JURNAL

ILMU PENGETAHUAN & TEKNOLOGI KOMPUTER

Vol. 8. No. 2 February 2023

ISSN: 2685-8223 (Printed)

ISSN: 2527-4864 (Online)

LPPM nusamandiri

Publisher:
Lembaga Penelitian dan Pengabdian Masyarakat Universitas Nusa Mandiri
Jl. Jatiwaringin Raya No. 02 RT 08 RW 013 Kelurahan
Cipinang Melayu Kecamatan Makassar Jakarta Timur 13620
Phone: 021 28534471
http://ejournal.nusamandiri.ac.id/index.php/jitk/index

EDITORIAL BOARD

Advisors	:	Chairman of Universitas Nusa Mandiri
Responsible Person	:	Chairman of LPPM Universitas Nusa Mandiri
Editor In Chief	:	Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom Profile: ID Scopus: 56242366200 | Universitas Nusa Mandiri
Associate Editors	:	Ruhul Amin, M.Kom Profile: Scopus ID: 57224841038 | Universitas Nusa Mandiri
Board of Editors	:	Prof. Dr. Eng. Rendy Thamrin Profile: Scopus ID: 16048432600 | Universitas Andalas
		Dr. Rice Novita Profile: Scopus ID: 57202978035 | Universitas Islam Negeri Sultan Syarif Kasim Riau
		Dr. Karmila Suryani Profile: Scopus ID: 57194111577 | Universitas Bung Hatta
		Dr. Titin Winarti Profile: Scopus ID: 57192922506 | Universitas Semarang
		Dr. Yenny Desnelita Profile: Scopus ID : 57200088180 | Institut Bisnis dan Teknologi Pelita Indonesia
		Dr. Raimon Efendi,S.A.B,M.Kom Profile: Scopus ID : 57208861363 | Universitas Dharmas Indonesia
		Riandy Mardhika Adif, MM Profile: Scopus ID 57209549911 | Universitas Islam Negeri Imam Bonjol Padang
		Idria Maita, M.Sc Profile: Scopus ID 57204425016 | Universitas Islam Negeri Sultan Syarif Kasim Riau
		Rosda Syelly, M.Kom. Profile: Scopus ID 57209684900 | Sekolah Tinggi Teknologi Payakumbuh
		Alfa Saleh, M.Kom Profile: Scopus ID 57200210551 | Universitas Potensi Utama
		Dinar Ajeng Kristiyanti, M.Kom Profile: Scopus ID 57191189495 | Universitas Multimedia Nusantara

Vol. 8. No. 2 February 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v8i2

Ninuk Wiliani, M.Kom
Profile: Scopus ID 57212350451 | Institut Teknologi & Bisnis BRI

Anggi Oktaviani, M.Kom
Profile: Scopus ID 57205210550 | Universitas Nusa Mandiri

Eni Irfiani, M. Kom
Scopus ID: 57210461376 | Universitas Bina Sarana Informatika

Esron Rikardo Nainggolan, M.Kom
ID Scopus: 57208281339 | Universitas Nusa Mandiri

Ari Pambudi, M.Kom
Profile: Scopus ID - | Universitas Indonusa Esa Unggul

Taufik Hidayatulloh, M.Kom
Profile: ID Scopus - | Universitas Bina Sarana Informatika

Nurajijah, M.Kom
Profile: ID Scopus: 57208282684 | Universitas Nusa Mandiri

Daning Nur Sulistyowati, M.Kom
Profile: Scopus ID: 57208283626 | Universitas Nusa Mandiri

Sanjaya Pinem, S.Kom, M.Sc
Profile: Scopus ID: -| Politeknik Negeri Media Kreatif

Muhammad Fadlan, S.Kom., M.Kom.
Profile: Scopus ID: 57221493210 | STMIK PPKIA Tarakanita Rahmawati

Anita Hidayati, S.Kom, M.Kom
Profile: Scopus ID: 57201446053 | Politeknik Negeri Jakarta

Evita Fitri, M.Kom
Profile: Scopus ID: -| Universitas Nusa Mandiri

Reviewers : Prof. Dr. Jufriadif Na`am
Profile: ID Scopus: 57189371499 | SAFE-Network - INDONESIA

Prof. Dr. dr. Johan Harlan
Profile: ID Scopus: 56156931200 | Universitas Gunadarma Jakarta

Prof. Dr. Novizar Nazir
Profile: Scopus ID: 35503777500 | Universitas Andalas

Dr. Didi Rosiyadi, M.Kom
Profile: ID Scopus: 55355563700 | Pusat Penelitian Informatika LIPI

Dr Hilman Ferdinandus Pardede, S.T, MEICT
Profile: ID Scopus: 55351397700 | Pusat Penelitian Informatika LIPI

Dr. Foni Agus Setiawan, M.Kom
Profile: ID Scopus: 56669695700 | Universitas Ibnu Khaldun Bogor, Indonesia

Dr. Windu Gata, M.Kom
Profile: ID Scopus: 57193213766 | Universitas Nusa Mandiri

Dr. Aida Firdaus Muhammad Nurul Azmi
Profile: ID Scopus: 55037673000 | Universiti Teknologi MARA

Dr. Donny Maha Putra
Profile: ID Scopus: 57195753813 | Kementerian Keuangan Republik Indonesia

Dr. Muhardi
Profile: ID Scopus: 57218939131 | STMIK Hang Tuah Pekanbaru

Dr. Asrul Huda
Profile: ID Scopus: 57002969300 | Universitas Negeri Padang

Dr. Ing. Insannul Kamil
Profile: ID Scopus: 56149407000 | Universitas Andalas

Dr. Dolly Indra
Profile: Scopus ID: 57196343003| Universitas Muslim Indonesia Makassar

Dr. Wahyudi
Profile: Scopus ID: 57196343003| STMIK Indonesia Padang

Dr. Yusra
Profile: Scopus ID: 57196343003| Universitas Bung Hatta

Dr. Vitriani
Profile: Scopus ID: 57196343003| Universitas Muhammadiyah Riau

Dr. Faiza Rini
Profile: Scopus ID: 57196343003| Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) PGRI Sumatera Barat

Dr. Maidiawati
Profile: Scopus ID: 36246392000| Institut Teknologi Padang

Dr. Supratman
Profile: Scopus ID: 57205129622| Institut Agama Islam Negeri (IAIN) Bukittinggi

Dr. Sumijan
Profile: Scopus ID: 57194787076| Universitas Putra Indonesia YPTK Padang

Dr. Tati Erlina
Profile: Scopus ID: 57205103093| Nara Institute of Science and Technology (NAIST)

Vol. 8. No. 2 February 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v8i2

Dr. Mujiono Sadikin
Profile: Scopus ID: 55440223300| Universitas Bhayangkara Jakarta Raya

