# COMPARISON OF DATA MINING CLASSIFICATION ALGORITHM FOR PREDICTING THE PERFORMANCE OF HIGH SCHOOL STUDENTS

Tiska Pattiasina<sup>1</sup>; Didi Rosiyadi<sup>2</sup>

<sup>1, 2</sup> Masters in Computer Science STMIK Nusa Mandiri Jakarta, Indonesia <sup>1</sup>elleoratiska07@gmail.com, <sup>2</sup> didi.rosiyadi@gmail.com

Abstract— Data Mining is a series of processes to explore added value in the form of unknown information manually from the database. In the world of data mining education can be used to obtain information about student performance. In this study the researchers took research samples from class XI (eleven) students at SMAN 3 Ambon by classifying student performance based on thirteen attributes, namely: age, sex, school organization, extracurricular activities, pocket money, duration of study at home, duration of social media, online game duration, attendance, illness, permits, semester 1 and semester 2 grades. Using the KDD (Knowledge Discovery Database) method and classification algorithm that will be used, namely, decision tree, Naïve Bayes and K-Nearest Neighbor. And then do the test using k-fold cross validation.

**Essence** — In the world of data mining education can be utilized to obtain student performance information. This research presents the results of new decision making using data mining techniques. This study aims to compare the Decision Tree, Naive Bayes, and k-Nearets Neighbor (k-NN) algorithms to improve accuracy in student performance at Ambon State High School 3. And the research method used is the classification method by comparing two algorithms, Naive Bayes and K-nearest neighbor. In this research, the highest accuracy is obtained in the Decision Tree algorithm, which is 99.6047%.

*Keywords*: Data mining, classification, Decision *Tree*, Naive Bayes, KNN.

#### **INTRODUCTION**

Data Mining is the extraction of important or interesting information or patterns from existing data in a large database (Siregar & Pusphabuana, 2017). And data mining is a mixture of statistics, artificial intelligence, and database research that is still developing (Gorunescu, 2011). Education Data Mining also referred to as EDM is defined as a field of scientific inquiry centered around developing methods to make discoveries in unique types of data that originate from educational settings, and use these methods to better understand students and the settings they learn (Peterson, Penelope L; Baker, Eva; McGaw, 2010)

In the world of education research using data mining has been carried out by several researchers to provide increased excellence in higher education, which creates human resources is an ongoing subject. Therefore, the prediction of high school level student performance is very important for continuing education, because the quality of the teaching process can be provided according to student needs. In this case the data and information collected to be able to maintain the quality of students.

Some research in the world of education has been carried out, namely the first research Prediction of Student Performance Using Decision Tree C 4.5 Algorithm, the study aims to predict achievement based parents' student on socioeconomic status, student discipline and student achievement using data mining methods with the Decision Tree, CHAID, Regression algorithm Multiple. Research subjects were students at SD Negeri 4 Trimuloyo. after recording data obtained 352record. Based on data analysis using a decision tree, data mining to predict student achievement based on the socioeconomic status of parents, student discipline and student achievement using data mining methods obtained the following results: economic variables are variables that determine the potential for student success. or not learn achievements in the future. This is evidenced by the variables that become the root node in the decision tree formed. Student achievement variable, is the second important variable in the success of student studies. This shows that aspects of students' knowledge or intelligence are very influential on the success of their learning. Conversely, even though students have less predictable knowledge, high willpower can still be achieved at least in the b or c category. The average C4.5 success algorithm in carrying out classification data reaches 99.43% in accuracy. This shows that this algorithm has a reliable performance in doing classification (Kuntoro & Sudarwanto, 2017)

The second study with the title using data mining to predict secondary school student

performance, the beginning of this study discusses the level of education of the Portuguese population which has increased in the last decade but it turns out statistical results make Portugal remain at the tail end of Europe because of the high rate of student failure. In particular, the lack of success in the Mathematics and Potugic core classes is very serious. On the other hand, the field of Business Intelligence (BI) or / Data Mining (DM), which aims to extract high-level knowledge from raw data, offers interesting algorithmic automation that can help the education field. This study aims to approach student achievement in secondary education using the BI or / Data Mining technique. Current real-world data (eg student grades, demographics, social and school features) are collected using school reports and questionnaires. Two core classes (ie Mathematics and Portuguese) are modeled under binary / five-level classification and regression tasks. Also, four DM models (namely Decision Tree, Random Forest, Neural Network and Support Vector Machine) and three input choices (for example with and without previous values) are tested. The results show that good predictive accuracy can be achieved, provided the grades for the first and / or second school periods are available. Although student performance has been strongly influenced by past evaluations, explanatory analysis has shown that there are also other relevant features such as the amount of absenteeism, work and parental education, alcohol consumption (Cortez & Silva, 2008).

The third study, entitled the K-Nearest Neighbor (k-NN) Algorithm Model For Student Graduation Prediction, predicting the rise in the ability of students to complete studies on time is one element of university accreditation assessment This research was conducted to predict student graduation using data mining classifications with processing student data obtained from 1633 student databases, with attributes nim, name, age, faculty, semester 1 to IP semester 8 using the K-Nearest Neighbor algorithm by clustering data k = 1, k-2, k-3, k = 4, and k = 5 (Rohman, 2015).

In this study the researchers took a sample of research from class XI students of SMA Negeri 3 Ambon by classifying student learning outcomes based on age, organization and extracurricular activities followed by school students, pocket money earned by students from parents, duration or the amount of time students used to study, social media, play online games, attendance, permits and illness during 2 semesters, average semester 1 and 2. Using the KDD (Knowledge Discovery Database) method and classification algorithm to be used, namely Decision Tree, Naïve Bayes and K-Nearest Neighbor use weka tools. And then testing is done using k-fold cross validation, then evaluating and validating with a confusion matrix, precision, recall, ROC.

The study aims to compare the Decision Tree, Naive Bayes, and k-Nearets Neighbor (k-NN) algorithms to improve accuracy in student performance at Ambon State High School 3.

