomo81	MMRE
[17]	*Software Effort Estimation in the Early Stages of the Software Life Cycle Using a Cascade Correlation Neural Network Model*	-CNN -*Multiple Linear Regression*	*industrial and educational projects*	MMRE
[19]	*Improving the Accuracy in Software Effort Estimation*	-ANN+PSO	COCOMO I Nasa93	MMRE
[5]	*Data Mining Techniques for Software Effort Estimation: A Comparative Study*	-OLS -RBF -MLP	ISBSG Nasa93 Cocomo81 Desharnais Maxwell	MMRE

MATERIALS AND METHODS

This research had been done using several machine learning algorithms, namely LR, MLP, RBF, and DTRF. To improve accuracy, we use feature selection. The feature selection we use is PSO. The research process is shown in Figure 1.

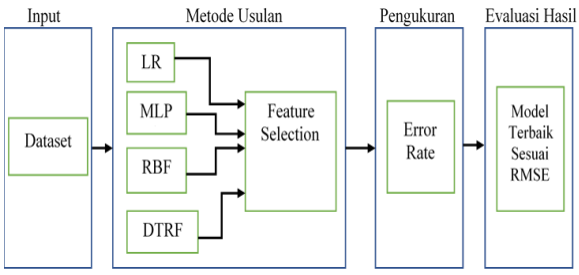


Figure 1: Step of the research process.

A. Dataset Description

In this research, the dataset to be used is the Albercth dataset with 24 projects and 8 attributes, Kemerer with 15 projects and 8 attributes, China with 499 and 19 attributes, Cocomonasa_2 with 101 and 24 attributes, Cocomonasa_v1 with 60 and 17 attributes of data, Desharnais with 81 and 12 attributes, Kitchenham with 145 and 10 attributes, Maxwell with 62 and 27 attributes, Miyazaki94 with 48 and 9 attributes, and cocomo81 with 63 projects and 17 attributes.

Data collection is using the dataset from many resources. The datasets used are Cocomo81 (1981), Desharnais (1989), Miyazaki (1994), Maxwell (2002), Kitchenham CSC (2002), Cocomo NASA v1 (2005), Cocomo NASA 2 (2006), China (2007), Albrecht (2009), and Kemerer.

The entire dataset is then calculated by comparing the performance results between the Linear Regression (LR) algorithm, Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Decision Tree Random Forest (DTRF) using Waikato Environment for Knowledge Analysis (Weka) under the GNU General Public License.

Then the experiment was carried out using the addition of the PSO feature selection algorithm in all algorithms used. The resulting performance is based on the Root Mean Squared Error (RMSE). The algorithm with the lowest RMSE value is the best method for software effort estimation. The collected dataset has many variations of attributes and instances, as shown in Table 2.

Table 2. Dataset Description

Dataset	*Attribute*	*Instance*
Albrecht	8	24
Desharnais	8	15
Kemerer	19	499
Cocomonasa1	17	60
Cocomonasa2	24	101
China	12	81
Cocomo81	10	145
Miyazaki94	27	62
Kitchenham	9	48
Maxwell	17	63

B. Linear Regression (LR) Model

This section presents the proposed linear regression model is presented, linear regression is the method most often used in effort estimation software and always gets high accuracy values [8]. According to Harlan [20], the dependent variable in linear regression is also called response or criterion, while the independent variable is also known as a predictor or regressor. The model used for simple linear regression can be described as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \ ; \ i = 1, 2, \ldots, n \quad \ldots\ldots\ldots\ldots (1)$$

Where :

Y_i : Response for subject i

X_i : Predictor for subject i

ε_i : error for subject i

β_0 and β_1 are the parameters in the population to be estimated in the fitting model. fitting the model with sample data will produce the equation:

$$Y_i = b_0 + b_1 X_i \ ; \ i = 1, 2, \ldots, n \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

C. Multi-Layer Perceptron (MLP) Model

Multilayer Perceptron (MLP) is a class of Artificial Neural Networks (ANN). This section presents the Multi-Layer Perceptron neural network model. The neural network structure is very suitable for calculating software effort estimates. The network will stop training when the number of epochs reaches 250 or when the Mean Squared Error (MSE) becomes zero or when the MU value exceeds 1e+10. The time is set to "infinity" which indicates that the training time has no control over when the exercise should be stopped. Two common activation functions have historically been both sigmoids, and are described by the equation:

$$y\,(v_i) = \tanh(v_i)\ and$$
$$y\,(v_i) = (1 + e^{-v_i})^{-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Backpropagation works through an iterative process using training data, comparing the predicted value of each network with each data contained in the training. In each process, the weight of the relation in the network is modified to minimize the Mean Squared Error (MSE) value between the predicted value of the network and the real value.

D. Radial Basis Function (RBF) Model

The radial basis Layer contains different types of neurons, which contains the Radial Basis Function (RBF) as an activation function. A single (same) radial basis layer may contain neurons with different radial basis functions. Radial Basis Function (RBF) artificial neural network is an artificial neural network model with one unit in the hidden layer, where the activation function is a basic function and a linear function in the output layer. This method is suitable for predicting software effort estimates. RBF is usually used to build a functional approach from an equation:

$$y\,(x) = \sum_{i=1}^{N} w_i\, \varphi(\|x - x_i\|) \quad \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Where the approximation function $y(\mathbf{x})$ is represented as the sum of **N** radial basis functions,

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1968

each corresponding to a different center x_i, and weighted by the corresponding coefficient w_i.

D. Decision Tree Random Forest (DTRF) Model

The DTRF model consists of a collection of decision trees that grow in parallel. According to Nassif, the tree predictions are combined to make the overall tree prediction for the forest [4].

DTRF can be defined as an ensemble learning method for classification, regression, and other tasks that operate by building multiple decision trees at the time of training and issuing classes which are class (classification) mode or average/average (regression) prediction of each tree.

If the classifying ensemble is h_1 (x), h_2 (x),. . . , h_K (x), and with the training set randomly drawn from the random vector distribution Y, X, then to determine the margin function as follows:

$$mg(X,Y) = av_k I(h_k(X) = Y) - max_{j \neq Y}\, av_k I\left(h_{k(X)=j}\right) \quad (5)$$

where *I(·)* is the indicator function. The margin measures the extent to which the average number of votes at *X, Y* for the right class exceeds the average vote for any other class.

E. Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a research-based on population, and this population includes lots of particles where each particle represents a solution of an optimization problem. During every iteration, each particle is updated by following two "best" values, pbest and gbest. After finding the two best values, the position and velocity of the particles are updated by the following two equations:

$$v_i^k = vw_i^k + c_1 r_1\left(pbest_i^k - x_i^k\right) + c_2 r_2\left(gbest_i^k - x_i^k\right)$$
$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad \text{.....(6)}$$

where v_i^k is the velocity of the *i*th particle at the *k*th iteration, and x_i^k is the current solution (or position) of the ith particle at the kth iteration. c_1, c_2 are positive constants, and r_1, r_2 are two random variables with a uniform distribution between 0 and 1.

In this equation, w is the inertia weight which shows the effect of the previous velocity vector on the new vector. An upper bound is placed on the velocity in all dimensions v_{max}. This limitation prevents the particle from moving too rapidly from one region in the search space to another. This value is usually initialized as a function of the range of the problem.

RESULTS AND DISCUSSION

A. Evaluation Method

The dataset is calculated by comparing the performance results between LR, MLP, RBF, and DTRF. To evaluate the performance of the model or algorithm the data is separated into two subsets, namely learning process data and validation/evaluation data. The model or algorithm is trained by the learning subset and validated by the validation subset then 10-fold cross-validation is applied. Antoni Wibowo recommends 10-Fold Cross-validation is recommended as the best model selection that can provide a clearer estimate of accuracy than other cross-validations [21].

The researcher uses cross-validation to select the appropriate model by comparing the value of the Root Mean Squared Error Cross-Validation (RMSECV) with the following formula:

$$RMSECV = 1/10\ 1/N\_cv\ \sum\_(k = 1)\text{^}10 \text{▒}\ \llbracket\ \sum\_(i = 1)\text{^}(N\_cv)(t\_l\text{^}k - y\_l\text{^}k\ )\text{^}2\ \rrbracket \quad \text{.....(7)}$$

In this section, the researcher will explain the use of machine learning methods to calculate estimated software effort estimation.The result from RMSE values of comparative analysis between machine learning algorithms namely LR, MLP, RBF, and DTRF are compared with the machine learning algorithm with the feature selection PSO shown in Table 3 and Table 4.

Table 3. The Results Obtained From LR, MLP, RBF, and DTRF

Dataset	Method			
	LR	MLP	RBF	DTRF
Albercth	13.007	22.24	12.42	12.92
Desharnais	2988.93	5992.12	4052.81	3348.23
Kemerer	281.06	303.82	256.10	238.54
Cocomonasa1	431.76	310.36	516.99	403.44
Cocomonasa2	1142.46	1025.20	1004.62	825.10
China	968.62	1444.62	4816.94	2088.45
Cocomo81	1480.80	1651.88	1616.80	1288.80
Miyazaki94	155.92	192.67	238.83	198.90
Kitchenham	2302.34	8407.07	9297.63	8528.53
Maxwell	6306.54	6849.10	6507.27	7197.51

The table above shows the RMSE results through the calculation of all the selected methods. Of the 10 data used to have different RMSE values. This shows that the method used gives different results according to the function of the algorithm for what calculation. However, the LR method gives the best value from the average results obtained.

To try and find the maximum result, this research adds the PSO selection feature to the calculation. The results can be seen in Table 4.

Table 4. The Results Obtained From LR, MLP, RBF, and DTRF with Feature Selection PSO

Dataset	Method+ PSO			
	LR	MLP	RBF	DTRF
Albercth	10.691	11.045	8.946	10.413
Desharnais	2920.137	5254.074	3316.407	3195.768
Kemerer	261.052	389.714	257.341	255.030
Cocomonasa1	659.657	358.027	493.059	419.651
Cocomonasa2	976.800	1826.769	969.472	809.119
China	1077.627	1468.317	1265.420	1907.217
Cocomo81	1442.859	2160.937	1634.587	1273.597
Miyazaki94	157.504	249.214	214.061	198.814
Kitchenham	2329.636	7871.083	9693.927	8452.060
Maxwell	5762.562	7510.228	6246.617	6593.583

The table above shows some data that displays the results with a significant change in value after adding the feature selection. However, there are still data that give the same results. The addition of the PSO feature selection can increase the value for the difference from the RMSE value obtained.

C. Result Evaluation

In this section, the significance level is used to determine how influential the use of the PSO feature selection is in increasing accuracy or reducing the RMSE value.

The level of significance is the threshold used to determine significance. If the p-value is less than or equal to the level of significance, the data are considered statistically significant.

- As a general rule, the level of significance (alpha) is set at 0.05, meaning that the probability of the two data groups being equal is only 5%.
- Using a higher confidence level (lower p-value) means that the experimental results will be considered more significant.
- If you want to increase the confidence level of your data, decrease the p-value even more to 0.01. Lower p values are commonly used in manufacturing when detecting product defects. A high level of confidence is essential to ensure that every part produced works according to its function.
- For hypothesis testing experiments, a significance level of 0.05 is acceptable.

A test that uses F-distribution, named by Sir Ronald Fisher, is called an F-Test. F-distribution or the Fischer-Snedecor distribution is a continuous statistical distribution used to test whether two observed samples have the same variance.

F-Test compares two variances, s_X and s_Y, by dividing them. Since variances are positive, the result is always a positive number. The critical value for F-Test is determined by the equation.

$$"F" = (S\_X\text{^}2)/(S\_Y\text{^}2\ ) \quad \ldots\ldots (8)$$

To find the sample variations s_X and s_Y, by using the formulas:

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - \bar{X})^2 \ and\ S_y^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(y_i - \bar{Y})^2 \quad \ldots\ldots (9)$$

If the variances are the same, then the variance ratio is 1. The larger sample variance must be in the F-ratio numerator and the smaller sample variance in the denominator. Thus, this ratio is always greater than 1 and makes hypothesis testing easier.

After being calculated, then compared, the most optimal machine learning method between LR, MLP, RBF, and DTRF with the PSO Selection Feature can be seen in table 5 below.

Table 5. Comparison Of The Results Obtained From LR, MLP, RBF, and DTRF with Feature Selection PSO

Method	Mean of RMSE		Improvement	
	Without PSO	With PSO	Percentage	Status
LR	1607.149	1559.852	2.94%	enhanced
MLP	2619.912	2709.941	-3.44%	Not enhanced
RBF	2832.040	2409.984	14.90%	enhanced
DTRF	2413.043	2311.525	4.21%	enhanced

The sample standard deviation (S) of population X (without PSO) is considered to be less than or equal to the sample standard deviation (S) of population Y (with PSO). The p-value equals 0.4425, ($p(x \le F) = 0.5575$). Standard deviation X is 535.32, and Y is 488.94. The test statistic F equals 1.1987, which is in the 95% region of acceptance: [-∞ : 9.2766]. $F=S_X/S_Y=1.09$. The result of the F-test is shown in Figure 2.

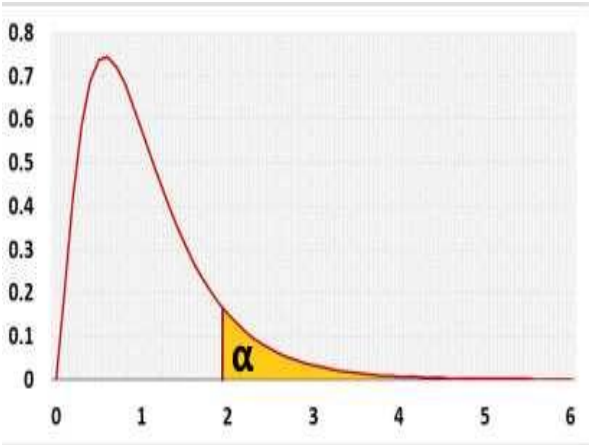


Figure 2: F Distribution

Based on the table above shows that using the PSO selection feature can significantly reduce

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1968

the RMSE value with a significant level (alpha) = 0.05. The results show that the value of F =1.094 exceeds the significant level (alpha) value.

CONCLUSION

Time, cost, and labor are important factors in the software development process, an effective software development process that can be achieved by evaluating these parameters at an early stage of the project. The estimated evaluation of software efforts will lead to an increase in the efficiency of software development and increase its success rate. Based on this research, the data mining algorithm used to calculate the most optimal software effort estimate is the Linear Regression algorithm with an average RMSE value of 1607.149 for the 10 datasets tested. Then using the PSO feature selection can increase the accuracy or reduce the RMSE average value to 1559.852. The result indicates that, compared with the original regression linear model, the accuracy or error rate of software effort estimation has increased by 2.94% by applying PSO feature selection. Some other computation technologies such as genetic algorithms with another method to increase the accuracy can be explored and applied on software effort estimation models in the future.

REFERENCE

[1] B. Jeng, D. Yeh, D. Wang, S. L. Lhu, and C. M. Chen, "A Specific Effort Estimation Method Using Function Point," *J. Inf. Sci. Eng.*, vol. 27, no. 4, pp. 1363–1376, 2011.

[2] D. Pratiwi, "Implementation of Function Point Analysis in Measuring the Volume Estimation of Software System in Object Oriented and Structural Model of Academic System," *Int. J. Comput. Appl.*, 2013.

[3] G. R. Finnie, G. E. Wittig, and J.-M. Desharnais, "A comparison of software effort estimation techniques: Using function points with neural networks, case-based reasoning and regression models," *J. Syst. Softw.*, vol. 39, no. 3, pp. 281–289, Dec. 1997.

[4] A. B. Nassif, M. Azzeh, L. F. Capretz, and D. Ho, "A comparison between decision trees and decision tree forest models for software development effort estimation," in *2013 Third International Conference on Communications and Information Technology (ICCIT)*, 2013, pp. 220–224.

[5] K. Dejaeger, W. Verbeke, D. Martens, and B. Baesens, "Data mining techniques for software effort estimation: A comparative study," *IEEE Trans. Softw. Eng.*, vol. 38, no. 2, pp. 375–397, 2012.

[6] M. O. Elish, "Improved estimation of software project effort using multiple additive regression trees," *Expert Syst. Appl.*, 2009.

[7] A. BaniMustafa, "Predicting Software Effort Estimation Using Machine Learning Techniques," in *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, 2018, pp. 249–256.

[8] A. B. Nassif, D. Ho, and L. F. Capretz, "Towards an early software estimation using log-linear regression and a multilayer perceptron model," *J. Syst. Softw.*, vol. 86, no. 1, pp. 144–160, Jan. 2013.

[9] I. Wieczorek, "Improved Software Cost Estimation – A Robust and Interpretable Modelling Method and a Comprehensive Empirical Investigation," *Empir. Softw. Eng.*, vol. 7, no. 2, pp. 177–180, 2002.

[10] B. W. Boehm, *Software Engineering Economics*. Prentice Hall, 1981.

[11] Putnam and L. H, "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," *IEEE Trans. Softw. Eng.*, vol. 4, no. 4, pp. 345–361, 1987.

[12] D. D. Galorath and M. W. Evans, *Software Sizing, Estimation, and Risk Management*. Auerbach Publication, 2006.

[13] A. B. Nassif, M. Azzeh, A. Idri, and A. Abran, "Software Development Effort Estimation Using Regression Fuzzy Models," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–17, Feb. 2019.

[14] I. Kurniawan, "Kombinasi Median Weighted Information Gain Dengan K-Nearest Neighbor Pada Dataset Label Months Software Effort Estimation," *J. Teknoinfo*, vol. 14, no. 2, p. 138, Jul. 2020.

[15] A. Panda, S. M. Satapathy, and S. K. Rath, "Empirical Validation of Neural Network Models for Agile Software Effort Estimation based on Story Points," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 772–781, 2015.

[16] I. Attarzadeh, A. Mehranzadeh, and A. Barati, "Proposing an enhanced artificial neural network prediction model to improve the accuracy in software effort estimation," *Proc. - 2012 4th Int. Conf. Comput. Intell. Commun. Syst. Networks, CICSyN 2012*, pp. 167–172, 2012.

[17] A. B. Nassif, L. F. Capretz, and D. Ho, "Software effort estimation in the early stages of the software life cycle using a cascade correlation neural network model," *Proc. - 13th ACIS Int. Conf. Softw. Eng. Artif.*

Intell. Networking, Parallel/Distributed Comput. SNPD 2012, pp. 589–594, 2012.

[18] P. V. G. D. P. Reddy, K. R. Sudha, P. R. Sree, and S. N. S. V. S. C. Ramesh, "Software Effort Estimation using Radial Basis and Generalized Regression Neural Networks," vol. 2, no. 5, pp. 87–92, 2010.

[19] Z. Dan, "Improving the Accuracy in Software Effort estimation using ann model based on PSO," in *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics*, 2013, pp. 180–185.

[20] J. Harlan, *Analisis Regresi Linear*. Yogyakarta: Penerbit Gunadarma, 2018.

[21] A. Wibowo, "10-Fold Cross Validation," *MTI Binus*, 2017. [Online]. Available: https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/. [Accessed: 02-Jan-2021].

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1935

IMPLEMENTATION OF ENSEMBLE TECHNIQUES FOR DIARRHEA CASES CLASSIFICATION OF UNDER-FIVE CHILDREN IN INDONESIA

Andriansyah Muqiit Wardoyo Saputra [1]; Arie Wahyu Wijayanto [2*]

Program Studi D4 Statistika[1]; Program Studi D4 Komputasi Statistik[2*]
Politeknik Statistika STIS
www.stis.ac.id
211709554@stis.ac.id [1]; ariewahyu@stis.ac.id [2*]
(*) Corresponding Author

Abstract—Diarrhea is an endemic disease in Indonesia with symptoms of three or more defecations with the consistency of liquid stool. According to WHO, diarrhea is the second largest contributor to the death of under-five children. Data and cases of children under five years who have diarrhea are very difficult to find, so the data analysis process becomes difficult due to the lack of information obtained. Difficulties in the data analysis process can be overcome by rebalancing, so the category ratios are balanced. The method that is popularly used is SMOTE. To solve imbalanced data and improve classification performance, this study implements the combination of SMOTE with several ensemble techniques in diarrhea cases of under-five children in Indonesia. Ensemble models that are used in this study are Random Forest, Adaptive Boosting, and XGBoost with Decision Tree as a baseline method. The results show that all SMOTE-based methods demonstrate a competitive performance whereas SMOTE-XGB gains a slightly higher accuracy (0.88), precision (0.96), and f1-score (0.86). The implementation of the SMOTE strategy improved the recall, precision, and f1-score metrics and give higher AUC of all methods (DT, RF, ADA, and XGB). This study is useful to solve the imbalanced problems in official statistics data provided by BPS Statistics Indonesia.

Keywords: Diarrhea, Ensemble techniques, Imbalanced class, XGBoost.

Abstrak—*Penyakit diare merupakan penyakit endemis Indonesia dengan gejala buang air besar sebanyak tiga kali atau lebih dengan konsistensi tinja cair. Menurut WHO, diare menjadi penyumbang terbesar kedua kematian anak di bawah lima tahun. Data dan kasus balita yang mengalami diare sangat sulit ditemukan sehingga proses analisis data menjadi sulit karena kurangnya informasi yang didapatkan. Kesulitan dalam proses analisis data dapat diatasi dengan melakukan rebalancing agar rasio kategori menjadi berimbang. Metode yang populer digunakan adalah SMOTE. Untuk menyelesaikan masalah data tidak berimbang dan meningkatkan kinerja klasifikasi, penelitian ini mengimplementasikan kombinasi SMOTE dengan beberapa teknik ansambel pada kasus diare balita di Indonesia. Model ansambel yang digunakan adalah Random Forest, Adaptive Boosting, dan XGBoost dengan Decision Tree sebagai metode baseline. Hasil penelitian menunjukkan bahwa semua metode berbasis SMOTE menunjukkan kinerja kompetitif dengan SMOTE-XGB mendapatkan akurasi yang sedikit lebih tinggi (0,88), presisi (0,96), dan f1-score (0,86). Implementasi metode SMOTE jelas meningkatkan ukuran metrik recall, presisi, dan f1-score dan memberikan AUC yang lebih tinggi dari semua metode (DT, RF, ADA, dan XGB). Penelitian ini berguna sebagai pembelajaran dalam mengatasi data tidak berimbang pada data statistik resmi yang disediakan oleh BPS Statistics Indonesia.*

Kata Kunci: *Diare, Ensemble techniques, Imbalanced class, XGBoost*

INTRODUCTION

Diarrheal disease is the second-largest cause of death in children under the age of five and contributes to the death of approximately 525,000 children each year. Diarrhea is defined as a condition in which an individual experiences defecation with a frequency of three times or more with the consistency of liquid stool, which can be accompanied by blood and or mucus [1].

Diarrhea is one of the endemic diseases in Indonesia. The disease is also classified into potential diseases of Extraordinary Events (KLB) which are often followed by death [2]. So, cases of diarrhea need to be of particular concern to be able to be prevented.

Data about toddlers with diarrhea is very difficult to obtain and cases are rare. The process of data analysis becomes difficult due to the lack of information obtained. When the statistical analysis

is done, information about a little data will cause the data to have an imbalanced class.

The application of a simple classification algorithm on data that has an imbalanced class will cause bias in larger classes/categories and treat fewer classes as noise [3]. Therefore, it is necessary to do special handling of the data. The most common way to handle imbalanced data is balancing the data using resampling techniques. This technique is also known as the data level approach to class imbalance learning. A popular method used in this resampling technique is the Synthetic Minority Oversampling Technique or SMOTE [4]. However, this method can raise some problems such as bias that negatively affect the model performance [5], overfitting, loss of important information from data, and so on [6]. In research conducted by [7], it is proven that applying oversampling to training set samples before learning can improve the recall rate of minority categories significantly. Thus, this approach can solve problems in the minority categories as well.

Another approach used to overcome the imbalanced class is to apply the ensemble method. In this method, an algorithmic approach is carried out by increasing the weight on samples undergoing misclassification to improve classification performance [8]. The ensemble methods proposed in this study are Random Forest [9, 10], AdaBoost [11], and XGBoost [12].

According to the previous studies [12, 13], applying a hybrid method (data level approach and ensemble) for imbalance learning was superior to other models. In [6], concluded that a hybrid method can increase AUC gradually with an increase in the percentage of oversampling. Related research was also conducted [15] by comparing SMOTE Random Forest and SMOTE XGBoost models on the HCV dataset. The results obtained that models with SMOTE can significantly increase the recall value from under 2% to more than 70%.

In this study, researchers aimed to overcome misclassification in cases of diarrhea of toddlers in Indonesia by combining SMOTE and ensemble-based classification methods. SMOTE is applied so that the data ratio becomes balanced and then the classification results are improved by several ensemble methods.

MATERIALS AND METHODS

This study is a further research from a scientific article entitled "Penerapan Metode Resampling dalam Mengatasi *Imbalanced* Data pada Determinan Kasus Diare pada Balita di Indonesia (Analisis Data SDKI 2017)" [16].

Data and Variables

The source of data used in this study is derived from Indonesia Demographic and Health Surveys (IDHS) 2017 by BPS Statistics Indonesia in collaboration with the National Population and Family Planning Board (BKKBN), Ministry of Health, and ICF [17].

The units of analysis in this study are households that had under-five children, formerly 11,340 households. Within 11,340 households there were 1,560 households had under-five children with diarrhea cases in two weeks recently during the survey. The research used eight predictor variables that can be seen in Table 1 below.

Table 1. Research variables [16]

Name	Symbol	Type	Category
Diarrhea case (Target)	Y	Categoric	0, No 1, Yes, 2 weeks recently
Predictor:			
Children's age	X1	Numeric	-
Mother's age	X2	Numeric	-
Children's sex	X3	Categoric	0, Female 1, Male
Residence's type	X4	Categoric	0, Rural 1, Urban
Educational attainment	X5	Categoric	0, No education 1, Complete elementary 2, Complete junior high school 3, Complete senior high school
Main floor material	X6	Categoric	0, Natural floor 1, Material floor 2, Finished floor
Type of toilet facility	X7	Categoric	0, No toilet 1, Good toilet 2, Shared toilet 3, Bad toilet
Source of drinking water	X8	Categoric	0, Proper 1, Not proper

Software Specification

The analysis and modeling process in this study was done using the Python programming language with Google Colaboratory backend. The specification of the Google Colaboratory backend shown in Table 2 below.

Table 2. Software specification

Disk	RAM	CPU
107,77 GB (31.94 used)	13 GB (0.75 used)	Intel Xeon Processor 2 cores @ 2.30 GHz

The machine learning tool that is used in this study for classification algorithms is scikit-learn API in python [18].

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1935

Research Framework

This section will explain the framework of the research to get the results. First, the imbalanced dataset is carried out to the data preprocessing step. In this step, we prepare the data before performing the analysis process. Then, create a balanced dataset that has gone through by SMOTE. This dataset is used to compare to the original data. The two datasets perform in the training process using several ensemble methods. Last, this study applies the evaluation and validation process to obtain how SMOTE and ensemble methods affect the classification result. The process is shown in Figure 1 as follows.

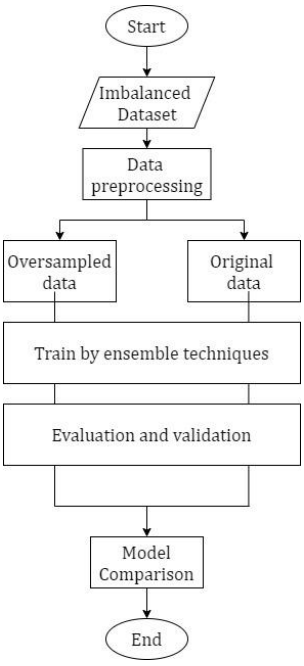


Figure 1. Research framework

Methods

This section explains the brief method that is used in this study. Based on the research framework, the explanation of the analysis steps are as follows:

1. Data preprocessing

The initial stage before an in-depth analysis is the initial processing step or preprocessing. This step aims to prepare the data ready to be processed and provide results with maximum performance. The process carried out is as follows:

a. Standardization

Numeric variables are standardized using the Z-score method with the formula:

$$Z = \frac{X-\mu}{\sigma} \quad \text{.....(1)}$$

b. Creating dummy variables

Categorical variables that have more than two categories are converted into dummy variables with the first category as the reference.

2. Rebalancing with SMOTE

The SMOTE method is applied in this step to overcome imbalanced class in diarrhea cases data in toddlers. The resampling process is done only to the minority class of category 1 (Yes, two weeks recently). The result of this process is new data with a balanced ratio of 50:50.

3. Data modeling

Binary classification is applied in this research as a learning method. Because the learning process focuses to solve the problem in minority classes (cases of toddler's diarrhea), then the algorithm used is ensemble-based methods. The methods used are Random Forest [9, 10], AdaBoost [11], and XGBoost [12]. All parameter settings in Random Forest and AdaBoost are set to default. While in XGBoost are set n_estimator to 100 and max_depth to 8. Also, the Decision Tree is applied in training as a baseline algorithm with default parameter setting [19].

Model training with Decision Tree, Random Forest, AdaBoost, and XGBoost algorithms are also applied to the dataset without SMOTE. This is an assessment to prove that the hybrid method (data level approach and ensemble) can overcome imbalanced class well.

4. Data evaluation and validation

The final step in this research is the evaluation and validation of the proposed model at the modeling step. The data is divided into two sets, namely the training set and the test set. Learning will be done using a training set while evaluation and validation using the test set.

The evaluation of the model in this study used a confusion matrix [20] with details in Table 3 below.

Table 3. Confusion matrix

	Predicted positive	Predicted negative
Actual positive	TP	FN
Actual negative	FP	TN

Based on the confusion matrix can be derived into several metrics as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{.....(2)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{.....(3)}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{.....(4)}$$

$$F1-score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \quad \text{.....(5)}$$

In addition to the confusion matrix and some of its derivative metrics, the study also used ROC to look at the comparison of models based on

the Area Under Curve (AUC). This curve is proven to be more sensitive to compare different learning models [20, 21]. Then, validation is also carried out by looking at the learning curve of cross-validation results as mitigation in the case of model overfitting.

RESULTS AND DISCUSSION

Oversampling Results

Table 4 is shown regarding the number of samples used in the study. Diarrhea cases data in toddlers have an imbalanced ratio of 6.27:1. This imbalance is necessary to be rebalanced by SMOTE.

Table 4. Information of data sample

# Total	# Majority	# Minority	IR
11,340	9,780	1,560	6.27:1

The comparison of the number of cases before and after resampling is shown through the bar chart in Figure 2. It appears that the result of oversampling has obtained new data with a balanced ratio (50:50).

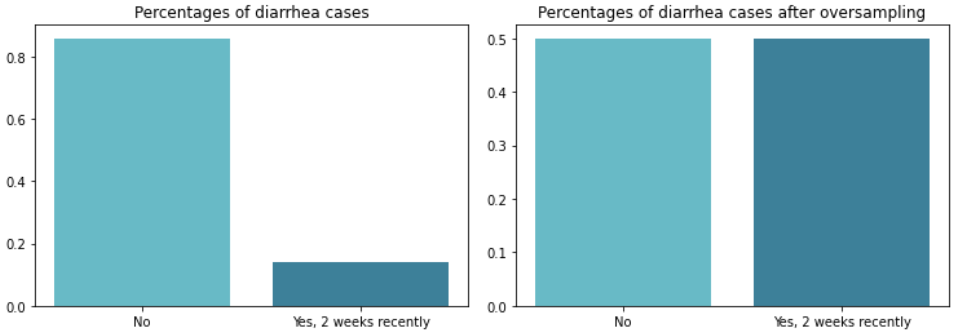


Figure 2. Percentages of diarrhea cases in each category before and after oversampling

Table 5. Model's performance comparison

Model	Accuracy	Recall	Precision	F1-score
DT	0.80	0.13	0.20	0.15
RF	0.83	0.07	0.20	0.33
ADA	0.81	0.11	0.20	0.14
XGB	0.86	0.01	0.19	0.02
SMOTE-DT	0.85	**0.83**	0.86	0.84
SMOTE-RF	0.82	0.78	0.86	0.82
SMOTE-ADA	0.85	0.82	0.87	0.85
SMOTE-XGB	**0.88**	0.79	**0.96**	**0.86**

Model Comparison

After modeling with the proposed classification method, the evaluation of each model is carried out. The evaluation is done by making predictions on the test data based on several measures that have been described in the research method.

Table 5 showing the comparison of models' performance using eight classifiers with four metrics evaluation derived from the confusion matrix. All SMOTE-based methods demonstrate a competitive performance whereas SMOTE-XGB gains a slightly higher accuracy (0.88), precision (0.96), and f1-score (0.86). Overall, the implementation of the SMOTE strategy improved the recall, precision, and F1-score metrics of all methods (DT, RF, ADA, and XGB).

Based on the ROC curve shown in Figure 3, all SMOTE-based methods give higher AUC than other methods (DT, RF, ADA, and XGB). This result is related to the previous study [6] that explained applying SMOTE to imbalanced data can improve the AUC value.