Dr. Ida Astuti
Profile: Scopus ID: 57210469776| Universitas Gunadarma Jakarta

Dr. Elfizar
Profile: Scopus ID: 57035411800| Universitas Riau

Dr. Setiawan Assegaff
Profile: Scopus ID: 54929915000| Universitas Dinamika Bangsa

Dr. Ahmad Sufril Azlan Mohamed
Profile: Scopus ID: 57190968285| Universiti Sains Malaysia

Dr. Mohd Nadhir Ab Wahab
Profile: Scopus ID: 57190968285| Universiti Sains Malaysia

Dr. Jasmir
Profile: Scopus ID: 57206727457| Universitas Dinamika Bangsa

Dr. (can) Bagus Haryadi, MT
Profile: Scopus ID: 56161996000| National Dong Hwa University

Agustinus Fritz Wijaya, S.Kom., M.Cs.
Profile: Scopus ID: 56461093400| Universitas Kristen Satya Wacana

Yulia, M.Kom
Profile: Scopus ID: 57221517936| Universitas Putera Batam

Ilham Tri Maulana, M.Pd.T
Profile: Scopus ID: 57193430920| STMIK Indonesia Padang

Catur Suharinto, M.Kom
Profile: Scopus ID: 57203434293| Universitas Baiturrahmah Padang

Widianto, MT
Profile: Scopus ID: 57194554583| Univ. Muhammadiyah Malang

Adil Setiawan, M.Kom
Profile: ID Scopus: 57200211356 | Universitas potensi utama

Lusiana, M.Sc
Profile: ID Scopus: 36607684800 | STMIK AMIK Riau

Meddiati Fajri Putri, M.Sc.
Profile: ID Scopus: 57201502821 |Universitas Negeri Semarang

Indra Laksmana, M.Kom
Profile: ID Scopus: 56192111300 |Politeknik Pertanian Negeri Payakumbuh

Rezagi Meilano, M.Kom
Profile: ID Scopus: 57222624559|Politeknik Jambi

Juliza Hidayati, MT
Profile: ID Scopus: 57189370093|Universitas Sumatera Utara

Hariyadi, M.Kom
Profile: ID Scopus: 57213172288|Universitas Muhammadiyah Sumatera Barat

Mohammad Hafiz Hersyah, M.Kom
Profile: ID Scopus: 55902415300|Universitas Andalas

Taufik Baidawi, M.Kom
Profile: ID Scopus: 57200216518 | Universitas Bina Sarana Informatika

Sukmawati Anggraeni Putri, M.Kom
Profile: ID Scopus: 57200216199 | Universitas Nusa Mandiri

Sopiyan Dalis, M.Kom
Profile: Scopus ID: 57208282541 | Universitas Bina Sarana Informatika

Bambang Krismono Triwijoyo, M.Kom
Profile: ID Scopus: 57193868829 | STMIK Bumi Gora Mataram

Nurmalasari, M.Kom
Profile: ID Scopus: - | Universitas Nusa Mandiri

Sri Hadianti, M.Kom
Profile: ID Scopus: -| Universitas Nusa Mandiri

Agung Wibowo, M.Kom
Profile: ID Scopus: 57203100012 | Universitas Bina Sarana Informatika

Bahtiar Imran, S.ST., M.TI
Profile: ID Scopus: 57198810908 | Universitas Teknologi Mataram

Dimas Sasongko, S.Kom., M.Eng.
Profile: ID Scopus: 57193503355 | Universitas Muhammadiyah Magelang

Rohmat Gunawan, M.T.
Profile: ID Scopus: 57211267688 | Universitas Siliwangi

Rohmat Indra Borman, M.Kom.
Profile: ID Scopus: 57205615323 | Universitas Teknokrat Indonesia

Pressa Perdana Surya Saputra, ST., MT.
Profile: ID Scopus: 57210972475 | Universitas Muhammadiyah Gresik

Galih Abdul Fatah Maulani, S.Kom., M.Kom
Profile: ID Scopus: 57202308014 | Universitas Garut

Linda Marlinda, S.Kom,M.M.,M.Kom
Profile: ID Scopus: 57200210746 | Universitas Nusa Mandiri

Ardhin Primadewi, S. Si, M. TI
Profile: ID Scopus: 57210971067 | Universitas Muhammadiyah Magelang

Nicodias Palasara, M.Kom
Profile: ID Scopus: 57220181348 | Universitas Nusa Mandiri

Vol. 8. No. 2 February 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v8i2

Soleman, S.Kom, M.Kom, MCE
Profile: ID Scopus: 57212621666 **|** Universitas Borobudur

Amandus Jong Tallo, S.T., M.Eng.
Profile: ID Scopus: 57200984216 **|** Politeknik Negeri Kupang

Muhammad Sholeh, ST,MT
Profile: ID Scopus: 57220037895 **|** Institut Sains & Teknologi AKPRIND Yogyakarta

Harya Bima Dirgantara, S. Kom., M.T.I.
Profile: ID Scopus: 57191222989 **|** Institut Teknologi dan Bisnis Kalbis

Agus Nursikuwagus, MT.,MM.,MOS.,MTA
Profile: ID Scopus: 57194782786 **|** Universitas Komputer Indonesia

Yuwan Jumaryadi, S.Kom, MM, M.Kom
Profile: ID Scopus: 57204735588 **|** Universitas Mercu Buana

Wahyuddin S, S.Kom, M. Kom
Profile: ID Scopus: 57214590873 **|** STMIK Amika Soppeng

Editorial Address : Kampus Universitas Nusa Mandiri Tower Jatiwaringin
Jl. Jatiwaringin Raya No. 02 RT 08 RW 013 Kelurahan Cipinang Melayu
Kecamatan Makassar Jakarta Timur 13620

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

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

Vol. 8. No. 2 February 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v8i2

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. 8. No. 2 February 2023, 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. 8. No. 2 February 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk/v8i2

TABLE OF CONTENTS

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.3335

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

THE DECISION MAKING METHOD FOR AWARDING SCHOLARSHIPS TO STUDENTS USING COMPOSITE PERFORMANCE INDEX ALGORITHM

Budy Satria[1*]; Acihmah Sidauruk[2]; Muhammad Tofa Nurcholis[3]; Raditya Wardhana[4]; Bister Purba[5]

Department of Computer Engineering[1]
Institut Teknologi Mitra Gama
http://www.amikmitragama.ac.id
budysatriadeveloper@gmail.com[1]

Department of Computer Science[2, 3, 4]
Universitas Amikom Yogyakarta
https://home.amikom.ac.id
acihmah@amikom.ac.id[2], tofa@amikom.ac.id[3], raditic@amikom.ac.id[4]

Department of Graphic Multimedia Engineering Technology[5]
Politeknik Negeri Medan
https://polmed.ac.id
bisterpurba36@gmail.com[5]