#### **MATERIALS AND METHODS**

#### A. Datasets

The data used in this research is real-world data collected from class XI report cards (eleven) with a total of 253 students and questionnaire results that have been filled out by students of SMA Negeri 3 Ambon. Data taken in the form of name, class, age, gender, organization at school, extracurricular, pocket money, duration of study, duration of social media, duration of online games, information on attendance, illness, permission, semester value 1, semester value 2 and label sample raw data can be seen in table 1. The data consists of two labels namely; Very Good (A) with a total of 81 and Good (B) with a total of 172.

No	Name	Age	Gen	School organization	Extra- curricular	Pocket money	Duration of study at home	Social media duration	Duration of online	Total present	Total sick	Tot Permi- ssion	Sem = 1	Sem =2	L
1	Dini Erna Karlina	15	2	2	1	15000	3 at nonne	2	games 2	240	0	0	85	86	В
2	Shanty	14	2	2	2	35000	2	3	0	237	3	0	81	81	В
3	Muhammad Joseph	14	1	2	2	20000	1	4	1	240	0	0	81	81	В
4	Fatahilah Bayu	16	1	2	1	15000	1	1	1	240	0	0	85	86	В
5	Akmal A	15	1	2	1	10000	1	2	2	240	0	0	80	81	В
6	Taftaniza	16	2	1	2	12000	4	1	2	236	4	0	83	84	В
7	Giansun	15	2	2	1	20000	1	3	0	237	2	1	80	83	В
8	Mirna	16	2	2	2	10000	1	2	0	240	0	0	81	82	В

Table 1. Data SMA Negeri 3 Ambon

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Data Mining...* Techno Nusa Mandiri : Journal of Computing and Information Technology As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018** 

9	Ananda	15	1	1	1	15000	2	2	2	237	2	1	80	80	В
10	Ilhamni Zein	15	2	2	2	50000	3	2	0	240	0	0	91	92	А
11	M Rival	16	1	1	2	30000	1	1	1	240	0	0	81	81	В
12	Nissaeryn Z	15	2	2	1	15000	2	2	0	240	0	0	91	91	А
13	Umayyah Novita	15	2	2	2	25000	2	0	0	240	0	0	91	92	А
14	Giannisa R	16	2	2	2	20000	2	1	0	240	0	0	81	82	В
15	Safni Tuara	16	2	1	2	20000	2	1	0	240	0	0	81	82	В
16	Veni	15	2	2	2	20000	1	2	3	240	0	0	91	92	А
17	Armianti Juenda Anaci	15	2	2	2	15000	1	2	0	240	0	0	81	83	В
18	Aldo Sahetapy	15	1	2	1	15000	2	1	1	240	0	0	81	82	В
19	Roy A	16	1	2	1	25000	1	2	2	240	0	0	83	84	В
20	Souhoka Michelle Manane	15	2	1	1	35000	2	1	0	240	0	0	80	81	В
253	Estrella L Mararessy	15	2	2	2	15000	2	2	0	240	0	0	70	76	В

Source : (Pattiasina & Rosiyadi, 2020)

Following are the attributes that are inputted on the Weka tools can be seen in Figure 1.





#### B. Method

This study is to classify the learning outcomes of class XI (eleven) students with the research methodology used in this study is to use the Knowledge Discovery in Database (KDD) method consisting of 5 stages: (Sugianto, 2015):



Source : (Sugianto, 2015)

Figure 2. Research Methods

- 1. Determine the target data which includes data selection and focus more on the subset of data.
- 2. Cleaning and integration, for data cleaning, which removes noise and inconsistent data, while Integration is combining data from a variety of different sources.
- 3. Selection and transformation, in data selection that is taking data in accordance with the task of analysis from the database, while Data transformation, which combines data into a form or model suitable for excavation through summary or aggregation operations.
- 4. Data mining is an important and primary process for extracting patterns from data with more recent methods.
- 5. Pattern evaluation, which identifies interesting patterns and represents knowledge based on interestingness measures.
- 6. Knowledge presentation, in the form of presentation of knowledge that is extracted and presented to users using visualization and

knowledge representation techniques (Sugianto, 2015).

Then the research steps that will be carried out can be seen in the picture below:





In the research steps above can be explained as follows:

- a. In accordance with the objectives of the study, which is to implement and compare the Decision Tree, Naïve Bayes and K-Nearest Neighbor algorithms to improve accuracy in student performance at Ambon 3 High Schools.
- b. Then do a search of data to the curriculum section of Ambon 3 High School as data to be processed and look for literature studies that are in accordance with research.
- c. Data obtained from the curriculum is then processed using the classification method with the Decision Tree algorithm, naive bayes and knearest neighbor.
- d. Data were tested using validation to get the best accuracy from the algorithm.
- e. The results of the algorithm are compared in order to get the best algorithm results, which will be used as a reference in knowing student learning outcomes.

The results of the implementation of the – data classification are evaluated and validated so that it can be known how accurate the prediction results will be. Evaluation and validation can be done using confunsion matrix and receiver operating characteristics (ROC).

Confusion matrix is a method used to calculate accuracy in the concept of data mining.

Evaluation with confunsion matrix produces accuracy, precision, and recall values. Accuracy in classification is the presentation of the accuracy of data records that are classified correctly after testing the classification results. Precision or confidence is the proportion of positively predicted mattresses that is also positive true to the actual data. Recall or sensitivity is the proportion of positive cases that are actually correctly predicted correctly (Hand, 2007). This method uses a matrix table as in table 3.

Class Pred	Act	tual
	+	-
+	True positive(A)	False negative(B)
-	False positive(C)	True negative(D)
Source: (Hand, 2	007)	

The ROC (Receiver Operating Characteristic) curve is another way to evaluate the accuracy of the classification visually (Gorunescu, 2011). The accuracy can be diagnosed as follows: Accuracy 0.90 - 1.00 = Excellent classification Accuracy 0.80 - 0.90 = Good classification Accuracy 0.70 - 0.80 = Fair classification Accuracy 0.60 - 0.70 = Poor classification Accuracy 0.50 - 0.60 = Failure

# **RESULTS AND DISCUSSION**

Classification is done to discuss the performance of high school students using weka with the decision tree algorithm, naive bayes, k-nearest neighbor. The classification test is shown by cross validation using 10 fold.