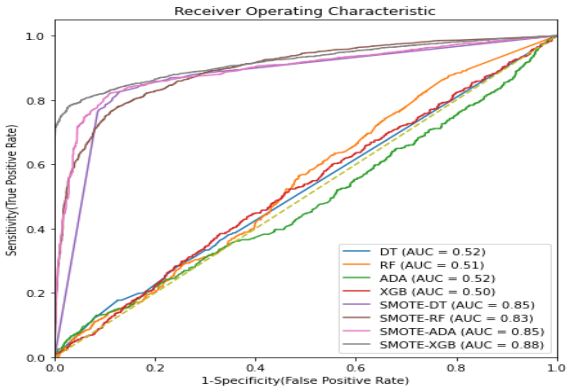


Figure 3. ROC Curve

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1935

Model Validation

To avoid overfitting issues, the researcher applies a learning curve into the SMOTE-based model by compare training scores and cross-validation scores. The cross-validation option is set to 100 splits and 30% samples for the test set. Based on Figure 4, the learning curve will converge into a specific score as the training size increases. SMOTE-XGB converges faster than another model. It happens because XGBoost has scalability that runs faster than other models [12].

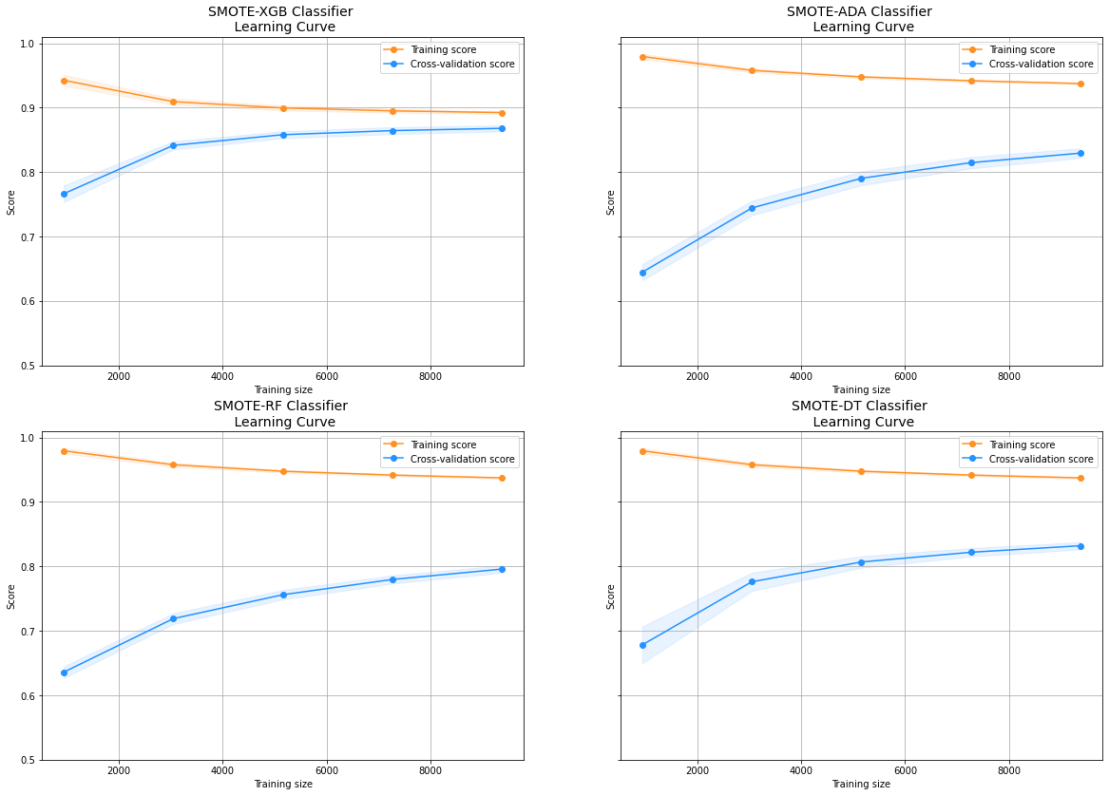


Figure 4. Learning curves

CONCLUSION

From the results and discussions, it can be concluded that the application of the hybrid method can handle an imbalanced class. This explains by the implementation of the SMOTE strategy improved the recall, precision, and F1-score metrics of all methods (DT, RF, ADA, and XGB) with the SMOTE-XGB gains slightly higher accuracy. SMOTE-based methods also get higher AUC values than others. Belonging to the learning curve, SMOTE-XGB runs competitively to other methods that converge fast. This study is useful for doing assessments and solve the imbalanced data in official statistics data provided by BPS Statistics Indonesia. For further studies, this study suggests applying a more complex algorithm to solve imbalanced data. Furthermore, statistical testing (parametric or nonparametric) with a specific level of significance is needed to choose whether the best model.

REFERENCE

[1] World Health Organization, "Diarrhoeal disease," 2017. https://www.who.int/en/news-room/fact-sheets/detail/diarrhoeal-disease (accessed Dec. 16, 2020).

[2] Kementerian Kesehatan RI, "Profil Kesehatan Indonesia Tahun 2017," 2017. [Online]. Available: https://www.kemkes.go.id/.

[3] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, 2012, DOI: 10.1109/TSMCC.2011.2161285.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif.*

Intell. Res., vol. 16, pp. 321–357, 2002.

[5] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, 2019, DOI: 10.1016/j.ins.2019.07.070.

[6] U. R. Salunkhe and S. N. Mali, "Classifier Ensemble Design for Imbalanced Data Classification: A Hybrid Approach," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 725–732, 2016, DOI: 10.1016/j.procs.2016.05.259.

[7] H. Dong, D. He, and F. Wang, "SMOTE-XGBoost using Tree Parzen Estimator optimization for copper flotation method classification," *Powder Technol.*, vol. 375, pp. 174–181, 2020, DOI: 10.1016/j.powtec.2020.07.065.

[8] K. Li, G. Zhou, J. Zhai, F. Li, and M. Shao, "Improved PSO_AdaBoost ensemble algorithm for imbalanced data," *Sensors (Switzerland)*, vol. 19, no. 6, 2019, DOI: 10.3390/s19061476.

[9] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, DOI: 10.1023/A:1010933404324.

[10] R. Polikar, "Ensemble Learning," in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 1–34.

[11] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997, DOI: https://doi.org/10.1006/jcss.1997.1504.

[12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, DOI: 10.1145/2939672.2939785.

[13] Z. Chen, J. Duan, L. Kang, and G. Qiu, "A Hybrid Data-Level Ensemble to Enable Learning from Highly Imbalanced Dataset," *Inf. Sci. (Ny).*, 2020, DOI: 10.1016/j.ins.2020.12.023.

[14] S. Wang *et al.*, "A new method of diesel fuel brands identification: SMOTE oversampling combined with XGBoost ensemble learning," *Fuel*, vol. 282, no. May, p. 118848, 2020, DOI: 10.1016/j.fuel.2020.118848.

[15] M. Syukron, R. Santoso, and T. Widiharih, "Perbandingan Metode SMOTE Random Forest dan SMOTE XGBoost untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data," *J. Gaussian*, vol. 9, no. 3, pp. 227–236, 2020.

[16] A. M. W. Saputra, I. P. Ananda, M. A. Rizki, Z. D. Hapsari, and R. Nooraeni, "Penerapan Metode Resampling dalam Mengatasi Imbalanced Data pada Determinan Kasus Diare pada Balita di Indonesia (Analisis Data SDKI 2017)," *J. Mat. dan Stat. serta Apl.*, vol. 8, pp. 19–27, 2020, DOI: 10.24252/msa.v8i1.13452.

[17] National Population and Family Planning Board - BKKBN, Statistical Indonesia - BPS, Ministry of Health - Kemenkes, and ICF, "Indonesia Demographic and Health Survey 2017 [Dataset]." 2018. Distributed by ICF. Available: http://dhsprogram.com/pubs/pdf/FR342/FR342.pdf.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[19] C. Gao and H. Elzarka, "The use of decision tree based predictive models for improving the culvert inspection process," *Adv. Eng. Informatics*, vol. 47, no. October 2020, p. 101203, 2020, DOI: 10.1016/j.aei.2020.101203.

[20] J. Han, M. Kamber, and J. Pei, "8 - Classification: Basic Concepts," in *Data Mining (Third Edition)*, Third Edit., J. Han, M. Kamber, and J. Pei, Eds. Boston: Morgan Kaufmann, 2012, pp. 327–391.

[21] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases.," *Radiology*, vol. 148, no. 3, pp. 839–843, Sep. 1983, DOI: 10.1148/radiology.148.3.6878708.

[22] K. Hajian-Tilaki, "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation.," *Casp. J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

K-MEANS CLUSTERING AREAS PRONE TO TRAFFIC ACCIDENTS IN ASAHAN REGENCY

Nurul Rahmadani[1*]; Elly Rahayu[2]; Ayu Lestari[3]

Manajemen Informatika [1]; Sistem Informasi [2,3]
Sekolah Tinggi Manajemen Informatika dan Komputer Royal
www.stmikroyal.ac.id
cloudyrara@gmail.com[1]; ellyrahayu68@gmail.com[2]; ayulesstarii27@gmail.com[3]
(*) Corresponding Author

Abstract— Traffic accidents on the highway still contribute to the high mortality rate in Indonesia, so it is of particular concern to the police in this country. Accidents occur in various places with different time events, this makes it difficult to determine which areas have a high level of traffic accident vulnerability. Information about traffic accident-prone areas is needed by the community and law enforcement. This information can be taken into consideration for supervision and anticipatory action, especially for the police. The initial stage of traffic accident prevention is to know the factors that cause traffic accidents obtained through traffic accident data analysis. The information system in this study analyzed traffic accident-prone areas in Asahan Regency. The analysis can be done with data mining, namely K-Means Clustering which can group data into several groups according to the characteristics of the data. The results of this study are the Asahan District Police Satlantas can find out the accident-prone areas in the most vulnerable categories, quite vulnerable and not vulnerable.

Keywords: Accident Prone Areas, K-Means Clustering, Traffic Accident.

Abstrak— *Kecelakaan lalu lintas di jalan raya masih menjadi penyumbang tingginya angka kematian di Indonesia, sehingga menjadi perhatian khusus bagi kepolisian di negara ini. Kecelakaan terjadi di berbagai tempat dengan waktu kejadian yang berbeda, hal ini menyebabkan sulitnya menentukan daerah mana yang memiliki tingkat kerawanan kecelakaan lalu lintas. Informasi mengenai daerah rawan kecelakaan sangat dibutuhkan oleh masyarakat dan penegak hukum. Informasi tersebut dapat dijadikan bahan pertimbangan untuk pengawasan maupun tindakan antisipasi khususnya bagi kepolisian. Tahapan awal pencegahan kecelakaan lalu lintas adalah dengan mengetahui faktor-faktor penyebab kecelakaan lalu lintas yang diperoleh melalui analisa data kecelakaan. Sistem Informasi pada penelitian ini melakukan analisa terhadap wilayah rawan kecelakaan di wilayah Kabupaten Asahan. Analisa tersebut dapat dilakukan dengan data mining, yaitu K-Means Clustering. K-Means Clustering mengelompokkan data menjadi beberapa cluster sesuai karakteristik data tersebut. Hasil dari penelitian ini ialah Satlantas Polres Kabupaten Asahan dapat mengetahui daerah-daerah rawan kecelakaan dalam katagori paling rawan, cukup rawan dan tidak rawan.*

Kata Kunci*: Daerah Rawan Kecelakaan, K-Means Clustering, Kecelakaan Lalu Lintas.*

INTRODUCTION

The rate of population growth and the amount of traffic flow in the Asahan Regency is increasing rapidly [1], so the need for transportation infrastructure continues to grow. This situation greatly affects the service so that if it is not balanced with an increase in inadequate transportation infrastructure, the resulting impact is the emergence of problems with traffic, such as traffic jams and accidents. The development of transportation will indirectly increase the risk of growing traffic problems. Traffic accidents according to RI Law No. 22 of 2009 Article 1 paragraph 24 are an unexpected and unintentional event on a road involving vehicles with or without other road users resulting in human casualties [2] and/or property losses.

All the developments and growth that occur naturally emerge several transformation problems that exist. One of the problems that are most often highlighted is the issue of traffic safety or can be called safety life. Based on Table 1, accidents involve the number of accident victims, starting with death, heavy injuries, minor injuries.

Table 1. Number of Accident Victims in 2015-2018

No	Year	Total Accident	Death	Heavy Injuries	Minor Injuries
1	2015	390	129	223	414
2	2016	453	148	162	542
3	2017	377	122	85	547
4	2018	395	108	91	500
Total		1615	507	561	2003

In order for the resulting policy to be relevant to the problems encountered to prevent accidents, the policy must be supported by information derived from traffic accident data that has been occurring so far. As in Table 2, which contains data on accident victims in Asahan Regency from 2015 to 2018.

Table 2. Accident Victim Data for 2015-2018

No	Sub-district	Total of Accident Victims			
		2015	2016	2017	2018
1	Aek Kuasan	5	7	6	6
2	Aek Ledong	4	6	5	5
3	Aek Songsongan	5	7	4	6
4	Air Batu	24	30	30	24
5	Air Joman	22	32	29	22
6	Bandar Pasir Mandoge	14	18	12	14
7	Bandar Pulau	32	38	27	32
8	Buntu Pane	6	5	4	6
9	Kisaran Barat Kota	37	47	31	37
10	Kisaran Timur Kota	36	35	36	36
11	Meranti	31	44	33	31
12	Pulau Rakyat	6	6	4	7
13	Pulo Bandring	33	29	29	33
14	Rahuning	3	4	4	4
15	Rawang Panca Arga	18	18	14	18
16	Sei Dadap	13	16	15	13
17	Sei Kepayang	12	13	15	12
18	Sei Kepayang Barat	10	16	10	10
19	Sei Kepayang Timur	11	15	11	11
20	Setia Janji	5	6	5	5
21	Silau Laut	6	7	5	6
22	Simpang Empat	13	13	11	13
23	Tanjung Balai	36	30	29	36
24	Teluk Dalam	3	5	4	3
25	Tinggi Raja	5	6	4	5

Therefore, a grouping of areas is carried out, namely the most vulnerable areas (areas that have the highest number of accidents, high risk, and potential accidents on a road segment), areas that are quite vulnerable (areas that have high enough accident rates) and areas that are not vulnerable. This can be seen from the number of various victims of accidents that are accident-prone areas and makes all parties feel the need to take preventative measures and also to find out the factors that trigger accidents [3].

Researches related to the theme of traffic accidents have been carried out. The first journal from Aljofey & Alwagih [4] is to analyze the time of the frequency of traffic accidents for the location of the highway. The k-means algorithm is applied to find out when and where accidents often occur within 24 hours. The second is the journal from Anshori & Nuraini [5] where the research was conducted in Tasikmalaya with 4 clusters based on time grouping, namely night, daytime, evening' and morning. And the results obtained are the most traffic accidents occur in the morning in Tasikmalaya. The third is a journal from Purwaningsih [6] who analyzed traffic accidents in Jakarta City in 2013 by grouping them into 3 clusters through RapidMiner tools as a medium for calculating K-Means Clustering. The fourth journal from Wicaksono, Kusrini, & Lutfi [7] by analyzing the data on traffic accident vulnerability in Bantul Regional Police using K-Means, found that the most vulnerable time for traffic accidents is at night from 19.30 to 23.59 WIB.

Data mining can be used to identify patterns and predict future behavior [3]. One method of data mining is K-Means Clustering which is a method of grouping data into clusters [8] based on the similarity of each of the existing clusters [9], [10], [11]. The purpose of grouping traffic accidents in the Asahan Regency is to find out the accident-prone areas. An accident-prone area is an area where the accident rate is high with repeated accidents occurring in the same period and relatively active space [12].

The application of K-Means Clustering in traffic accident data in this study will determine the initial centroid, K-Means Clustering processing, and display the resulting clusters. Then an analysis of clusters produced by accident-prone areas will be carried out to help reduce the risk of accidents in the Asahan Regency.

MATERIALS AND METHODS

Data Mining is a method of processing data on a large scale, where the data will be stored in a database, data warehouse, or information storage. Data mining plays an important role in several fields, including economics, industry, science and technology, and weather [13]. Data Mining is the process of finding patterns and relationships hidden in a large amount of data to classify, estimate, forecasting, associate rules, sequential patterns, clustering, regression, description, and visualization [12]. Besides the data processed using data mining techniques will produce new knowledge derived from old data, so the results obtained from the data mining process can be used to determine future decisions [14].

K-Means is a method of grouping that is partial [15], [16]. This method partitioned data into groups (clusters) that have the same characteristics [17]. Clustering is one of the non-clustering methods hierarchy which divides data into groups so that data that has the same characteristics are grouped into the same cluster and data that has different characteristics are grouped into clusters [18], [11]. The purpose of grouping this data is to minimize the objective functions established in the grouping process, which generally try to minimize variations within a group and maximize variation between groups [19].

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

The following are data grouping techniques using the K-Means algorithm [20]:

1) Determine the number of K clusters.
2) Initialization of the center point of the K cluster (centroid) can be done randomly and used as the initial cluster.
3) Allocate each data to the closest centroid with the specified matrix distance. To calculate this, the Euclidean distance theory is formulated as follows:

$$(T_{(x,y)}) = \sqrt{(T_{1x} - T_{1y})^2 + (T_{2x} - T_{2y})^2 + (T_{kx} - T_{ky})^2} \quad \text{.....(1)}$$

Where, $T_{(x,y)}$ is the data of x distance to the center of the clustery; T_{kx} is i-data in the k attribute data; T_{ky} is the center of j in the k attribute.

4) Recalculate the cluster center with the new cluster membership. This is calculated by determining the centroid/cluster center.
5) Set each object as the center of the new cluster, if the cluster center is changed, then return to the third step, otherwise, the grouping is complete.
6) Analyze the results in the grouping process.

One of the characteristics of the K-Means algorithm is that it is very sensitive in determining the initial center point of the cluster because K-Means generates random center points of the initial cluster. When the initial random center point generation approaches the final center cluster solution, K-Means has a high possibility to find the right center point of the cluster. Conversely, if the start of the central point is far from the final solution of the center of the cluster, then this is most likely to cause incorrect clustering results. As a result, K-Means does not guarantee unique clustering results. This is what makes the K-Means method difficult to achieve global optimum, but only a local minimum. Besides, the K-Means algorithm can only be used for data whose attributes are numeric.

RESULTS AND DISCUSSION

The purpose of this study is to classify the area of traffic accidents in the Asahan Regency. This research was made based on data from traffic accidents in Asahan Regency from 2015 to 2018 in table 2.

This stage of the analysis is carried out by the K-Means Clustering method in grouping data against a fact or rule. The following is an example of the data used for the calculation of the K-Means Clustering method. For initial determination the following data are assumed:

Taken 23rd data as the center of the 1st cluster.

C1 = 36 30 29 36

Taken the 15th data as the center of the 2nd cluster.

C2 = 18 18 14 18

Taken the 3rd data as the center of the 3rd cluster.

C3 = 5 7 4 6

After knowing the number of clusters and initial centroids, then measuring between centroids using equation 2 and then the distance matrix will be obtained namely C1, C2, C3.

The example of calculating the 1st data distance ie Aek Kuasan in each cluster is the following equation (1), so tha5t the results are [54,03 22,32 2,00]. The same equation and calculation will be implemented in 24 other data to get the distance of each data in each cluster as in table 3.

Table 3. Cluster Center Distance Calculation Results

Data to	C1	C2	C3	Shortest Distance
1	54,03	22,32	2,00	2,00
2	56,01	24,29	2,00	2,00
3	54,91	23,11	0,00	0,00
4	17.00	21,73	43,47	17,00
5	19,90	21,28	42,37	19,90
6	37,43	6,00	18,17	6,00
7	10,00	31,00	53,81	10,00
8	55,23	23,60	2,24	2,24
9	17,18	43,03	65,68	17,18
10	8,60	37,70	60,57	8,60
11	16,19	37,08	59,25	16,19
12	54,24	22,56	1,73	1,73
13	4,36	28,21	51,21	4,36
14	58,43	26,78	4,12	4,12
15	31,89	0,00	23,11	0,00
16	38,08	7,42	17,75	7,42
17	40,46	9,90	15,56	9,90
18	43,69	12,17	12,57	12,17
19	42,41	10,77	13,19	10,77
20	55,44	23,73	1,73	1,73
21	53,90	22,14	1,41	1,41
22	40,88	9,17	14,07	9,17
23	0,00	31,89	54,91	0,00
24	58,55	26,81	4,12	4,12
25	55,88	24,12	1,41	1,41

The next step is to create a grouping table where the smallest value of 3 groups (clusters) is given a value of 1 (one) while the remainder is given a value of 0 (zero), as in Table 4.

Table 4. Iteration-1 Grouping

Data to	C1	C2	C3
1	0	0	1
2	0	0	1
3	0	0	1
4	1	0	0
5	1	0	0
6	0	1	0
7	1	0	0

Data to	C1	C2	C3
8	0	0	1
9	1	0	0
10	1	0	0
11	1	0	0
12	0	0	1
13	1	0	0
14	0	0	1
15	0	1	0
16	0	1	0
17	0	1	0
18	0	1	0
19	0	1	0
20	0	0	1
21	0	0	1
22	0	1	0
23	1	0	0
24	0	0	1
25	0	0	1

After that, the process of this iteration-2 will be calculated by its centroid which is no longer based on the previous sample but from the following new centroid.

C1= 31,38 35,63 30,50 31,38
C2 = 13,00 15,57 12,57 13,00
C3 = 4,80 5,90 4,50 5,30

Repeat the calculation process as in the previous example, until the last grouping of data is equal to the value of the previous data set from clustering results. The results of the calculation after the iteration has stopped (2nd iteration), can be seen in Table 5.

Table 5. Results Distance Between Data and Centroid

Data to	C1	C2	C3	Shortest Distance
1	52,53	15,15	2,00	2,00
2	54,52	17,14	0,99	0,99
3	53,49	16,12	1,41	1,41
4	11,86	27,46	44,15	11,86
5	13,83	26,49	43,08	13,83
6	35,45	2,87	19,05	2,87
7	4,32	37,86	54,68	4,32
8	54,11	16,83	1,73	1,73
9	13,89	49,79	66,58	13,89
10	8,57	44,55	61,28	8,57
11	8,76	43,28	60,09	8,76
12	53,08	15,81	2,14	2,14
13	7,17	35,36	51,93	7,17
14	57,07	19,71	2,96	2,96
15	30,67	7,61	23,92	7,61
16	36,06	2,47	18,41	2,47
17	38,77	3,81	16,04	3,81
18	41,46	4,98	13,47	4,98
19	40,45	3,29	14,00	3,29
20	54,03	16,64	0,62	0,62
21	52,51	15,13	1,84	1,84
22	39,59	3,01	14,81	3,01
23	8,76	39,19	55,65	8,76
24	57,01	19,63	3,10	3,10
25	54,51	17,12	0,62	0,62

The grouping results obtained from the last iteration are iteration-2, where the results are the same as grouping iteration-1 (see Table 4).

To be easier to use, this research was implemented using the Visual Studio 2010 programming language as a tool for the K-Means Clustering system in traffic accident decision making.

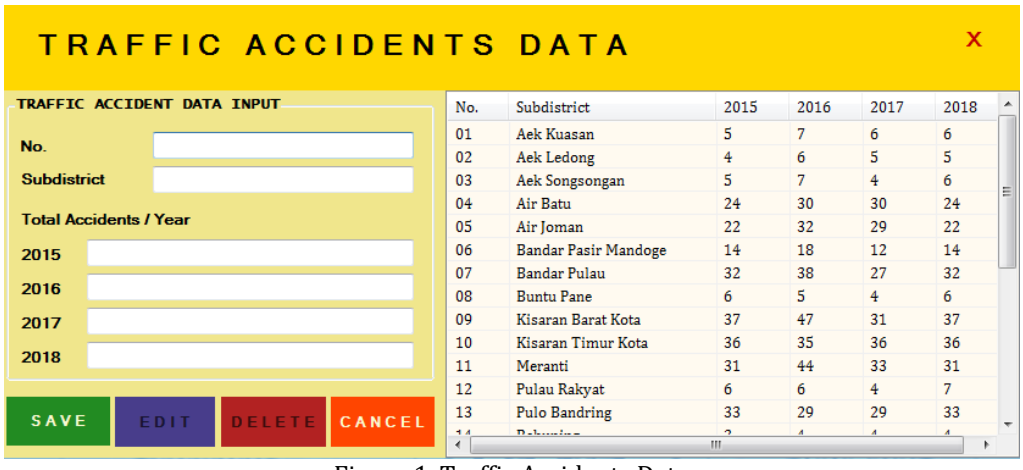


Figure 1. Traffic Accidents Data

In Figure 1, there is a menu for traffic accident data that contains the initial data to be clustered. Users can also save, modify, delete, and cancel traffic accident data in the menu.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1519

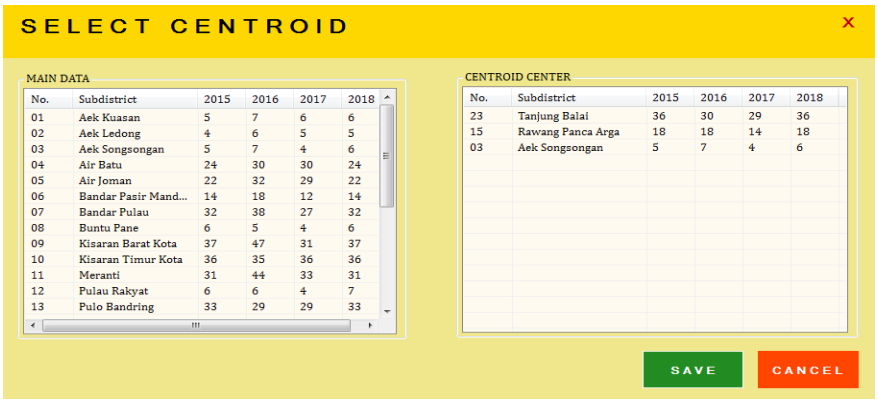


Figure 2. Centroid Data

In Figure 2, the user is intended to select the starting center point of the cluster as many as three central points only and after that, the data will be stored to continue the clustering process. The center point that the user chooses is data 24, 15, and 03.

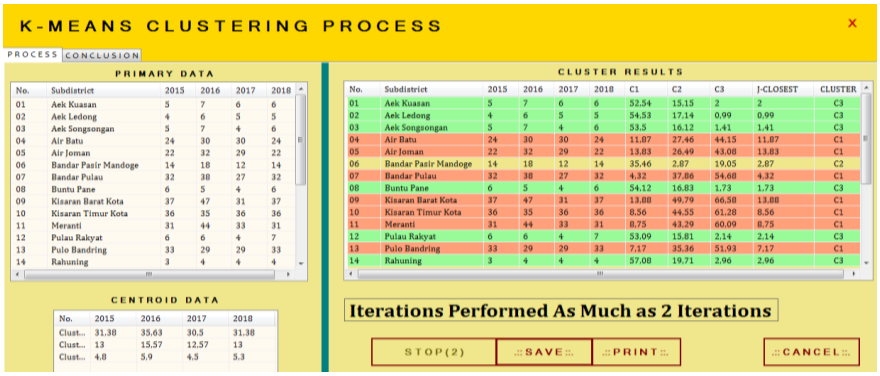


Figure 3. K-Means Clustering Process

In Figure 3, there are 3 tables: initial data containing traffic accident data, initial centroid data tables, and grouped data result in tables. From Figure 3, it can be seen that: "Iterations Performed As Much as 2 Iterations". Besides, users can also see conclusions from the calculation results of traffic accidents by pressing the conclusion button. The following is the display of clustered data (clusters):

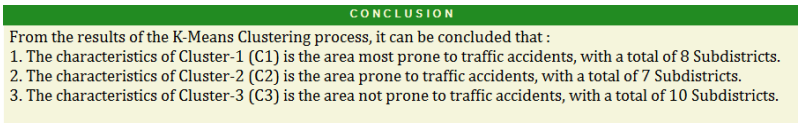


Figure 4. Conclusion of Clustering Results

Based on the results of the calculation and implementation of the group of traffic accident-prone areas in Asahan Regency, the same results can be found:

1. In Cluster-1 (C1), which is the area most prone to traffic accidents, there are 8 Sub-districts: Air Batu, Air Joman, Bandar Pulau, Kisaran Barat, Kisaran Timur, Meranti, Pulo Bandring, and Tanjung Balai.
2. In Cluster-2 (C2), which is an area prone to traffic accidents, there are 7 Sub-districts: Bandar Pasir Mandoge, Rawang Panca Arga, Sei Dadap, Sei Kepayang, Sei Kepayang Barat, Sei Kepayang Timur, and Simpang Empat.
3. In Cluster-3 (C3), which is the area not prone to traffic accidents, there are 10 Sub-districts: Aek Kuasan, Aek Ledong, Aek Songsongan, Buntu Pane, Pulau Rakyat, Rahuning, Setia Janji, Silau Laut, Teluk Dalam, and Tinggi Raja.

CONCLUSION

From the results and discussions that have been carried out, it can be concluded that the K-Means Clustering method can classify traffic accident-prone areas in Asahan Regency with 3 clusters, namely the most accident-prone areas with a total of 8 Subdistricts, the area is quite prone to traffic

accidents with results totaling 7 Subdistricts, and areas not prone to traffic accidents with results totaling 10 Subdistricts.

REFERENCE

[1] F. A. Arisandi, M. Lubis, and M. H. M. Hasibuan, "Penerapan Managemen Lalu Lintas Pada Kabupaten Asahan," vol. 15, no. 2, 2020.

[2] S. Hussain, L. J. Muhammad, F. S. Ishaq, A. Yakubu, and I. A. Mohammed, "Performance evaluation of various data mining algorithms on road traffic accident dataset," in *Smart Innovation, Systems and Technologies*, 2019, vol. 106, pp. 67–78.

[3] A. Almjewail, A. Almjewail, S. Alsenaydi, H. ALSudairy, and I. Al-Turaiki, "Analysis of traffic accident in Riyadh using clustering algorithms," in *Advances in Intelligent Systems and Computing*, 2018, vol. 753, pp. 12–25.

[4] A. M. Aljofey and K. Alwagih, "Analysis of Accident Times for Highway Locations Using K-Means Clustering and Decision Rules Extracted from Decision Trees," 2018.

[5] I. Fardian Anshori and Y. Nuraini, "Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Tasikmalaya Menggunakan Algoritma K-Means," *J. Responsif*, vol. 2, no. 1, pp. 118–127, 2020.

[6] E. Purwaningsih, "Analisis Kecelakaan Berlalu Lintas Di Kota Jakarta Dengan Menggunakan Metode K-Means," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 5, no. 1, pp. 1–6, 2019.

[7] E. A. Wicaksono, K. Kusrini, and E. T. Lutfi, "Analisis Data Kerawanan Kecelakaan Lalu Lintas Menggunakan Metode K-Means (Studi Kasus Polres Bantul)," in *Seminar Nasional Teknologi Informasi dan Multimedia*, 2017, pp. 109–113.

[8] S. Nawrin, M. Rahatur, and S. Akhter, "Exploring K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017.

[9] A. A. Mousa, M. A. El-Shorbagy, and M. A. Farag, "K-means-clustering based evolutionary algorithm for multi-objective resource allocation problems," *Appl. Math. Inf. Sci.*, vol. 11, no. 6, pp. 1681–1692, Nov. 2017.

[10] U. R. Raval and C. Jani, "Implementing & Improvisation of K-means Clustering Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 55, no. 5, pp. 191–203, 2016.

[11] V. Soundarya, "Recommendation System for Criminal Behavioral Analysis on Social Network using Genetic Weighted K-Means Clustering," *J. Comput.*, pp. 212–220, 2017.

[12] R. R. Fiska, "SATIN-Sains dan Teknologi Informasi Penerapan Teknik Data Mining Dengan Metode Support Vector Machine (SVM) Untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)," 2017.

[13] R. Wulan Sari, A. Wanto, and A. Perdana Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2018, vol. 2, no. 1, pp. 224–230.

[14] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.

[15] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, Oct. 2017.

[16] D. Xia, F. Ning, and W. He, "Research on Parallel Adaptive Canopy-K-Means Clustering Algorithm for Big Data Mining Based on Cloud Platform," *J. Grid Comput.*, Jun. 2020.

[17] N. Ramadhani, A. F. Rahman, and D. Riskiyati, "Aplikasi Cluster Data Perkara Lalu Lintas Mingguan Di Pengadilan Negeri Pamekasan," *J. LINK*, vol. 26, no. 2, pp. 18–24, 2017.

[18] W. Astuti, D. Djoko, and A. Widodo, "Pemetaan Tindak Kejahatan Jalanan di Kota Semarang Menggunakan Algoritma K-Means Clustering," *J. Tek. Elektro*, vol. 8, no. 1, pp. 5–7, 2016.

[19] H. Priyatman, F. Sajid, and D. Haldivany, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Waktu Kelulusan Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.