(*) Corresponding Author

Abstract— *Higher education in Indonesia has several programs to help reduce the burden on students, one of which is through a scholarship program. Scholarships given can be obtained with the terms and conditions that apply at each university. Mitra Gama Institute of Technology is one of the private universities in the province of Riau which always runs a scholarship aid program. The problem that has been happening so far is that the procedures carried out are still using a document checking system without involving a weighting system and the right criteria and time constraints have always been an obstacle in determining scholarship recipients. This research was conducted as a solution to create an innovation in the form of making a computerized decision support system using criteria and weight values so that scholarship recipients are on target. Composite performance index is the method used in this study. The purpose of this research is to create a decision support system for the selection of scholarship recipients to be more systematic and time efficient in the process. There are 5 alternatives used and 4 criteria, namely parents' income, GPA, electricity consumption and semester. The results of the research carried out were obtained the 5 highest composite index values, namely MHS4 with a value of 200.00, MHS1 with a value of 134.14, MHS5 with a value of 120.00, MHS3 with a value of 87.00 and MHS2 with a value of 85.71.*

Keywords: *composite performance index, dss, positive and negative trend, scholarship*

Intisari— *Pendidikan tinggi di Indonesia memiliki beberapa program untuk membantu mengurangi beban kepada mahasiswanya salah satunya melalui program beasiswa. Beasiswa yang diberikan bisa didapatkan dengan syarat dan ketentuan yang berlaku di masing-masing perguruan tinggi. Institut Teknologi Mitra Gama merupakan salah satu perguruan tinggi swasta di provinsi riau yang selalu menjalankan program bantuan beasiswa. Permasalahan yang selama ini terjadi adalah prosedur yang dilakukan masih menggunakan sistem pemeriksaan dokumen tanpa melibatkan sistem pembobotan dan Criteria yang tepat dan keterbatasan waktu selalu menjadi kendala dalam menentukan penerima bantuan beasiswa. Penelitian ini dilakukan sebagai solusi untuk membuat suatu inovasi berupa pembuatan sistem pendukung keputusan berbasis komputerisasi menggunakan Criteria dan nilai bobot agar penerima beasiswa tepat sasaran. Composite performance index adalah metode yang digunakan pada penelitian ini. Tujuan dari penelitian ini adalah untuk membuat suatu sistem pendukung keputusan untuk seleksi penerima beasiswa agar lebih sistematis dan efisiensi waktu dalam prosesnya. Terdapat 5 alternatif yang digunakan dan 4 Criteria yaitu penghasilan orang tua, IPK, pemakaian listrik dan semester. Hasil penelitian adalah diperoleh 5 Nilai composite index tertinggi yaitu MHS4 dengan nilai 200.00, MHS1 dengan nilai 134.14, MHS5 dengan nilai 120.00, MHS3 dengan nilai 87.00 dan MHS2 dengan nilai 85.71.*

Kata Kunci: *beasiswa, indeks gabungan, spk, seleksi, tren positif dan negatif*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3335

INTRODUCTION

Scholarship is a form of appreciation that appears to students and students while undergoing education in educational institutions [1]. Scholarship is a financing that does not come from individuals, but scholarships are given by the government, companies, embassies, universities, and educational institutions or from the office where they work for the reason of an achievement in order to support the improvement of human resource capacity through education [2]. Assistance provided through a scholarship program with the aim of reducing the cost of studying [3]. The problem that occurs in this study is that the scholarship aid selection system has not been computerized because it has to look at the scholarship requirements document based on the criteria and eligibility of the recipient of the assistance so that it creates difficulties in making a decision[4]. The limited time they have often makes it difficult for the team to determine the right students to receive scholarships [5]. Therefore, not all students who apply to receive scholarships can be granted because the number of submissions is very large [6].

To overcome this problem, it is necessary to have a decision support system. A decision support system is a collection of interconnected systems that produce information for user decision making [7]. The method used in this research is the composite performance index. This method is a method by using a combined index to determine the assessment or ranking of several alternatives based on predetermined criteria [8]. The decision support system only provides alternative decisions, while the final decision is still determined by the decision maker [9]. Decision support systems are not intended to replace the role of decision makers, but to assist and support decision systems [10].

In previous studies, the composite performance index method was used to determine the placement of mentors at development centers [11]. This method is also used to determine the selection of food in patients with ulcer disease, so that the results obtained that ulcer sufferers are helped in choosing foods that are suitable for consumption [12]. In addition, the CPI method has also been used to assess class promotion for PTPN employees of the South Solok business unit with the best alternative results with the most group increases from 1A/4 to 1B/0 in the 2019 period with a value of 108.78[13]. The composite performance index method also helps in determining the chairman of the student council at the Kavri Talun Kenas Private Junior High School [14]. The application of the composite performance index method was also applied by the Social Service of North Sumatra for recipients of joint business group (KUBE) assistance[15].

Research related to the composite performance index method has also been carried out for the selection of the best student thesis [16] and the results obtained that the first rank is A1 with a value of 218.75, the second rank is A2 with a value of 198.75 and the third rank is A4 with a value of 178.75. Another research is in determining the priority selection decision for village infrastructure development using the composite performance index method[17].

The purpose of this study is to create a decision support system using the composite performance index method for the selection of scholarship recipients at the Mitra Gama Institute of Technology. With a decision support system, data processing becomes faster [18].

MATERIALS AND METHODS

Some of the stages carried out in this research are as follows:

1. Identify the problem, which is a problem found when selecting new employees based on the information obtained.
2. Data collection, namely collecting the data needed in this study by means of observation, interviews and literature on the object of research.
3. Determine the criteria, namely to be a reference in the calculation process using the composite performance index (CPI) method for the selection of scholarship recipients.
4. Data analysis, namely data that has been obtained on the object of research will be processed and given a weighted value for each criterion [19]
5. Implementation of the composite performance index (cpi) in order to obtain the best results in determining a decision.
6. Alternative ranking, which is to carry out a ranking process to get the highest value from all alternative data [20].
7. Conclusion, namely taking a conclusion on the data that has been analyzed and processed previously so that it becomes the final result in this study.