#### A. Decision Tree

Decision Tree can be seen from the probability that will affect student performance. In the Decision Tree the data is expressed in tabular form with attributes and records. Decision tree process flow is to change the form of table data into a model tree, change the tree into a rule and simplify it. The decision tree algorithm stages are; prepare training data, determine roots and trees, calculate the gain value. The results can be seen in table 3.

Table 3. Entrophy and gain values

			1 2	0	
The Knot		В	Α	ENTROPY	GAIN
TOTAL	253	172	81	0.90455330	
Age					1.7544107
14 Year	7	6	1	0.59167277	
15 Year	120	85	35	0.87086446	
16 Year	114	73	41	0.94239154	
17 Year	10	8	2	0.72192809	

		В	Α	ENTROPY	GAIN	The Knot		В	Α	ENTROPY	GAIN
18 Year	2	0	2	0		Total sick					0.0066497
total	253	172	81			0-4	250	169	81	0.90867838	
Gender					1.0682747	5 – 9	3	3	0	0	
1. Man	102	68	34	0.91829583		Total	253	172	81		
2. Woman	151	104	47	0.89462059		Total permiss	ion				0.0065187
total	253	172	81			0 - 6 Kali	252	172	80	0.901598235	
School organ	ization				1.5041040	7 - 13 kali	1	0	1	0	
1 Yes	42	28	14	0.91829583		total	253	172	81		
2. No	211	144	67	0.90168129		Semester 1 g	rades				1.5505393
Total	253	172	81			70 – 79	40	40	0	0	
Extracurricu	lar				1.3034146	80 - 89	187	132	55	0.873981048	
1 Yes	72	50	22	0.88797632		90 -100	26	0	26	0	
2. No	181	122	59	0.91075254		Total	253	172	81		
Total	253	172	81			Semester 2 gi	ades				1.518359
Pocket money	/				1.1154092	70 – 79	13	13	0	0	
5000 -	91	57	34	0.95341587		80 – 89	204	159	45	0.76124015	
10.000						90 -100	36	0	36	0	
11.000- 20.000	122	89	33	0.84216948		Total	253	172	81	Ŭ	
21.000 up	40	26	14	0.93406805					-		
Total	253	172	81			Source: (Pa	ttiasina	a & Ro	siyad	1,2020)	
Duration of st	udy at h	ome			1.0789076	As	can be	seen	in ta	able 3 that t	he attribu
1 Hour	100	68	32	0.90438148		with the	highes	st gaiı	n val	ue is the to	otal prese
						which is 1	L.7758	2, the	n th	e total prese	nt becom
2 Hour	115	82	33	0.86475726		which is 1 the root n	l.7758 10de. 7	2, the The fo	en the llowi	e total prese ng accuracy	ent becom results ca
2 Hour 3 Hour	115 25	82 12	33 13	0.86475726 0.99884556		which is 1 the root n be seen in	L.7758 Iode. 7 Figur	2, the The fo e 4, u	n the llowi sing	e total prese	ent becom results ca
2 Hour 3 Hour 4 Hour	115 25 13	82 12 10	33 13 3	0.86475726		which is 1 the root n	L.7758 Iode. 7 Figur	2, the The fo e 4, u	n the llowi sing	e total prese ng accuracy	ent becom results ca
2 Hour 3 Hour 4 Hour Total	115 25 13 253	82 12	33 13	0.86475726 0.99884556	1 7225611	which is 1 the root n be seen in decision th @ Weka Explorer	1.7758 iode. 1 Figur ree alg	2, the The fo e 4, u orithr	en the llowi sing n.	e total prese ng accuracy the Weka to	ent becom results ca
2 Hour 3 Hour 4 Hour Total Social media o	115 25 13 253 duration	82 12 10 172	33 13 3 81	0.86475726 0.99884556 0.77934987	1.7335611	which is 1 the root n be seen in decision tr	1.7758 iode. 1 Figur ree alg	2, the The fo e 4, u orithr	en the llowi sing n.	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No	115 25 13 253 duration 9	82 12 10 172 5	33 13 3 81 4	0.86475726 0.99884556 0.77934987 0.99107606	1.7335611	which is 1 the root n be seen in decision tr @ Wela Explorer Preprocess Classify	L.7758 lode. T Figur ree alg	2, the The fo e 4, u orithr	en the llowi sing n.	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour	115 25 13 253 duration 9 89	82 12 10 172 5 61	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977	1.7335611	which is 1 the root n be seen in decision tr	L.7758 lode. T Figur ree alg	2, the The fo e 4, u orithr	en the llowi sing n.	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour	115 25 13 253 duration 9 89 84	82 12 10 172 5 61 59	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099	1.7335611	which is 1 the root n be seen in decision th Weba Explorer Preprocess Classify Classifier	L.7758 lode. T Figur ree alg	2, the The fo e 4, u orithr	en the llowi sing n.	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour 3 Hour	115 25 13 253 duration 9 89 84 39	82 12 10 172 5 61 59 24	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665	1.7335611	which is a the root m be seen in decision th wear Explorer Preprocess Classifier Classifier Choose J48-C0.25 Use training set Supplied test set	L.7758 node. 1 Figur cee alg Cluster Associ	2, the The for e 4, u orithr ate Select attr Classifier o	en the llowi sing n.	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour 3 Hour 4 Hour	115 25 13 253 duration 9 89 84 39 32	82 12 10 172 5 61 59	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099	1.7335611	which is a the root m be seen in decision th wear Explorer Preprocess Classify Classifier Choose Use training set Supplied test set e Cross-validation For	L.7758 node. 1 Figur ree alg Cluster Associ	2, the 'he fo e 4, u orithr ale Select atr Classifier o Time ta 	en the llowi sing n. butes Visu	e total prese ng accuracy the Weka to	nt becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour 3 Hour 4 Hour Total	115 25 13 253 duration 9 89 84 39 32 253	82 12 10 172 5 61 59 24 23 172	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665	1.7335611	which is 1 the root m be seen in decision th Webs Explorer Preprocess Classify Classifier Choose J48-C0.25 Test options Use training set Suppled test set © Cross-validation Fe	L.7758 node. 1 Figur cee alg Cluster Associ	2, the Che fo e 4, u orithr ale Selectatr Classifier o Time ta Sup Correct	en the llowi sing n. butes Visu utput ken to buil atified cro mary === ly Classifi	e total prese ng accuracy the Weka to wa wa d model: 0.27 seconds ss-validation === ed Instances 252	ent becom results ca ols with J4