[20] Purnawansyah, Haviluddin, A. F. O. Gafar, and I. Tahyudin, "Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities," *Proc. Elev. Int. Conf. Manag. Sci. Eng. Manag.*, pp. 300–310, 2018.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1969

THE EFFECT OF STUDENT BEHAVIOR ON THE USE OF CAMPUS JOURNALS BY ADOPTING THE TAM METHOD

Asrul Sani[1*]; Agusta Pratama Wibawa[2]; Nur Nawaningtyas Pusparini[3]; Adrial Rusyid [4]; Anton Hindardjo[5]

Informatics
STMIK Widuri
www.kampuswiduri.ac.id
asrulsani@kampuswiduri.ac.id; agustapratama17@gmail.com; tyaspusparini@kampuswiduri.ac.id

Management
Universitas Muhammadiyah Tangerang
www.umt.ac.id
adrial@umt.ac.id

Management
Sekolah Tinggi Ekonomi Islam SEBI
www.sebi.ac.id
anton.hindardjo@sebi.ac.id

(*) Corresponding Author

Abstract— This research is a quantitative study using statistical analysis based on SEM-PLS in the analysis process. Collecting data using random sampling where respondents fill out a questionnaire simultaneously by conducting interviews. The number of respondents produced was 203 students from various universities who used website-based campus journals. The Technology Acceptance Model (TAM) is the reference in the model used, on the grounds that TAM is a technology acceptance model that is widely adopted in research based on acceptance technology, in this case, the use of campus journals. Initial data processing, coding according to the data, using SPSS, then using SmartPLS 3.0. The results obtained based on the analysis of measurement models and structural models are that student behavior in using journals gives positive results on the scientific development that they want to obtain from the journal. This is shown from the measurement model testing which shows all indicators accompanying the variables, giving positive results based on convergent validity, discriminant validity, and model reliability. Meanwhile, the structural model testing is calculated based on the path coefficient, determination, predictive relevance, and t-test. The number of path analyses obtained is four, the result is that all paths have a significant path coefficient, strong model determination, good predictive relevance and the results of the t-test for all variables that are interconnected based on the path formed are acceptable. These results prove that the behavioral effect of using campus journals by adopting TAM as the model used produces positive results.

Keywords: TAM, SmartPLS, SEM, Quantitative, Journal.

Abstrak— *Penelitian ini merupakan penelitian kuantitatif dengan menggunakan analisa statistika berbasis SEM-PLS dalam proses analisis. Pengambilan data menggunakan random sampling dimana responden mengisi kuesioner bersamaan dengan melakukan wawancara. Jumlah responden yang dihasilkan sebanyak 203 mahasiswa dari berbagai perguruan tinggi yang menggunakan jurnal kampus berbasis website. Technology Acceptance Model (TAM) menjadi acuan dalam model yang digunakan, dengan alasan TAM merupakan model penerimaan teknologi yang banyak di adopsi dalam penelitian yang berbasis teknologi penerimaan dalam hal ini penggunaan jurnal kampus. Pengolahan data awal, dilakukan pengkodean yang disesuaikan dengan data, menggunakan SPSS, yang selanjutnya menggunakan SmartPLS 3.0. Hasil yang didapatkan berdasarkan analisis measurement model dan structural model adalah perilaku mahasiswa dalam penggunaan jurnal memberikan hasil yang positif terhadap perkembangan keilmuan yang ingin diperoleh dari jurnal tersebut. Hal ini ditunjukkan dari pengujian measurement model yang menunjukkan semua indikator yang menyertai variabel, memberikan hasil yang positif berdasarkan validitas convergent, validitas discriminat dan realibilitas dari model. Sedangkan dari pengujian structural model dihitung berdasarkan path coefficient, determinasi, predictive relevance dan t-test. Jumlah analisis jalur yang diperoleh ada empat, hasilnya semua jalur memperoleh path coefficient yang signifikan, determinasi model*

yang kuat, predictive relevance yang baik dan hasil t-test semua variabel yang saling berhubungan berdasarkan jalur yang terbentuk dapat diterima. Hasil ini membuktikan bahwa efek perilaku penggunaan jurnal kampus dengan mengadopsi TAM sebagai model yang digunakan, menghasilkan hasil yang positif.

***Kata Kunci**: TAM, SmartPLS, SEM, Kuantitatif, Jurnal*

INTRODUCTION

Student behavior is an activity carried out by students towards each individual and group that can be observed directly or indirectly covering several aspects of daily life, in the learning process of lectures and the campus environment to provide responses or reactions to achieve a goal. So that you get a concrete assessment as well as support for the results of student activeness that will be obtained as well as a benchmark for the development of science in a complex and structured manner, which is then identified in the context of the parts of learning [1]

According to Skinner [2] behavior change can be seen from the form of response which includes stimulus, behavior is divided into 2 types of grouping, first closed behavior, namely stimulus-response that occurs covertly which is still limited to aspects of attention, perception, knowledge, awareness, and attitudes that occur in individuals and have not been seen clearly or randomly, while the second is open behavior, namely the individual's response to the form of reality that occurs is clear in an action that is implemented easily and can be studied flexibly. Based on the point of view of known student behavior, it has an important role in the learning process such as campus journal visits. [3][4]

Campus journal visits are activities carried out from observations including surveys or directly examining campus journals within the scope of higher education to obtain relevant information while utilizing available services then analyzing and considering in factual form to provide an overview as knowledge where there are various kinds of branching knowledge around certain aspects and can be used as a reference to understand the meaning of this knowledge, the purpose of continuous campus journal visits to journal accreditation is to increase the potential of the journal [5][6]

The method in this research is carried out by using quantitative research where this research is more likely to analyze a student's behavior which will get results following those in the field and the Technology Acceptance Model (TAM) method, TAM is a technology acceptance model from certain aspects that affect and most widely used in the field of Information Technology or information systems. Therefore, TAM extensively uses various elements of the model intending to get a viewpoint from the TAM side and a more concrete or valid explanation of the existing technology acceptance process both individually and in groups [7][8][9]

The relationship between student behavior and the TAM method on visits to campus journal websites is very important for students, where they will know how much their perception of acceptance is. The TAM variable used consists of 4 variables, namely, perceived ease of use, perceived usefulness, attitude toward using, and behavioral intention to use which is poured into a questionnaire to find out and evaluate the results of student behavior based on the calculations in the TAM so that at the end of the process TAM concludes to meet existing information on-campus journal visits [10][11][12]

Table 1. Research Literature

Research Problem		Literature
RP1	The environment that affects students in knowledge development, as well as behavioral changes that occur	The environment that makes students have to change, student behavior in the use of technology, the importance of technological knowledge to support student activities [1] [2] [3][4]
RP2	How students can use campus-based journals as a source of information to find references for writing scientific papers Readiness of students to use website-based campus journals	Campus journals as a means of student reference, student readiness in web journals, use of web-based journals, and knowledge of student behavior in the use of these journals, as well as viewing the model used. [5][6][7]
RP3	Technology Acceptance Model is one of the technologies developed to see student behavior using technology.	The use of TAM as a behavior regulator in using technology is used in various organizations. Measurement using TAM can use several methodologies [8][9] [10][11]

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1969

From the information sources in the literature in the table above, it can be seen that the use of technology in an organization is absolute. The problem is how we measure the user by using it. Many researchers have previously measured the use of technology using the TAM model, but only 2 studies have been discussed specifically for student journals in Indonesia. The novelty of this research is that the analysis used is SmartPLS 3.0 by testing measurement models and structural models. In the meantime, the two previous studies did not use the method used in this study. In addition, the study took samples from several universities in Jakarta and Depok as a source of respondents so that they had a good impact on this research.

This study aims to analyze and evaluate student behavior towards campus journals so that it is known the perception of acceptance of campus journals among students. This study uses four variables in TAM, namely perceived ease of use, perceived usefulness, attitude in use, and behavioral intention. Calculation analysis using SmartPLS 3.0 software. The questionnaire used in this study uses the Linkert scale so that later can be determined the value of the measurement and structural analysis obtained. The two questions posed as a guide in conducting this research are;

Q1: What is the level of acceptable behavior of campus journals against the conditions of existing journals?
Q2: What factors influence the level of student behavior in existing campus journals?

From the results of the research questions above, a hypothesis will be formed which will explain the relationship between the variables in the TAM model. This paper will be divided into four parts, where the first part will explain the background of taking the title, problem, research objectives, and research questions. The second part will explain several related theories that refer to previous research and related to the Technology Acceptance Model (TAM) theory. This section will also describe aspects of the research methodology that will provide an overview of the research phases, starting with model development, data collection, and, finally, data processing. The initial test will measure the validity level of the questionnaire, especially how the prospective respondents understand the contents of the research questions. The research will use a special sampling technique based on key informational aspects [13][14][15]

A total of 203 respondents were obtained in this study, and the collected data would be edited and coded using Ms. software Excel and SPSS. Meanwhile, the quantitative analysis process uses the SEM-PLS approach with SmartPLS 3.0. The third part will explain the analysis obtained after processing the data in the form of discussions and the next stages of research.

The interpretation will be carried out on the basis of the results of the analysis, with the main consideration of the results of the quantitative analysis supported by the results of the descriptive analysis. And the last part will provide conclusions that will later be used as a reference for further research. Theoretically, this study will provide an overview of how students perceive journal acceptance, especially final year students. The involvement of interested parties in using campus journals can represent the validation of measurements based on the definition of TAM, namely how technology can be utilized by those who use it.

MATERIALS AND METHODS

The study of the influence of student behavior on campus journals using the TAM method is carried out into seven main stages (Figure 1). The initial study (1) was carried out by making initial preparations, namely looking at the scope of the research, where the literature study and the preparation of the model that will be used as the research base include looking at the types of campus journals, and the social aspects of the IS study, for example, usability, satisfaction, and readiness [16][17]. Model initiation (2) becomes part of the research, namely making preparations such as the tools and tools used. The third stage is conducting design research, finding indicators that match the research theme based on TAM theory. The measurement model test (4) is the next stage to ensure the level of reliability and validity of the model is achieved. After the reliability and validity meet the required requirements, the structural model test is continued, which simultaneously performs descriptive analysis and inferential analysis.

Refers to the research points identified in the early stages of the research as described in the introduction, in particular the research questions section; this research was conducted to answer the above questions using a quantitative approach [13][18] by applying the researcher's objective point of view on how to determine the status of student behavior in progress and examining the relationship between variable variables in TAM. In particular, following the predetermined approach, the next phases of research also adjust to the approach using quantitative methods, techniques, and tools as demonstrated by the research process. For example, the data collection technique was carried out by means of a survey with a

questionnaire research tool, the data analysis was carried out statistically with related computer software [19][20]

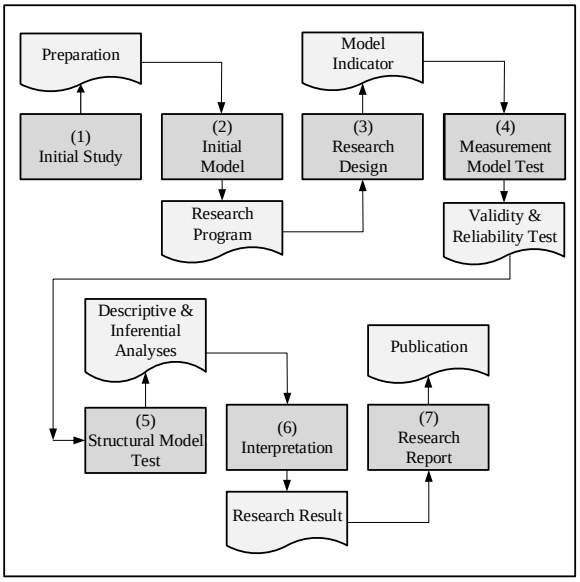


Figure 1. Research Procedure

The SEM equation model is a popular analysis technique used in social science research and computer science based on computer behavior or social computers. This analysis technique is a combination of two scientific discipline methodologies, namely the econometric perspective that focuses on prediction and psychometrics which can describe the model concept. Latent variables (variables that cannot be measured directly) but are measured through the indicators (manifest variables). SEM essentially offers the ability to perform path analysis with latent variables.

Partial Least Square Analysis Structural Equation Model (PLS-SEM) usually consists of two sub-models, namely the measurement model or often called the outer model, and the structural model or often called the inner model. The measurement model shows how the manifest or observed variable represents the latent variable to be measured, while the structural model shows the power of estimation between latent variables. The latent variables formed in PLS-SEM have indicators that can be reflective or formative. Reflective indicators are indicators that are manifestations of the construct and are following the classical test theory which assumes that the variance in the measurement of the latent variable score is a function of the true score plus the error. Meanwhile, the construct and are following the classical test theory which assumes that the variance in the measurement of the latent variable score is a function of the true score plus the error. Meanwhile, formative indicators are indicators that define characteristics or explain constructs. In TAM there are elemental constructions that have not been modified, of which construction has the main component, namely, perceived ease of use, perceived usefulness, an attitude of use (attitude towards using), and finally behavioral intention to use. And then the structure of the image which includes the construction elements of TAM is as follows 8

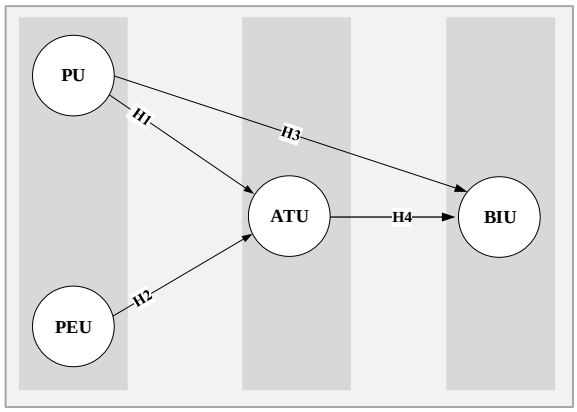


Figure 2. Research Model

The TAM model used in this study uses the variable Perceived Usefulness (PU), Perceived Ease of Use (PEU), Attitude Toward Using (APU), and Behavioral Intention to Use (BIU). All variables were adopted from previous research. Furthermore, a hypothesis is made to explain the relationship between the relationships in the research question. The hypothesis that is formed is:

H1: Perceived usefulness has a significant effect on Attitude Toward Using
H2: Perceived Ease of Use has a significant effect on Attitude Toward Using
H3: Perceived usefulness has a significant effect on Behavioral Intention to Use
H4: Attitude Toward Using has a significant effect on Behavioral Intention to Use.

RESULTS AND DISCUSSION

The results of the respondent profile analysis can be seen in Table 1 below, where this table shows the importance of reliability and validity of the data sources obtained [21][22]. From the number of respondents as many as 203, it is known that the number of respondents with female gender is 114 people (56.2%) and this number is more than the number of respondents with male gender as many as 89 people (43.8%).

For the age category of respondents, namely 18-20 years, there were 12 people, 21-24 years old totaling 155 people and for the age category above 24 years, there were 36 people.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1969

Analysis of the measurement model

The method most often used by researchers in the field of SEM for measurement model analysis through confirmatory factor analysis is to use the Multi Trait Multi-Method approach with convergent validity, discriminant validity, and reliability testing.

Convergent validity testing, where this test is done by looking at the standardized loading factor value. This value illustrates the magnitude of the correlation between each measurement item or indicator and its variables. The limit value is above 0.7 if the indicator is said to be valid as an indicator that measures the variable. In Figure 2 and Table 1 below, it can be seen that all indicators show values above 0.7, and all indicators are declared valid. Besides, for convergent validity testing, the Average Variance Extracted (AVE) value must be above 0.5. In table 1, it can be seen that all the variables show that they are acceptable because the limiting value exceeds 0.5.

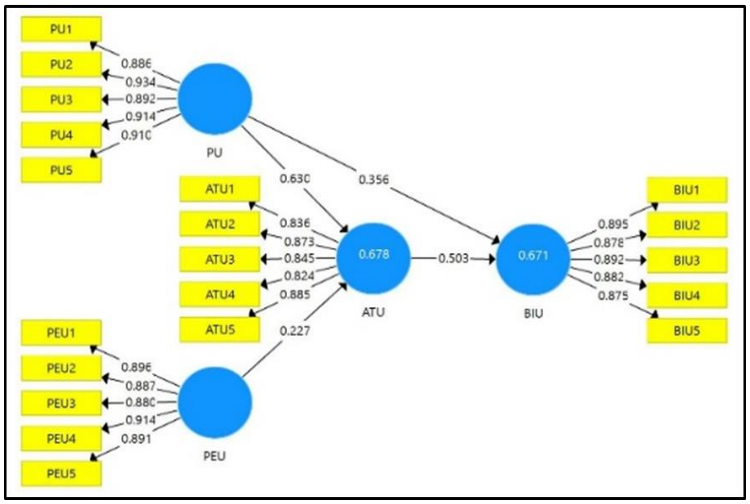


Figure 3. Path diagram of the measurement model test.

Table 2. Statistical results of the measurement model

VAR	IND	OL	CROSSLOADING (CL)				AVE	CA	CR	R^2
			PU	PEU	BIU	ATU				
PU	PU1	0.886	**0.886**	0.741	0.727	0.793	0.823	0.946	0.959	
	PU2	0.934	**0.934**	0.729	0.705	0.713				
	PU3	0.892	**0.892**	0.736	0.669	0.700				
	PU4	0.914	**0.914**	0.709	0.662	0.719				
	PU5	0910	**0.910**	0.745	0.703	0.754				
PEU	PEU1	0.896	0.685	**0.896**	0.665	0.679	0.799	0.937	0.952	
	PEU 2	0.887	0.741	**0.887**	0.648	0.666				
	PEU 3	0.880	0.752	**0.880**	0.670	0.679				
	PEU 4	0.914	0.740	**0.914**	0.607	0.636				
	PEU 5	0.891	0.688	**0.891**	0.608	0.619				
BIU	BIU 1	0.895	0.731	0.675	**0.895**	0.741	0.782	0.930	0.947	0.671
	BIU 2	0.878	0.677	0.629	**0.878**	0.722				
	BIU 3	0.892	0.665	0.609	**0.892**	0.709				
	BIU 4	0.882	0.661	0.635	**0.882**	0.677				
	BIU 5	0.875	0.643	0.620	**0.875**	0.650				
ATU	ATU1	0.836	0.769	0.743	0.715	**0.836**	0.727	0.907	0.930	0.678
	ATU 2	0.873	0.767	0.690	0.713	**0.873**				
	ATU 3	0.845	0.585	0.514	0.587	**0.845**				
	ATU 4	0.824	0.556	0.465	0.571	**0.824**				
	ATU 5	0.885	0.736	0.662	0.753	**0.885**				

Discriminant validity testing, carried out by two stages of cross loading testing, namely, cross-loading between indicators and cross-loading Fornell-Lacker. Cross loading indicator is checked by comparing the correlation of indicator variables with one another. If the correlation between indicators with variables is higher with other variables, it can be said that these variables predict the size of the block better than other blocks. Checking Fornell-Lacker's cross-loading is by looking at the root value of AVE. The root value of AVE must be higher than the correlation between one variable and another. Table 1 shows that all cross-loading indicator values with the variable are

higher with the correlation to other variables. In table 2, it shows that the root value of AVE is higher when compared to the correlation between variables on other variables.

Table 3. Discriminant validity of Fornell-Lacker's

	ATU	BIU	PEU	PU
ATU	**0.853**			
BIU	0.793	**0.884**		
PEU	0.735	0.717	**0.894**	
PU	0.813	0.765	0.807	**0.907**

In addition to the validity test, model measurement is also carried out to test the reliability of a variable. A reliability test is conducted to prove the accuracy, consistency, and accuracy of the instrument in measuring variables. To measure the reliability of a variable with reflective indicators, it can be done by looking at the value of Cronbach's Alpha (CA) and Composite Reliability (CR). The limit value used for both tests is a value above 0.7. Based on table 1 above, it can be seen that all variables have a value above 0.7 and this shows that the model has reliable variables.

The value of the test shows that all the results of the student questionnaire give a positive response, the resulting indicator values for the PU, PEU, ATU, and BIU variables are all above 0.82 even though the limit of the indicator value for exploratory research is above 0.7. This shows that the questionnaire filled out is well understood by the respondents. Based on the results of testing the outer model, both indicators and variables, it can be concluded that for variables, PU, PEU, ATU, and BIU can be continued for inner model testing.

Analysis of the structural model (inner model)

This analysis uses two testing procedures, namely by using bootstrapping testing and blindfolding testing. Bootstrapping test analysis was carried out on the path coefficient (β), determination coefficient (R^2), effect size (f^2) based on the R^2 value and t-test, while for blindfolding it was carried out by testing predictive relevance (Q^2) and the q^2 value calculated based on the value Q^2. Figure 3 shows the path diagram for the results of the research model for testing.

The path coefficient test (β) is carried out to test the β value, where the value is> 0.1 which indicates a significant effect of the model. The coefficient of determination (R^2) test was carried out to explain the independent variable with a standard measurement value of about 0.670 for the strong variant, 0.333 for the moderate variant, and 0.190 for the weak variant.

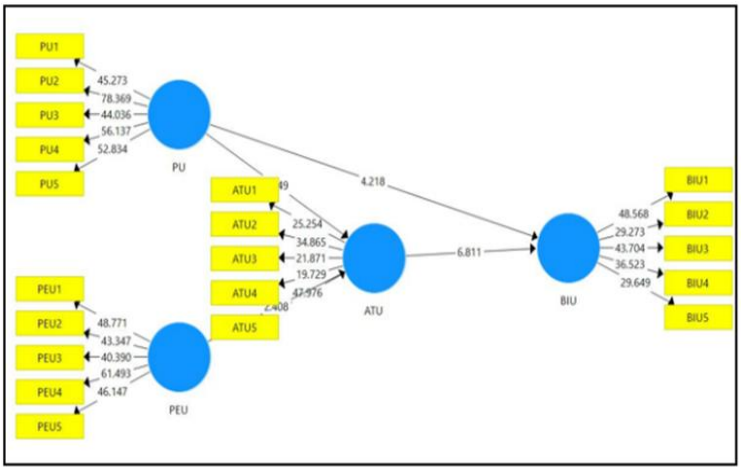


Figure 4. Path diagram of the structural model test.

To calculate the effect size f 2, the formula is needed:

$$f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}} \quad \text{.....(1)}$$

$$q^2 = \frac{Q^2_{included} - Q^2_{excluded}}{1 - Q^2_{included}} \quad \text{.....(2)}$$

From the formula above, the values of f^2 and q^2 can be obtained, where the standard measurement values are 0.02 for small effect sizes, 0.15 for medium effect sizes, and 0.35 for large effect sizes. The predictive relevance (Q^2) test is conducted to determine the predictive relationship between certain variables and other variables with measurement limits above zero. Hypothesis testing (t-test) uses a two tailed test with a significance of 5%, meaning that the hypothesis will be accepted if it has a t-test> 1.96 [19][20]

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1969

Table 4. Structural Model Analysis Results

Path	β	R^2	Q^2	f^2	q^2	t-test	Analysis					
							β	R^2	Q^2	f^2	q^2	t-test
PU → ATU	0,63	0,68	0,47	0,31	0,11	6,31	Sign	Strong	Predictive	Strong	Moderate	Accept
PEU → ATU	0,23	0,68	0,47	0,05	0,01	2,29	Sign	Strong	Predictive	Moderate	Weak	Accept
PU → BIU	0,35	0,67	0,52	0,09	0,04	4,25	Sign	Strong	Predictive	Moderate	Strong	Accept
ATU → BIU	0,50	0,67	0,52	0,00	0,00	6,56	Sign	Strong	Predictive	Weak	Weak	Accept

In table 4 above, it can be seen that all paths have a significant effect because they have a value of β > 0.1, namely the PU → ATU, PEU → ATU, PU → BIU, and ATU → BIU pathways. The resulting R^2 test for the PU → ATU line is strong with a value of 67.8%, the PEU → ATU line is strong with a value of 67.8%, the PU → BIU line is also strong with a value of 67.1%, and finally the ATU → BIU line also strong with a value of 67.1%. The resulting Q^2 test for all pathways shows Q^2> 0 and this shows that it has predictive relevance for all variables. For the f^2 test, the PU → ATU is strong with a value of 31%, the PEU → ATU is moderate with a value of 5%, the PU → BIU is moderate with a value of 9%, and the ATU → BIU is weak with a value of 0. Hypothesis testing (t-test) for all lines in an accepted position because for all paths namely PU → ATU, PEU → ATU, PU → BIU, and ATU → BIU have a value> 1.96.

The independent variables in this study namely perceived usefulness (PU) and perceived ease of use (PEU), each have a major influence on student behavior and student habits using website-based journals. This is also seen from the age of using the journal. The biggest users are those aged 21-24 years, which are indeed students who are in the process of doing practical work or the process of writing a thesis. The habit shown by students proves that website-based campus journals are very supportive of student work in doing assignments and final assignments. When viewed from the large percentage, female students dominate the use of campus journals. The habit of using the internet also affects students to view existing campus journal journals. This is because the level of ease in accessing journals is felt by students to be easy to access. The dependent variable also has a big influence on this research. This can be seen from the determination value of R^2 which shows that all pathways are above 67%. This possibility is due to the positive response obtained from students using campus journals. Experience the ease of downloading or getting access, has a very positive impact when filling out the questionnaire. In the behavior intention (BIU) variable, it can be seen that the R^2 value is 67.1%, which indicates a good response from the user, and the ease of accessing the journal is also a consideration.

CONCLUSION

The behavior of using internet technology (behavior intention) to use campus journals is very necessary for students who are working on assignments, especially final assignments. This behavior is also influenced by perceived usefulness and perceived ease of use in utilizing this technology so that there is a desire to use campus journals. The age factor affects students in using campus journals, but the gender factor does not influence the interest in using campus journals. The results of the measurement model analysis show that all indicators and variables meet the threshold requirements. Testing of convergent validity, discriminant validity, and reliability shows that the value of cross loading, Cronbach's Alpha, average variance extracted, and composite reliability meets the predetermined threshold values, and based on this, the model is continued to analyze the structural model. When viewed from the influencing variables, the independent variable, and the dependent variable influence each other. This can be seen from the structural model analysis by testing the path coefficient values for all paths giving a significant value. Likewise, the coefficient of determination test gives strong results for these lines. Another test is predictive relevance, which provides that all pathways have predictive value. The final test is to see the t-test value which shows that all paths in the TAM model are acceptable. All the results above gave a positive response to the questionnaire conducted, especially when filling out the questionnaire conducted by students, in this case, the behavior of using campus journals.

This research is preliminary research for the TAM model with the subject of behavior towards campus journals. In the future, further research will be carried out using other reference models that may have a modified model to improve information systems research.

REFERENCES

[1] Á. Zsóka, Z. M. Szerényi, A. Széchy, and T. Kocsis, "Greening due to environmental education? Environmental knowledge, attitudes, consumer behavior and everyday pro-environmental activities of Hungarian high school and university students," in *Journal of Cleaner Production*, 2013, vol. 48, pp. 126–138.

[2] E. S. Djatikusuma and H. Widagdo, "Analisa Perilaku Calon Mahasiswa Terhadap Minat Untuk Menjadi Mahasiswa Di Sekolah Tinggi Ilmu Ekonomi Multi Data Palembang Menggunakan Metode Unifield Theory Of Acceptance And Use Of Technology (UTAUT)," *J. Ilm. STIE MDP*, vol. 5, no. 1, pp. 46–55, Sep. 2015.

[3] K. Perilaku, J. Surya, and K. Amri, "Kajian Perilaku Mahasiswa Dalam Menggunakan Internet Dengan Pendekatan Technology Acceptance Model (TAM)," *J. Penelit. Pos dan Inform.*, vol. 3, no. 1, pp. 67–80, 2013.

[4] S. T. Muntianah, S. Astuti, and D. F. Azizah, "Pengaruh Minat Perilaku Terhadap Actual Use Teknologi Informasi dengan Pendekatan Technology Acceptance Model (TAM) (Studi Kasus Pada Kegiatan Belajar Mahasiswa Fakultas Ilmu Administrasi Universitas Brawijaya Malang)," *PROFIT J. Adm. Bisnis*, vol. 6, no. 1, pp. 88–113, 2012.

[5] A. H. Harahap and I. M. Pasaribu, "Layanan Pemustaka Terhadap Tingkat Kunjungan Pada Perpustakaan Universitas Tjut Nyak Dhien Medan," *J. Net. Libr. Inf.*, vol. 1, no. 1, pp. 30–39, 2018.

[6] "Analisis pemanfaatan jurnal elektronik oleh mahasiswa menggunakan Technology Acceptance Model (TAM)," *Harmon. Sos. J. Pendidik. IPS*, vol. 6, no. 2, pp. 175–182, 2019.

[7] M. Al-Emran, V. Mezhuyev, and A. Kamaludin, "Technology Acceptance Model in M-learning context: A systematic review," *Comput. Educ.*, vol. 125, pp. 389–412, 2018.

[8] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q. Manag. Inf. Syst.*, vol. 13, no. 3, pp. 319–339, 1989.

[9] A. Granić and N. Marangunić, "Technology acceptance model in educational context: A systematic literature review," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2572–2593, Sep. 2019.

[10] R. Estriegana, J. A. Medina-Merodio, and R. Barchino, "Student acceptance of virtual laboratory and practical work: An extension of the technology acceptance model," *Comput. Educ.*, vol. 135, pp. 1–14, Jul. 2019.

[11] K. L. Hakim, "Analisis Penggunaan Unnes Journals Berdasarkan Pendekatan Technology Acceptance Model (TAM) Pada Mahasiswa Program Studi Pendidikan Akuntansi Universitas Negeri Semarang," Universitas Negeri Semarang, Semarang, 2016.

[12] A. Sani, N. Wiliani, A. Budiyantara, and N. Nawaningtyas, "Pengembangan Model Adopsi Teknologi Informasi terhadap Model Penerimaan Teknologi diantara UMKM," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 151–158, Feb. 2020.

[13] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles: SAGE Publications, 2017.

[14] A. Subiyakto and A. R. Ahlan, "Implementation of input-process-output model for measuring information system project success," *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 12, no. 7, pp. 5603–5612, 2014.

[15] A. Sani, N. Nawangtyas, A. Budiyantara, and N. Wiliani, "Measurement Of Readiness And Information Technology Adoption Based On Organizational Context Among," *J. Pilar Nusa Mandiri*, vol. 16, no. 2, pp. 225–232, Sep. 2020.

[16] R. P. Aritonang, S. Sumarlin, and R. Kaban, "Kajian Tingkat Kepuasan Mahasiswa Terhadap Sistem Informasi Berbasis Web dengan Metode Technology Acceptance Model(TAM)," *J. TEKESNOS*, vol. 1, no. 1, pp. 40–47, Dec. 2019.

[17] S. S. Alam, M. Y. Ali, and M. F. M. Jani, "An empirical study of factors affecting electronic commerce adoption among SMEs in Malaysia," *J. Bus. Econ. Manag.*, vol. 12, no. 2, pp. 375–399, 2011.

[18] Sugiyono, *Metode Penelitian Bisnis (Pendekatan Kuantitatif, Kualitatif, dan R&D)*. Bandung: Alfabeta, 2018.

[19] K. K.-K. Wong, "Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS," *Mark. Bull.*, vol. 24, no. 1, pp. 1–32, 2013.

[20] J. F. Hair, M. Sarstedt, C. M. Ringle, and J. A. Mena, "An assessment of the use of partial least squares structural equation modeling in marketing research," *J. Acad. Mark. Sci.*, vol. 40, no. 3, pp. 414–433, May 2012.

[21] A. Subiyakto, A. R. Ahlan, M. Kartiwi, N. Hakiem, M. Q. Huda, and A. Susanto, "The Information System Project Profiles among Universities in Indonesia," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 2, pp. 865–872, 2018.

[22] A. Subiyakto, "Development of the Readiness and Success Model for Assessing the Information System Integration," in *Conference: The 2nd International Conference on Science and Technology (ICOSAT)*, 2017, pp. 110–115.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1695

IMPLEMENTATION OF PARTICLE SWARM OPTIMIZATION BASED MACHINE LEARNING ALGORITHM FOR STUDENT PERFORMANCE PREDICTION

Muhammad Iqbal[1*]; Irwan Herliawan[2]; Ridwansyah[3]; Windu Gata[4]; Abdul Hamid[5]; Jajang Jaya Purnama[6]; Yudhistira[7]

Ilmu Komputer[1,2,4], Teknik Informatika[3], Sistem Informasi[6]
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri[1,2,3,4,6]
www.nusamandiri.ac.id[1,2,3,4,6]
14002390@nusamandiri.ac.id [1*]; 14002397@nusamandiri.ac.id [2]; ridwansyah.rid@nusamandiri.ac.id [3]; windu@nusamandiri.ac.id [4]; jajangja2412@bsi.ac.id [6]

Teknologi Informasi[5], Teknologi Komputer[7]
Universitas Bina Sarana Informatika[5,7]
www.bsi.ac.id[5,7]
hamid.adh@bsi.ac.id [5]; yudhistira.yht@bsi.ac.id [7]

(*) Corresponding Author

Abstract—*Education plays an important role in the development of a country, especially educational institutions as places where the educational process has an important goal to create quality education in improving student performance. Based on research conducted in the last few decades the quality of education in Portugal has improved, but statistics show that the failure rate of students in Portugal is high, especially in the fields of Mathematics and Portuguese. On the other hand, machine learning which is part of Artificial Intelligence is considered to be helpful in the field of education, one of which is in predicting student performance. However, measuring student performance becomes a challenge since student performance has several factors, one of which is the relationship of variables and factors for predicting the performance of participating in an orderly manner. This study aims to find out how the application of machine learning algorithms based on particle swarm optimization to predict student performance. By using experimental research methods and the results of empirical studies shown in each model, namely random forest, decision tree, support vector machine, and particle swarm optimization based neural network can improve the accuracy of student performance predictions.*

Keywords: *Student Performance, Machine Learning, Particle Swarm Optimization, Prediction.*

Intisari—Pendidikan memegang peranan penting terhadap perkembangan suatu negara terutama lembaga pendidikan sebagai tempat proses pendidikan mempunyai tujuan penting untuk menciptakan pendidikan yang berkualitas dalam meningkatkan kinerja siswa. Berdasarkan penelitian yang dilakukan dalam beberapa dekade terakhir kualitas pendidikan di Portugal mengalami peningkatan, namun secara statistik menunjukkan bahwa tingkat kegagalan siswa di Portugal tinggi khususnya pada bidang studi Matematika dan Bahasa Portugis. Disisi lain, *machine learning* yang merupakan bagian dari Kecerdasan Buatan dinilai dapat membantu dalam bidang pendidikan salah satunya dalam memprediksi kinerja siswa. Namun pengukur kinerja siswa menjadi sebuah tantangan sejak kinerja siswa memiliki beberapa faktor salah satunya hubungan variabel dan faktor untuk memprediksi kinerja berpartisipasi dengan cara tidak berurut. Penelitian ini bertujuan untuk mengetahui bagaimana penerapan algoritma *machine learning* berbasis *particle sworm optimization* untuk memprediksi kinerja siswa. Dengan menggunakan metode penelitian eksperimen dan hasil studi empiris ditunjukkan pada masing-masing model yaitu *random forest*, *decision tree*, *support vector machine* dan *neural network* berbasis *particle swarm optimization* dapat meningkatkan akurasi prediksi kinerja siswa.