The data collection techniques used are:

1. Observation, namely the collection of data sourced from the object of research.
2. Literature study is an approach with references such as journals or books that are in accordance with the research topic
3. Interviews, namely conducting discussions with related parties to be able to obtain information on what is needed as research material.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.3335

The following are the steps for calculating using the Composite Performance Index (CPI) equation as follows:

1. Identification of trend criteria is divided into 2, namely positive and negative. The positive trend (benefit) is that the higher the value, the better. The negative trend (cost) is that the lower the value, the better.
2. Formation of the decision matrix (X)

$$X=\begin{bmatrix} X_{01} & \cdots & X_{0j} & \cdots & X_{On} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{i1} & \cdots & X_{ij} & \cdots & X_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{mj} & \cdots & X_{mn} \end{bmatrix}$$

$$i=\overline{(0,1,2,\dots,m}; j=\overline{1,2,\dots,n)} \qquad (1)$$

The above formula can be given information, namely:

X_{ij} = The performance value of the i alternative on criteria-j
m = Many alternatives
n = Number of criteria

3. Formation of the normalization matrix (R)

In the positive trend criteria, the minimum value for each criterion is transformed to one hundred, while the other values are transformed proportionally higher.

$$r_{ij} = \frac{x_{ij}*100}{x_{ij}(min)} \qquad (2)$$

In the negative trend criteria, the minimum value for each criterion is transformed to one hundred, while the other values are transformed lower.

$$r_{ij} = \frac{x_{ij}(min)*100}{x_{ij}} \qquad (3)$$

The formula above can be given information, namely:

x_{ij} = Alternative value to i on criterion j
$x_{ij\,(min)}$ = Minimum value on criterion - j
r_{ij} = Normalization value for alternative i on criterion j

4. Composite Index Weighting (I)

$$i_i = \sum_{j=1}^{n}(r_{ij} * w_j) \qquad (4)$$

The calculation of the alternative value or composite index (i) is the sum of the multiplication between the criteria values and the criteria weights for each alternative.

5. Ranking

Determination of the best alternative is obtained from the ranking of alternative values or Composite Index (I) from the largest to the smallest. The value with the highest alternative value is the best.

RESULTS AND DISCUSSION

1. Alternative Data

Alternative data (Ai) can be seen in Table 1.

Table 1. Alternative Data

Code	Full Name	Institution
B1	MHS1	Institut Teknologi Mitra Gama
B2	MHS2	Institut Teknologi Mitra Gama
B3	MHS3	Institut Teknologi Mitra Gama
B4	MHS4	Institut Teknologi Mitra Gama
B5	MHS5	Institut Teknologi Mitra Gama

In Table 1 there are 5 prospective scholarship recipients who are at the Mitra Gama Institute of Technology.

2. Determine Criteria and Weights

Determination of criteria (Ci) and weights can be seen in Table 2

Table 2. Criteria and Weights

Code	Criterias	Trend	Weights
C1	Penghasilan Orang Tua	Negatif	0.200
C2	IPK	Positif	0.500
C3	Pemakaian Listrik	Negatif	0.100
C4	Semester	Negatif	0.200

In Table 2 there are criteria consisting of parental income (Negative) with a weight of 0.20, GPA (Positive) with a weight of 0.50, electricity consumption (Negative) with a weight of 0.10 and semester (Negative) 0.20.

3. Evaluation Data

Evaluation data is a collection of data that has been processed based on the weight value of all alternative data (Ai) and criteria (Ci). The evaluation data in this study can be seen as in Table 3.

Table 3. Evaluation Data

Code	Criteria	C1	C2	C3	C4
B1	MHS1	50	40	80	70
B2	MHS2	40	20	70	80
B3	MHS3	50	20	40	80
B4	MHS4	30	60	40	60
B5	MHS5	30	30	80	60

In Table 3 there are 5 alternative data, 4 criteria and weighted values.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3335

1. **Decision Matrix (X)**

The decision matrix is derived from evaluation data in tabulated form and then remade into a decision matrix form (X).

$$X=\begin{bmatrix}50 & 40 & 80 & 70\\ 40 & 20 & 70 & 80\\ 50 & 20 & 40 & 80\\ 30 & 60 & 40 & 60\\ 30 & 30 & 80 & 60\end{bmatrix}$$

The decision matrix (X) above consists of 5 alternative data in the top-down order and 4 criteria in the left-to-right order.

2. **Determine the Minimum Value of Each Criteria**

In the composite performance index method, there are positive trends and negative trends. Determining the minimum value for each Criteria is used to determine the lowest value for each Criteria. The minimum value for each Criteria can be seen as in Table 4.

Table 4. Minimum Value of Each Criteria

Code	Criteria	C1	C2	C3	C4
B1	MHS1	50	40	80	70
B2	MHS2	40	**20**	70	80
B3	MHS3	50	**20**	**40**	80
B4	MHS4	**30**	60	**40**	**60**
B5	MHS5	**30**	30	80	**60**

In Table 4 there are minimum values for each criterion, namely C1 with 2 minimum values, C2 with 2 minimum values, C3 with 2 minimum values and C4 with 2 minimum values. The minimum value of each Criteria is used when calculating the equations in the normalized matrix (R).

3. **Normalization Matrix (R)**

In this study, there are 1 Criteria with a positive trend value and 3 negative trends. The calculation of the value of the normalization matrix (R) is based on the equations in the positive and negative trend formulas.

The calculation of the normalization matrix on Parental Income (C1) has a negative trend value which can be calculated as follows:

$$\boldsymbol{r_{ij}} = \frac{\boldsymbol{x_{ij}(min) * 100}}{x_{ij}}$$

$$= \frac{\mathbf{30*100}}{\mathbf{50}}$$

$$= \frac{\mathbf{3000}}{\mathbf{50}}$$

$$\mathbf{= 60}$$

The calculation of the normalization matrix on the GPA (C2) having a positive trend value can be calculated as follows:

$$\boldsymbol{r_{ij}} = \frac{\boldsymbol{x_{ij} * 100}}{x_{ij}(min)}$$

$$= \frac{\mathbf{40*100}}{\mathbf{20}}$$

$$= \frac{\mathbf{4000}}{\mathbf{20}}$$

$$\mathbf{= 200}$$

So that the normalization matrix for all criteria is obtained as follows:

$$R=\begin{bmatrix}60 & 200 & 50 & 85{,}71\\ 75 & 100 & 57{,}14 & 75\\ 60 & 100 & 100 & 75\\ 100 & 300 & 100 & 100\\ 100 & 150 & 50 & 100\end{bmatrix}$$

The normalization matrix above is the result of calculations for each Criteria that has a negative and positive trend. The normalized matrix values in tabulated form are shown in Table 5.

Table 5. Normalization Matrix

Code	Criteria	C1	C2	C3	C4
B1	MHS1	60	200	50	85,71
B2	MHS2	75	100	57,14	75
B3	MHS3	60	100	100	75
B4	MHS4	100	300	100	100
B5	MHS5	100	150	50	100

In Table 5 there are normalized matrix values resulting from calculations using the equations in the negative and positive trend formulas.

4. **Composite Index Value (I)**

The alternative composite index value is calculated based on the equation in the applicable formula. The calculation of the composite index value for each alternative can be calculated as follows:

$$i_i = \sum_{j=1}^{n} (r_{ij} * w_j)$$

= 60 * 0.20 + 200 * 0.50 + 50 * 0.10 + 85.71 * 0.20
= 12 + 100 + 5 + 17.142
= 134.14

So that the composite index values for all alternative data are obtained as follows:

I =[134.14, 85.71, 87.00, 200.00, 120.00]

The Composite Index value can be seen in Table 6.