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour 3 Hour 4 Hour Total	115 25 13 253 duration 9 89 84 39 32 253	82 12 10 172 5 61 59 24 23 172	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665	0.7057835	which is 1 the root m be seen in decision th weak Explore Preprocess Classify Classifier Choose J48-C0.25 Test options Use training set Supplied test set e Cross-validation Fo Percentage split	L.7758 node. 1 Figur cee alg Cluster Associ	2, the Che fo e 4, u orithr	en the llowi sing m. butes Visu utput ken to buil atified cro mary === ly Classific totly Classific totly classific	e total prese ng accuracy the Weka to alter d model: 0.27 seconds se-validation === ed Instances 252 fied Instances 1 0.39	99.6047 %
2 Hour 3 Hour 4 Hour Total Social media o No 1 Hour 2 Hour 3 Hour 4 Hour Total	115 25 13 253 duration 9 89 84 39 32 253	82 12 10 172 5 61 59 24 23 172	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665		which is 1 the root m be seen in decision th weak Explore Preprocess Classify Classifier Choose J48-C0.25 Test options Use training set Supplied test set e Cross-validation Fo Percentage split	L.7758 node. 1 Figur cee alg Cluster Associ	2, the Che fo e 4, u orithr ate Select atr Time ta Sum Correct Incorre Kappa s Kort mean ab	en thi Billowi sing n. butes Visu utput ken to bull atified crcc many == ly Classifi tatistic solute error an quared	e total prese ng accuracy the Weka to alize d model: 0.27 seconds ss-validation === ed Instances 252 fied instances 1 0.99 r 0.000 error 0.06	99.6047 k 0.3953 k
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total	115 25 13 253 duration 9 89 84 39 32 253 ame onl	82 12 10 172 5 61 59 24 23 172 ine	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847		which is 1 the root m be seen in decision tr version tr	L.7758 node. T Figur ree alg Cluster Associ -M 2 Set. No Stop	2, the Che fo e 4, u orithr	en the sing of the sing of the sing of the sing of the sing of the upper the sing of the s	e total prese ng accuracy the Weka to wile d model: 0.27 seconds s=-validation === ed Instances 1 fied Instances 1 fied Instances 1 fied Instances 0.06 error 0.06 error 0.06	99.6047 % 0.3953 %
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151	82 12 10 172 5 61 59 24 23 172 ine 105	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847		which is 1 the root m be seen in decision tr version tr	L.7758 node. 1 Figur cee alg Cluster Associ -M2 Set. Nds 10 % 66 Stop ptionsy	2, the Che fo e 4, u orithr ale Selectatr Time ta e Str Cassifier o Time ta e Str correct kappa s Mean ab Relativ Root me Relativ	en the sing of n. utput utput ken to buill atified core ranzy	e total prese ng accuracy the Weka to withe d model: 0.27 seconds se-validation === ed Instances 19 r 0.99 r 0.00 error 0.06 error 1.76 error 1.76 stances 253	99.6047 % 0.3953 %
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No 1 Hour	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48	82 12 10 172 5 61 59 24 23 172 ine 105 33	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823		which is 1 the root m be seen in decision tr version tr	L.7758 node. 1 Figur ree alg Cluster Associ -M2 Set set stop ptions) tttiasina	2, the Che fo e 4, u orithr ate Select atr Classifier o Time ta Select atr Time ta Select atr Correct Ropt me Relative Root me Root me	en thi sing m. butes visue utput tken to buill atified crc solute error solute error solute error an anguarde abolute lative squa mumber of In siyad	e total prese ng accuracy the Weka to withe d model: 0.27 seconds se-validation === ed Instances 19 r 0.99 r 0.00 error 0.06 error 1.76 error 1.76 stances 253	99.6047 % 0.3953 %
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 4 Hour Total Duration of g No 1 Hour 2 Hour 2 Hour	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38	82 12 10 172 5 61 59 24 23 172 ine 105 33 30	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.88688405 0.89603823 0.74248757		which is a the root m be seen in decision tr weatsplore Preprocess Classifier Choose J48 - C 0.26 Test options Use training set Supplied test set © Cross-validation For Percentage split More option (Nom) LABEL Start Result list (right-click for oo Source: (Pa Figure	L.7758 node. T Figur ree alg Cluster Associ Stop stor Stop ptons ttiasina re 4. A	2, the Che for e 4, u orithr ale Select attr Classifier o Time ta Select attr Classifier o Time ta Select attr Classifier o Time ta Select attr Rean ab Rean ab	en thi sing on the sing on the bules were the utput were to built atified error were the site of the s	e total prese ng accuracy the Weka to alter d model: 0.27 seconds se-validation === ed Instances 252 fied Instances 1 r 0.000 error 0.060 error 1.3.56 stances 253 (i,2020) sults on weka	99.6047 % 0.3953 %
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour Total Duration of g No 1 Hour 2 Hour 3 Hour 3 Hour 3 Hour 3 Hour 1 Hour 3 Hour	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38 9	82 12 10 172 5 61 59 24 23 172 ine 105 33 30 3 30 3	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> <li>6</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823 0.74248757 0.91829583		which is a the root m be seen in decision th weatspore Preprocess Classifier Choose Jule - Co 22 Test options Use training set Cross-validation Fo Percentage split Result list (right-click for Source: (Pa Figure In figure	Cluster Associations Cluster Associations Cluster Association Stop Stop ptons) ttiasina re 4. A 4 the	2, the Che for e 4, u orithr ate Select attr Classifier o Time ta Sum Classifier o Time ta Sum Classifier o Time ta Sum Classifier o Time ta Sum Classifier o Time ta Sum Classifier o Classifier o Classifi	en thi sing m. 	e total prese ng accuracy the Weka to atte d model: 0.27 seconds es-validation == ed Instances 252 fied Instances 253 fied Instances 253 r 0.009 r 0.000 stances 253 (j.2020) sults on weka in j48 is 9	99.6047 t a tools 9.6047, ti