Kata Kunci: *Kinerja Siswa, Machine Learning, Particle Swarm Optimization, Prediksi.*

INTRODUCTION

Education plays an important role in the development of a country in improving standards and ensuring the continuity of the life of an intelligent and cultured nation [1]. Meanwhile, the implementation of education is a shared responsibility, one of which plays an educational

role. Educational institutions as the place where the educational process takes place have an important goal to create quality education in improving student performance [2]. Student performance is an illustration of the ability of students to complete specific tasks given based on predetermined criteria [3]. Based on research by Paulo Cortez and Alice Silva that although the last few decades the quality of education in Portugal has improved, statistically it shows that the failure rate of students in Portugal is high. In 2006 the initial starting school level in Portugal was 40% for ages 18-24 years, while the European Union had an average grade of only 15%. Specifically, student failure is caused by a lack of student success in completing several fields of study, namely Mathematics and Portuguese [4].

In education predicting student performance becomes an important research topic that uses machine learning to explore data from the field of education [5]. Machine Learning is a computer science that studies how computers or machines have intelligence. Machine Learning also functions to optimize criteria or groupings by using data or experience [6]. On the other hand, Machine Learning which is part of Artificial Intelligence which aims to extract knowledge from data offers interesting automated tools that can analyze raw data for decision-makers who can help in the field of education, one of which is in predicting student performance [7]. Accurately predicting student performance in the early stages of learning helps in identifying weak students and enables management to take corrective actions to prevent student failure [8].

Prediction of student performance in higher education has been identified as one of the important research topics in machine learning because it is considered to be able to analyze student academic behavior effectively and estimate its performance [9]. However, measuring student performance becomes a challenge since student performance depends on several factors, the relationship of variables and factors to predict the performance of participating in complicated non-sequential ways [5]. Information on student learning progress is also one of the factors of an educator's assessment in analyzing student performance, but the method is considered insufficient as an indicator of students and educators to make improvements in teaching and learning [10].

The main problem that existed in the previous research was that the value was still not optimal because there was no addition of existing attributes so that the student's assessment performance was still lacking, therefore the renewal of this study will improve these attributes with Particle Swarm Optimization Method. PSO is a population-based looping algorithm with lots of randomly initialized particles that can solve the optimization problem [11]. Particles will represent each candidate's solution and move towards an optimal position through search space that is adjusted for historical behavior [11]. The research literature used is as follows:

Table 1. Research Literature

	Research Problem (RP)	Literatur Supports
RP1	Studen performance problems are an indicator of the assessment of accreditation results	Too many inactive students will influence the accreditation assessment [12]. Student performance problems are very complex because they can affect academic results [13].
RP2	The condition of student performance that is too early and unclear makes learning methods weak and slow	The unclear state of student performance results in slow and weak learning methods [14]. Student performance cannot be investigated early, causing delays in student graduation [15].
RP3	The number of factors and large data becomes the measurement of student performance	Large data sources are a difficult task for institutions to measure student performance [16]. Internal and external factors that can affect student performance in graduation [17].

Based on the above background, this study aims to determine how the application of machine learning algorithms in predicting PSO-based student performance using research methods in the form of experiments. To achieve this goal, it is necessary to do what factors affect student performance [18].

MATERIALS AND METHODS

In this study, the methodology used to develop predictive models using data mining is implemented following the CRISP-DM (Cross Industry Standard Process for Data Mining) model, where the process involves transforming business problems that predict student performance into

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1695

data mining problems. Furthermore, it involves data analysis including the collection and introduction of raw data and data preparation. Then, data modeling involving several prediction algorithms was developed, including Decision Tree, Random Forest, Neural Network, and Support Vector Machine. After the model is developed, the final step is to evaluate and test the model [13].

Figure 1. CRISP-DM Model

1. Business Understanding
   By utilizing existing data sources, it can be analyzed and predicted using data mining techniques whose business objective is to make a classification on a machine learning algorithm based on particle swarm optimization to predict student performance.
2. Data Understanding
   The data source used is student performance which is a dataset from the UCI Machine Learning Repository. The dataset was provided by the University of Minho in Portugal, collected during the 2005-2006 school year, using reports and questionnaires from two Portuguese high schools, determined in Mathematics. During the year, students are evaluated in two periods (G1, G2) and the third period combines the two previous periods to obtain the final results.

Table 1. Attributes, Data, and Data Descriptions

No	Attribute	Data	Data Description
1	School	MS/GP	MS: Mousinho da Silveira GP: Gabriel Pereira
2	Sex	M/F	M: Male F: Female
3	Age	15-22	
4	Address	R/U	R: Rural U: Urban
5	Famsize	LE3/GT3	LE3: <=3 GT: >3
6	Pstatus	A/T	A: Apart T: Together
7	Medu	0/1/2/3/4	0: Nothing 1: Elementary School 2: Middle School 3: High School 4: Higher Education
8	Fedu	0/1/2/3/4	0: Nothing 1: Elementary School 2: Middle School 3: High School 4: Higher Education
9	Mjob	Teacher/Health/Service/ At home/Other	
10	Fjob	Teacher/Health/Service/ At home/Other	
11	Reason	Home/Reputation/ Course/Other	
12	Guardian	Mother/Father/Other	
13	Traveltime	1/2/3/4	1: <15 minutes 2: 15-30 minutes 3: 30 minutes–1 Hour 4: >1 Hour
14	Study-time	1/2/3/4	1: <2 hours 2: 2-5 hours 3: 5-10 hours 4: >10 hours
15	Failures	1/2/3/4	1: 1 time 2: 2 times 3: 3 times 4: 4 times
16	School-up	Yes/No	
17	Famsup	Yes/No	
18	Paid	Yes/No	
19	Activities	Yes/No	
20	Nursery	Yes/No	
21	Higher	Yes/No	
22	Internet	Yes/No	
23	Romantic	Yes/No	
24	Famrel	1/2/3/4/5	1: Very bad 2: Bad 3: Good 4: Great 5: Excellent
25	Freetime	1/2/3/4/5	1: Very low 2: Low 3: Normal 4: High 5: Very high
26	Goout	1/2/3/4/5	1: Very low 2: Low 3: Normal 4: High 5: Very high
27	Dalc	1/2/3/4/5	1: Very low 2: Low 3: Normal 4: High 5: Very high
28	Walc	1/2/3/4/5	1: Very low

No	Attribute	Data	Data Description
			2: Low 3: Normal 4: High 5: Very high
29	Health	1/2/3/4/5	1: Very bad 2: Bad 3: Normal 4: Good 5: Very good
30	Absences	0-75	
31	Results	Pass/Fail	

The assessment method used in Portuguese education is a 20 point rating scale where zero indicates the lowest grade and 20 is the highest grade. The student performance dataset consists of 30 and 1 class.

3. Data Preparation
   The preparation of the data aims that the data source can be applied in the modeling phase then it needs to be transformed. Models to be used are Random Forest, Decision Tree, Support Vector Machine, Neural Network.
4. Modeling
   Modeling is done by developing models that have formed classification using binary classification approach and 5-level classification with Random Forest (RF) algorithm, Decision Tree (DT), Support Vector Machine (SVM) and Neural Network (NN) and apply Particle Swarm Optimization (PSO) to improve the accuracy of the student performance prediction model.

The binary classification and 5-level classification approach with the Random Forest algorithm

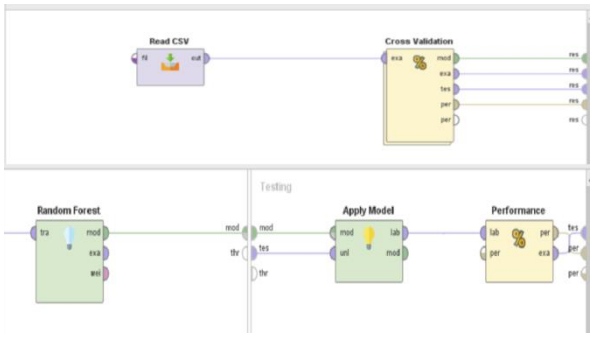


Figure 2. Testing binary classification and 5-level classification approaches with RF algorithm

From Figure 2 it can be explained that the student achievement dataset is validated and used by the Random Forest method. Binary classification approach and 5-level classification with PSO-based Random Forest algorithm

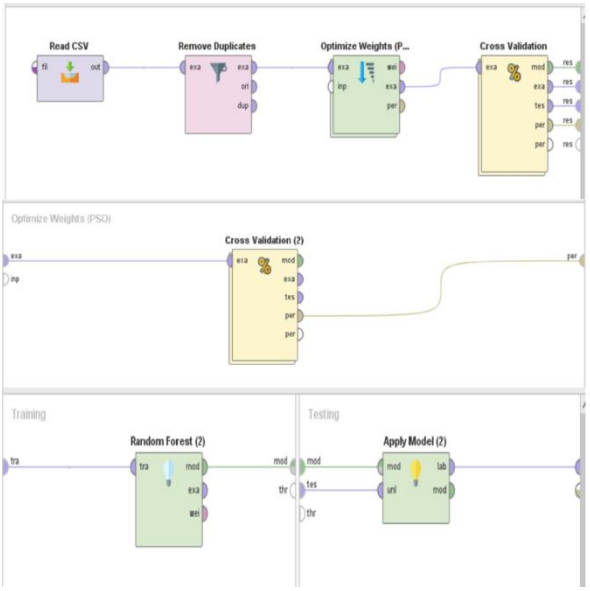


Figure 3. Testing binary classification and 5-level classification approaches with PSO-based RF algorithm

From Figure 3 it can be explained that the student performance dataset is the subject for duplicate data processing, after that the deletion process is carried out using the PSO method and validation using the Random Forest method with a binary classification approach and 5 level classification. Binary classification and 5-level classification approaches with the Decision Tree algorithm

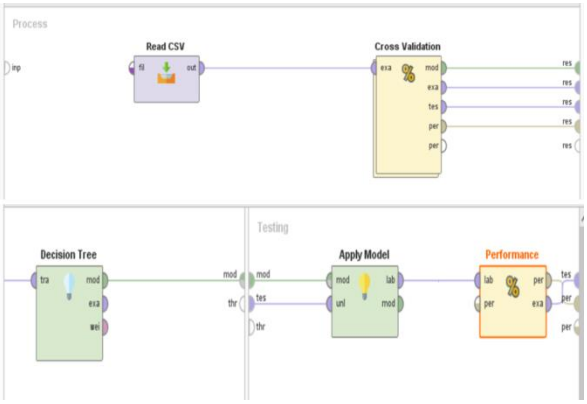


Figure 4. Testing binary classification and 5-level classification approaches with DT algorithm

From Figure 4 it can be explained that the student performance dataset is validated and uses the Decision Tree method with a binary classification approach and 5-level classification. The binary classification and 5-level classification approaches using the Particle Swarm Optimization-based Decision Tree algorithm

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1695

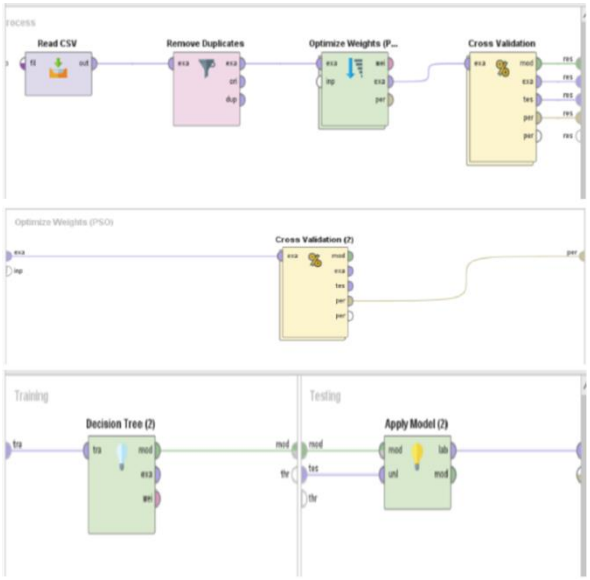


Figure 5. Testing binary classification and 5-level classification approaches with DT based PSO algorithm

From Figure 5, it can be explained that the duplicate data is removed from the student performance dataset after the deletion process is done using the PSO method and validation using the Decision Tree method with a binary classification approach and 5-level classification.

The binary classification and 5-level classification approach with the Support Vector Machine algorithm

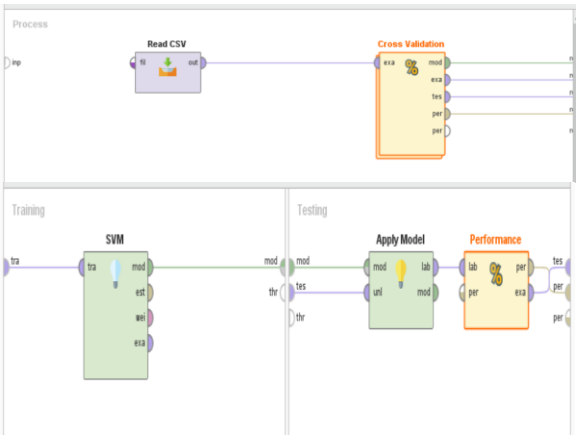


Figure 6. Testing binary classification and 5-level classification approaches with the SVM algorithm

From Figure 6 it can be explained that the student performance dataset is validated and uses the Support Vector Machine method with a binary classification approach and 5-level classification.

The binary classification and 5-level classification approaches using the Particle Swarm Optimization Support Vector Machine algorithm

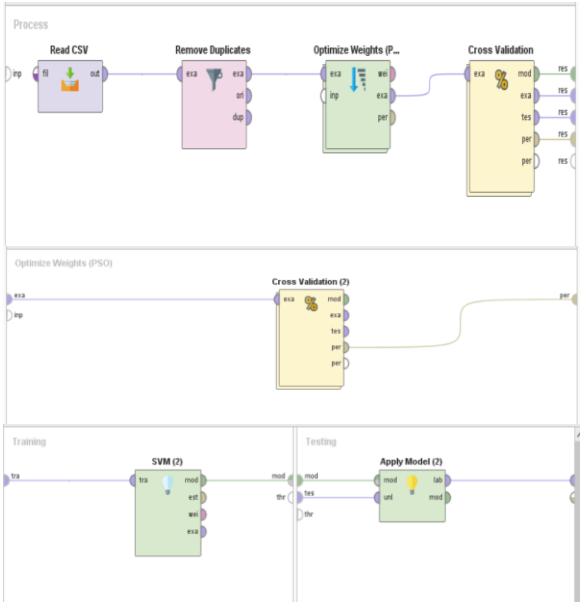


Figure 7. Testing binary classification and 5-level classification approaches with PSO-based SVM algorithm

From Figure 7 it can be explained that duplicate data is removed from the student performance dataset after the deletion process is carried out using the PSO method and validation using the Support Vector Machine method with a binary classification approach and 5 level classification. Binary classification and 5-level classification approaches with Neural Network algorithm

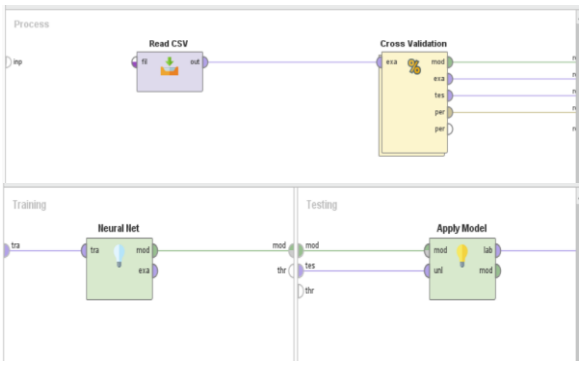


Figure 8. Testing binary classification and 5-level classification approaches with the NN algorithm

From Figure 8 it can be explained that the student performance dataset is validated and uses the Neural Network method with a binary classification approach and 5 level classification. The binary classification approach and 5-level classification with Neural Network algorithm based on Particle Swarm Optimization

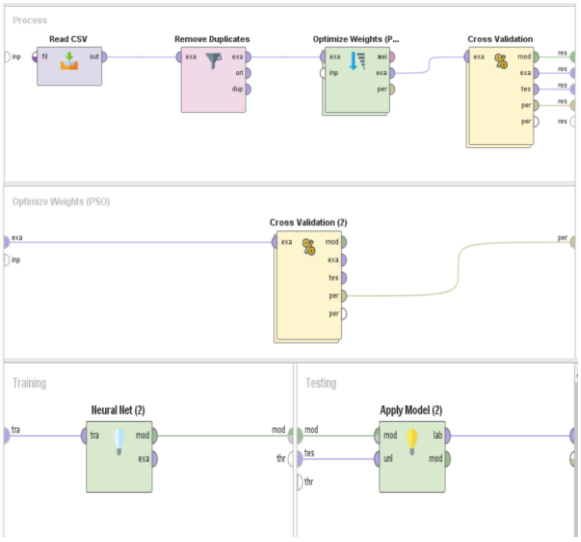


Figure 9. Testing the binary classification approach and 5-level classification with the PSO-based NN algorithm

From Figure 9, it can be explained that the duplicate data is removed from the student performance dataset after the deletion process is carried out using the PSO method and validation using the Neural Network method with a binary classification approach and 5-level classification.

RESULTS AND DISCUSSION

1. Evaluation

Evaluation is done using confusion matrix and Area Under Curve (AUC) techniques. The experimental results of testing the binary classification approach with the Random Forest (RF) algorithm based on Particle Swarm Optimization

PerformanceVector

```
PerformanceVector:
accuracy: 92.91% +/- 1.22% (micro average: 92.91%)
ConfusionMatrix:
True:   LULUS   GAGAL
LULUS:  538     35
GAGAL:  11      65
```

Figure 10. The experimental results of testing the binary classification approach with the PSO-based Random Forest (RF) algorithm

Figure 10 shows the results of the confusion matrix experiment testing the binary classification approach with the Random Forest algorithm based on Particle Swarm Optimization (PSO). Based on Figure 10, it can be seen that from 649 data, 538 predicted PASS, following predictions made by the RF method, then 35 data predicted PASS, but it turns out the prediction results are FAILED. Then 65 data were predicted to FAIL, according to predictions made by the RF method, and 11 were predicted to be FAIL but the prediction was PASS.

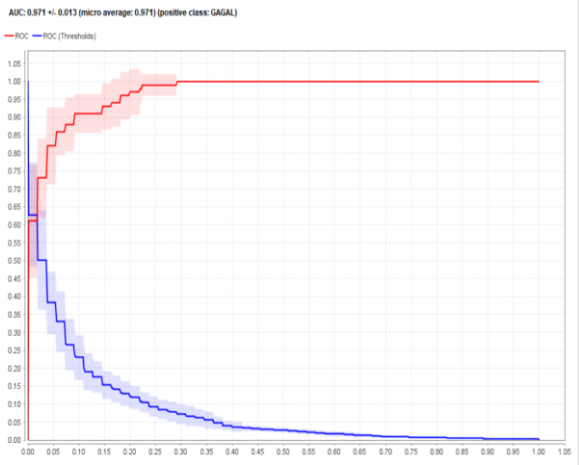


Figure 11. *Area Under Curve* (AUC) binary classification approach uses the PSO-based RF method

From Figure 11, the binary classification approach with the PSO-based Random Forest method produces an AUC of 0.971 with a very good classification. The experimental results of testing the binary classification approach with the Decision Tree (DT) algorithm based on Particle Swarm Optimization

PerformanceVector

```
PerformanceVector:
accuracy: 92.14% +/- 1.47% (micro average: 92.14%)
ConfusionMatrix:
True:   LULUS   GAGAL
LULUS:  529     31
GAGAL:  20      69
```

Figure 12. The results of the experimental testing of the binary classification approach with the PSO-based Decision Tree (DT) algorithm

Figure 12 shows the results of the confusion matrix experiment testing the binary classification approach with the Decision Tree algorithm based on Particle Swarm Optimization (PSO). Based on Figure 12, it can be seen that from 649 data, 529 predicted PASS, by predictions made by the DT method, then 31 data predicted PASS, but it turns out the prediction result is FAIL. Then 69 data were predicted to FAIL, according to predictions made by the DT method, and 20 were predicted to be FAIL but the prediction was PASS.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1695

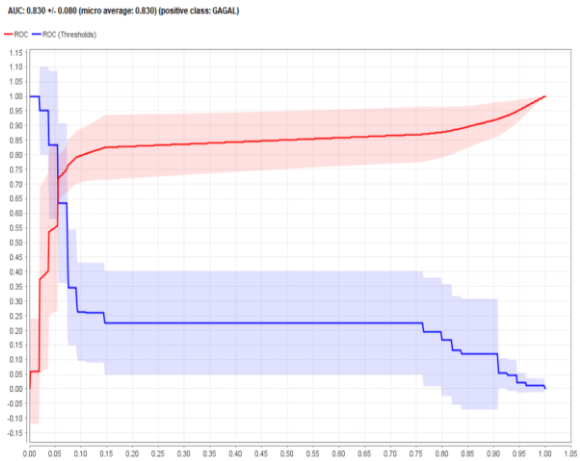


Figure 13. *Area Under Curve* (AUC) binary classification approach uses the PSO-based DT method

From Figure 13, the binary classification approach using the PSO-based Decision Tree method produces an AUC of 0.830 with a very good classification. The experimental results of testing the binary classification approach with the Support Vector Machine (SVM) algorithm based on Particle Swarm Optimization

PerformanceVector

PerformanceVector:
accuracy: 92.61% +/- 2.90% (micro average: 92.60%)
ConfusionMatrix:

True:	1	0
1:	536	35
0:	13	65

Figure 14. The experimental results of testing the binary classification approach with the PSO-based Support Vector Machine (SVM) algorithm

Figure 14 shows the results of the confusion matrix experiment that tested the binary classification approach with the Particle Swarm Optimization (PSO) algorithm based on Support Vector Machine. Based on Figure 14, it can be seen that out of 649 data, 536 data were predicted to be PASS, according to the predictions made by the SVM method, then 35 data were predicted to be PASS, but the prediction results were FAIL. Then 65 data were predicted to FAIL, according to the predictions made by the SVM method, and 13 data were predicted to FAIL but in fact, the predictions were PASS.

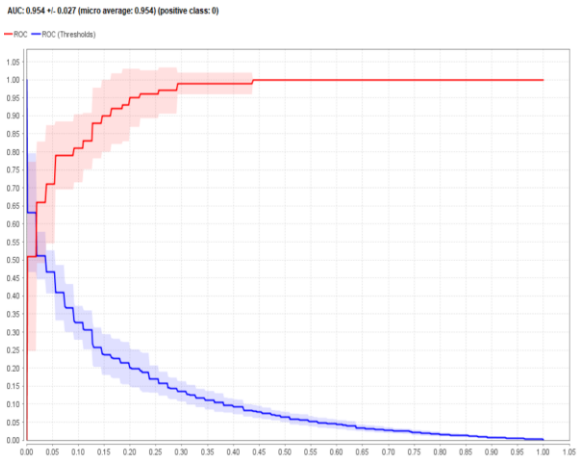


Figure 15 *Area Under Curve* (AUC) binary classification approach uses the PSO-based SVM method

From Figure 15, the binary classification approach with the PSO-based Support Vector Machine method produces an AUC of 0.954 with a very good classification. The experimental results of testing the binary classification approach with the Neural Network (NN) algorithm based on Particle Swarm Optimization

PerformanceVector

PerformanceVector:
accuracy: 89.98% +/- 2.44% (micro average: 89.98%)
ConfusionMatrix:

True:	1	0
1:	518	34
0:	31	66

Figure 16. The experimental results of testing the binary classification approach with PSO-based Neural Network (NN) algorithm

Figure 16 shows the results of the confusion matrix experiment testing the binary classification approach with the Particle Swarm Optimization (PSO) Neural Network algorithm. Based on Figure 16, it can be seen that from 649 data, 518 predicted PASS, by predictions made by the NN method, then 34 data predicted PASS, but it turns out the prediction result is FAIL. Then 66 data were predicted to FAIL, according to predictions made by the NN method, and 31 were predicted to FAIL but the prediction was PASS.

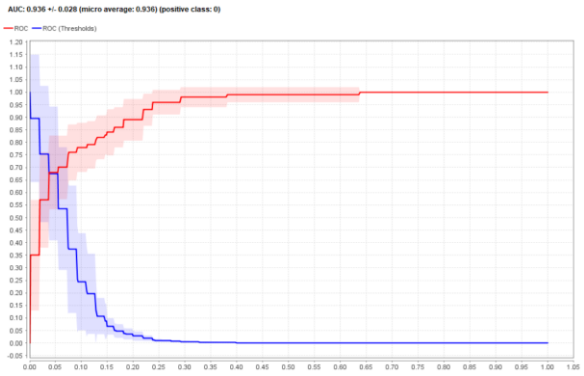


Figure 17. *Area Under Curve* (AUC) binary classification approach uses the PSO-based NN method

From Figure 17, the PSO-based Neural Network method binary classification approach produces an AUC of 0.936 with excellent classification. The experimental results of testing the 5-level classification approach with the Random Forest (RF) algorithm based on Particle Swarm Optimization

PerformanceVector

PerformanceVector:
accuracy: 75.20% +/- 6.15% (micro average: 75.19%)
ConfusionMatrix:

True:	D	C	B	A	E
D:	185	34	2	1	36
C:	3	117	40	0	1
B:	0	3	62	20	0
A:	0	0	8	61	0
E:	13	0	0	0	63

Figure 18. The experimental results of testing the 5-level classification approach with PSO-based Random Forest (RF) algorithm

Figure 18 shows the results of the confusion matrix experiment testing a 5-level classification approach with a Particle Swarm Optimization (PSO) based Random Forest algorithm.

The results of the experimental testing of the 5-level classification approach with the Decision Tree (DT) algorithm based on Particle Swarm Optimization

PerformanceVector

PerformanceVector:
accuracy: 72.26% +/- 4.99% (micro average: 72.27%)
ConfusionMatrix:

True:	D	C	B	A	E
D:	171	28	2	0	34
C:	10	111	37	0	1
B:	0	15	63	22	0
A:	0	0	10	59	0
E:	20	0	0	1	65

Figure 19. The experimental results of testing the 5-level classification approach with PSO-based Decision Tree (DT) algorithm

Figure 19 shows the results of a confusion matrix experimental testing of the 5-level classification approach with the Decision Tree algorithm based on Particle Swarm Optimization (PSO).

The results of experimental testing of the 5-level classification approach with the Particle Swarm Optimization-based Support Vector Machine (SVM) algorithm

PerformanceVector

PerformanceVector:
accuracy: 71.19% +/- 6.54% (micro average: 71.19%)
ConfusionMatrix:

True:	3	2	1	0	4
3:	177	29	6	1	46
2:	20	115	39	1	1
1:	0	10	65	28	0
0:	0	0	2	52	0
4:	4	0	0	0	53

Figure 20. The experimental results of testing the 5-level classification approach with the PSO-based Support Vector Machine (SVM) algorithm

Figure 20 shows the results of the confusion matrix experiment testing the binary classification approach with the Support Vector Machine based on Particle Swarm Optimization (PSO).

The results of experimental testing of the 5-level classification approach with the Neural Network (NN) algorithm based on Particle Swarm Optimization

PerformanceVector

PerformanceVector:
accuracy: 64.40% +/- 4.03% (micro average: 64.41%)
ConfusionMatrix:

True:	3	2	1	0	4
3:	149	35	5	1	29
2:	28	80	30	2	1
1:	1	36	56	16	0
0:	0	3	21	63	0
4:	23	0	0	0	70

Figure 21. The experimental results of testing the binary classification approach with the Neural Network (NN) algorithm based on Particle Swarm Optimization

Figure 21 shows the results of the confusion matrix experimental testing of the 5-level classification approach with the Particle Swarm Optimization (PSO) Neural Network algorithm.

2. Deployment

From the results of modeling and evaluation stages that have been obtained, it can be used to predict student performance.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1695

CONCLUSION

In this study, binary classification and 5-level classification were tested using Random Forest, Decision Tree, Support Vector Machine, and Particle Optimization-based Neural Network models using rapid tools 9.0 on student performance datasets obtained from UCI Repository. The research results tested with a model with a binary classification approach using PSO-based Random Forest obtained an accuracy of 92.91% and AUC 0.971. While testing with the binary classification approach using Decision Tree based on Particle Swarm Optimization obtained an accuracy value of 92.14% with an AUC value of 0.830. Furthermore, testing using the binary classification approach using Support Vector Machine based on Particle Swarm Optimization obtained an accuracy value of 92.61% with an AUC value of 0.954. Finally testing using the binary classification approach using Neural Network based on Particle Swarm Optimization obtained an accuracy value of 89.98% with an AUC value of 0.936. The resulting model was tested to get the accuracy value of each algorithm so that testing with a 5-level classification approach using Random Forest based on Particle Swarm Optimization obtained an accuracy value of 75.20%. While testing with the 5-level classification approach using Decision Tree based on Particle Swarm Optimization obtained an accuracy value of 72.26%. Furthermore, testing with a 5-level classification approach using Support Vector Machine based on Particle Swarm Optimization obtained an accuracy value of 71.19%. Finally, testing using the 5-level classification approach using Neural Network based on Particle Swarm Optimization obtained an accuracy value of 64.40%. Based on these results which show that the results are quite good so that it can improve the accuracy of predictions of student performance.

REFERENCE

[1] R. H. Pambudi, B. D. Setiawan, and Indriati, "Penerapan Algoritma C4.5 Dalam Program Untuk Memprediksi Kinerja Siswa Sekolah Menengah," *Look. Forward, Look. Back Draw. Past To Shape Futur. Mark.*, 2018.

[2] H. Hamsa, S. Indiradevi, and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technol.*, vol. 25, pp. 326–332, 2016.

[3] U. P. Indonesia, "Peningkatan Kinerja Siswa Melaluipendekatan Saintifik Pada Pembelajaran Ipa Terpadumodel Webbed," *J. Penelit. Pendidik.*, vol. 17, no. 1, 2017.

[4] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in *15th European Concurrent Engineering Conference 2008, ECEC 2008 - 5th Future Business Technology Conference, FUBUTEC 2008*, 2008.

[5] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, "Predicting Students Performance in Educational Data Mining," in *Proceedings - 2015 International Symposium on Educational Technology, ISET 2015*, 2016.

[6] T. Ginting and Y. E. Rohmadi, "Machine Learning untuk Localization Berbasis RSS Menggunakan CELL-ID GSM," *Teknomatika*, vol. 7, no. 2, p. 79, 2015.

[7] A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017.

[8] M. Pandey and S. Taruna, "Towards the integration of multiple classifiers pertaining to the Student's performance prediction," *Perspect. Sci.*, vol. 8, pp. 364–366, 2016.

[9] G. Kostopoulos, A. D. Lipitakis, S. Kotsiantis, and G. Gravvanis, "Predicting student performance in distance higher education using active learning," in *Communications in Computer and Information Science*, 2017.

[10] F. Yang and F. W. B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Comput. Educ.*, vol. 123, no. October 2017, pp. 97–108, 2018.

[11] A. Rachman and S. Wasiyanti, "Pengukuran Kualitas E-Commerce Shopee Terhadap Kepuasan Pengguna," *Paradig. J. Komput. dan Inform. Univ. Bina Sarana Inform.*, vol. 21, no. 2, pp. 143–148, 2019.

[12] S. Wiyono and T. Abidin, "IMPLEMENTATION OF K-NEAREST NEIGHBOUR (KNN) ALGORITHM TO PREDICT STUDENT'S PERFORMANCE," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, 2018.

[13] W. F. W. Yaacob, S. A. M. Nasir, W. F. W. Yaacob, and N. M. Sobri, "Supervised data mining approach for predicting student performance," *Indones. J. Electr. Eng. Comput. Sci.*, 2019.

[14] M. Mohammadi, M. Dawodi, W. Tomohisa, and N. Ahmadi, "Comparative study of supervised learning algorithms for student performance prediction," in *2019*

International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 124–127.

[15] E. P. Rohmawan, "Menggunakan Metode Desicion Tree," *J. Ilm. MATRIK Vol.20 No.1, April 201821-30*, pp. 21–30, 2018.