Table 6. *Composite Index* (I)

Code	Criteria	Composite Index (I)
B1	MHS1	134,14
B2	MHS2	85,71

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

B3	MHS3	87,00
B4	MHS4	200,00
B5	MHS5	120,00

In Table 6 there is a composite index value of 5 alternative data.

5. Rangking

The ranking data is obtained from composite index data starting from the largest to the smallest order as shown in Table 7.

Table 7. Ranking Results

Code	Full Name	Composite Index (I)	Ranking
B4	MHS4	200,00	1
B1	MHS1	134,14	2
B5	MHS5	120,00	3
B3	MHS3	87,00	4
B2	MHS2	85,71	5

Based on Table 7, the ranking results have been calculated. Obtained 5 highest composite index values starting from MHS4, MHS1, MHS5, MHS3 and MHS2.

CONCLUSION

This study uses the composite performance index method as a way to assist the process of determining scholarship recipients at the Mitra Gama Institute of Technology. The implementation of the composite performance index method is very helpful in decision making. There are 5 alternative data used, using negative and positive trends and using 4 predetermined criteria. Based on the ranking results, the 5 highest composite index values were obtained, namely 200.00, 134.14, 120.00, 87.00 and 85.71.

REFERENCES

[1] T. S. Rinaldi and Hutrianto, "Implementation of The Oreste Method for Scholarship Recipients at The Institute of Technology & Business (ITBis) Lembah Dempo," *J. Inform.*, vol. 10, no. 1, pp. 24–34, 2022.

[2] Nurmansyah, U. Lestari, and R. Y. Ariyana, "Sistem Pendukung Keputusan Pemberian Beasiswa Dengan Metode Simple Additive Weighting (SAW) Studi Kasus: SMK Negeri 4 Kepahiang Provinsi Bengkulu," *J. Scr.*, vol. 9, no. 2, 2021.

[3] A. Adikvika, N. Merlina, and N. A. Mayangky, "Sistem Pendukung Keputusan Pemilihan Penerima Beasiswa Pendidikan Dengan Menggunakan Metode Weighted Product Di Yatim Mandiri," *Indones. J. Softw. Eng.*, vol. 7, no. 2, pp. 148–158, 2021.

[4] D. Alita, I. Sari, A. R. Isnain, and Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *JDMSI*, vol. 2, no. 1, pp. 17–23, 2021.

[5] E. Pawan, W. W. Widianto, and P. Hasan, "Implementasi Metode Profile Matching Untuk Menentukan Penerima Beasiswa Bidikmisi," *Creat. Inf. Technol. J.*, vol. 8, no. 1, pp. 54–62, 2021, doi: 10.24076/citec.2021v8i1.257.

[6] D. Ayudia, G. W. Nurcahyo, and Sumijan, "Optimalisasi Penentuan Criteria Penerima Bantuan Program Indonesia Pintar dengan Metode TOPSIS," *J. Sistim Inf. dan Teknol.*, vol. 3, no. 3, pp. 142–149, 2021, doi: 10.37034/jsisfotek.v3i3.58.

[7] M. Iqbal, B. Satria, and T. Radillah, "Implementasi SPK Menggunakan Metode ARAS Untuk Penentuan SMA dan SMK Terbaik Berbasis Website," *Indones. J. Comput. Sci. A*, vol. 10, no. 2, pp. 425–435, 2021.

[8] M. Walid, B. Satria, and M. Makruf, "Seleksi Karyawan Baru Menggunakan Metode Composite Perfomence Index (CPI) dan Rank Order Centroid (ROC)," *J. Ilm. Ilk. - Ilmu Komput. Inform.*, vol. 5, no. 1, pp. 11–18, 2022, doi: 10.47324/ilkominfo.v5i1.137.

[9] I. N. Okta and B. Satria, "Sistem Pendukung Keputusan Dalam Menentukan Perbaikan Jalan Rusak Dengan Menggunakan Metode Simple Additive Weighting (SAW) (Studi Kasus : Kabupaten Kuantan Singingi)," *Jar. Sist. Inf. Robot.*, vol. 3, no. 1, p. 194202, 2019.

[10] R. Siahaan, "Perancangan Sistem Pendukung Keputusan Penerimaan Siswa Baru Dengan Menerapkan Metode Composite Performance Index(cpi) (studi kasus: SDN 101864 Gunung Rintih)," *J. Multimed. dan Teknol. Inf.*, vol. 3, no. 2, pp. 38–46, 2021, doi: 10.54209/jatilima.v3i02.148.

[11] S. Lukas, E. Buulolu, A. M. H. Sihite, and D. P. Utomo, "Sistem Pendukung Keputusan Penempatan Mentor Pada Pusat Pengembangan Anak IO 558 Sangkakala Medan Menggunakan Metode CPI dan ROC," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, pp. 315–321, 2020, doi: 10.30865/komik.v4i1.2713.

[12] C. A. Putri, J. Minardi, and N. Azizah, "Sistem Pendukung Keputusan Pemilihan Makanan Pada Penderita Maag Menggunakan Metode Composite Performance Index (CPI)," *J. Imiah Inform. dan Komput.*, vol. 1, no. 2, pp. 115–121, 2022.

[13] D. E. Dewi, R. Aprilia, and N. H. Prasetya, "Sistem Pendukung Keputusan Penilaian Kenaikan Golongan Pada Karyawan PTPN VI