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No 1 Hour 2 Hour 3 Hour 4 Hour 2 Hour 3 Hour 4 Hour 7 Total	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38 9 7 253	82 12 10 172 5 61 59 24 23 172 ine 105 33 30 3 1	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> <li>6</li> <li>6</li> <li>6</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823 0.74248757 0.91829583		which is 1 the root m be seen in decision th wear Explorer Preprocess Classifier Choose Jue - Core Supplied test set Cross-validation For Percentage split More option (Nom) LABEL Start Figure Choose Jue - Core Source: (Pa Figure	Cluster Association Cluster Association Cluster Association Cluster Association Stop ptons) tttiasina re 4. A 4 the ion mail	2, the che for e 4, u orithr ate Select attr Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Classifier o Time ta e Sum Classifier o Classifier o	en thi llowi sing n. ubutes [ Visu utput ken to buil atified oro many ly Classified a bablute the sister of t	e total prese ng accuracy the Weka to atte d model: 0.27 seconds as-validation == ed instances 252 fied instances 1 r 0.000 pr 0.000 prore 1.75 red error 1.75 stances 253 tiances 253 tiances 255 tiances 2555 tiances 2555	99.6047 k 0.3953 k 0.3953 k 0.3953 k 14 14 14 15 14 14 15 15 14 15 15 15 15 15 15 15 15 15 15
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No 1 Hour 2 Hour 3 Hour 4 Hour 2 Hour 3 Hour 1 Hour 2 Hour 1 Hour 2 Hour 1 Total Total	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38 9 7 253	82 12 10 172 5 61 59 24 23 172 ine 105 33 30 3 1	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> <li>6</li> <li>6</li> <li>6</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823 0.74248757 0.91829583	0.7057835	which is 1 the root m be seen in decision th wear Explorer Preprocess Classifier Choose Jue - Core Supplied test set Cross-validation For Percentage split More option (Nom) LABEL Start Figure Choose Jue - Core Source: (Pa Figure	L.7758 node. T Figur ree alg Cluster Associ 	2, the che for e 4, u orithr ate Select attr Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Time ta e Sum Classifier o Classifier o Time ta e Sum Classifier o Classifier o	en thé llowi sing n. uput uput ken to buil atified cre tatistic alutie squared e abolute atified colute error silvad cy res: racy racy racy an bé urati	e total prese ng accuracy the Weka to atte d model: 0.27 seconds es-validation == ed Instances 252 fied Instances 253 fied Instances 253 r 0.009 r 0.000 stances 253 (j.2020) sults on weka in j48 is 9	99.6047 k 0.3953 k 0.3953 k 0.3953 k 14 14 14 15 14 14 15 15 14 15 15 15 15 15 15 15 15 15 15
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No 1 Hour 2 Hour 3 Hour 4 Hour Total Total Total Total Total	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38 9 7 253	82 12 10 172 5 61 59 24 23 172 ine 105 33 30 3 1 172	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> <li>6</li> <li>6</li> <li>81</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823 0.74248757 0.91829583 0.59167279	0.7057835	which is 1 the root m be seen in decision tr weat Explore Preprocess Classify Classifier Choose Jue-C0.25 Test options Use training set Supplied test set Percentage split More option (Nom) LABEL Start Result list (right-click foro Source: (Pa Figure configurat Tal	Cluster Associations Cluster Associations Cluster Associations Stop ptions) ttiasina re 4. A 4 the cion matching ble 4. (Control of the triangle of the the triangle of the triangle of the the triangle of the the triangle of the the triangle of the triangle of the	2, the Che for e 4, u orithr ate Select attr Classifier of Time ta e Select attr Classifier of Time ta e Str e	en thé llowi sing n. uput uput ken to buil atified cre tatistic alutie squared e abolute atified colute error silvad cy res: racy racy racy an bé urati	e total prese ng accuracy the Weka to d model: 0.27 seconds s=-validation == ed Instances 1 fied Instances 2 fied Instances 1 fied Instances 1 fied Instances 2 fied Instance	99.6047 k 0.3953 k 0.3954 k 0.3955 k 0.395
2 Hour 3 Hour 4 Hour Total Social media of No 1 Hour 2 Hour 3 Hour 4 Hour Total Duration of g No 1 Hour 2 Hour 3 Hour 4 Hour 4 Hour 3 Hour 4 Hour	115 25 13 253 duration 9 89 84 39 32 253 ame onl 151 48 38 9 7 253 3	82 12 10 172 5 61 59 24 23 172 ine 105 33 30 3 1 172 2	<ul> <li>33</li> <li>13</li> <li>3</li> <li>81</li> <li>4</li> <li>28</li> <li>25</li> <li>15</li> <li>9</li> <li>81</li> <li>46</li> <li>15</li> <li>8</li> <li>6</li> <li>81</li> <li>1</li> </ul>	0.86475726 0.99884556 0.77934987 0.99107606 0.89841977 0.87836099 0.96123665 0.85714847 0.88688405 0.89603823 0.74248757 0.91829583 0.59167279	0.7057835	which is 1 the root m be seen in decision th wear Explorer Preprocess Classifier Choose Jue - Core Supplied test set Cross-validation For Percentage split More option (Nom) LABEL Start Figure Choose Jue - Core Source: (Pa Figure	Cluster Association Cluster Association Cluster Association Cluster Association Set Cluster	2, the Che for e 4, u orithr ate Select attr Classifier on Time ta Select attr Correct Incorrect Incorrect Reata ba Root me Relativ Root me Accurate accurate Config	en thé llowi sing n. uput uput ken to buil atified cre tatistic alutie squared e abolute atified colute error silvad cy res: racy racy racy an bé urati	e total prese ng accuracy the Weka to and a couracy the Weka to and a couracy and a co	99.6047 k 09.6047 k 0.3953 k 0.3954 k 0.39