[16] N. V. Krishna Rao, N. Mangathayaru, and M. Sreenivasa Rao, "Evolution and prediction of a radical multi-dimensional e-learning system with cluster-based data mining techniques," *Proc. - Int. Conf. Trends Electron. Informatics, ICEI 2017*, vol. 2018-Janua, pp. 701–707, 2018.

[17] A. Rakhman, "Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimation (PSO)," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 6, no. 1, pp. 193–197, 2017.

[18] Z. Ibrahim and D. Rusli, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision tree, And Linear Regression," *Proc. 21st Annu. SAS Malaysia Forum*, 2007.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1664

DEVELOPMENT DASHBOARD HR LEARNING BASED OLAP IN LIFE INSURANCE COMPANY

William Frado Pattipeilohy[1]; Sanwani [2*]; Ade Priyatna[3]

Program Study Teknologi Informasi[1]
Universitas Budi Luhur
budiluhur.ac.id
william.frado@budiluhur.ac.id

Sistem Informasi[2]
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
nusamandiri.ac.id
sanwani.swq@nusamandiri.ac.id [2*], ade.aqe@nusamandiri.ac.id [3]
(*) Corresponding Author

Abstract— *Currently in the business world, companies have the challenge to survive and even thrive in the face of business changes that are so fast and flexible both in terms of markets, consumers, products to competitors. Companies are required to be able to take quick and appropriate decisions in the face of all these business changes. Implementation of Business Intelligence in a company becomes a real solution to facing all those challenges. Whereby using Business Intelligence, management within a company can make appropriate decisions based on valid data that has been processed into knowledge and presented from various perspectives needed. This research developed an integrated system in the form of an OLAP-based dashboard to present data reports, especially in the Human Resources Division. Various data in the excel file upload are automatically generated by ETL to the database using SSIS and consolidated into a data warehouse. Processing using SSAS, which is displayed in the form of a dashboard using Reporting Services with a more interesting summary form. With the development of the HR Learning and Development dashboard, top-level management companies get reports for quick, precise, and accurate decision making.*

Keywords: *Human Resources, HRIS, Business Intelligence, OLAP, Dashboard*

Abstrak— *Dalam dunia bisnis sekarang, perusahaan mempunyai tantangan untuk tetap bisa bertahan bahkan berkembang dalam menghadapi perubahan bisnis yang begitu cepat dan flexible baik dari segi pasar, konsumen, produk hingga kompetitor. Perusahaan dituntut bisa mengambil keputusan cepat dan tepat dalam menghadapi segala perubahan bisnis tersebut. Implementasi Business Intelligence pada suatu perusahaan menjadi solusi yang nyata dalam menghadapi semua tantangan itu. Dimana dengan menggunakan Business Intelligence, management di dalam suatu perusahaan dapat mengambil keputusan yang tepat berdasarkan data valid yang sudah diolah menjadi pengetahuan dan disajikan dari berbagai sudut pandang yang dibutuhkan. Penelitian ini dikembangkan sistem terintegrasi dalam bentuk dashboard berbasis OLAP untuk menyajikan laporan data khususnya pada Divisi Sumber Daya Manusia. Berbagai data dalam unggahan file excel yang otomatis tergenerate oleh ETL ke basis data menggunakan SSIS dan dikonsolidasikan menjadi data warehouse. Pengolahan menggunakan SSAS yang hasilkan tampilkan dalam bentuk dashboard menggunakan Reporting Service dengan bentuk summary yang lebih menarik. Adanya pengembangan dashboard HR Learning and Development ini, para top level management perusahaan mendapatkan laporan untuk pengambilan keputusan secara cepat, tepat dan akurat.*

Kata Kunci: *Sumber Daya Manusia, HRIS, Business Intelligence, OLAP, Dashboard*

INTRODUCTION

The development of information technology has made business competition even tighter. The development of information technology also makes the resulting data increase. The Statista.com site predicts that by 2024 the volume of data and information generated in the world of information generated from data and technology will be an important factor for the development of a company. With the information, the company can improve its performance in making accurate decisions with the minimum possible risk[1] to maximize the company's profits and progress [2]. Abundant data will be useless if there is no good

management and analysis of business activities is a non-negotiable need.

The speed of data processing and information retrieval affects the speed in deciding on the company[3]. The speed and accuracy in making decisions about a problem or topic in the company make the company able to cope with and even avoid events that can cause losses to the company[4]. Herein lies the importance of Business Intelligence (BI).

In life assurance companies currently, the decision-making process is still using conventional methods, namely processing data using the help of an excel application so that it takes longer to view reports or dashboards [5]. Companies need a fast and real-time decision-making application to assist in the decision-making process so that companies can determine future strategies [6] for business progress in the insurance sector.

There are several options for dashboard development, one of which is the Business Intelligence (BI) application as a concept and method of how to improve the quality of business decision making based on data-based systems [7]. BI is often equated with the terms briefing books, reports and query tools, and executive information systems. BI is a data-based decision support system [8]. Apart from Business Intelligence, there are also web-based development applications combined with programming using python[9]. From the choice of the two applications, Business Intelligence (BI) was selected with the consideration that there is already a complete platform for the dashboard tool to make it easier in the floating process.

The purpose of research on developing HR learning dashboards using business intelligence in life insurance companies is to assist companies in the decision-making process generated from the BI (business intelligence) application dashboard so that companies can make decisions quickly and in real-time[10] so that the company can determine business strategies for progress life insurance business.

MATERIALS AND METHODS

In this study, there are several stages in the development of business intelligence applications, including analysis, system development, testing, and system implementation. In the analysis phase Based on business needs, the Human Resources (HR) workflow for implementing Business Intelligence has been proposed to be developed to meet the needs of the Human Resources (HR) team. This research stage begins with an analysis of existing or running systems[11]. The system applied to the HR division still uses a manual system because they have to query the system database and then convert it through Microsoft Excel and then submit it to the TOP level of management[12]. Top-level management here is the decision-making stakeholders. After that, a requirements specification analysis will be carried out which will produce several designs, namely conceptual design, logic design, and physical design[13].

The process begins at the beginning of the data source obtained from an excel file with .xlsx format obtained from the human resources (HR) database, using the multidimensional analysis method on OLAP[14]. OLAP can help query quickly, easily, and efficiently and supports business intelligence. The purpose of OLAP, among others, is to make it easier for stakeholders to make decisions based on existing transactional data because[13] OLAP will display data from various dimensions or sides (multidimensional).

The system development process begins with creating a data warehouse. The first step is to create a staging database, a flat database where it is needed to create a fact table & dimension table[15]. What is meant by a fact table is information containing the size or something that can be calculated, while the dimension table is in the form of information that supports the fact table. After that, the cube and dimensions are made. Cube is a multidimensional data set that allows it to be presented quickly[16]. Cube is the core of the OLAP (Online Analytical Processing) concept in Analysis Services and the data used is HR data in this life insurance company.

At this stage, it is carried out after the creation of the data warehouse is complete[13]. This is done to test whether the basic OLAP operation is implemented successfully according to the output of user data. The implementation process will be carried out after all system development processes are complete then testing.

RESULTS AND DISCUSSION

The high-level process architecture structure and data warehouse in this study can be seen in Figure 1 describe the process architecture and data warehouse.

Figure 1 describes High-level architecture with the implementation of the BI system, this dependence can be overcome because top-level management can immediately get human resource reports directly, either by looking at the application portal or by receiving an automated email. The database used for this BI System is sourced from a data warehouse where the data is obtained from an operational database (OLTP) which is integrated through the ETL (Extract,

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1664

Transformation, Loading) process using SQL Server Integration Services and run for a certain period with automating the job.

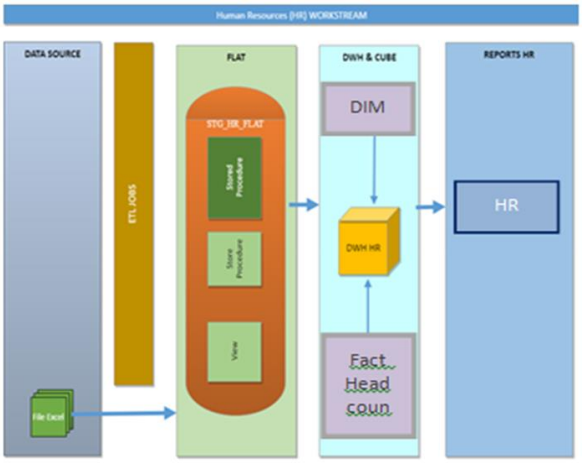


Figure 1. *High-Level* Architecture Process and Data warehouse

The first step in designing a data warehouse is designing dimension tables and fact tables which will later be analyzed using the BI system. The first stage is to create dimensions. The data warehouse is designed to consist of several dimension tables that are used to provide the measurement context (subject).

The second stage is to create a fact table that represents a business process, namely a business process model as an artifact in the data model. This table contains "measurement" elements or metrics or facts on a business process. In the data warehouse that is built, there is a fact table for employee data, namely fact_headcount, fact_movement, and fact_mpp which consists of several attributes. In this fact table, there is a foreign key for the dimension tables that have been designed.

After the fact table creation is complete, then create a star schema in the star schema containing ten dimension tables and three fact tables. The Fact Table contains the foreign keys of each dimension table. In the star schema, one dimension is represented by one table and each table is represented by several attributes

The next step is making a cube or data cube. Cube is grouping data into dimensional cubes to make data analysis easier. At this stage, a database is established where the data is stored in the form of cells, and the position of these cells is determined by several variables called dimensions. The result of this stage is a data warehouse visualization that has been made previously.

Cube creation is done with the help of Microsoft SQL Service's analysis tools. Before creating a dashboard we must create a mapping view.

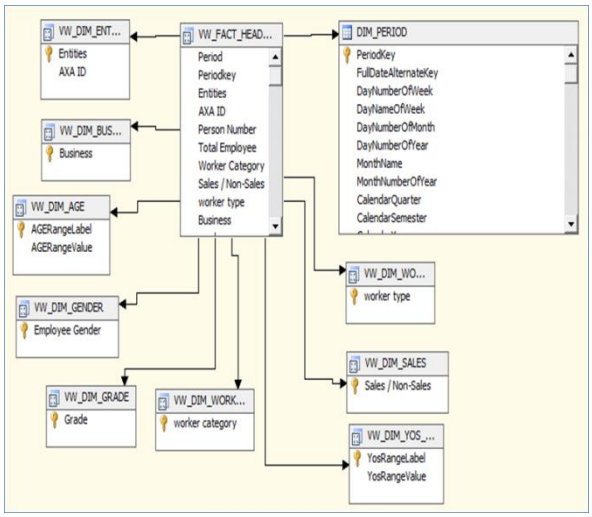


Figure 2. Star Schema about Table

Figure 2 describes stars schema mapping this used for relation view in SQL database. Mapping relation can create manual by key field or automatic with the primary key in database SQL server.

Table 1 Mapping Report Talent Acquisition

KPI	Dashboard Status
Total Position	Sheet TA Select Month Column T = Current Month And Entity Column M = 'Entity Selected by User' Group by Dashboard Status Column U Count (No Request Column A) Per Dashboard Status Sum(No Request Column A) All Dashboard Status
KPI	**SLA**
Total Hired	Sheet TA Select Month Column T = Current Month And Entity Column M = 'Entity Selected by User' And Dashboard Status Column U = 'Hired' Group by SLA Column V Count (No Request Column A) Per SLA Column V

Table 1 describes the Mapping report talent acquisition used to calculate an aggregation value. The report has calculated to summary KPI Total Position and Total Hired.

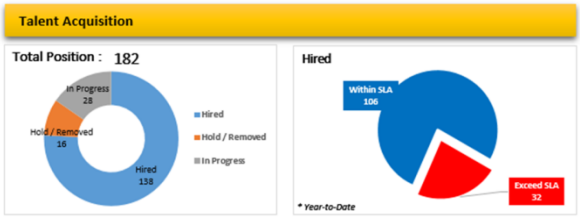


Figure 3. Report Talent Acquisition

Figure 3 describes the talent acquisition report to display the total position and total hired per dashboard status in the current month. The report shows, a summary status description in progress, hold and hired. In the display pie chart Hired describe within SLA and exceed SLA with value and present period year to date.

Table 2 Mapping data to KPI and Dimension report employee by age

KPI	Male	Female
Headcount	Sheet Headcount Select Date (Column A) = 'Current Month' and Worker Type (Column Y) = 'FTE' and Sales / Non Sales (Column V) = 'Non Sales' and Adcomm / Mancom (Column AO) = 'Adcomm' and Employee Gender (Column AE) = 'Male' and Entities (Column AG) = 'Entity Selected by User' Count(Total Record)	Sheet Headcount Select Date (Column A) = 'Current Month' and Worker Type (Column Y) = 'FTE' and Sales / Non Sales (Column V) = 'Non Sales' and Adcomm / Mancom (Column AO) = 'Adcomm' and Employee Gender (Column AE) = 'Female' and Entities (Column AG) = 'Entity Selected by User' Count(Total Record)

Table 2 describes mapping data to KPI and dimension this mapping used report Employee by age. To create a business intelligence dashboard report, you need a report mapping data mapping to KPI and Dimensions. Mapping data KPI in there any calculation and query data from database SQL server for generating data with condition Male a female

This stage is a continuation of the system design stage. The result of this implementation is a system that is ready to be tested and used. Implementing a BI system requires hardware and software. For hardware, the need is a server unit for the system and database. At this stage, several reports will be displayed on the results of the development of the business intelligence system, including reports, namely Report Employee by age

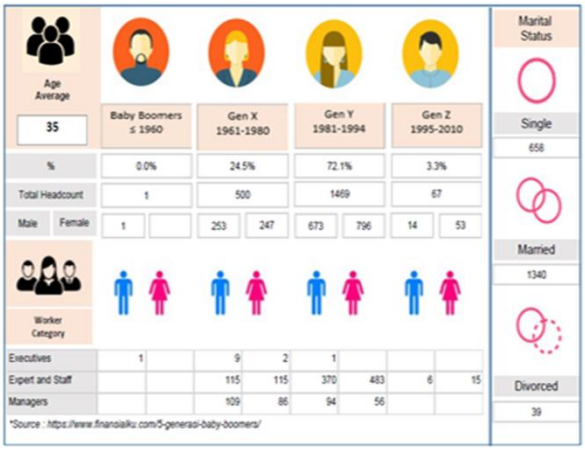


Figure 4. Report Employee by Age

Figure 4 describe report employee by age the report displays data on all total employees based on age and generation. Also displays the total number of male and female employees, the employee's marital status, and the total number of employees based on their position in the company.

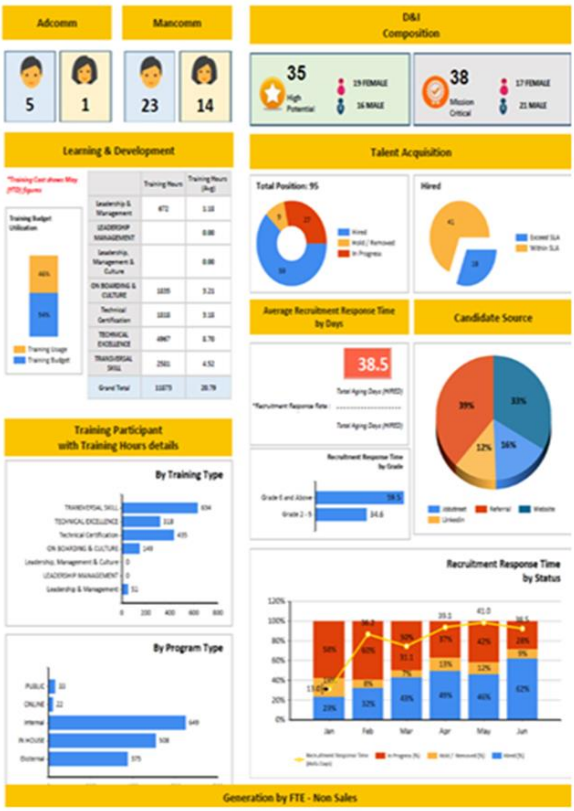


Figure 5. Report Training Employee

Figure 5 describes the employee training report displays data on all reports of employees who participated and training as well as training materials that were attended. Also displays the percentage of employees who have attended training and who have not attended the training.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1664

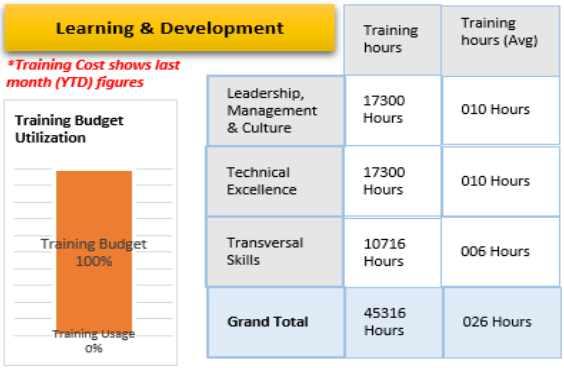


Figure 6. Report Learning and Development

Figure 6 describes the Report learning and development of the total training budget and training hours per type of training in the current year. The component training budget is Leadership, technical excellence, and transversal skill.

Table 3 Table Mapping Report Learning and Development

KPI	YTD
Training Hours	Sheet L&D Training Hours Select Date Column A from First Month until 'Current Month' and Worker Status column V = 'FTE' and Entities column E = 'Entity Selected by User' group by Training Type (Column J) Count(Training Hours Column N) per Training Type (Column J)
Headcount	Sheet Headcount Select Date Column A from First Month until 'Current Month' and Worker Type column Y = 'FTE' and Entities column E = 'Entity Selected by User' Count(Person Number)
% Avg Headcount	Headcount YTD / Total Month
% Avg Training Hours	Training Hours / Avg Headcount
Budget	Sheet Budget Select Entities column A = 'Entity Selected by User' Sum (Budget)
Total Training Cost	Sheet L&D Training Cost Select Month to use Column A from First Month until 'Current Month' and Entity column B = 'Entity Selected by User' Sum (IDR Column I)
% Training Budget	(Budget – Total Training Cost) / Budget
% Training Usage	Total Training Cost / Budget

Table 3 describes mapping report learning and development used to calculate field-training hours, headcount, average headcount, presents training, budget, and training cost.

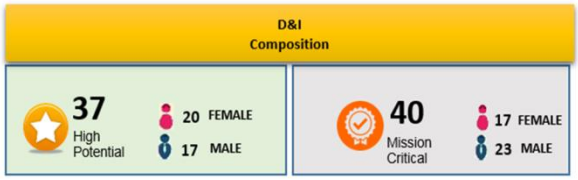


Figure 7. Report D & I Composition

Figure 7 describes D & I Composition report, this report shows how many high potencies and important tasks for every man and woman this month.

Table 4 Data Mapping D & I Composition

KPI	High Potential	Mission Critical
Headcount D&I (Male)	Sheet D&I Select Date (Column A) = 'Current Month' And D&I Composiion (Column B) = 'High Potential' and Entity (Column C) = 'Entity Selected by User' Get (Male (Column D))	Sheet D&I Select Date (Column A) = 'Current Month' And D&I Composiion (Column B) = 'Mission Critical' and Entity (Column C) = 'Entity Selected by User' Get (Male (Column D))
Headcount D&I (Female)	Sheet D&I Select Date (Column A) = 'Current Month' And D&I Composiion (Column B) = 'High Potential' and Entity (Column C) = 'Entity Selected by User' Get (Female (Column E))	Sheet D&I Select Date (Column A) = 'Current Month' And D&I Composiion (Column B) = 'Mission Critical' and Entity (Column C) = 'Entity Selected by User' Get (Female (Column E))

Table 4 describe data mapping D & I Composition for report D & I Composition this mapping is used to calculate field headcount D&I male and headcount D&I female.

CONCLUSION

From the results of research on dashboard development using OLAP, the results obtained for each analysis report, namely top-level management or decision-making stakeholders can enter the portal system and view the report dashboard, Presentation of reports from Human Resources data, no need to ask the MIS team. To send reports via email or hardcopy to top-level management, the implication of this Business Intelligence system has proven to make it easier for top-level management or decision-making stakeholders to be able to see employee needs based on existing KPIs because data is presented in a more attractive form, namely graphics and interfaces which is user friendly, by looking at employee data in a top-level management company can quickly make decisions which can later help determine the business

direction, make business strategies and make decisions to increase company needs, After going through the user acceptance phase e testing which shows the similarity with user data based on existing summary reports and KPIs as well as carrying out the implementation verification testing process where it shows that the system successfully displays data according to the data required by the user. Suggestion In developing the next report it is expected not only for HR needs but for all department I of the Life insurance company so that in the future all analysis of company activities can be monitored from one report portal that has been integrated with realtime data.

REFERENCE

[1] Z. Nur and I. Mukhlash, "Implementasi Business Intelligence Pada Manajemen Report Bank XYZ," *J. Sains Dan Senni Pomits*, vol. 3, no. 3, pp. A16–A21, 2014.

[2] W. S. Prasetya, Kusrini, and H. Al Fatta, "Sistem Pendukung Keputusan Pemilihan Produk Asuransi Studi Kasus : PT Commonwealth Life Pontianak Decision Support System Insurance Product Selection Case Study : PT Commonwealth Life Pontianak," *J. STMIK AMIKOM*, vol. 8, no. 1, pp. 105–116, 2018.

[3] A. Zikri, J. Adrian, A. Soniawan, R. Azim, R. Dinur, and R. Akbar, "Implementasi Business Intelligence untuk Menganalisis Data Persalinan Anak di Klinik Ani Padang dengan Menggunakan Aplikasi Tableau Public," *J. Online Inform.*, vol. 2, no. 1, pp. 20–24, 2017.

[4] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making*. 2009.

[5] N. S. Ambarwati, G. A. Yuniarta, and N. K. Sinarwati, "Pengaruh Modal Kerja, Likuiditas, Aktivitas dan Ukuran Perusahaan Terhadap Profitabilitas Pada Perusahaan Manufaktur yang Terdaftar di Bursa Efek Indonesia," *e-journal S1 Ak Univ. Pendidik. Ganesha*, vol. 3, no. 1, 2015.

[6] G. Ade and Okfalisa, "Sistem Pendukung Keputusan Untuk Produk Asuransi Jiwa Bagi Nasabah Menggunakan Metode SMARTER," *J. Sains, Teknol. dan Ind.*, vol. 12, no. 1, pp. 73–79, 2014.

[7] Kusnawi, "Implementasi OLAP dan Reporting Services sebagai Bagian Proses Business Intelligence," *Bisnis Intelejen*, vol. 12, no. 2, pp. 43–48, 2010.

[8] E. Turban, J. E. Aronson, and T.-P. Liang, "Decision Support Systems and Business Intelligence," *Decis. Support Bus. Intell. Syst. 7/E*, pp. 1–35, 2007.

[9] H. Kurniawan, B. Setiyono, and R. R. Isnanto, "Aplikasi Penjawab Pesan Singkat Automatis Dengan Bahasa Python," *Dr. Diss.*, 2011.

[10] T. Kusnandar, "Penerapan Crm Dengan Sistem Informasi Berbasis Web Untuk Kepuasan Pelanggan," *J. Comput. BISNIS*, vol. 5, no. 1, pp. 6–13, 2011.

[11] S. Kasus, D. Pt, I. Adi, and P. Semarang, "Implementasi Sistem Business Intelligence Untuk Melakukan Analisis Data Guna Mendukung Pembuatan Keputusan Manajer," *Techno.COM*, vol. XIV, no. 02, pp. 83–92, 2008.

[12] A. Priyatna, "Implementasi Sistem Penunjang Keputusan Menggunakan Business Intelligence Untuk Umkm Di Gunung Putri Kab.Bogor," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 7–12, 2019.

[13] L. Hasim, "Penerapan Business Intelligence Pada Manajemen Dashboard Report Perusahaan Asuransi," *Syntax Lit. ; J. Ilm. Indones.*, vol. 4, no. 12, pp. 123–136, 2019.

[14] N. Ardista, P. Purbandini, and T. Taufik, "Rancang Bangun Data Warehouse Untuk Pembuatan Laporan dan Analisis pada Data Kunjungan Pasien Rawat Jalan Rumah Sakit Universitas Airlangga Berbasis Online Analytical Processing (OLAP)," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 1, pp. 40–51, 2017.

[15] A. Dwiputra Wijaya and T. Gantini, "Analisis Forecasting dengan Implementasi Dashboard Business Intelligence Untuk Data Penjualan Pada PT. 'X,'" *Prof. Drg. Surya Sumantri*, vol. 1, no. 2, pp. 457–470, 2019.

[16] Maimunah, S. F. Saadah, and R. Andrian, "Rancang Bangun Aplikasi Data Warehouse Untuk Business Intelligence," *CSRID J.*, vol. 5, no. 1, pp. 1–10, 2012.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1740

DETERMINATION OF PERMANENT LECTURERS IN IBM ASMI INFORMATION SYSTEM PRODUCT WITH SAW AND ARAS METHOD

Rendi Septian[1*]; Istiqal Hadi[2]; Ridwansyah[3]; Windu Gata[4]; Widiastuti[5]; Muhammad Iqbal[6]

Computer Science Study Program[1,2,4,5,6], Informatics Engineering Study Program[3]
STMIK Nusa Mandiri[1,2,3,4,5,6]
www.nusamandiri.ac.id[1,2,3,4,5,6]
14002348@nusamandiri.ac.id [1*]; 14002343@nusamandiri.ac.id [2]; ridwansyah.rid@nusamandiri.ac.id [3]; windu@nusamandiri.ac.id [4]; 14002380@nusamandiri.ac.id [5]; 14002390@nusamandiri.ac.id [6]

(*) Corresponding Author

Abstract— *Determining the quality standards of lecturers refers to the criteria of education, research, and community service. The campus can carry out the first process for selecting permanent lecturers effectively by looking at several criteria. By using a Decision Support System (DSS), the four categories can be used as guidelines for decision-makers to choose permanent lecturers. The goal of writing this journal is to support the effectiveness of the time for decision-makers to choose permanent lecturers in the early stages by combining the Additive Ratio Assessment (ARAS) and Simple Additive Weighting (SAW) methods. Where the SAW method has the advantage of accurate assessment because the value of criteria and weights have been determined, while the ARAS method compares each criterion value to each optimal alternative as a whole to get an ideal alternative. The result of combining the two methods can describe the prospective lecturers who are suitable to be used as permanent lecturer criteria. Judging from the ranking results in calculations, the values obtained are 0.146341, 0.134146, and 0.121951. These results prove that ranking with an assessment using the combination of SAW and ARAS methods results in an effective, accurate, and efficient assessment.*

Keywords: *Additive Ratio Assessment, Decision Support Systems, Permanent lecturer, Simple Additive Weighting.*

Abstrak— *Menentukan standar kualitas dosen mengacu kepada kriteria pendidikan, penelitian, dan pengabdian kepada masyarakat. Pihak kampus dapat melakukan proses pertama untuk pemilihan dosen tetap secara efektif dengan melihat beberapa kriteria. Dengan menggunakan Sistem Penunjang Keputusan (SPK), keempat kategori tersebut bisa dijadikan pedoman bagi pengambil keputusan untuk memilih dosen tetap. Tujuan penulisan jurnal ini adalah untuk membantu keefektifan waktu bagi pengambil keputusan untuk memilih dosen tetap tahap awal dengan penggabungan metodeAdditive Ratio Assessment (ARAS) dan Simple Additive Weighting (SAW).Dimana metode SAW mempunyai keunggulan penilaian akurat karena untuk nilai kriteria dan bobot telah ditentukan, sementara metode ARASmelakukan perbandingan setiap nilai kriteria terhadap masing alternatif optimal secara keseluruhan untuk mendapatkan alternatif yang ideal. Hasil penggabungan dua metode tersebut dapat menggambarkan calon dosen yang sesuai untuk dijadikan kriteria dosen tetap.Dilihat dari hasil perangkingan dalam perhitungan, nilai yang didapat 0,146341 ,0,134146 dan 0,121951. Hasil ini membuktikan bahwa perangkingan dengan penilaian menggunakan penggabungan metode SAW dan ARAS menghasilkan penilaian yang efektif, akurat dan efisien..*

Kata Kunci: *Additive Ratio Assessment, Sistem Penunjang Keputusan, Dosen Tetap, Simple Additive Weighting.*

INTRODUCTION

Education is a learning process activity carried out by a nation and a country. Educational instruments are in an important spotlight because they support quality in an educational institution. The quality supporting instruments include teaching staff, one of which is the lecturer. Lecturers are professional educators assigned to disseminate knowledge and technology as well as community service [1].

However, in determining lecturers, of course, there are several factors to be an assessment including the involvement of lecturers in the teaching, research, and community service process. The main problem in previous research was that decision making was still done manually so that it was difficult to conduct research and evaluate data to consider the final result as a

decision making. By doing this manually, it is feared that there will be decisions that are not right on target. For that, we need a method of decision support systems that can help in making decisions that are right on target and accurate.

Related research conducted by Nadeak Application of the ARAS Method to assess the best teachers, that with the Additive Ratio Assessment (ARAS) Method, it is hoped that this decision-making will be able to select teacher performance effectively. The decision-making criteria are expected to be able to choose teacher performance effectively [2]. Previous related research was also carried out by Guterres, namely by implementing SAW and TOPSIS so that it could produce an output in the form of priority provision of livable housing assistance to the poor, especially those in East Kupang District, Kupang Regency [3]. Research conducted by Rahmat et al, shows the SAW and TOPSIS methods are two methods that can be applied in finding new locations because the SAW and TOPSIS methods can produce a better decision support system than using one of them [4]. Further research conducted by Halimah et al, used the ARAS and Shannon Entropy methods for weighting based on the weight value criteria generated from the calculation of alternative data and the ARAS method carried out a ranking process based on utility functions [5]. The next research conducted by Dadang and Sri is using the ARAS method to provide the best teacher recommendations by determining the highest ranking based on Pedagogic, Personality, Social, Professional, and responsibility [6]. The research literature used is as follows:

Table 1. Research Literature

Research Problem (RP1)		Literatur Supports
RP1	Manual determination results in not being on target in determining decisions	The work is still manual for recruitment so that it is difficult to do research and evaluate data [7].
		The assessment process in determining teacher performance is still done manually and is not yet detailed [2].
		Uncomputed processing is still done manually so that it can cause errors in determining decisions [8].
RP2	The determination process is less than optimal so it takes a lot of time in making decisions	Activities in determining decisions are still carried out randomly and are not fully computerized with the application [3].
		The work carried out is not optimal because it is not optimal in determining decisions [9].
		The process of determining the lecturer is not appropriate and takes a lot of time so that it can experience errors in making decisions [10].
RP3	Subjective determination of the assessment causes social jealousy and is not transparent in making decisions	Decisions are made based on subjectivity so that there is a concern that social jealousy towards abilities is not much different from the others [11].
		Assessment is subjective in nature based on distributing questionnaires so that it should not be used as a reference for decision making [12].
		The assessment is carried out quantitatively based on subjective selection resulting in a lack of transparency in the decision making process [13].

From the above problems, the writer in this paper uses a combination of SAW and ARAS methods [14]. Where in the alternative data process, the alternative data conversion and the alternative data normalization use the SAW method [14], then determine the normalized matrix weight, determine the value of the optimization function, and determine the highest level of the alternative using the ARAS method [14]. The purpose of this paper is to help decision-makers in universities, in this case, the ASMI Institute of Business Management, an undergraduate program majoring in information systems to find qualified lecturers to be considered as permanent lecturers. The criteria seen are academic qualifications, teaching experience, relevant to the study program, and courses that have been taught.

MATERIALS AND METHODS

a. Simple Additive Weighting

The SAW method is often also known as a weighted addition method. The basic concept of the SAW method is to find the weighted sum of the performance ratings for each alternative on all attributes. The SAW method is recommended to solve selection problems in multi-process decision-

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1740

making systems. The Simple Additive Weighting method is a method that is widely used in decision making which has many attributes. The SAW method requires a decision matrix normalization process (X) to a scale that can be compared with all existing alternative ratings. Nofriansyah in [15].

$$Rij = \begin{cases} \frac{Xij}{Max.Xij} \\ \frac{Min.ij}{Xij} \end{cases} \quad \text{.....(1)}$$

= if j is the (benefit) attribute
= if j is the (cost) attribute

r_{ij} is the normalized performance rating of the alternatives Ai for attribute Cj; i = 1,2,3,4 ..., m and j = 1,2,3,4 ..., n.

The option value for each (Vi) is determined as:

$$Vi = \sum Wjrij \quad \text{.....(2)}$$

The largest Vi value indicates that the alternative Ai is preferred.

b. Additive Ratio Assessment (ARAS)

The ARAS method is part of the DSS that is used to rank a criterion, in ranking it, this method has several steps to calculate it [7].

The steps of the Additive Ratio Assessment (ARAS) method are as follows:

1. Establishment of Decision Making Matrix

$$X = \begin{bmatrix} X01 & X0j & \dots & X0n \\ Xi1 & Xij & \dots & Xin \\ \dots & \dots & \dots & \dots \\ Xn1 & Xmj & \dots & Xmn \end{bmatrix} (i = 0, m; j = 1, n) \quad \text{.....(3)}$$

Which one
n = total criteria
m = total alternatives

X_{ij} = performance value of alternative $_i$ against criterion $_jX_{0j}$ = optimum value of criterion $_j$

If the Optimal Value of Criterion $_j$ X_{0j} is not known, then:

$$X0j = \frac{\max}{i} . Xij \text{ , if } \frac{\max}{i} . Xij \text{ is preferable} \quad \text{.....(4)}$$

$$X0j = \frac{\min}{i} . Xij \text{ , if } \frac{\min}{i} . Xij \text{ is prefable} \quad \text{.....(5)}$$

2. Normalize the decision matrix for all criteria

If the criteria are Beneficial (Benefit), normalization is carried out following:

$$X_{ij} = \frac{Xij}{\sum_{i=0}^{m} Xij} \quad \text{.....(6)}$$

Where X_{ij} is the normalized value.