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3335

Unit Usaha Solok Selatan Menggunakan Metode Composite Performance Index (CPI),” *Math. Appl. J.*, pp. 1–7, 2021.
[14] S. A. Tarigan, “Perancangan Sistem Pendukung Keputusan Penentuan Ketua Osis Dengan Menerapkan Metode Composite Performance Index (CPI) (Studi Kasus : SMP Swasta Kavri Talun Kenas),” *J. Multimed. dan Teknol. Inf.*, vol. 3, no. 1, pp. 31–37, 2021, doi: doi.org/10.54209/jatilima. V3i1.147.
[15] Z. Sipahutar, B. Nadeak, and P. Ramadhani, “Penerapan Metode Composite Performance Index (CPI) Dalam Penerima Bantuan Kelompok Usaha Bersama (KUBE),” *J. Sist. Komput. dan Inform.*, vol. 2, no. 3, pp. 255–260, 2021, doi: 10.30865/json.v2i3.2627.
[16] T. T. Muryono, I. K. Sudaryana, and Irwansyah, “Pemilihan Skripsi Mahasiswa Terbaik Menggunakan Metode Composite Performance Index (CPI),” *INFOTECH J. Technol. Inf.*, vol. 7, no. 2, pp. 99–104, 2021, doi: https://doi.org/10.37365/jti.v7i2.119.
[17] F. R. Tampubolon, “Penerapan Metode CPI dan ROC Dalam Keputusan Pemilihan,” *BEES Bull. Electr. Electron. Eng.*, vol. 2, no. 2, pp. 67–73, 2021.
[18] B. Satria, “Implementation of Additive Ratio Assessment (ARAS) Method on Decision Support System for Recipient of Inhabitable House,” *JITK (Jurnal Ilmu Pengetah. Dan Teknol. Komputer)*, vol. 6, no. 1, pp. 121–128, 2020, doi: 10.33480/jitk.v6i1.1389.
[19] Supriatin, A. R. Saputra, and B. Satria, “Implementation of Aras Algorithm on Decision Support System to Determine The Best Lecturer,” *JITK (Jurnal Ilmu Pengetah. Dan Teknol. Komputer)*, vol. 8, no. 1, pp. 25–32, 2022, doi: 10.33480/jitk.v8i1.3152.
[20] B. Satria, M. Iqbal, and T. Radillah, “Additive Ratio Assessment Algorithm on Decision Support System for Selecting The Best SMA and SMK,” *JITK (Jurnal Ilmu Pengetah. Dan Teknol. Komputer)*, vol. 7, no. 1, pp. 29–36, 2021, doi: 10.33480/JITK.V7i1.2217.

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.3997

APPLICATION OF THE K-NEAREST NEIGHBOR (KNN) ALGORITHM IN SENTIMENT ANALYSIS OF THE OVO E-WALLET APPLICATION

Siti Masturoh[1*]; Risca Lusiana Pratiwi[2]; M Rangga Ramadhan Saelan[3], Ummu Radiyah[4]

Sistem Informasi[1, 2]
Sains Data[3]
Informatika[4]
Universitas Nusa Mandiri
https://nusamandiri.ac.id/
siti.uro@nusamandiri.ac.id[1], risca.ral@nusamandiri.ac.id[2], rangga.mgg@nusamandiri.ac.id[3], ummu.urd@nusamandiri.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *The OVO application can be downloaded on the Android platform via Google Play, Google play has a review feature on the application product to be downloaded, so that the review can be viewed or accessed by anyone, With these reviews, potential users of the application will see how important it is to consider using an application, problems regarding reviews or sentiment analysis of applications processed using text mining. The purpose of this study is to provide information to prospective OVO application users before using the application which can be seen from the results of giving reviews based on rating or stars (*) in the OVO application review column on Google Play and the authors categorize them into 3 classes, the first class (1 to 5 stars, second class (1 and 5 stars) third class by providing labeling grouping (1&2 stars are negative labels, 3 stars are neutral labels and 4&5 stars are positive labels) testing using the k-nearest neighbor method by finding the value of k from the k value of 1-10 to get the highest accuracy value, in order to obtain the highest accuracy value of 84.86% in the 2nd class test and giving a value of k 1 which means that the 1st and 5th star tests get positive values so that they can give a good impression to prospective application users OVO.*

Keywords: *E-Wallet, OVO, K-Nearest Neighbor*

Intisari— *Aplikasi OVO bisa di unduh pada platform android melalui google play, Google play terdapat fitur ulasan terhadap produk aplikasi yang akan diunduh, sehingga ulasan tersebut dapat dilihat atau diakses oleh siapa saja, dengan ulasan tersebut calon pengguna aplikasi akan melihat betapa pentingnya mempertimbangkan dalam menggunakan suatu aplikasi, permasalahan mengenai ulasan atau analisa sentimen terhadap aplikasi diolah menggunakan text mining. Metode text minning yang digunakan pada penelitia ini yaitu menggunakan algoritma K-Nearest Neighbor (KNN). Tujuan penelitian ini untuk memberikan informasi kepada para calon pengguna aplikasi OVO sebelum menggunakan aplikasi tersebut yang dapat dilihat dari hasil pemberian ulasan berdasarkan pemberian rating atau bintang (*) pada kolom ulasan aplikasi OVO di google play lalu penulis mengkategorikan ke dalam 3 kelas, kelas pertama (bintang 1 sampai bintang 5, kelas ke dua (bintang 1 dan bintang 5) kelas ke tiga dengan memberikan pengelompokan pelabelan (bintang 1&2 label negatif, bintang 3 label netral dan bintang 4&5 label positif) penggujian menggunakan metode k-nearest neighbor dengan mencari nilai k dari nilai k 1-10 untuk mendapatkan nilai accuracy tertinggi, sehingga didapatkan nilai accuracy tertinggi 84.86% pada penggujian kelas ke 2 dan pemberian nilai k 1 yang artinya pengujian bintang 1 dan bintang 5 mendapatkan nilai positif sehingga bisa memberikan kesan yang baik bagi para calon pengguna aplikasi OVO.*

Kata Kunci: *E-Wallet, OVO, K-Nearest Neighbor*

INTRODUCTION

The industrial revolution 4.0 is a phenomenon that collaborates cyber technology and automation technology which is marked by the digitization of almost all economic sectors including the financial sector (fintech). [1]. One of the fintech products is a digital wallet (E-wallet). [2], Currently, many people have switched to making payment transactions using e-wallets, moreover, there are many benefits and offers that can be obtained [3]. Apart from that, for the security of the existence of

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

digital wallets (e-wallets) in Indonesia, of course, you must obtain permission from Bank Indonesia (BI) so that the data of e-wallet users is maintained. [4].

E-wallets also have a positive impact, apart from being more efficient they can also reduce cash circulation [5], as well as being the best solution when there is a covid-19 virus to reduce the spread of the virus [6]. In Indonesia, the e-wallet products that we can use are OVO, DANA, LinkAja and Go-Pay [7]. From several e-wallet products, researchers took case studies on the OVO application.

We can download the OVO application on the Android platform on Google Play. Google Play has a review feature for application products that will be downloaded, so that these reviews can be seen or accessed by anyone [8], These reviews can affect potential users as a material consideration in using an application [9].

We can also process review data to see how important the sustainability of an application including OVO is, so it would be good if sentiment analysis was carried out on OVO application review data on Google Play. Sentiment analysis can determine whether an opinion or opinion on an entity has a positive or negative tendency so that it can be used as a reference for improving service and product quality. [10].