Following are the results of the ROC graph on the J48 decision tree algorithm in the Weka tools can be seen in Figure 5.



Source: (Pattiasina & Rosiyadi,2020) Figure 5. ROC results from decision tree j48

The results of ROC processing for j48 decision tree algorithm is 0.9946 with a diagnosis level of Excellent classification.

# B. Naive Bayes

The following classification results using Naive Bayes can be seen in Table 5 using Weka tools with Naive Bayes algorithm.

Table 5. Results of the Naive Bayes classification

Attribute	Class			
	В	Α		
	(0.68)	(0.32)		
Age				
Mean	15.4826	15.6173		
std. dev.	0.6423	0.6771		
weight sum	172	81		
Precision	1	1		
Gender				
Mean	1.6047	1.5802		
std. dev.	0.4889	0.4935		
weight sum	172	81		
Precision	1	1		
School organization				
Mean	1.8372	1.8272		
std. dev.	0.3692	0.3781		
weight sum	172	81		
Precision	1	1		
Extracurricular				
Mean	1.6977	1.7284		
std. dev.	0.4718	0.4448		
weight sum	172	81		
Precision	1	1		
Pocket money				
Mean	16882.69	16372.55		
std. dev.	7686.855	7522.873		
weight sum	172	81		
precision	2647.059	2647.059		
Duration of study at home				
mean	1.7907	1.8395		
std. dev.	0.8086	0.8234		
weight sum	172	81		
precision	1	1		
Social media duration	_	_		
mean	1.9942	1.963		
std. dev.	1.0702	1.0823		