If the criteria are Non-Beneficial (Cost), normalization is carried out following:

$$\text{Step 1 : } Xij^* = \frac{1}{Xij} \text{ and Step 2 : } R = \frac{Xij*}{\sum_{i=0}^{m} Xij*} \quad \text{.....(7)}$$

3. Determination of the normalized matrix weights

Where W_j Criteria Weight $_jD = [D_{ij}]$ m.n = $R_{ij}.W_{ij}$

4. Determine the value of the optimization function (S_i)

$$Si = \sum_{j=1}^{n} Dij; (i=1,2,..,m; j=1,2,..,n) \quad \text{.....(8)}$$

Where S_i is the value of the alternative optimality function i. The greatest value is the best, and the least value is the worst. Taking into account the process, the proportional relationship with the value and weight of the criterion understudy influences the final result.

5. Determines the highest ranking level of the alternatives

$$Ki = \frac{Si}{S0} \quad \text{.....(9)}$$

Where S_i and S0 are the optimal criteria values, obtained from the equation. The value, calculated on Uidan, is in the interval [0.1].

RESULTS AND DISCUSSION

In this paper, using a combination of 2 methods, namely the merger of SAW with ARAS.

The data used in the journal comes from the IBM ASMI Jakarta campus, namely 8 lecturers who teach in the undergraduate program majoring in information systems. We obtained data related to academic qualifications, academic relevance to the study program, number of years of teaching, and the suitability of courses that have been taught, which we obtained from the academic team at the IBM ASMI Jakarta campus. Here are alternative data:

Process Data with the SAW Method

1) Input Value obtained by each candidate

The process of combining SAW with ARAS can be seen in the following Figure 1.

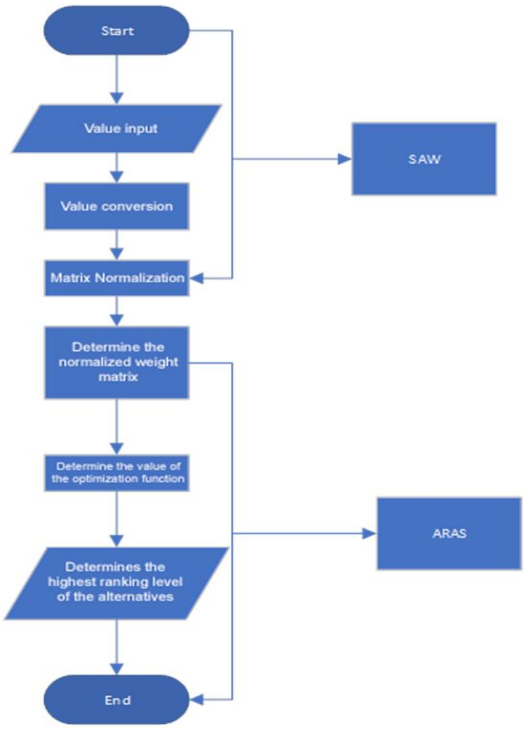


Figure 1. Flowchart of the Merger of SAW and ARAS

Table 1. Alternative Data for Permanent Lecturer Recommendations

Name	Academic Qualifications	Academic Relevance to the Study Program	Number of Teaching Years	Number of Subject Matches that have been taught
Adya Satyapuspita	S2	Not	1	1
Aristarkus Didimus Rumpak	S2	Not	16	1
Aston Freddy	S2	Not	6	1
Imam Jayadi	S2	Not	8	1
Jerry M. Logahan	S2	Relevant	16	1
Jones Zenas Rante	S2	Relevant	16	1
Rolty Glendy Wowiling	S2	Relevant	2	1
Yakub M. Saleh	S2	Not	10	1

2) Convert values based on a range of values

Table 2. Criteria Table

Criteria	Information	Type	Weight
K1	Academic Qualifications	Benefit	30%
K2	Relevance of Study Program	Benefit	25%
K3	Teaching Year	Benefit	30%
K4	Subjects taught	Benefit	15%

These numbers can be converted to crips numbers K1 = 0.3; K2 = 0.25; K3 = 0.3; K4 = 0.15 ;.

Table 3. Alternative Value Weight Table

Type of Criteria	Name	Conversion Weight
Academic Qualifications	S2	1
	S3	2
Relevance of Study Program	Not	1
	Relevant	2
Teaching Year	<5 years	1
	>=5 years	2
Subjects Taught	<5	1
	>=5	2

After the weights are determined and can be converted into crips numbers; 1 (0.4), 2 (0.6).

Table 4. Alternative Value Conversion Table

Name	Value Criteria			
	K1	K2	K3	K4
Adya Satyapuspita	0,4	0,4	0,4	0,4
Aristarkus Didimus Rumpak	0,4	0,4	0,6	0,4
Aston Freddy	0,4	0,4	0,6	0,4
Imam Jaya	0,4	0,4	0,6	0,4
Jerry M. Logahan	0,6	0,6	0,6	0,4
Jones Zenas Rante	0,4	0,6	0,6	0,4
Rolty Glendy Wowiling	0,4	0,6	0,4	0,4
Yakub M. Saleh	0,4	0,4	0,6	0,4

3) Normalization Matrix R

Based on the suitability rating table, a decision matrix can be formed, as follows:

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1740

$$Rij = \begin{cases} \frac{Xij}{Max.Xij} \\ \frac{Min.ij}{Xij} \end{cases} \quad \text{..... (7)}$$

= if j is the (benefit) attribute
= if j is the (cost) attribute

For Academic Qualification Criteria (C1) Included in the benefit attribute:

Table 5. Result of Normalization of Decision Matrix R

C1	C2	C3	C4
0,67	0,67	0,67	1
0,67	0,67	1	1
0,67	0,67	1	1
0,67	0,67	1	1
1	1	1	1
0,67	1	1	1
0,67	1	0,67	1
0,67	0,67	1	1

Process with the ARAS Method

1) Determine the normalized weight

The next process is determining the normalized weight of the matrix

The calculation above produces a matrix which can be seen in table 6:

Table 6. Results of Normalized Weights

0,20	0,20	0,20	0,30
0,20	0,20	0,30	0,30
0,20	0,20	0,30	0,30
0,20	0,20	0,30	0,30
0,30	0,30	0,30	0,30
0,20	0,30	0,30	0,30
0,20	0,30	0,20	0,30
0,20	0,20	0,30	0,30

2) The next step is to determine the value of the optimization function

Label Name	Result
S_1	0,90
S_2	1,00
S_3	1,00
S_4	1,00
S_5	1,20
S_6	1,10
S_7	1,00
S_8	1,00
Total	8,20

3) Determine the highest ranking level of the alternative

Label Name	Result
K_1	0,109756098
K_2	0,12195122
K_3	0,12195122
K_4	0,12195122
K_5	0,146341463
K_6	0,134146341
K_7	0,12195122
K_8	0,12195122

From the above calculations, the highest-ranking level is obtained, where each prospective lecturer has high to low scores.

Table 7. Highest Rank Ranking Results

Alternative	Value (K)	Ranking
Jerry M. Logahan	0,146341	1
Jones Zenas Rante	0,134146	2
Aristarkus Didimus Rumpak	0,121951	3
Aston Freddy	0,121951	4
Imam Jayadi	0,121951	5
Rolty Glendy Wowiling	0,121951	6
Yakub M. Saleh	0,121951	7
Adya Satyapuspita	0,109756	8

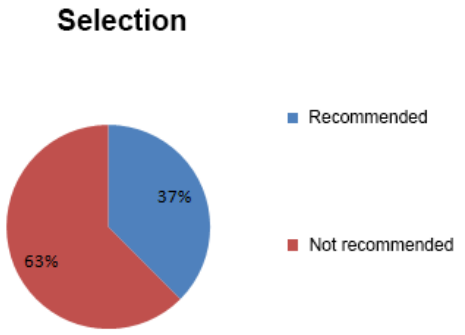


Figure 2. Assessment Criteria Graph

From the calculation above, the results show that Jerry M. Logahan is the most recommended to become a permanent lecturer with a score of 0.146341. In the second and third positions, for SAW, Brother Jones Zenas Rante and Brother Aristarkus Didimus Rumpak with each score of 0.134146 and 0.121951. The score for the first rank has a big difference with the second rank because of different academic qualifications, Brother Jerry M. Logahan has an S3 academic qualification while the second rank is an S2 academic qualification. The difference in academic qualifications is the most significant among the other criteria because lecturer standards are the minimum criteria for qualifications for education in order to achieve graduate learning.

= jika j adalah atribut keuntungan (*benefit)*
=jika j adalah atribut biaya *(cost)*

CONCLUSION

In accordance with the results of the research, the determination of honorary lecturers to become permanent lecturers at the IBM ASMI Information Systems Study Program by combining the SAW and ARAS methods, the authors draw the following conclusions: Determination of honorary lecturers to become permanent lecturers at the IBM ASMI Information System Study Program with several criteria, namely academic qualifications, academic relevance, years of teaching, and courses taught using manual methods which are still inaccurate and effective because there is no weight on each predetermined criterion. By combining the SAW and ARAS methods of calculating the value for the criteria for determining honorary lecturers to become permanent lecturers, it produces accurate analysis and information compared to only one of the calculation methods so that IBM ASMI can use it as a tool to make the right decisions. Judging from the results of the ranking in the calculations, the values obtained are 0.146341, 0.134146, and 0.121951. These results prove that the ranking with the assessment using the SAW and ARAS methods produces an effective, accurate, and efficient assessment.

REFERENCE

[1] D. M. Rajagukguk, "Implementasi Metode Simple Additive Weighting pada Sistem Pendukung Keputusan Penerima Beasiswa," *Edu Komputika J.*, vol. 4, no. 1, pp. 55–55, 2017.

[2] A. S. Nadeak, "Penerapan Metode Aras (Additive Ratio Assessment) Dalam Penilaian Guru Terbaik," *Semin. Nas. Teknol. Komput. Sains*, vol. 2, no. 2010, pp. 571–578, 2019.

[3] J. A. D. Guterres, "Pemanfaatan Metode Saw Dan Topsis Sebagai Media Pendukung Keputusan Pemberian Bantuan Rumah Layak Huni,"*Pros. SINTAK 2017*, pp. 51–56, 2017.

[4] M. A. Rahmat, B. Pramono, and R. A. Saputra, "Pemilihan lokasi baru bts telkomsel cabang kota kendari menggunakan metode saw dan topsis berbasis web gis *1,2,3," vol. 3, no. 1, pp. 47–54, 2017.

[5] D. Kartini *et al.*, "the Selection of Prospective Employees Using the Shannon Entropy Weighting Method and Aras Method," vol. 01, no. 1, pp. 23–32, 2020.

[6] A. Aras and U. Mendukung, "Penerapan Metode Additive Ratio," vol. 8106, pp. 1–13, 2020.

[7] R. Fachrizal, "IMPLEMENTASI ARAS (Additive Ratio Assessment) DALAM PEMILIHAN KASIR TERBAIK STUDI KASUS OUTLET CARDINAL STORE PLAZA MEDAN FAIR," *Sainteks*, no. Januari, pp. 501–510, 2019.

[8] A. S. Yogha Radhitya, Fitro Nur Hakim, "Rancang Bangun Sistem Pendukung Keputusan Penentuan Penerima Beasiswa Dengan Metode SAW," *J. Speed – Sentra Penelit. Eng. dan Edukasi*, vol. 8, pp. 23–32, 2016.

[9] I. Septiana, M. Irfan, A. R. Atmadja, and B. Subaeki, "Sistem Pendukung Keputusan Penentu Dosen Penguji Dan Pembimbing Tugas Akhir Menggunakan Fuzzy Multiple Attribute Decision Making dengan Simple Additive Weighting (Studi Kasus: Jurusan Teknik Informatika UIN SGD Bandung)," *J. Online Inform.*, vol. 1, no. 1, p. 43, 2016.

[10] M. F. Asnawi and M. A. M. Baihaqy, "Implementasi Metode Simple Additive Weigth (SAW) pada Sistem Pendukung Keputusan Penentuan Dosen Berprestasi," *J. Penelit. dan Pengabdi. Kpd. Masy. UNSIQ*, vol. 6, no. 3, pp. 198–205, 2019.

[11] A. Apandi, "Sistem Pendukung Keputusan Penilaian Guru Terbaik Menggunakan Metode Additive Ratio Assessment (Aras)," *Semin. Nas. Teknol. Komput. Sains*, pp. 476–483, 2020.

[12] P. Stmik and B. Mulia, "Dewi Kusumawati , Siti Nurhayati dan Fitriyanti Andi Masse . Implementasi Metode Simple Additive Weighting (SAW) Untuk Evaluasi Kinerja Dosen Pada Stmik Bina Mulia Palu IJ Jurnal Sains Komputer dan Teknologi Informasi Page," pp. 116–122, 2020.

[13] L. A. Yunus, "Sistem Pendukung Keputusan Penilaian Kinerja Dosen Dengan Menggunakan Metode Additive Ratio Assesment," *J. Sist. Inf. Dan Tek. Komput.*, vol. 5, no. 1, pp. 31–35, 2020.

[14] R. Septian, I. Hadi, Ridwansyah, W. Gata, Widiastuti, and M. Iqbal, "Penelitian Dosen Yayasan," 2021.

[15] D. M. D. Utami Putra and I. P. A. Pratama, "Rancang Bangun Sistem Pendukung Keputusan Penentuan Penerima Beasiswa Dengan Menggunakan Metode TOPSIS," *S@Cies*, vol. 7, no. 1, pp. 30–36, 2016.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1971

THE DEPLOYMENT OF REVENUE CYCLE ACCOUNTING INFORMATION SYSTEM IN THE SERVICE BUSINESSES OWNED BY THE REGIONAL GOVERNMENT OF WEST NUSA TENGGARA PROVINCE

Rafa Karimah[1]; Izzari Amperaningrum[2]; Ninuk Wiliani[3]

Magister Manajemen Sistem Informasi[1,2]
Gunadarma University[1,2]
Gunadarma.ac.id[1,2]
12afa.karimah@gmail.com [1]; izzati@staff.gunadarma.ac.id[2]

Sistem Teknologi dan Informasi[3]
BRI Institute[3]
bri-institute.ac.id[3]
ninukwiliani@bri-institute.ac.id [3]

Abstract— Based on Law Number 25/2009 on Public Services, the West Nusa Tenggara Government established Wisma NTB. However, the lack of interest from the government to stay at Wisma NTB while they are in Jakarta is the main cause of the inability of the NTB Liaison Agency to reach its local revenue target. In connection with the reason and the latest technological development, information systems are urgently needed by the agency to assist its operational and decision-making activities. This study aims to determine the current income cycle provided at Wisma NTB, to evaluate the effectiveness and efficiency of the ongoing cycle, and to determine whether or not to develop the available system. The research methods used for data compilation consisted of field observations, interviews, and literature studies. The system development method used is Extreme Programming (XP), software development procedures launched through practical principles and techniques. Meanwhile, the internal control evaluation applies COSO. The results reveal the prevailing control system has been proper and in line with prevailing environmental control and monitoring standards. However, weaknesses can be found in control, risk assessment, information, and communication activities. The elements are related to technology because the current income cycle worked less than optimal. Thus, it is necessary to develop the system. The developments include the provision of a room reservation system, a payment gateway system, and a system for recording online transactions so such transactions can be accessed anytime, anywhere.

Keywords: accounting information system, income cycles, service business, room reservations, local government

Abstrak— *Sebagai bentuk pengamalan Undang-Undang Nomor 25 Tahun 2009 tentang Pelayanan Publik, Pemerintah Provinsi Nusa Tenggara Barat berusaha memberikan pelayanan publik melalui keberadaan Wisma NTB. Namun kurangnya minat pegawai dan pejabat menginap di Wisma NTB saat perjalanan dinas ke Jakarta menjadi penyebab utama tidak mampunya Badan Penghubung Provinsi NTB mencapai target pendapatan asli daerah. Sehubungan dengan itu serta perkembangan teknologi yang terus mengalami perubahan dan kemajuan, sistem informasi sangat dibutuhkan oleh instansi untuk membantu kegiatan operasional dan pengambilan keputusan. Penelitian ini bertujuan untuk mengetahui sistem informasi akuntansi siklus pendapatan yang sedang berjalan pada Wisma NTB, mengevaluasi efektifitas dan efisiensi sistem informasi akuntansi siklus pendapatan yang sedang berjalan serta untuk mengetahui perlu tidaknya dilakukan pengembangan terhadap sistem. Metode penelitian yang dilakukan dalam penyusunan data terdiri atas observasi lapangan, wawancara, dan studi pustaka. Metode pengembangan sistem yang digunakan adalah EXtreme Programming (XP), yaitu pengembangan perangkat lunak melalui berbagai prinsip dan teknik praktis pengembangan perangkat lunak. Sedangkan metode evaluasi pengendalian internal yang dilakukan menggunakan COSO (Committee Of Sponsoring Organizations). Hasil penelitian menunjukan penerapan sistem pengendalian di Wisma NTB sudah dijalankan dengan baik dan sesuai dari segi pengendalian lingkungan dan monitoring. Namun masih terdapat kelemahan pada aktivitas pengendalian, penilaian resiko serta informasi dan komunikasi. Ketiga unsur tersebut erat kaitannya dengan teknologi, Karena kurang optimalnya Sistem Informasi Akuntansi siklus pendapatan yang ada saat ini, maka diperlukan pengembangan pada sistem. Pengembangan tersebut antara lain pada sistem pemesanan*

kamar, sistem pembayaran dengan metode payment gateway, serta sistem pencatatan transaksi yang dilakukan secara online sehingga dapat diakses kapanpun dimanapun.

***Kata Kunci**: sistem informasi akuntansi, siklus pendapatan, penerimaan kas, usaha jasa, pemesanan kamar, pemerintah daerah.*

INTRODUCTION

Based on the 2019 Ministry of Communication and Information Technology's Performance Report, the development of digital technology has driven Indonesia's economic growth from conventional to digital[1]. The Indonesian economy has also experienced a shift from an economy focused on commodities to service-based economic activities. The dynamics of this digital global era have also led to new ways to take advantage of technological changes, especially digital in several sectors, which have emerged as a solution to overcoming inefficiencies[2]. To encourage the growth of the digital economy, the government is required to carry out a transformation in providing better public services to the society by utilizing digital technology. One form of service provided by the West Nusa Tenggara Provincial Liaison Agency in Jakarta to Officials, Provincial Government Employees, Leaders, and Members of the West Nusa Tenggara Provincial DPRD who carry out official duties to Jakarta is through the Wisma NTB facility located on Jl. Garut No.5 Menteng, Central Jakarta. As a form of practicing Law Number 25 of 2009[3] on Public Services, the West Nusa Tenggara Provincial Government strives to provide the best public services to all stakeholders and the wider community through the existence of this Wisma NTB [4].

In connection with the foregoing as well as technological developments that continue to experience changes and advances, at this time information systems are needed by an agency to assist its operational activities[5] and assist agencies in making decisions or policies based on the information obtained. One of the information systems that play an important role in operational activities is the accounting information system[6].

One of the cycles in the accounting information system is the income cycle. The revenue cycle consists of sales, accounts receivable, and cash receipts[7]. The revenue cycle itself is a direct exchange of end products and services into cash in one transaction between buyers and sellers[7]. Thus, a series of business activities and related information processing activities are repeated by providing goods and services to customers and collecting cash as payment for these sales[8].

Based on the identification of the problems mentioned above, the objectives of this study are as follows: To determine the current income cycle accounting information system at Wisma NTB, to evaluate the effectiveness and efficiency[9] of the income cycle accounting information system at Wisma NTB which is currently running and to find out whether or not it is necessary to develop an ongoing income cycle accounting information system at Wisma NTB[4].

Table 1. The following Research Literature summarizes a review of previous researches on Accounting Information System Development which are used as the references for the development of the research conducted by the researchers.

Table 1. Research Problems

Research Problems (RP1)		Literature Supports
RP 1	The income cycle accounting information system at Wisma NTB was recorded manually	Nowadays' increasingly tough competition forces all companies to integrate more sophisticated technology to support all of their activities. Along with current technological developments, the use of computers is a must to expedite the companies' business activities. [10]
		Sales accounting information system puts its concerns on how the company organization can plan, coordinate, master, or control various sales activities carried out by the organization. [11]
		Without the presence of the accounting reports, managers and supervisors cannot monitor the company's financial position which is very useful to track the company's development and to determine the company's policies which shall be adopted in the future. [12]

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1971

RP 2	The income cycle accounting information system at Wisma NTB is not yet effective and efficient	Every business organization always does its best to meet its information demand. The companies maximize their technological resources because they need relevant, fast, timely, and accurate information that reflects the company's physical condition to assist them in planning, coordinating, and controlling their operations. [10]
		Adequate application of information will support effective internal control, resulting in the effective information required by management to support their decision-making process. [11]
		Every business organization always does its best to meet its information demand. The companies maximize their technological resources because they need relevant, fast, timely, and accurate information that reflects the company's physical condition to assist them in planning, coordinating, and controlling their operations. [10]
RP 3	No development has been made to the ongoing income cycle accounting information system at Wisma NTB	The purpose of the sales accounting information system is to create strong controls in a situation where no one unit can complete a transaction without involving other people's responsibilities. [11]
		A computerized accounting information system connected to the head office implements company activities well controlled even though some weaknesses in internal control can still be recorded. [13]
		Nowadays' increasingly tough competition forces all companies to integrate more sophisticated technology to support all of their activities. Along with current technological developments, the use of computers is a must to expedite the companies' business activities[10].

Table 1 explains some of the problems that exist in Wisma NTB, so it is necessary to develop an accounting information system especially for the existing income system in Wisma NTB. There are a lot of things that need to be prepared so that the purpose of building a system can work by the desired reality.

MATERIALS AND METHODS

The system development method approach used by researchers is one of the agile methods, namely Extreme Programming *(XP)*. The Extreme Programming (XP) method according to Prabowo (2018) is a software engineering process that tends to approach an object-oriented approach and the goal of this method is a team formed on a small to medium scale and this method is also suitable if the team is faced with changing requirements - very fast change. Extreme Programming is known as a method or technical how-to, how a technical team develops software efficiently through various practical principles and techniques of software development. Extreme Programming (XP) was chosen because the software to be made is not too complex and is classified as small-scale software and also requires less development time[9].

Extreme Programming (XP) was chosen because the software that will be built is not too complex and can be classified into small-scale software. Furthermore, the software also requires less development time[12]. There are four stages [14] in the Extreme Programming method, namely:

1. *Planning*

The XP methodology approach at the planning stage is the first step in system development where in this stage several planning activities are carried out, namely identifying problems, analyzing needs to determine the schedule for implementing system development. In this stage, the initial needs of the user are collected or in XP it is called user stories. This is necessary so that developers understand the content business, system output requirements, and the main features of the software being developed[15].

2. *Design*

The design stage is where modeling activities are carried out starting from system modeling, architectural modeling to database modeling. System and architecture modeling uses Unified Modeling Language (UML) diagrams while database modeling uses Entity-Relationship Diagrams (ERD). The design of the system in this study is described by the UML model in the form of use case diagrams, activity diagrams, class diagrams, and sequence diagrams. XP design still prioritizes the principle of Keep it Simple (KIS). The design here is a representation of the system to make it easier for developers to build the system[16].

3. *Coding*

This stage is an activity of implementing modeling that has been made into a user interface using a programming language. At this stage, an Income Cycle Accounting Information System is built which is the object of research. The system is built based on designs that have been made in the previous stage. The programming language used is Hypertext Preprocessor (PHP) and for database management systems using MySQL software[17].

4. *Testing*

The final stage is carried out after the coding is complete, then the testing stage is carried out on the system to find out what errors arise when the application is running and find out whether the system built is by user needs. The test method used in this stage is the BlackBox testing method, where tests are carried out on several input forms, whether they are running according to their respective functions[9].

RESULTS AND DISCUSSION

Based on the system analysis, the results of the observation and interview methods show that the current income cycle accounting information system is not yet optimal. It is said that it is not optimal because the data processing in the financial statements has not been integrated directly with the income process at Wisma NTB, so the two processes that should have been carried out simultaneously in one system had to be carried out twice on a separate system.

As a whole, promotional activities and room reservations at Wisma NTB take place manually. Even proof of room rental is only a receipt written manually. This is a weakness because in addition to being ineffective [18]. This is a weakness because, in addition to being ineffective and inefficient, proof of the transaction can also be fabricated[13] so that the overnight civil servants cannot be held accountable. Figure 1. below is the Flowchart for Room Reservation Transactions, Current System Procedures.

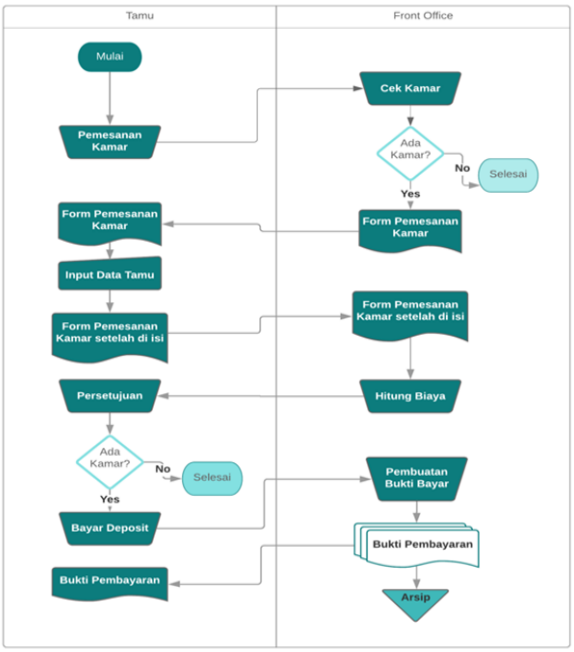


Figure 1. Flowchart of Room Reservation Transactions, Current System Procedures

A. Planning

Flowcharts can develop an understanding of how a process is carried out. To illustrate the development of the proposed Income Cycle Accounting Information System, a flowchart is made which can be seen in Figure 2. The proposed Income Cycle Accounting Information System Flowchart is as follows:

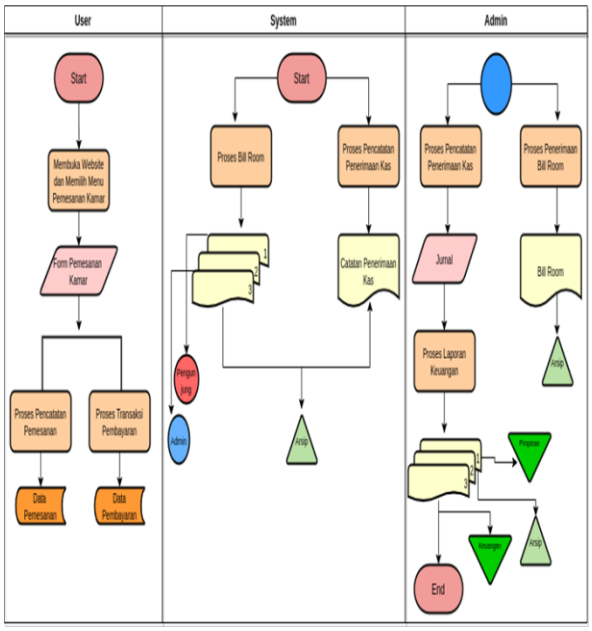


Figure 2. The Proposed Revenue Cycle Accounting Information System Flowchart

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1971

B. Design

1. System Modeling Design

The use case diagram describes an interaction between one or more actors and the system to be created. Use case diagrams are used to find out what functions are in a system and who has the right to use these functions. Figure 3. below is the Use Case Diagram of the Proposed System at Wisma NTB

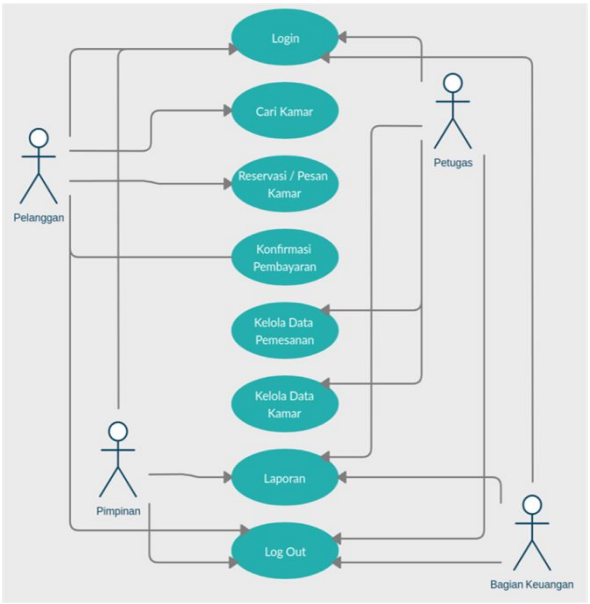
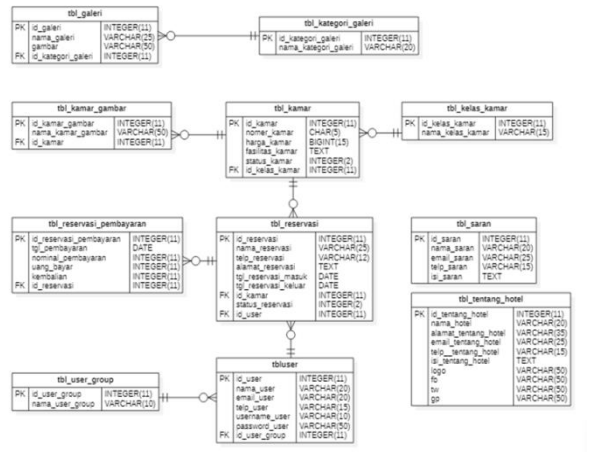


Figure 3. Use Case Diagram of the Proposed System at Wisma NTB

2. Database Design

Entity Relationship Diagram (ERD) is a diagram used to design a database to describe related data in a database. Figure 4 below is the ERD on the Proposed System at Wisma NTB:



Gambar 4. ERD on the Proposed System at Wisma NTB

In general, after the ERD design is completed, the next step is to physically design the database, namely making tables, indexes while still considering performance.

3. User Interface Design

a. The Main Page of the System

Figure 5. The Main Page of Wisma NTB Income Cycle Accounting Information System is a page that functions as the homepage, the initial display that displays the overall profile of the website. The Main Menu consists of the home menu, the room list menu, the about us menu, the complaints and advice menu, the gallery menu, and then the login/sign up menu.

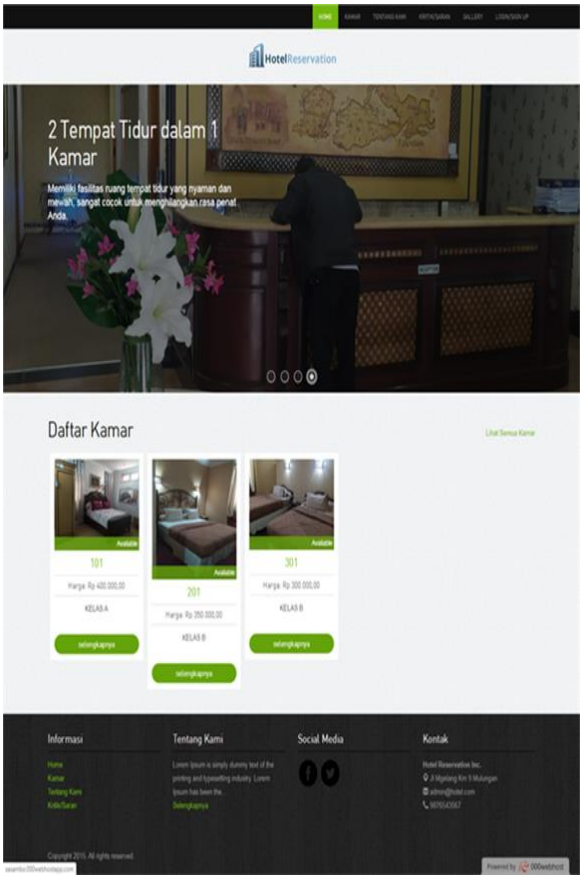


Figure 5. The Main Page of Wisma NTB Income Cycle Accounting Information System

b. The Room List Page

Figure 6. The Room List Page of Wisma NTB Income Cycle Accounting Information System below is a layer capture of the Room List Page of the proposed Wisma NTB Income Cycle Accounting Information System. On this page, users can see the details of the room to be booked.

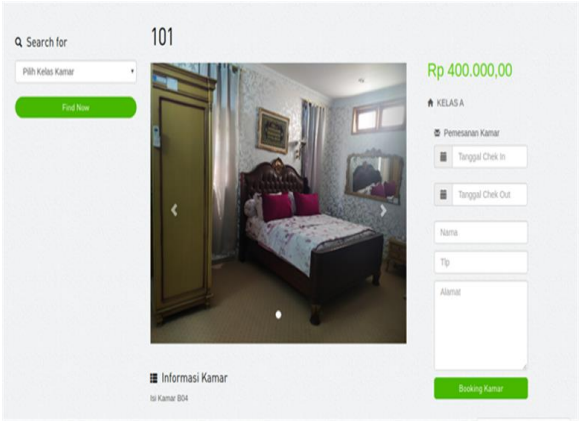


Figure 6. The Room List Page of Wisma NTB Income Cycle Accounting Information System

c. The Administrator Page

Figure 7. The Main Page of Wisma NTB Income Cycle Accounting Information System Administrator below is a screenshot of the layer Main Page Administrator of the proposed Wisma NTB Income Cycle Accounting Information System. On this page, users at the administrator level manage the system that has been designed.