Research on e-wallet sentiment analysis has been carried out by previous researchers, and the following are several reference papers which are described in table.1 Reference paper is as follows :

Table. 1 Reference Paper

Research Title	Research Titles
Implementation of Naïve Bayes Classifier and Information Gain for sentiment analysis of e-wallet users during a pandemic	Using the naïve Bayes classifier algorithm and information gain by researching e-wallet users and showing classification results for 92% accuracy, 92% precision and 100% recall
Sentiment analysis based on marketing mix aspects of access to digital wallet application reviews (case study: the LinkAja application on Twitter)	Using the SVM algorithm with the undersampling method to balance classes into datasets. The classification results get negative sentiment on the product aspect with a value of 98% of the total reviews and from the aspect of the place with 100% of the total reviews, while neutral sentiment on the product aspect gets a value of 89% of the total reviews, and positive aspects as much as 98% of the total review
The effect of perceived convenience and service features on the interest in using e-wallets in the DANA application in Surabaya	significant and positive influence on the significant and positive interest in using (Y) with a value of Sig. 0.000 <0.05, while simultaneously (X1) and service features (X2) have an influence on interest (Y) with a presentation value of 26.2%, but 73.8% affect other variables that are not in the research model

Based on Table 1. Referral Paper, we can see research on sentiment analysis of e-wallet users using the naïve Bayes and information gain method. get the classification results using naïve bayes classifier produce 84% accuracy, 91% precision, and 91% recall. Meanwhile, the classification results using naïve Bayes and information gain are 92% accuracy, 92% precision, 100% recall [11]. Furthermore, for sentiment analysis research based on marketing mix aspects of access to digital wallet application reviews (case study: the LinkAja application on Twitter) this research focuses on aspect-based sentiment analysis to determine positive, negative or neutral aspects of the reviews given by consumers.

To classify aspects using string matching and using the thefuzz library, then for sentiment classification using the SVM algorithm then using undersampling to balance classes in the dataset. The classification results show that the LinkAja

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

application gets a negative sentiment value on the product aspect with a value of 98%, 89% neutral and 98% positive from the total reviews [12].

Subsequent research entitled The effect of perceived convenience and service features on the interest in using e-wallets in the DANA application in Surabaya, in his research using multiple linear regression analysis techniques which aim to determine the effect of convenience and service features on the interest in using e-wallets in applications with a sample that used a total of 214 respondents with a population of all DANA application users.

The results of this study show that ease (X1) has a significant and positive influence on the significant and positive interest in using (Y) with a Sig value. 0.000 <0.05, while simultaneously (X1) and service features (X2) have an influence on interest (Y) with a presentation value of 26.2%, but 73.8% affect other variables that are not in the research model [13].

Based on the references above, the author tries to conduct research with a different object, namely the OVO application and tries to use a different method, namely in this study using the KNN method, where data is collected based on reviews on Google Play to provide information to potential OVO application users.

MATERIALS AND METHODS

This study uses the results of user reviews of the OVO application on Google Play which gives a star rating (*) on the link https://play.google.com/store/apps/details?id=ovo.id. The data used in this study were 5000 data which were divided into two types of data with a ratio of 90% for training data and 10% for test data. The dataset The data used in this study were 5000 data which were divided into two types of data with a ratio of 90% for training data and 10% for test data [14], the number of reviews giving stars in the comments column on the OVO application.

Table 2 Data Testing Number of OVO Reviews

	Number of Stars (*)				
	1	2	3	4	5
Amount Data	100	100	100	100	100
Total					500

Based on Table 2. Data Testing Number of OVO Reviews, researchers took data testing based on star rating in the Google Play comments column with each amount of data taken, namely from 1 to 5 stars of 100 data. And the following is picture 1. An example of an OVO user review in the Google Play review column.

Fig.1 Example of an OVO User Review

Based on figure 1, we can see the reviews and stars (*) of the OVO application users regarding their opinion of the OVO application. We get the data from manual processes one by one from OVO application user reviews, after that the data is pre-processed using the web gata framework on the link http://www.gataframework.com/ selanjutnya managed with the help of rapidminer application. The research stages are described based on Figure 2. The research stages are explained as follows:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

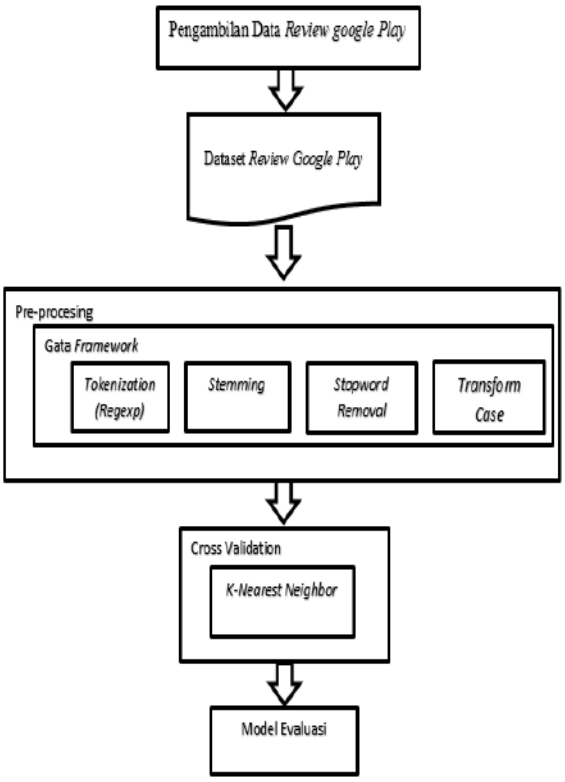


Fig. 2 Research Stages

Based on Figure 2 above, it explains how the stages of the research were carried out, first the dataset was taken from Google Play reviews on the OVO application based on giving stars (*) from each star 1-5, 100 data were taken then grouped into 3 classes, namely :

a. the first class consists of (stars 1 to 5).

b. class 2 consists of (1 and 5 stars)

c. class 3 is obtained from grouping stars 1 and 2 as negative labels, stars 3 are labeled neutral and stars 4 and 5 are labeled positive.

Furthermore, after data collection, data preprocessing is carried out which aims to prepare the text into data that will be ready to undergo processing at the next stage by first cleaning the entities that can interfere with the analysis process. [15]

The data preprocessing stage is carried out using the web gata framework with several stages such as :

1. 1. Tokenization regexp which has the function of breaking sentences in a file into words while removing unnecessary characters [16]
2. 2. Indonesian Stemming whose job is to find the basic words of a word. By eliminating all good affixes consisting of prefixes, infixes, endings and confixes (combinations of prefixes and suffixes) in derived words [17]
3. 3. Indonesian Stopword Removal has the function of removing connecting words in an inserted sentence [14]
4. 4. Transform Case which functions to change capital letters to all lowercase letters. This is intended so that when classifying data there is no diversity of letters and no errors occur in the tokenize process [14]

After the data has been preprocessed, the next stage is the classifier process using the K-Nearest Neighbor (KNN) algorithm. The K-Nearest Neighbor (KNN) algorithm is used because it has simplicity for a process because the process is carried out based on a simple weighting approach and is easy to implement, adapt and elerning process and has a high accuracy value [18]. K-Nearest Neighbor (KNN) is also a process for grouping data based on the closest distance or the level of similarity of the data with the existing training dataset/data so that later the data will be grouped into a class by looking at the number of k values closest to the training data [19].