B         A           (0.68)         (0.32)           weight sum precision         172         81           precision         1         1           Duration of game online mean         0.6163         0.9012           std. dev.         0.8716         1.2727           weight sum         172         81           precision         1         1           Total present         1         1           mean         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total present         172         81           precision         1.1818         1.1818           Total sick         0         0.4622         0.0741           std. dev.         1.2459         0.4015         weight sum         172         81           precision         1.5         1.5         1.5         1.5           Total permission         172         81         172         81           precision         2.1667         2.1667         2.1667           Semester 1 grades         1.	Attribute	Class	
weight sum         172         81           precision         1         1           Duration of game online	—	В	Α
Interface         Interface           precision         1         1           Duration of game online         0.6163         0.9012           std. dev.         0.8716         1.2727           weight sum         172         81           precision         1         1           Total present         172         81           mean         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total present         0.4622         0.0741           weight sum         172         81           precision         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         mean         80.5465         88.9037		(0.68)	(0.32)
Duration of game online         0.6163         0.9012           std. dev.         0.8716         1.2727           weight sum         172         81           precision         1         1           Total present         172         81           mean         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total sick         172         81           mean         0.4622         0.0741           std. dev.         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total sick         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 gr	weight sum	172	81
mean         0.6163         0.9012           std. dev.         0.8716         1.2727           weight sum         172         81           precision         1         1           Total present         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total sick         mean         0.4622         0.0741           std. dev.         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         mean         80.5465         88.9037           std. dev.         3.1834         3.3689         weight sum         172         81           precision         1.2         1.2         12         81           precision         1	precision	1	1
std. dev.       0.8716       1.2727         weight sum       172       81         precision       1       1         Total present       239.222       239.6027         std. dev.       1.7283       1.5401         weight sum       172       81         precision       1.1818       1.5401         weight sum       172       81         precision       1.1818       1.1818         Total sick       mean       0.4622       0.0741         std. dev.       1.2459       0.4015       9         weight sum       172       81       9         precision       1.5       1.5       1.5         Total permission       1.5       1.5       1.5         mean       0.189       0.2407       std. dev.       0.7694       1.5226         weight sum       172       81       9       1.5226       81         precision       2.1667       2.1667       2.667       2.667       2.667       2.667       2.667       2.667       2.667       2.667       3.689       3.689       3.689       3.689       3.689       3.689       3.689       3.689       3.689       3.689<	Duration of game online		
weight sum       172       81         precision       1       1         Total present       239.222       239.6027         std. dev.       1.7283       1.5401         weight sum       172       81         precision       1.1818       1.5401         weight sum       172       81         precision       1.1818       1.1818         Total sick       0.4622       0.0741         std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission       1.5       1.5         mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades       mean       80.5465       88.9037         std. dev.       3.1834       3.3689       3.1834       3.3689         weight sum       172       81       172       81         precision       1.2       1.2       1.2         Semester 2 grades       mean       82.1618       89.7557	mean	0.6163	0.9012
Initial         Initial         Initial           precision         1         1           Total present         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total sick         0         0           mean         0.4622         0.0741           std. dev.         1.2459         0.4011           std. dev.         1.2459         0.4011           weight sum         172         81           precision         1.5         1.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         mean         80.5465         88.9037           std. dev.         3.1834         3.3689           weight sum         172         81           precision         1.2         1.2           Semester 2 grades         mean         82.1618	std. dev.		1.2727
Total present       2       239.222       239.6027         std. dev.       1.7283       1.5401         weight sum       172       81         precision       1.1818       1.1818         Total sick       0.4622       0.0741         mean       0.4622       0.0741         std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission       1.5       1.5         mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades       mean       80.5465       88.9037         std. dev.       3.1834       3.3689       3.1834       3.3689         weight sum       172       81       1.2       1.2         Semester 2 grades       1.2       1.2       1.2         mean       82.1618       89.7757       std. dev.       2.513       3.248	weight sum	172	81
mean         239.222         239.6027           std. dev.         1.7283         1.5401           weight sum         172         81           precision         1.1818         1.1818           Total sick             mean         0.4622         0.0741           std. dev.         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades             mean         80.5465         88.9037           std. dev.         3.1834         3.3689           weight sum         172         81           precision         1.2         1.2           Semester 2 grades          1.2           mean         82.1618         89.757           std. dev.         2.513         3.248	precision	1	1
std. dev.       1.7283       1.5401         weight sum       172       81         precision       1.1818       1.1818         Total sick           mean       0.4622       0.0741         std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission       1.5       1.5         mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades       mean       80.5465       88.9037         std. dev.       3.1834       3.3689       weight sum       172       81         precision       1.2       1.2       12       12         Semester 1 grades       1.2       1.2       12         mean       80.5465       88.9037       3.1834       3.3689         weight sum       172       81       12       12         Semester 2 grades       1.2       1.2       12         mean       82.1618       89.7557	Total present		
weight sum       172       81         precision       1.1818       1.1818         Total sick           mean       0.4622       0.0741         std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission       1.5       1.5         mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades           mean       80.5465       88.9037         std. dev.       3.1834       3.3689         weight sum       172       81         precision       1.2       1.2         Semester 2 grades           mean       82.1618       89.7577         std. dev.       2.513       3.248	mean	239.222	239.6027
precision         1.1818         1.1818           Total sick	std. dev.	1.7283	1.5401
Total sick       0.4622       0.0741         mean       0.4622       0.0741         std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades       mean       80.5465       88.9037         std. dev.       3.1834       3.3689       weight sum       172       81         precision       1.2       1.2       12       81         mean       80.5465       88.9037       std. dev.       3.1834       3.3689         weight sum       172       81       172       81         precision       1.2       1.2       1.2         Semester 2 grades       mean       82.1618       89.7577         std. dev.       2.513       3.248	weight sum	172	81
mean         0.4622         0.0741           std. dev.         1.2459         0.4015           weight sum         172         81           precision         1.5         1.5           Total permission         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         mean         80.5465         88.9037           std. dev.         3.1834         3.3689         3.689           weight sum         172         81         172         81           precision         2.163         3.248         3.689	precision	1.1818	1.1818
std. dev.       1.2459       0.4015         weight sum       172       81         precision       1.5       1.5         Total permission           mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades           mean       80.5465       88.9037         std. dev.       3.1834       3.3689         weight sum       172       81         precision       1.2       12         Semester 2 grades           mean       82.1618       89.7757         std. dev.       2.513       3.248	Total sick		
weight sum     172     81       precision     1.5     1.5       Total permission     1.5     1.5       mean     0.189     0.2407       std. dev.     0.7694     1.5226       weight sum     172     81       precision     2.1667     2.1667       Semester 1 grades     80.5465     88.9037       std. dev.     3.1834     3.3689       weight sum     172     81       precision     1.2     1.2       Semester 2 grades     1.2     1.2       mean     82.1618     89.7757       std. dev.     2.513     3.248	mean	0.4622	0.0741
Interprecision         I.5         I.5           Total permission         1.5         1.5           mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         80.5465         88.9037           std. dev.         3.1834         3.3689           weight sum         172         81           precision         1.2         1.2           Semester 2 grades         1.2         1.2           mean         82.1618         89.7757           std. dev.         2.513         3.248	std. dev.	1.2459	0.4015
Total permission         mean       0.189       0.2407         std. dev.       0.7694       1.5226         weight sum       172       81         precision       2.1667       2.1667         Semester 1 grades       80.5465       88.9037         std. dev.       3.1834       3.3689         weight sum       172       81         precision       1.2       1.2         Semester 2 grades       1.2       1.2         mean       82.1618       89.7757         std. dev.       2.513       3.248	weight sum	172	81
mean         0.189         0.2407           std. dev.         0.7694         1.5226           weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades         80.5465         88.9037           std. dev.         3.1834         3.3689           weight sum         172         81           precision         1.2         12           Semester 2 grades         12         12           mean         82.1618         89.7757           std. dev.         2.513         3.248	precision	1.5	1.5
std. dev.     0.7694     1.5226       weight sum     172     81       precision     2.1667     2.1667       Semester 1 grades         mean     80.5465     88.9037       std. dev.     3.1834     3.3689       weight sum     172     81       precision     1.2     1.2       Semester 2 grades         mean     82.1618     89.7757       std. dev.     2.513     3.248	Total permission		
weight sum         172         81           precision         2.1667         2.1667           Semester 1 grades	mean	0.189	0.2407
precision         2.1667         2.1667           Semester 1 grades	std. dev.	0.7694	1.5226
Semester 1 grades         mean       80.5465       88.9037         std. dev.       3.1834       3.3689         weight sum       172       81         precision       1.2       1.2         Semester 2 grades       82.1618       89.7757         std. dev.       2.513       3.248	weight sum	172	81
mean         80.5465         88.9037           std. dev.         3.1834         3.3689           weight sum         172         81           precision         1.2         1.2           Semester 2 grades          89.7757           std. dev.         2.513         3.248	precision	2.1667	2.1667
std. dev.     3.1834     3.3689       weight sum     172     81       precision     1.2     1.2       Semester 2 grades     82.1618     89.7757       std. dev.     2.513     3.248	Semester 1 grades		
weight sum         172         81           precision         1.2         1.2           Semester 2 grades          12           mean         82.1618         89.7757           std. dev.         2.513         3.248	mean	80.5465	88.9037
Image: precision         1.2         1.2           Semester 2 grades         100         100         100           mean         82.1618         89.7757         89.7757           std. dev.         2.513         3.248	std. dev.	3.1834	3.3689
precision         1.2         1.2           Semester 2 grades         82.1618         89.7757           std. dev.         2.513         3.248	weight sum	172	81
Semester 2 grades         82.1618         89.7757           std. dev.         2.513         3.248	-	1.2	1.2
std. dev. 2.513 3.248	-		
	mean	82.1618	89.7757
weight sum 172 81	std. dev.	2.513	3.248
	weight sum	172	81
precision 1.1667 1.1667	•	1.1667	1.1667

Source: (Pattiasina & Rosiyadi,2020)

The following accuracy results for the Naive Bayes algorithm can be seen in Figure 5.

Time taken to build model: 0.03 seconds

 Stratified	cross-validation	===
 Summary ===		

Correctly Classified Instances	238		94.0711 %
Incorrectly Classified Instances	15		5.9289 %
Kappa statistic	0.8677		
Mean absolute error	0.0787		
Root mean squared error	0.2147		
Relative absolute error	18.0463	8	
Root relative squared error	46.013	§.	
Total Number of Instances	253		

Source: (Pattiasina & Rosiyadi,2020) Figure 6. Results of NB accuracy on weka tools

In Figure 6 the accuracy of Naive Bayes is 94.0711, the details of the configuration matrix can be seen in Table 6.