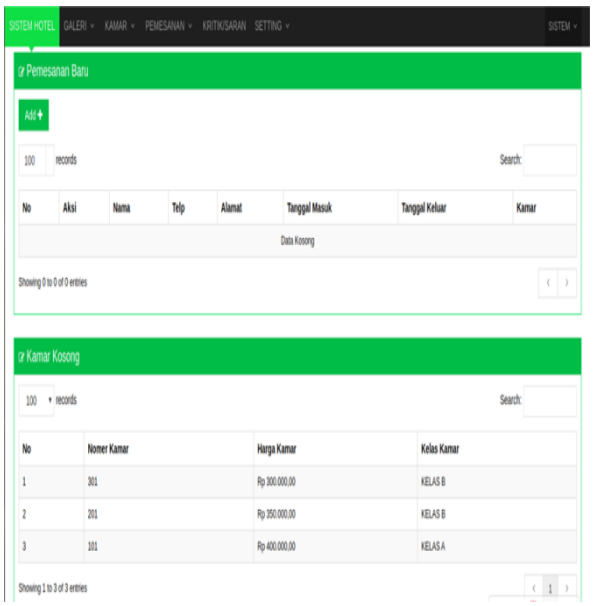


Figure 7. The Main Page of Wisma NTB Income Cycle Accounting Information System Administrator

d. The Payment Page

Figure 8. The following page for the Payment System of Wisma NTB Income Cycle Accounting Information System is a screenshot of the payment page layer of the proposed Wisma NTB Income Cycle Accounting Information System. On this page, the users complete the room booking transaction and make a payment.

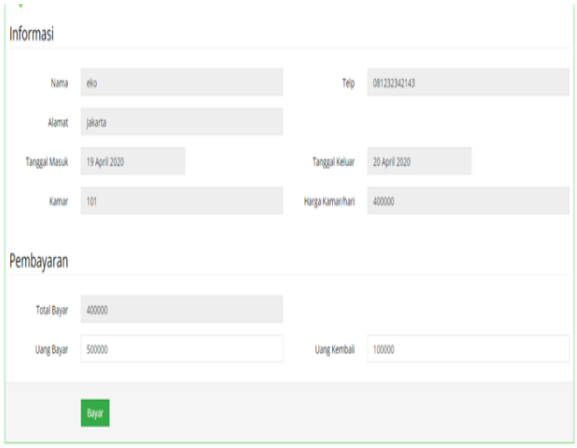


Figure 8. The Payment Page of Wisma NTB Income Cycle Accounting Information System

e. The Financial Report Page

Figure 9. The following page of the Financial Report Accounting Information System of Wisma NTB Income Cycle is a layer capture of the Financial Report Page of the proposed Wisma NTB Income Cycle Accounting Information System. On this page, users at the administrator and leadership levels can get information related to the financial reports recorded at Wisma NTB.

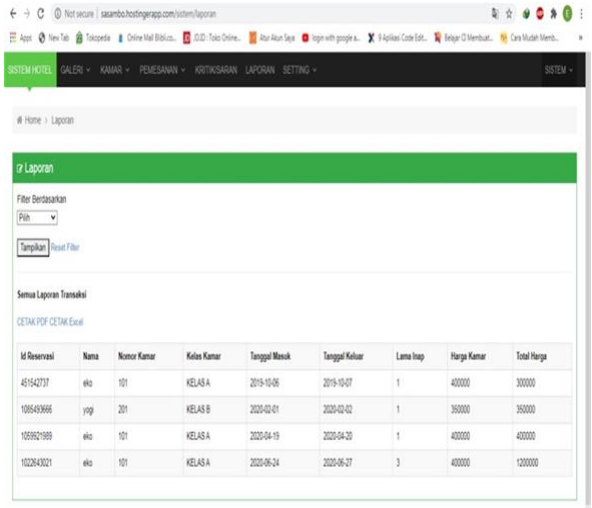


Figure 9. The Financial Report Page of Wisma NTB Income Cycle Accounting Information System

f. The Financial Reports Graph Page

Figure 10. The Graph Page of Financial Reports on the Income Cycle Accounting Information System of Wisma NTB. The following is a screenshot of the page of the Graph of Financial Reports of the proposed Income Cycle Accounting Information System of Wisma NTB. On this page, users at the administrator or leader level can view financial reports in graphical form.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1971

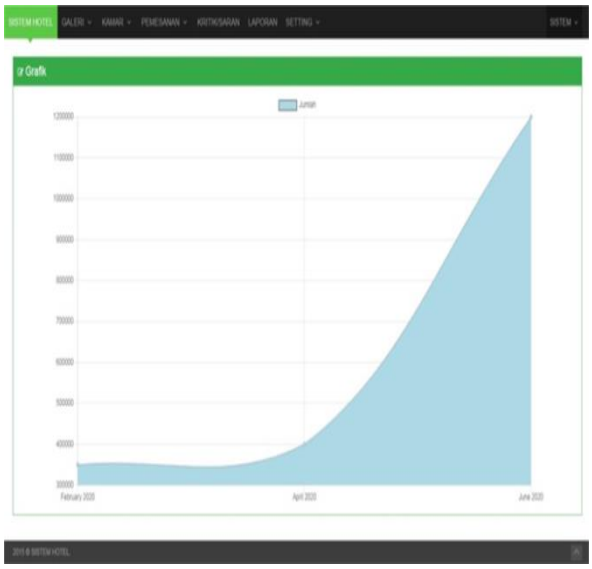


Figure 10. Graph Page of Financial Report on the Accounting Information System for the Income Cycle of Wisma NTB

g. The Financial Statement Page

Figure 11. The Monthly Financial Statement Page on the Income Cycle Accounting Information System of Wisma NTB is a layer capture of the Financial Statement page of the proposed Income Cycle Accounting Information System of Wisma NTB. On this page, the downloaded report results can then display the required Financial Statement data. Financial Statements can be obtained by day, month, or year.

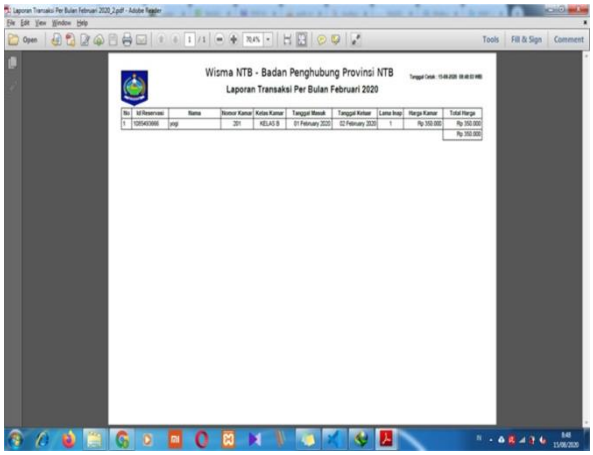


Figure 11. The Page of Financial Statement on Wisma NTB Income Cycle Accounting Information System

C. Testing

The technique of evaluating the effectiveness and efficiency of the Income Cycle at Wisma NTB is based on the COSO theory:

One of the components of policies and procedures that are designed and implemented to achieve internal control objectives is the control environment. Table 2. Results of the Control Environment Evaluation. The following is a table of results of the evaluation of the control environment at Wisma NTB:

Table 2. Results of the Control Environment Evaluation

No	COSO Theory	Wisma NTB	Evaluation	Result
1	Control Environment: Integrity, Ethical Values, Commitment, Organizational Structure	Every employee has good integrity values, has a spirit of responsibility, and is oriented towards public service	Has been carried out properly by their respective duties and responsibilities.	Matched

An agency or company must be aware of and manage the risks it faces so that the organization can run harmoniously. Table 3. Results of the Evaluation of Risk Assessment Elements below is a table that describes the results of the evaluation of risk assessments at Wisma NTB:

Table 3 Evaluation Results of Risk Assessment Elements

No	COSO Theory	Wisma NTB	Evaluation	Result
2	Risk Assessment: Based on activities in the cash receipt process	Physical matching of cash that is in the cashier every day, there is evidence of each transaction	The existence of calculations, matching, and reporting from the cashier to the cash storage department. However, everything is still done manually, transactions directly at the guesthouse (Wisma).	Not yet matched

Activity control is a policy and/or procedure required to deal with existing risks. Table 4. Evaluation Results for Activity Control. The following are the results of an evaluation of activity control at Wisma NTB:

Table 4. Evaluation Results for Activity Control

No	COSO Theory	Wisma NTB	Evaluation	Result
3	Control Activities: The level of control over the entity, transactions, and information technology	Transactions are by existing SOPs, but the use of technology is still minimal	The cashier and room reservations are not computerized	Not yet matched

With the existence of information and communication systems, it is possible to obtain and exchange the information needed to carry out and manage the operational activities of the agency. Table 5. The following results of the evaluation of information and communication elements are the current results of an evaluation of the use of the existing information and communication system at Wisma NTB:

Table 5. Evaluation Results of Information and Communication Elements

No	COSO Theory	Wisma NTB	Evaluation	Result
4	Information and Communication: Identity, understand, analyze, record, and report information exchange	Information can only be obtained by requesting data from the Finance and Treasury Division	Limited information access	Not yet matched

The whole process must be monitored, and evaluated. Some changes shall be made when necessary. Table 6. Evaluation Results of Monitoring/Supervising Elements. Hereinbelow are the results of the evaluation of the current monitoring and/or supervising process at Wisma NTB, to check whether the system has been run appropriately or not:

Table 6. Evaluation Results of Monitoring/Supervisoring Elements

No	COSO Theory	Wisma NTB	Evaluation	Result
5	Monitoring and Supervising	Every month checking and evaluating cash and bookkeeping are carried out	Each division is always monitored based on the information and every month a cash check is carried out	Matched

Based on the evaluation results from table 2 to table 5 above, it can be seen that the application of the internal control system implemented at Wisma NTB is according to the theory put forward by COSO (Committee Of Sponsoring Organizations), where according to the theory applied by COSO in the implementation of the control system at Wisma NTB has been implemented properly and by environmental control and monitoring. However, there are still weaknesses or discordance with control activities, risk assessment as well as information and communication. These three elements are closely related to technology, the cashier/receptionist has not used the User Computing system. Customers also cannot access information related to inventory and room bookings in real-time online. Based on the risk assessment, financial data manipulation is likely to occur with the current system. Then in terms of information and communication, it is limited to space and time, so that access in the decision-making process and policymaking can be delayed.

CONCLUSION

The Accounting Information System designed at Wisma NTB was made through several processes, namely planning, designing, coding, implementing, and testing. Based on the results of the research that has been done, it can be concluded that the current income cycle information system at Wisma NTB is currently running semi-computerized. Where the room reservation process is done manually while the recording of the transaction is via a computer. The current income cycle information system at Wisma NTB is said to be ineffective and inefficient because the separate room reservation and transaction recording system is not optimal for prospective customers and also for leaders in determining policies and making decisions. Due to the less than optimal income cycle Accounting Information System that is currently available at Wisma NTB, it is necessary to develop the system. These developments include a room reservation system that is done online through a website, a payment system with a payment gateway method, and a system for recording transactions made online through the website so that it can be accessed anytime, anywhere. The weakness of this system, among others, is the need to add several features to support interaction between users and administrators such as real-time chat features, besides that it still needs to be refined in terms of

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1971

accounting reporting such as adding expense features, not just income features.

Suggestions

Based on the findings of problems in the process of making this Accounting Information System, the following suggestions can be drawn first, for further research it is expected to apply a different institution as a comparison. Second, Regarding the income cycle understudy, subsystems other than the sales subsystem and cash receipts subsystem can be used. For the future development of this Accounting Information System, further, development is still needed, such as adding more features and functions to the system.

REFERENCES

[1] R. DEPKES, "Peraturan Presiden Republik Indonesia Nomor 81 Tahun 2010," *Peratur. Pemerintah Republik Indones. Nomor 26 Tahun 1985 Tentang Jalan*, vol. 1999, no. 1, pp. 1–5, 2004.

[2] S. R. Chambers, *Undang Undang Republik Indonesia Nomor 25 Tahun 2009 Tentang Pelayanan Publik*, no. April. 2009.

[3] B. Dan, L. Teknis, P. Nusa, and T. Barat, "Peraturan Gubernur Nusa Tenggara Barat Nomor 6 Tahun 2014," 2008.

[4] P. Daerah, P. Nusa, and T. Barat, "Peraturan Gubernur Nusa Tenggara Barat Nomor 29 Tahun 2018," pp. 1–13, 2012.

[5] J. A. Hall, "Transaction Cycles and Business Processes," *Account. Inf. Syst.*, pp. 154–192, 2011.

[6] R. S. Wazlawick, *Object-Oriented Analysis, and Design for Information System*. Elsevier Inc., 2013.

[7] P. S. Foote, *Accounting Information Systems, Fifth Edition*, vol. 22. 2007.

[8] I. Aleksander, "Analysis and Design of Information Systems," *J. Inf. Technol.*, vol. 1, no. 4, pp. 61–61, 1986, DOI: 10.1057/jit.1986.38.

[9] T. Pratiwi, "Perancangan dan Implementasi Sistem Informasi Akuntansi Penerimaan Kas Mahasiswa di Poiteknik TEDC Bandung (Dengan Menggunakan PHP Dan MySQL).pdf," *TEDC*, vol. 13, no. 2 Mei, pp. 195–200, 2019.

[10] G. Kusumaningdiyah, D. Saptantinah, and M. R. Sunarko, "Evaluasi Penerapan Sistem Informasi Akuntansi Siklus Pendapatan Dengan Program Iris Menggunakan Pendekatan Metode System Development Life Cycle," *J. Akunt. dan Sist. Teknol. Inf.*, vol. 12, pp. 186–192, 2016, [Online]. Available: http://ejurnal.unisri.ac.id/index.php/Akuntansi/article/view/1231.

[11] A. Mujahidah, Faridah, and Thanwain, "Analisis Sistem Informasi Akuntansi Penjualan Pada PT Hadji Kalla (Toyota) Cabang Pinrang," *J. Ris. Ed. V Unibos Makasar*, vol. 4, no. 007, pp. 88–100, 2016.

[12] N. S. Meme Susilowati, "Desain Sistem Informasi Akuntansi Terintegrasi Lima Siklus Pada Usaha Jasa Laundry," *J. Teknol. Inf. dan Ind.*, vol. I, no. 2, pp. 65–72, 2018.

[13] F. O. Voets, J. J. SOndakh, and A. Wangkar, "Analisis SIstem Informasi Akuntansi Siklus Penjualan dan Penerimaan Kas Untuk Meningkatkan Pengendalian Intern," *J. Berk. Ilm. Efisiensi*, vol. 16, no. 4, pp. 191–202, 2016.

[14] D. N. Anggreni, T. Lestari, and U. B. Surabaya, "Analisis Sistem Informasi Akuntandi Atas Siklus Pendapatan Untuk Meningkatkan Pengendalian Intern Pada AJB Bumiputera 1912 Kantor Pemasaran Asuransi Jiwa Korporasi Cabang Darmo Surabaya," *Equity*, vol. 4, no. 2, pp. 23–33, 2018.

[15] A. Tundung, T. K. Priyambodo, and A. Armawi, "Tingkat Ketahanan Sistem Informasi Administrasi Kependudukan (Studi pada Dinas Kependudukan dan Pencatatan Sipil Kota Yogyakarta)," *J. Ketahanan Nas.*, vol. 23, no. 2, p. 21, Aug. 2017, doi: 10.22146/jkn.26345.

[16] A. Sani and N. Wiliani, "Faktor Kesiapan Dan Adopsi Teknologi Informasi Dalam Konteks Teknologi Serta Lingkungan Pada Umkm Di Jakarta," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 49–56, 2019, doi: 10.33480/jitk.v5i1.616.

[17] N. Wiliani and S. Zambi, "Rancang Bangun Aplikasi Kasir Tiket Dengan Visual Basic 2010 dan MySql," *J. Rekayasa Inf.*, vol. 6, no. 2, p. 77, 2017, doi: 10.1017/CBO9781107415324.004.

[18] R. Valero-Fernandez, D. J. Collins, K. P. Lam, C. Rigby, and J. Bailey, "Towards Accurate Predictions of Customer Purchasing Patterns," in *IEEE CIT 2017 - 17th IEEE International Conference on Computer and Information Technology*, 2017, DOI: 10.1109/CIT.2017.58.

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.1971

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1929

SYSTEMATIC LITERATURE REVIEW: IMPLEMENTATION OF KNOWLEDGE MANAGEMENT IN THE ORGANIZATION

Azizah Nurfauziah Yusril[1]; Evy Nurmiati[2]

Program Studi Sistem Informasi[12]
UIN Syarif Hidayatullah Jakarta
www.uinjkt.ac.id
azizah.ny17@mhs.uinjkt.ac.id, evy.nurmiati@uinjkt.ac.id
(*) Corresponding Author

Abstract— Knowledge management is an activity that organizations use to achieve goals and gain competitive advantages. This study features a systematic literature review that discusses the implementation of knowledge management in organizations covering 39 articles published from 2015 to 2020. This study aims to answer four research questions. The results show that the trend of knowledge management research in Indonesia is dominated by research related to the designing of knowledge management systems. The application of knowledge management in Indonesia has been applied in various fields. Some various models and methods can be used in creating a knowledge management system.

Keywords*:* systematic literature review, knowledge management, knowledge management system.

Intisari—Manajemen pengetahuan merupakan suatu aktivitas pengelolaan pengetahuan yang digunakan organisasi untuk mencapai tujuan dan memperoleh keunggulan. Penelitian ini menampilkan sebuah *systematic literature review* yang membahas implementasi manajemen pengetahuan di dalam organisasi yang mencakup total 39 artikel yang diterbitkan dari tahun 2015 hingga 2020. Penelitian ini bertujuan untuk menjawab lima pertanyaan penelitian. Didapatkan hasil bahwa tren penelitian manajemen pengetahuan di Indonesia, didominasi dengan penelitian terkait perancangan *knowledge management system* (KMS). Penerapan manajemen pengetahuan di Indonesia telah diterapkan di berbagai bidang. Ada berbagai model dan metode yang bisa digunakan dalam membuat *knowledge management system*.

Kata Kunci*: systematic literature review*, manajemen pengetahuan, sistem manajemen pengetahuan.

INTRODUCTION

Recently there have been changes in all areas of life, as a result of which technology has grown faster. This condition requires how to respond to all these changes to survive. Of course, only with knowledge, all changes can be addressed appropriately. Knowledge is the result of a process through social interaction with other people and the environment to become the goal of truth [1].

By the development of organizations and information technology in organizations and using knowledge, knowledge becomes an important asset in the organization [2]. The knowledge should make progress of the organization itself. Therefore, it is necessary to have knowledge management in the organization. With knowledge management, knowledge can be put to good use within the organization, coupled with good infrastructure support that will help disseminate information within the organization.

A systematic literature review is a research method for evaluating, evaluating, and interpreting all relevant research results related to specific research questions, specific topics, or phenomena of concern [3]. This method is used by reviewers to search for research journal articles published through electronic databases[4].

This study features a systematic literature review that discusses the implementation of knowledge management in Indonesia. This study aims to determine how the implementation of organizational management, designing a knowledge management system, and its effects, and what factors affect knowledge management.

Knowledge Management

Knowledge management is a system designed to create, document, classify, and knowledge within the organization. According to [5], in carrying out knowledge management four things must be considered, namely: knowledge acquisition, knowledge conversion, knowledge application, and knowledge protection.

Organizational management can be seen operationally and strategically. Operational KM is an organizational activity based on efforts to develop and utilize knowledge within an

organization. Meanwhile, strategic KM is a step to strengthen the efforts of each individual and organization to become a knowledge-based organization[6].

Organization

Organization comes from the word *organon* which means tool in Greek. The organization can be defined as a group consisting of two or more people who work together to achieve certain goals together. Organizations use all resources and facilities and infrastructure that are used effectively and efficiently to achieve organizational goals [7]. One of the organizational effectiveness can be seen from the organizational culture [8].

Organizational culture refers to the norms of behavior, assumptions, and beliefs of an organization, while organizational climate refers to the perceptions of people in the organization that reflect those norms, assumptions, and beliefs. It can be said that organizational culture is a set of rules and regulations agreed upon and implemented by its members. Organizational culture can shape the behavior patterns of its members.

Knowledge Management and Organization

In implementing knowledge management several things need to be considered, including organizational culture. The importance of organizational existence in the implementation of knowledge management is described in the literature [9]. According to Walker, there are several elements of implementing knowledge management, namely:

1. Knowledge creation, the creation of knowledge is facilitated through job design, namely by assigning tasks to work teams, not individuals.
2. Knowledge Transfer / Knowledge Sharing, in organizations sharing knowledge, is crucial in organizations where this process involves individuals, teams, departments, or divisions who are influenced by the experiences of others.

According to [10],], knowledge management is how to manage organizational processes to create, store and reuse organizational knowledge.

MATERIALS AND METHODS

The methodology uses the systematic literature review method where the stages in this systematic review are adopted from [11].

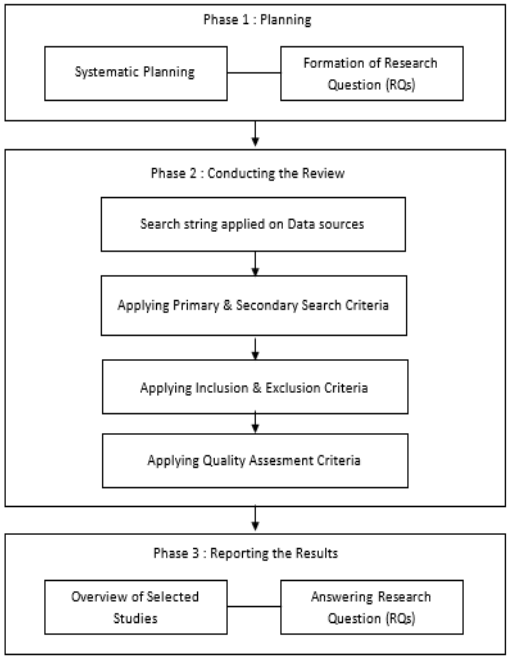


Figure 1.Systematic Literature Review Phase[11]

1. Planning phase

a. Research questions

To determine the extent of the concept of knowledge management in organizations in Indonesia, this study has several research questions:

- RQ1: What is the trend of knowledge management research in Indonesia?
- RQ2: What are the perspectives and applications of knowledge management in Indonesia?
- RQ3: How does knowledge management affect and what does it affect knowledge management?
- RQ4: What methods or models are used in forming a knowledge management system (KMS)?

2. *Conducting the review phase*

a. *Search strategy*

The keywords used in this literature search were "Knowledge Management in Organizations". The Source of literature data is obtained from Google Scholar which is limited by publications published from 2015-2020.

b. *Study selection criteria and procedures*

The list of primary literature that has been obtained will be selected and re-evaluated to separate the literature that fits the criteria and those that are excluded.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1929

Table1.Inclusion Criteria and Exclusion Criteria

No	Inclusion Criteria	Exclusion Criteria
1	Literature uses Indonesian or English	Literature that only provides theories related to knowledge management.
2	Literature where knowledge management concepts are applied in organizations	Literature that does not apply knowledge management in organizations
3	The literature answers at least one research question	Literature that only displays and provides abstracts or slideshows
4	Literature indexed under S1, S2, and S3 in Sinta (Science and technology index)	Literature indexed S4, S5, and S6 in Sinta (Science and technology index)

c. *Included and excluded studies*

Using a Google Scholar search, approximately 62,300 results were obtained. After the first stage selection through literature criteria selection, a total of 302 pieces of literature were obtained. Furthermore, the second stage of selection was carried out and 39 total kinds of literature were obtained.

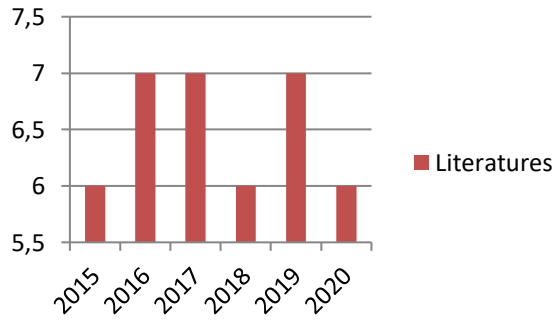


Figure 2.Selected literature statistics

d. *Study quality assessment*

The articles obtained were then evaluated for quality based on several criteria: (1) The articles matched the inclusion and exclusion criteria, (2) The literature search included all relevant studies, (3) The data/baseline studies were sufficiently explained.

e. *Data extraction strategy*

The following information is collected in data extraction: (1) Source (journal or conference) and complete references, (2) Classification of study types and research scope, (3) Main topics, (4) Author and institution, (5) Research summary, (6) Research questions, (7) Quality evaluation, (8) Practitioner-based and guided studies, (9) Number of major studies used.

RESULTS AND DISCUSSION

1. RQ1: What is the trend of knowledge management research in Indonesia?

From 2015 to 2020, the trend of research related to knowledge management, the literature that examines the design of KMS dominates the focus of the study with a percentage of 51%. The research theme of knowledge management perspective and its application as a percentage of 39%, and the theme of the influence of knowledge management and the factors that influence it has a percentage of 10%.

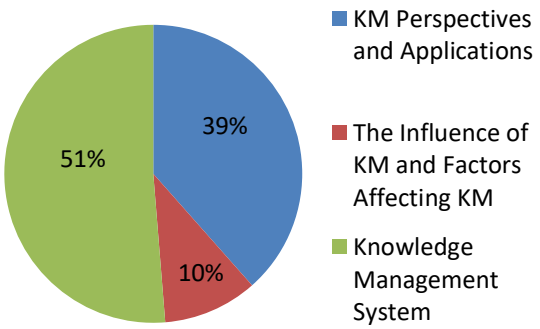


Figure3.Research Trends by Field, 2015 - 2020

2. RQ2: What are the perspectives and applications of knowledge management in organizations in Indonesia?

In public sector organizational reform, good organizational performance is needed. The effectiveness of knowledge and technology management is the key to increasing the competitiveness of public sector organizations [12]. One of the applications of knowledge management is applied in government institutions [13]. In government organizations, knowledge management is used to carry out regional development planning [14], [15]. Knowledge management is applied in the field of education, for example in the field of education such as libraries [16],[17],[18],[19]. Also, in higher education, knowledge creation and application of knowledge can improve lecturer performance[20]. Knowledge management is also applied in banking organizations [21]. The application of knowledge management is not only applied by large companies but can be applied in Small and Medium Enterprises (SME)[22].

Strategy-based knowledge management can be applied to gain a competitive advantage[23],[24],[25]. Knowledge management

implementation is carried out by establishing a knowledge management system along with a repository and database that begins with the SECI process (socialization, externalization, combination, and internalization)[26]. Therefore, it is necessary to have good internet network facilities and the quantity of computers in the organization to maximize the use of information technology to support a good knowledge management process.[27].

3. RQ3: How is knowledge management affected and what does it affect knowledge management?

Knowledge management affects employee performance. In [28] research, employee performance is determined by three independent variables, namely knowledge management, skills, and employee attitudes. Collaboration between employee empowerment and knowledge management has a positive effect on employee performance [28]. Determination, supervision, and knowledge management affect employee performance [29]. Knowledge management is influenced by organizational culture as described in the research [30]. There are 4 enablers in knowledge management, namely culture, technology, infrastructure, and measurement [31]. Trust influences positively on knowledge sharing[32].

4. RQ 4: What methods or models are used in forming a knowledge management system (KMS)?

The Nonaka model or what is called the SECI model, in addition to being used for knowledge creation, is used to identify and map knowledge within organizations [33], [34]. The SECI model can be combined with a soft system methodology that concentrates on stakeholder perspectives in system development [35]. Refinements of the SECI model resulted in the Inukshuk model. Inukshuk model can be used to map knowledge which also emphasizes aspects of leadership, technology, and culture [36][37], [38], [39].

The 10-step KM roadmap model is a framework developed by AmritTiwana, research [40][41], [42], [43][44]use this framework in developing KMS. In addition to the SECI model, knowledge mapping can use Zack's KM Cycle model. Knowledge Management System Life Cycle (KMSLC) can also be used in designing knowledge management systems [45]. Knowledge extraction can be done using a clustering approach [46]. Making business process-oriented KMS can use Strohmaier's organizational knowledge process framework [47].[48], [49], [50].System development methods such as SCRUM, Waterfall, and the System Development Life Cycle (SDLC) can also be used in the design of knowledge management systems.

CONCLUSION

There are 39 pieces of literature found from 2015 to 2020 that discussed the implementation of knowledge management in organizations. The research theme trend for the implementation of knowledge management in organizations is dominated by research related to the design of knowledge management systems (KMS). From 39 pieces of literature, it is found that knowledge management can be implemented in various fields such as government institutions, education, and banking. Not only in large institutions but can also be applied in small industries such as in Micro, Small, and Medium Enterprises (MSMEs). Knowledge management affects employee performance and the competitive advantage of an organization or company. Several methods and models are used in designing a knowledge management system, namely: the SECI model, the Inukshuk model, and the 10-step KM roadmap model. Meanwhile, the KMS development method can be done with SCRUM, Waterfall, and the System Development Life Cycle (SDLC).

Based on the review that has been carried out, the authors convey several suggestions, namely: (1) organizations should manage knowledge management optimally so that it can get maximum employee performance, and (2) increase research related to the influence of knowledge management and the factors that influence it.

REFERENCE

[1] I. at. a. Nonaka, *Managing Flow A Process Theory of the Knowledge-Based Firm.* New York: Palgrave Macmillan UK, 2008.

[2] K. Y. Tung, *Memahami Knowledge Management*. Jakarta: Penerbit Indeks, 2018.

[3] S. Siswanto, "Systematic Review Sebagai Metode Penelitian untuk Mensintesis Hasil-Hasil Penelitian.," *Bul. Penelit. Sist. Kesehat.*, vol. 13, no. 4, pp. 326-333., 2010.

[4] F. Mailani, "Kualitas Hidup Pasien Ginjal Kronik yang Menjalani Hemodialisis : Systematic Review," *Ners J. Keperawatan*, vol. 11, no. 1, pp. 1–8, 2015.

[5] T. H. Nguyen, Que Thi Nguyet, Philip A Neck and Nguyen., "The Critical Role of Knowledge Management in Achieving and Sustaining Organisational Competitive Advantage," *Int. Bus. Res.*

[6] J. H. Tjakraatmadja, "Mengapa perlu 'Boosting business results through knowledge

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2021
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i2.1929

management'.," in *Knowledge Management Summit Indonesia*, 2015.
[7] A. Ambarwati, *Perilaku dan Teori Organisasi*, 1st ed. Malang: Media Nusa Creative, 2018.
[8] R. L. Daft, *Organization theory and design*, 9th editio. Mason, Oklahoma: Thomson South-Western., 2007.
[9] K. Dalkir, *Knowledge Management in Theory and Practice*. United Kingdom: Elsevier Butterworth-Heinemann, 2005.
[10] A. Z. & D. F. Zhou, "The Intellectual Capital Web A Systematic Linking of Intellectual Capital and Knowledge Management:," *J. Intellect. Cap.*, 2003.
[11] S. Kitchenham, B., & Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.
[12] I. K. Kusumawijaya, "Upaya Meningkatkan Kinerja Organisasi Sektor Publik Berbasis Knowledge Management," in *PROSIDING SEMINAR NASIONAL & CALL FOR PAPERS 2015 Optimalisasi Peran Industri Kreatif dalam Menghadapi Masyarakat Ekonomi ASEAN*, 2015, pp. 163–175.
[13] M. A. Jabar, "Knowledge Management di Sub Direktorat Penilaian Kinerja dan Pengembangan Karier Guru Pendidikan Dasar," *J. Improv.*, vol. 4, no. 2, pp. 254–264, 2017.
[14] T. H. K. Harahap, Y. A., & Harahap, "Penerapan Knowledge Management pada Badan Perencanaan Pembangunan Daerah (Bappeda) Kota Pekanbaru," *JOM FISIP RIAU*, vol. 3, no. 2, pp. 1–15, 2016.
[15] J. Prayitno, "Knowledge Management Dalam Pengembangan Smart City," in *Konferensi Nasional Sistem Informasi 2018*, 2018, pp. 473–477.
[16] T. T. Prabowo, "Implementasi Manajemen Pengetahuan Di Perpustakaan Uin Sunan Kalijaga Yogyakarta," *J. Kaji. Inf. dan Perpust.*, vol. 4, no. 2, p. 161, 2016.
[17] J. Husna, "Implementasi Knowledge Management di Perpustakaan dalam Membangun Koleksi Warisan Budaya Batik," *Pustakaloka*, vol. 10, no. 2, p. 146, 2018.
[18] E. R. Wulandari and R. Nurisani, "Model Knowledge Management di Perpustakaan Universitas Padjadjaran," *Lentera Pustaka J. Kaji. Ilmu Perpustakaan, Inf. dan Kearsipan*, vol. 6, no. 1, pp. 23–36, 2020.
[19] T. T. Prabowo, "Penerapan Manajemen Pengetahuan dalam Pengelolaan Dokumen di Perpustakaan STIKes Guna Bangsa," *Khazanah J. Pengemb. Kearsipan*, vol. 13, no. 1, p. 17, 2020.
[20] I. G. A. D. Saryanti and N. L. G. P. Suwirmayanti, "Analisis Meningkatkan Kinerja Dosen Menggunakan Knowledge Management System," in *Konferensi Nasional Sistem & Informatika 2017*, 2017, pp. 152–157.
[21] A. Heksarini, "Knowledge Management melalui Pelaksanaan Manajemen Sumber Daya Manusia dan Pembelajaran Organisasi pada Perbankan Syariah," in *Prosiding Seminar Nasional Manajemen dan Ekonomi Bisnis*, 2017, vol. 1, pp. 265–273.
[22] S. M. Kuway, "Peranan Manajemen Pengetahuan Untuk Keunggulan Daya Saing UKM di Kota Pontianak," in *SENSITIF 2019: Seminar Nasional Sistem Informasi dan Teknik Informatika*, 2019, pp. 1119–1126.
[23] W. Wiyanto and U. Pamulang, "Analisis Penerapan Manajemen Pengetahuan dan Pengetahuan Berbasis Strategi Untuk Menciptakan Keunggukan Bersaing Berkelanjutan (Studi Kasus Pada SMK YPUI Parung)," in *Konferensi Nasional Riset Manajemen*, 2016, no. September, pp. 20–22.
[24] N. K. Devi, "Penerapan Knowledge Management Dalam Peningkatan Kinerja Pada PT . Indominco Bontang Kalimantan Timur," in *Festival Riset Manajemen & Akuntansi*, 2018, pp. 883–893.
[25] Wahyudi, "Analisa Penerapan Knowledge Management Pada Perusahaan Service (Studi Kasus Pt . Pit Elektronik)," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 4, no. 1, pp. 37–42, 2018.
[26] M. Fermilia, D. Taufiq, and W. Laksmono, "Penerapan Manajemen Pengetahuan di BAPETEN Sebagai Implementasi dari Amanah Reformasi Birokrasi," in *Seminar Keselamatan Nuklir*, 2016, pp. 1–5.
[27] U. Rusilowati, "Analisis Manajemen Pengetahuan Berbasis Teknologi Informasi: Studi Kasus pada Lemlitbang Pemerintah Pengambil Kebijakan," *J. Organ. dan Manaj.*, vol. 11, no. 1, pp. 44–61, 2015.
[28] P. Pragiwani, M., Alexandri, M. B., & Chandra, "Analisis Employee Empowerment dan Knowledge Management Pada Peningkatan Kinerja Karyawan di Sektor Pariwisata," *AdBispreneur J. Pemikir. dan Penelit. Adm. Bisnis dan Kewirausahaan*, vol. 3, no. 1, pp. 73–79, 2018.
[29] L. Baety, Jumi., and Taviyastuti, "Analisis Pengaruh Pengawasan dan Manajemen Pengetahuan Terhadap Kinerja Karyawan Menggunakan Metode Regresi Berganda," in *Seminar Nasional Terapan Riset Inovativ (SENTRINOV) Ke-6*, 2020, vol. 6, no. 2, pp. 155–163.
[30] D. J. Bantam, D. A. Nugraha, and N. Sa'adah, "Pengaruh Budaya Organisasi Dan Komitmen Organisasi Terhadap Knowledge Management

Pada Perusahaan Pengguna SAP," *PSIKOLOGIKA*, vol. 21, no. 1, pp. 12–24, 2016.
[31] R. Bose, *Knowledge Management Matrics*. 2004.
[32] E. F. Badar and A. N. L. Seniati, "Pengaruh Trust Terhadap Berbagi Pengetahuan Melalui Mediasi Komitmen Organisasi pada Dosen Perguruan Tinggi," *Indig. J. Ilm. Psikol.*, vol. 2, no. 1, pp. 7–27, 2017.
[33] S. Lusa and R. Zulham Ramadhani, "Pengembangan Knowledge Management Systemberbasis Web Pada Perusahaan Manufaktur: Studi Kasus PT. Malindo Feedmill, Tbk," in *Seminar Nasional Riset Inovatif III*, 2015, pp. 408–420.
[34] M. Y. Lopis, A. Iriani, and Y. Dwikurnaningsih, "Pengembangan Aplikasi Knowledge Management System Online Untuk Berbagi Pengetahuan Antara Alumni Dengan Program Studi," *J. Ilmu Sos. dan Hum.*, vol. 8, no. 1, pp. 35–45, 2019.
[35] A. Widi, N. Rachmat, and S. Selatan, "Pengembangan Knowledge Management System dengan model SECI dan pendekatan Soft System Methodology," in *Annual Research Seminar 2016*, 2016, vol. 2, no. 1, pp. 281–286.
[36] Y. Kurniawan, O. Sarwengga, T. F. Dewanto, and W. T. Hantoro, "Manajemen Pengetahuan Berbasis Perangkat Lunak Open Source (Studi Kasus)," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 1, pp. 81–90, 2017.
[37] O. Octaria and E. Ermatita, "Analisis Knowledge Management System dengan Metode Inukshuk," in *Prosiding Annual Research Seminar 2017 Computer Science and ICT*, 2017, vol. 3, no. 1, pp. 35–38.
[38] O. Octaria, E. Ermatita, and S. Sukemi, "Penerapan Knowledge Management System Berbasis Web Menggunakan Model Inukshuk dan Algoritma Levenshtein," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 11, no. 2, pp. 63–73, 2020.
[39] O. Octaria, E. Ermatita, and S. Sukemi, "Penerapan Knowledge Management System Menggunankan Algoritma Levenshtein," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 233–242, 2019.
[40] D. Cahyadi and W. Usino, "Knowledge Management System Berbasis Web Sebagai Sistem Pengembangan Sumber Daya Manusia Dalam Pengelolaan Sumber Daya Air : Studi Kasus Perusahaan Umum (Perum) Jasa Tirta II," *J. REKAYASA Teknol. NUSA PUTRA*, vol. 2, no. 3, pp. 1689–1699, 2016.
[41] Herlinda., I. Mutia, and Atikah, "Perancangan Knowledge Management System (KMS) Kurikulum 2013 Sekolah Menengah Atas Negeri di Jakarta Selatan.," in *Seminar Nasional TEKNOKA*, 2017, vol. 2, pp. 1–7.
[42] R. Hidayad, S. Fauziati, and E. Nugroho, "Perancangan Knowledge Management System di Direktorat Jenderal Bea dan Cukai Kementerian Keuangan," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2019, vol. 1, no. 1, pp. 813–821.
[43] Y. A. A. Sukma, N. Indrawati, and R. Z. Ruslam, "Perancangan Knowledge Management System pada Komite Teknis," *Pertem. dan Present. Ilm. Stand.*, pp. 209–218, 2019.
[44] I. Mutia, - Herlinda, and - Atikah, "Perancangan Knowledge Management System (KMS) Kurikulum 2013 Menggunakan Model Tiwana dan Zack," *J. Edukasi dan Penelit. Inform.*, vol. 3, no. 2, p. 152, 2017.
[45] T. Ahmad, I. Hermadi, and W. A. Kusuma, "Sistem Manajemen Pengetahuan Perlindungan Anak: Studi Kasus: Sakti Peksos di Kementerian Sosial," *SOSIO KONSEPSIA*, vol. 8, no. 3, pp. 213–226, 2019.
[46] D. R. Pulukadang, M. Mustafid, and F. Farikhin, "Pendekatan Clustering untuk Ekstraksi Pengetahuan pada Pembangunan Sistem Manajemen Pengetahuan," *J. Sist. Inf. Bisnis*, vol. 5, no. 2, pp. 79–83, 2015.
[47] A. W. A. Wibowo, Y. Priyandari, and R. W. Damayanti, "Perancangan Model Knowledge Management Pada Unit Layanan Pengadaan Barang dan Jasa Pemerintah," in *Proceeding Seminar Nasional Teknik & Manajemen Industri*, 2015, pp. 11–56.
[48] Y. Sonatha, I. Rahmayuni, A. Alanda, and I. Saputra, "Rancang Bangun Aplikasi Knowledge Management Berbasis Web," *INVOTEK J. Inov. Vokasional dan Teknol.*, vol. 18, no. 2, pp. 133–140, 2018.
[49] S. Karnila and Nurfiana., "Perancangan Model Knowledge Management Untuk Memonitoring Prestasi Akademik Mahasiswa Pada Perguruan Tinggi," in *Konferensi Nasional SIstem & Informatika 2015*, 2015, pp. 101–106.
[50] I. Kaniawulan *et al.*, "Sistem Manajemen Pengetahuan Forum Diskusi Dosen Knowledge Management System of Lecturer Discussion Forum," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 51–58, 2020.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.2118

ANALYSIS OF SHOE ORDERING PATTERN AT PT. PRATAMA ABADI INDUSTRI USING THE APRIORI ALGORITHM

Retno Wati[1*]; Imron Imron[2]; Albert Riyandi[3]

Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
www.nusamandiri.ac.id
retnowati383@gmail.com [1], albert.abe@nusamandiri.ac.id [3]

Sistem Informasi
Universitas Bina Sarana Informatika
www.bsi.ac.id
imron.imr@bsi.ac.id

(*) Corresponding Author

Abstract - Inventory is one of the most important factors in ensuring the continuity of the production process. PT. Pratama Abadi Industri is an industrial company in the field of shoe sales. In its activities, PT. Pratama Abadi Industri was unable to provide production materials appropriately, causing a shortage of shoe material stock for the production of the type of shoe that the customer is most interested in. The method used in this research is to use the a priori Algorithm and to get accurate results, this research uses the Tanagra 1.4.50 software. The purpose of this study was to analyze the ordering of NIKE shoes that are often ordered by customers by looking at their previous order patterns. This a priori algorithm can produce 7 final association rules between items with a minimum value of 30% of support and a minimum of 70% of confidence and from May 2020 to October 2020 it is known that the most popular order pattern for NIKE shoes is NIKE shoes with the Tailwind (100%), Revolution 5 (100%), Air Max Plus (100%) and Classic Cortez Leather (100%) which can be used as reference materials in determining the next production material inventory.

Keywords: Apriori algorithm, Data mining, Ordering

Abstrak *- Persediaan merupakan salah satu faktor terpenting dalam menjamin kelancaran proses produksi. PT. Pratama Abadi Industri merupakan perusahaan industri dalam bidang penjualan sepatu. Dalam kegiatannya, PT. Pratama Abadi Industri tidak dapat menyediakan bahan material produksi secara tepat sehingga menyebabkan terjadinya kekurangan stok persediaan material sepatu untuk dilakukannya produksi dari jenis sepatu yang paling diminati oleh customer. Metode yang digunakan dalam penelitian ini adalah dengan menggunakan Algoritma apriori dan untuk mendapatkan hasil yang akurat, penelitian ini menggunakan software Tanagra 1.4.50. Tujuan dari penelitian ini adalah untuk menganalisis pemesanan sepatu NIKE yang sering dipesan oleh customer dengan melihat pola pemesanan sebelumnya. Algoritma apriori ini dapat menghasilkan 7 aturan asosiasi final antar item dengan nilai minimum support 30% dan minimum confidence 70% serta pada bulan Mei 2020 sampai dengan Oktober 2020 diketahui pola pemesanan sepatu NIKE yang paling diminati adalah sepatu NIKE dengan jenis Tailwind (100%), Revolution 5 (100%), Air Max Plus (100%) dan Classic Cortez Leather (100%) yang dapat dijadikan bahan acuan dalam menentukan persediaan material produksi selanjutnya.*

Kata Kunci: *Algoritma Apriori, Data Mining, Pemesanan*

INTRODUCTION

Economic growth in Indonesia is currently entering a turning point due to the deterioration of the COVID-19 pandemic. The economic growth has begun to show an increasing trend and has even passed its critical phase. The Indonesian economy in the first quarter of 2020 against the first quarter of 2019 grew by 2.97 percent [1]. The current development of the shoe industry tends to increase. The increase in the shoe industry is a great opportunity for shoe producers considering that Indonesia is a country with one of the largest populations in the world [2].

Shoes are one of the most important fashions in supporting appearance. The percentage of

expenditure on clothing, footwear, headgear per capita of the Indonesian people is 5.80% in a month [3]. PT. Pratama Abadi Industri is a company engaged in the fashion footwear industry in producing athletic shoes for the NIKE brand. In its business activities, the company carries out production based on customer orders or make to order. Types of orders or orders from customers have variations and types and fluctuating amounts.

The inventory system applied in the company is that orders are made at the beginning of each month with a lead time of four days and [4] Order quantity is based on estimates only. Companies sometimes experience an excess or lack of production material when a customer changes an order request for certain types [5]. To solve this problem, a data mining technique is needed in analyzing the habits of ordering customers in knowing the relationship between an item and another item in order to find out what types of NIKE shoes are the most often ordered simultaneously by customers. [6] Data mining is the process of analyzing data to find a pattern from the hidden data sets. The application of the apriori algorithm to data mining techniques is very efficient and can speed up the process of forming item combination patterns [7] then the test is carried out to whether the combination meets the minimum support and confidence parameters, which are the threshold values given by the user [8].

This study aims to analyze shoe orders for 6 months using the apriori algorithm and association rules. So that the results can be seen what type of shoes are often purchased simultaneously by customers during the 6 months.

Literature Research

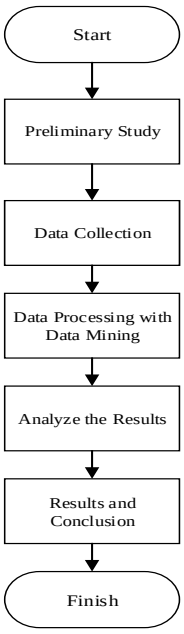
There are several previous studies related to the Apriori Algorithm. Research conducted by Anggraini, Putri, and Utami [9] Regarding the Implementation of the Apriori Algorithm in Determining the Most Interested Car Sales at Honda Permata Serpong, said that Data Mining is very useful to determine the relationship between the frequency pattern of Honda car sales that consumers are most interested in. The results of sales data for the Honda Permata Serpong 2018, with the highest support and confidence values are the Mobilio-Brio Satya (50% -100%), Jazz-HRV (33.33% -100%), and HRV-Mobilio-Brio Satya (33.33% -100%). With the results of the sales analysis at Honda Permata Serpong, it is hoped that it can regulate the stock of goods for marketing so that there is no accumulation of less desirable goods, which results in losses in the future. [10], suggested that the Apriori algorithm with minimum support of 30% and a minimum of 60% confidence can produce 6 final association rules which are very useful for making decisions in preparing stocks of what types of shoes are needed in the future. Subsequent research conducted by Lestari and Hafiz on the Application of the Apriori Algorithm to the Barbar Warehouse Sales Data [11], argues that Barbar Warehouse owners can find out which product categories are most sold each week in sales for 3 months by using apriori algorithm, namely the Bedding (bed sheet) and Watches (glasses) category with 100% confidence and support values. The a priori algorithm has a positive influence on the sales report of Barbar Warehouse because the owner can find out which products need to be reproduced or reduced in stock to increase sales turnover.

Based on the GAP analysis of the above research with this research using apriori algorithm is to determine the pattern of shoe orders by the customer which can be used as knowledge or a reference in determining efficient production material inventory so that the production process can run smoothly and can reduce additional costs such as costs of stock.

MATERIALS AND METHODS

Research Stages

Research is looking through a methodical process of adding to knowledge itself and with others, by the discovery of facts and unusual insights [12]. The following are the steps taken by the author in this research can be seen in figure 1 below:



Source: [13]

Figure 1. Stages of Research

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.2118

Explanation of the stages of the research:

a. Preliminary study
   The first step of this research is to find and study the problems that exist at PT. Pratama Abadi Industri then determines the scope of the problem, the background of the problem, and studies some literature related to the problem and how to find solutions to these problems.
b. Data Collection
   In this context, it is intended to find out important things related to this research. Data collection is done by making observations at PT. Pratama Abadi Industri. Also, researchers conducted interviews with parties related to this research. Then the authors took a sample of the order transaction database in May 2020 - October 2020 at PT. Pratama Abadi Industri to supporting this research.
c. Data Processing with Data Mining
   The data processing process is carried out to identify the problems and are often at PT. Pratama Abadi Industri, then an analysis of these problems is carried out for the author to get a solution, and then the algorithm can be used to obtain a solution. In the next stage, the authors use data mining techniques with a priori algorithms to get results as goals to be achieved.
d. Analyze the Results
   At this stage, the authors analyzed the results with the tools used as system testing, namely is Tanagra 1.4.50.
e. Results and Conclusion
   After being tested, the results of the analysis are between manual ways and testing using the software of data processing. The next step is to provide conclusions that can be made into new knowledge that has been obtained from the test results of a priori algorithm model.

Data collection technique

a. Observation
   The author made observations at PT. Pratama Abadi Industri. The author observes the business process and procedures that run in the company and collecting the data and information required in this research.
b. Interview
   The author collects the data by interacting and communicating directly with Mr. Jaelani as the Warehouse Staff at PT. Pratama Abadi Industri to provide the necessary information in writing this research.
c. Library Study
   This method is carried out by the author are reading and quoting books, journals, and other sources related to modeling using the data mining and a priori algorithms.

Data analysis method

To achieve the goals of this research, the author uses quantitative data analysis methods. [14] Quantitative data analysis is a data analysis that is used if the conclusions obtained can be proven by numbers and a formula that is related to the analysis is also used in the calculations. In analyzing the data, the author uses a priori algorithm calculation.

RESULTS AND DISCUSSION

Data analysis

In this research, the stages began with data analysis to apply the apriori algorithm in analyzing orders at PT. Pratama Abadi Industri. Data on this research are using secondary data on shoe orders obtained from PT. Pratama Abadi Industri as shown in table 1 below:

Table 1. List of Types NIKE Shoes

No.	Types of NIKE Shoes
1	Air Max Plus
2	Air Zoom
3	Classic Cortez Leather
4	Flyease
5	Revolution 5
6	SB Air Trainer I ISO
7	Speedrep
8	Tailwind
9	TN Supreme
10	Waffle Trainer
11	Zoom 2K

Source: [15]

The next stage is grouping the 3 types of NIKE shoes that are most ordered by customers based on the historical data of PT. Pratama Abadi Industri from May 2020 to October 2020, resulting in a transaction pattern as shown in table 2 below:

Table 2. Transaction Patterns of Ordering NIKE Shoes

Month	Itemset
5	Tailwind, Revolution 5, Air Max Plus
6	Waffle Trainer, Zoom 2K, Air Max Plus
7	Classic Cortez Leather, Air Max Plus, Revolution 5
8	Tailwind, Revolution 5, Air Max Plus
9	Zoom 2K, Classic Cortez Leather, Revolution 5
10	Speedrep, Zoom 2K, Revolution 5

Source: [15]

Then, the tabulation format of each month, if created will be like are in table 3 below:

Table 3. Tabulation Format of Data Transaction

Month	Tailwind	Revolution 5	Air Max Plus	Waffle Trainer	Zoom 2K	Classic Cortez	Speedrep
5	1	1	1	0	0	0	0
6	0	0	1	1	1	0	0
7	0	1	1	0	0	1	0
8	1	1	1	0	0	0	0
9	0	1	0	0	1	1	0
10	0	1	0	0	1	0	1

Source: [15]

Frequency Pattern Analysis

Itemset Formation

This stage looks for a combination of items that qualify the minimum requirements of the support value in the database. The process of forming C_1 or called is 1 itemset with a minimum value of support = 30% which produces the itemset as in table 4 below:

Table 4. Itemset with a Minimum Support of 30%

Itemset	Total	Support
Tailwind	2	33.33%
Revolution 5	5	83.33%
Air Max Plus	4	66.67%
Zoom 2K	3	50.00%
Classic Cortez Leather	2	33.33%

Source: [16]

After the calculation results from C_1 are obtained, it is followed by the process of forming C_2 or called is 2 itemset with a minimum support value = 30% which produces the itemset as in table 5 below:

Table 5. Itemset with a Minimum Support of 30%

Itemset	Total	Support
Tailwind, Revolution 5	2	33.33%
Tailwind, Air Max Plus	2	33.33%
Revolution 5, Air Max Plus	3	50.00%
Revolution 5, Zoom 2K	2	33.33%
Revolution 5, Classic Cortez Leather	2	33.33%

Source: [16]

The last stage is to continue the process of forming C3 or called is 3 itemset with a minimum support value = 30% which produces an itemset as in table 6 below:

Table 6. Itemset with a Minimum Support of 30%

Itemset	Total	Support
Tailwind, Revolution 5, Air Max Plus	2	33.33%

Source: [16]

Final Association Rules

After all high-frequency patterns are found (C_1, C_2, and C_3), then the association rules are looked for with the results of the frequency patterns shown in table 7 below:

Table 7. Results of Qualifying High-Frequency Patterns

Itemset	Support
Tailwind, Revolution 5	33.33%
Tailwind, Air Max Plus	33.33%
Revolution 5, Air Max Plus	50.00%
Revolution 5, Zoom 2K	33.33%
Revolution 5, Classic Cortez Leather	33.33%
Tailwind, Revolution 5, Air Max Plus	33.33%

Source: [16]

The next step is to look for association rules that qualify the minimum requirements for confidence, by calculating confidence or associative A → B, with a minimum value of confidence = 70% which results in the itemset as in table 8 below:

Table 8. Results of Confidence or Association

Rules	Confidence	
If you order Tailwind, it will order Revolution 5	2/2	100%
If you order Revolution 5, you order Tailwind	2/5	40%
If you order Tailwind, you will order Air Max Plus	2/2	100%
If you order Air Max Plus, you will order Tailwind	2/4	50%
If you order Revolution 5, you will order Air Max Plus	3/5	60%
If you order Air Max Plus, you will order Revolution 5	3/4	75%
If you order Revolution 5, it will order the Zoom 2K	2/5	40%
If you order Zoom 2K, it will order Revolution 5	2/3	66.67%
If you order Revolution 5, you will order Classic Cortez Leather	2/5	40%
If you order Classic Cortez Leather, you will order Revolution 5	2/2	100%

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.2118

Rules	Confidence	
If you order Tailwind, you will order Revolution 5 and Air Max Plus	2/2	100%
If you order Revolution 5, you will order Tailwind and Air Max Plus	2/5	40%
If you order Air Max Plus, you will order Revolution 5 and Tailwind	2/4	50%
If you order Tailwind and Revolution 5, you will order Air Max Plus	2/2	100%
If you order Tailwind and Air Max Plus, you will order Revolution 5	2/2	100%
If you order Air Max Plus and Revolution 5, you will order Tailwind	2/3	66.67%

Source: [16]

The final step is the formation of a final association rule that is obtained based on predetermined minimum support and minimum confidence, so can be seen in table 9 and the diagram below:

Table 9. Final Association Rules

Rules	*Support*	*Confidence*
If you order Tailwind, it will order Revolution 5	33.33%	100%
If you order Tailwind, you will order Air Max Plus	33.33%	100%
If you order Air Max Plus, you will order Revolution 5	50.00%	75%
If you order Classic Cortez Leather, you will order Revolution 5	33.33%	100%
If you order Tailwind, you will order Revolution 5 and Air Max Plus	33.33%	100%
If you order Tailwind and Revolution 5, you will order Air Max Plus	33.33%	100%
If you order Tailwind and Air Max Plus, you will order Revolution 5	33.33%	100%

Source: [16]

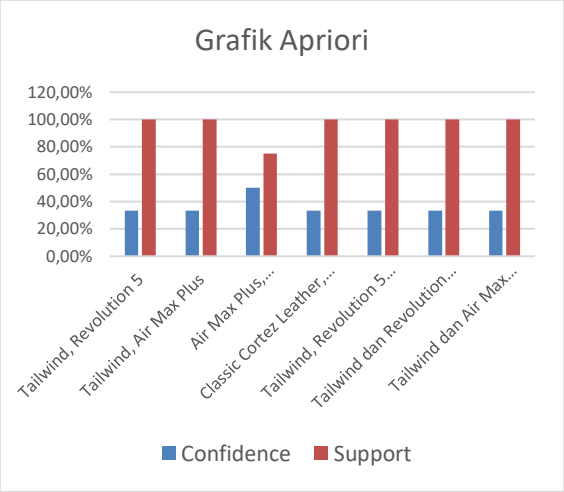


Figure 2. Final Association Rules Diagram for the Most Ordered Type of Shoes

Algorithm Calculation with Tanagra 1.4.50

1. Algorithm Support

In determining support which consists of input, process, and output. Following are the results of support based on the results of the analysis carried out and can be seen in figure 3 and 4 below:

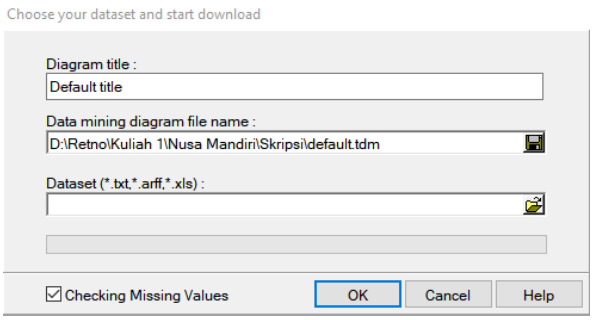


Figure 3. Data Retrieval Process

Figure 3 above is the process of retrieving data from Microsoft Excel to Tanagra 1.4.50 for the data processing.

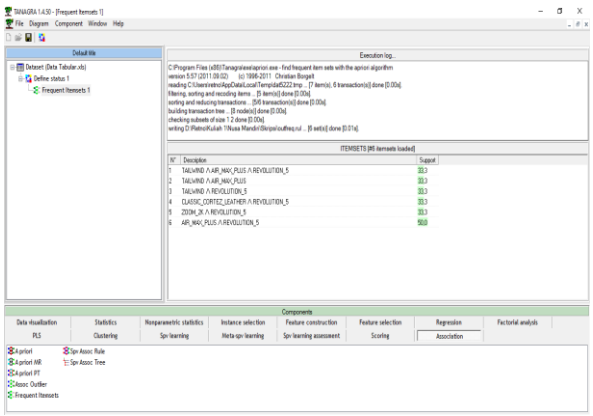


Figure 4. Display Support

Figure 4 above is the result of the support that has been previously determined by the author.

2. Algorithm Confidence

In determining confidence which consists of input, process, and output. Following are the results of the implementation using Tanagra 1.4.50 and can be seen in figure 5 and 6 below:

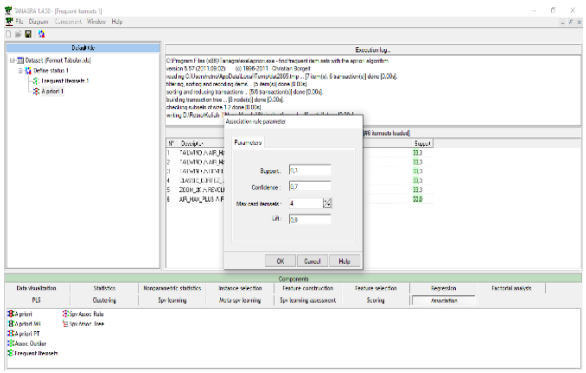


Figure 5. Determining the Value of Confidence

RULES

Number of rules : 7

N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"TAILWIND=true"	"REVOLUTION 5=true" - "AIR MAX PLUS=true"	2,00000	33,333	100,000
2	"TAILWIND=true"	"AIR MAX PLUS=true"	1,50000	33,333	100,000
3	"REVOLUTION 5=true" - "TAILWIND=true"	"AIR MAX PLUS=true"	1,50000	33,333	100,000
4	"TAILWIND=true"	"REVOLUTION 5=true"	1,20000	33,333	100,000
5	"CLASSIC CORTEZ LEATHER=true"	"REVOLUTION 5=true"	1,20000	33,333	100,000
6	"AIR MAX PLUS=true" - "TAILWIND=true"	"REVOLUTION 5=true"	1,20000	33,333	100,000
7	"AIR MAX PLUS=true"	"REVOLUTION 5=true"	0,90000	50,000	75,000

Computation time : 78 ms.
Created at 03/03/2021 02:48:53

Figure 6. Results of the Final Association Rule

Figure 6 above is the result of the final association of shoe ordering patterns at PT. Pratama Abadi Industri. So, based on the picture above, the application of the a priori algorithm can result in 7 strong rules associated with the most popular products by the customer are Tailwind, Revolution 5, Air Max Plus, and Classic Cortez Leather.

CONCLUSION

In this research, the a priori algorithm was successfully applied for the search of shoe ordering patterns at PT. Pratama Abadi Industri. This can be seen from the results of the research which shows that the most ordered types of shoes are ordered by customers at PT. Pratama Abadi Industri, by looking at products that qualify for minimum support and minimum confidence. From the highest final association rules, it is known that if you order Tailwind you will order Revolution 5 with 33.33% support and 100% confidence. If you order Tailwind, you will order Air Max Plus with 33.33% support and 100% confidence. If you order Classic Cortez leather, you will order Revolution 5 with 33.33% support and 100% confidence. If you order Tailwind, you will order Revolution 5 and Air Max Plus with 33.33% support and 100% confidence. If you order Tailwind and Revolution 5, you will order Air Max Plus with 33.33% support and 100% confidence. So based on the results of this research, knowing the type of NIKE shoes that customers are most interested in can be used as knowledge or reference for PT. Pratama Abadi Industri to determining the supply of production materials optimally and efficiently which can ensure the continuity of the production process on increasing the company's turnover in the future.

REFERENCE

[1] BPS, "Pertumbuhan Ekonomi Indonesia Triwulan I-2020," *Badan Pus. Stat.*, pp. 1–12, 2020.

[2] A. S. Rizal, N. L. W. S. Telagawathi, N. N. Yulianthini, and K. E. S. Putra, "Pengaruh Citra Merek Dan Harga Terhadap Loyalitas Pelanggan Sepatu Merek Nike Di Singaraja Futsal," *J. Prospek*, vol. 1, no. 2, pp. 34–40, 2019.

[3] S. Supriyanto, "Pengeluaran Rata-Rata Per Kapita Sebulan Menurut Kelompok Barang Indonesia," *Badan Pusat Statistik*, 2019. [Online]. Available: https://www.bps.go.id/statictable/2009/06/15/937/persentase-pengeluaran-rata-rata-per-kapita-sebulan-menurut-kelompok-barang-indonesia-1999-2002-2018.html.

[4] A. Bachtiar, "Inventory Control Indirect Material: EOQ Model, Efektivitas Produksi," *J. Ekon. dan Bisnis*, vol. 5, no. 2, pp. 103–113, 2017.

[5] J. Efendi, K. Hidayat, and R. Faridz, "Analisis Pengendalian Persediaan Bahan Baku Kerupuk Mentah Potato dan Kentang Keriting Menggunakan Metode Economic Order Quantity (EOQ)," *Junal Perfoma Media Ilm. Tek. Ind.*, vol. 18, no. 2, pp. 125–134, 2019.

[6] R. Yanto and H. Di Kesuma, "Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule," *J. Jatisi*, vol. 4, no. 1, pp. 1–10, 2017.

[7] E. L. Febrianti and A. Suryadi, "Penerapan Data Mining Dengan Algoritma Apriori Untuk Analisis Pola Pembelian Konsumen," *Semin. Nas. R.*, vol. 1, no. 1, pp. 173–176, 2018.

[8] M. Safii and A. Trydillah, "Implementasi Data

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti
No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.2118

Mining Dalam Menentukan Pola Pembelian Obat Dengan Metode Algoritma Apriori," *J. Manaj. Inform. Komputerisasi Akuntasi*, vol. 3, no. 1, pp. 66–71, 2019.

[9] D. Anggraini, S. A. Putri, and L. A. Utami, "Implementasi Algoritma Apriori Dalam Menentukan Penjualan Mobil Yang Paling Diminati Pada Honda Permata Serpong," *J. Media Inform. Budidarma*, vol. 4, no. April, pp. 302–308, 2020.

[10] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *J. Tek. Komput.*, vol. 4, no. 1, pp. 156–161, 2018.

[11] A. F. Lestari and M. Hafiz, "Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse," *J. Inovtek Polbeng*, vol. 5, no. 1, pp. 96–105, 2020.

[12] N. E. Putria, "Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel," *J. CBIS*, vol. 06, no. 01, pp. 29–39, 2018.

[13] A. F. Budiantara and C. Budihartanti, "Implementasi Data Mining Dalam Manajemen Inventory Pada PT. Mastersystem Infotama Menggunakan Metode Algoritma Apriori," *J. PROSISKO*, vol. 7, no. 1, pp. 26–31, 2020.

[14] M. A. Rahma, "Implementasi Metode Simple Additive Weighting (SAW) Sebagai Pendukung Keputusan Pembelian Sepeda Motor," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 4, no. 1, pp. 125–130, 2018.

[15] P. P. A. Industri, Laporan Transaksi Pemesanan Sepatu NIKE. 2020.

[16] R. Wati, I. Imron, and A. Riyandi, Laporan Akhir Penelitian: Analisis Pola Pemesanan Sepatu Pada PT. Pratama Abadi Industri Dengan Menggunakan Algoritma Apriori. Jakarta, 2020.

VOL. 6. NO. 2 FEBRUARY 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v6i2.2118