Table 6. Configuration matrix results						
	True B	True A	Class			
			Pecission			
Pred. B	160	12	98.20%			
Pred. A	3	78	86.70%			
Class recall	93.00%	96.30%				
с (р. ш.	· • • • •	1. 00000				

Source: (Pattiasina & Rosiyadi,2020)

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | Comparison of Data Mining... Techno Nusa Mandiri : Journal of Computing and Information Technology As an Accredited Journal Rank 4 based on SK Dirjen Risbang SK Nomor 21/E/KPT/2018 Following are the results of the ROC graph on the Naive Bayes algorithm on the Weka tools can be seen in Figure 7.





Figure 7. ROC results from naive bayes The results of ROC processing for naive bayes algorithm is 0.9686 with a diagnosis level of Excellent classification.

### C. K-Nearest Neighbor

The following results can be seen in the accuracy of Figure 8, using weka tools with k-NN algorithm.

Time taken to build model: 0.02 seconds

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

Correctly Classified Instances	210	83.004	ł
Incorrectly Classified Instances	43	16.996	ł
Kappa statistic	0.6108		
Mean absolute error	0.1728		
Root mean squared error	0.4105		
Relative absolute error	39.646 %		
Root relative squared error	87.9804 %		
Total Number of Instances	253		

Source: (Pattiasina & Rosiyadi,2020) Figure 8. Accuracy of k-NN on weka tools

In Figure 8 the accuracy of the K-NN is 83.004, the detailed configuration matrix can be seen in table 8.

Table 8. Configuration matrix results				
	True B	True A	Class	
			Pecission	
Pred. B	150	22	87.70%	
Pred. A	21	60	73.20%	
Class recall	87.20%	74.10%		

Source: (Pattiasina & Rosiyadi,2020)

Following are the results of the ROC graph on the K-NN algorithm on the Weka tools can be seen in Figure 9.



Source: (Pattiasina & Rosiyadi,2020) Figure 9. ROC results from K-NN

The results of ROC processing for K-NN algorithm is 0.8152 with a diagnosis level of Good classification.

#### D. Evaluation of Testing Results

After testing the tools for three algorithms, namely decision tree, naive bayes and k-nearest neighbor, showed different levels of different results. In testing the decision tree algorithm J48 has a higher accuracy that is 99.6047%. Even for other criteria such as Precision and recall, it is still above the two algorithms, namely Naive Bayes and K-NN.

94.10%

96.86%

83.00%

81.52%

Table 9. Testing DT, NB and k-NN Algorithms						
	Decision	Naive Bayes	K-NN	Ī		
	Tree	-				
Accuracy	99.6047%	94.0711%	83.004%			
Precission	99.60%	94.50%	83.10%			

Source: (Pattiasina & Rosiyadi,2020)

99.60%

99.46%

Recall ROC

The following can be seen in Figure 10, a comparison chart of the prediction results of decision tree, naive bayes and K-NN algorithms.



Source: (Pattiasina & Rosiyadi,2020) Figure 10. Graph of comparison of algorithm results.

Overall, the J48 decision tree algorithm is one type of algorithm that fits the researcher's data, so that it can obtain a high level of accuracy.

# CONCLUSION

This research was taken from class XI students of SMA Negeri 3 Ambon with a dataset of 253 students. There were fourteen attributes used in the form of age, sex, organization involved in school, extracurricular activities, pocket money, duration of study, duration of social media, duration of playing online games, information on attendance, illness, permission, semester grades one and two. After that, use the Knowledge Discovery in Databes (KDD) data mining method. with the classification algorithm Decision tree [48, Naive Bayes and K-Nearest Neighbor. After the data were processed and labeled, 172 students were classified as good (B) and 81 students as very good (A). For accuracy based on the application Weka can be seen as follows: Accuracy using the Decision Tree Algorithm is 99.6047%. The accuracy using Naive Bayes algorithm is 94.0711% and K-Nearest Neighbor is 83.004% so the Decision Tree J48 algorithm has a higher accuracy value than the Naive Bayes algorithm and K-Nearest Neighbor.

# REFERENCES

- Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. 15th European Concurrent Engineering Conference 2008, ECEC 2008 - 5th Future Business Technology Conference, FUBUTEC 2008, 2003(2000), 5–12.
- Gorunescu, F. (2011). No Title. In Data Mining: Concepts, Models and Techniques. New York:

P-ISSN: 1978-2136 | E-ISSN: 2527-676X | *Comparison of Data Mining...* Techno Nusa Mandiri : Journal of Computing and Information Technology As an Accredited Journal Rank 4 based on SK **Dirjen Risbang SK Nomor 21/E/KPT/2018** 

Springer-Verlag.

- Hand, D. J. (2007). Principles of data mining. In *Drug Safety* (Vol. 30, Issue 7). https://doi.org/10.2165/00002018-200730070-00010
- Kuntoro, R. K., & Sudarwanto, R. (2017). Prediction Of Student Performance Using Decision Tree C 4.5 Algorithm. 214–219.
- Peterson, Penelope L; Baker, Eva; McGaw, B. (2010). A Survey on Feature Selection Methods For Imbalanced Datasets. *International Encyclopedia of Education*. https://www.scholars.northwestern.edu/en/ publications/international-encyclopedia-ofeducation
- Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1), 1–9. https://doi.org/10.1017/CB0978110741532 4.004
- Siregar, M., & Pusphabuana, A. (2017). No Title. In Data Mining: Pengolahan Data Menjadi Informasi dengan Rapidminer. https://books.google.com/books?hl=en&lr= &id=rTlmDwAAQBAJ&oi=fnd&pg=PR7&dq= Data+Mining+adalah+serangkaian+proses+u ntuk+menggali+nilai+tambah+berupa+infor masi+yang+selama+ini+tidak+diketahui+sec ara+manual+dari+suatu+basis+data.+Inform asi+yang+dihasilkan+dip
- Sugianto, C. A. (2015). Penerapan Teknik Data Mining Untuk Menentukan Hasil Seleksi Masuk Sman 1 Cibeber Untuk Siswa Baru Menggunakan Decision Tree. *Tedc*, 9, 39–43.