Then the results of the research phase carried out after carrying out the data classification process using the k-nearest neighbor algorithm will get the best results from the 3 classes with the best accuracy value based on giving k values 1 to 10.

RESULTS AND DISCUSSION

Based on the results of research that has been carried out on the application of the KNN algorithm (K-Nearesrt Neighbor) in sentiment analysis of the OVO application with a total data of 500 user reviews based on star rating (*) with a total of 1 to 5 stars for each of 100 data which is then test using 3 classes and testing based on the value of k 1-10. The test results can be seen based on table 3 of the test results as follows :

Table 3. Test Results

value *k*	**Accuracy Value**		
	class 1 (1-5 stars)	**Class 2 (1 & 5 stars)**	**class 3 (positive, neutral, negative)**
1	76.56%	84.86%	82.45%
2	53.50%	74.71%	73.32%
3	46.88%	79.12%	72.63%
4	40.65%	71.33%	71.26%
5	41.78%	72.76%	72.12%
6	34.79%	65.48%	71.08%

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

value *k*	Accuracy Value		
	class 1 (1-5 stars)	Class 2 (1 & 5 stars)	class 3 (positive, neutral, negative)
7	34.40%	66.93%	71.25%
8	29.30%	60.17%	67.46%
9	28.72%	61.17%	69.52%
10	24.76%	58.76%	58.35%

Based on table 3, the test results explain that from testing 3 classes to get the highest accuracy value by giving a value of k 1 with a reduction in the first class consisting of a review giving 1 star to 5 stars the highest accuracy is 76.56%, then testing in the second class consisting of 1 star and 5 star reviews have the highest accuracy score of 84.86%, then for the third class which consists of positive, neutral and negative values the labeling results get the highest score of 82.45%. So that from the comparison of the three class tests, the highest accuracy value was obtained in the second class test, namely scoring 1 star and 5 star with the highest accuracy value of 84.86%.

CONCLUSION

The application of the KNN (K-Nearest Neighbor) algorithm to the OVO application sentiment analysis based on reviews on Google Play based on star rating has been grouped into 3 classes based on the number of stars (*) and is obtained with the highest accuracy value obtained by testing using class 2 on stars 1 and 5 stars and a value of k 1 is obtained by obtaining the best accuracy of 84.86%. So that it can give a good impression to potential OVO application users before they decide to use the application..

REFERENCE

[1] R. Sulistyowati, L. Paais, and R. Rina, “Persepsi Konsumen Terhadap Penggunaan Dompet Digital,” *ISOQUANT J. Ekon. Manaj. dan Akunt.*, vol. 4, no. 1, p. 17, 2020, doi: 10.24269/iso.v4i1.323.

[2] T. M. Kesumastuti, “The Process of Adoption Interest in Using Digital Wallet in Central Jakarta (Case Study on Go-Pay Users),” *Int. J. Multicult. Multireligious Underst.*, vol. 7, no. 2, pp. 277–286, 2020, [Online]. Available: https://ijmmu.com/index.php/ijmmu/article/download/1463/1159.

[3] A. B. P. Siti Masturoh, “Sentiment Analysis Against the Dana E-Wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm,” *Ejournal.Nusamandiri.Ac.Id*, pp. 53–58, 2020, [Online]. Available: www.bsi.ac.id.

[4] L. Angelica and U. Soebiantoro, “Analisa Menggunakan Dompet Digital,” *J. Manaj.*, vol. 14, no. 2, pp. 232–238, 2022, doi: 10.29264/jmmn.v14i2.11209.

[5] N. Inggiharti, “Pengaruh Electronic Wallet Terhadap Kegiatan Keuangan Indonesia (Perbandingan Aplikasi Electronic Wallet Milik Perusahaan Financial Technology terhadap Aplikasi Electronic Wallet Milik BUMN),” *Univ. Bengkulu Law J.*, vol. 5, no. 1, pp. 74–88, 2020, [Online]. Available: https://ejournal.unib.ac.id/index.php/ubelaj/article/view/10368.

[6] S. Aulia, “Pola Perilaku Konsumen Digital Dalam Memanfaatkan Aplikasi Dompet Digital,” *J. Komun.*, vol. 12, no. 2, p. 311, 2020, doi: 10.24912/jk.v12i2.9829.

[7] S. Watmah, S. Fauziah, and N. Herlinawati, “Identifikasi Faktor Pengaruh Penggunaan Dompet Digital Menggunakan Metode TAM Dan UTAUT2,” *Indones. J. Softw. Eng.*, vol. 6, no. 2, pp. 261–269, 2020, doi: 10.31294/ijse.v6i2.8833.

[8] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[9] M. R. Fahlevvi, “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine,” *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no. 1, pp. 1–13, 2022, [Online]. Available: http://ejournal.ipdn.ac.id/JTKP,.

[10] K. D. Adhine Salsabila, A. Ghiffarie, R. P. Baistama, M. I. Variadi, and M. D. Rhajendra, “Analisis Sentimen Terhadap Produk The Body Shop Tea Tree Oil,” *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 2, 2019, doi: 10.26905/jtmi.v5i2.3684.

[11] A. Isnanda, Y. Umaidah, and J. H. Jaman, “Implementasi Naïve Bayes Classifier Dan Information Gain Pada Analisis Sentimen Penggunaan E-Wallet Saat Pandemi,” *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, pp. 144–153, 2021, doi: 10.37012/jtik.v7i2.648.

[12] R. M. Turjaman and I. Budi, “Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter),” *J. Darma Agung*, vol. 30, no. 2, p. 266, 2022, doi:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3997

10.46930/ojsuda.v30i2.1672.

[13] N. D. Abrilia and S. Tri, “Pengaruh Persepsi Kemudahan Dan Fitur Layanan Terhadap Minat Menggunakan E-Wallet Pada Aplikasi Dana Di Surabaya,” *J. Pendidik. Tata Niaga*, vol. 8, no. 3, pp. 1006–1012, 2020.

[14] S. Aditya, D. Rosiyadi, W. Gata, and S. M. Husain, “Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization,” *J. Resti*, vol. 3, no. 3, pp. 377–382, 2019.

[15] S. Saidah and J. Mayary, “Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik Dengan Metode Lexicon Based Dan K – Nearest Neighbor,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020, doi: 10.35760/ik.2020.v25i1.2411.

[16] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[17] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 64–69, 2018, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/3802/2626.

[18] R. I. Pristiyanti, M. A. Fauzi, and L. Muflikhah, “Sentiment Analysis Peringkasan Review Film Menggunakan Metode Information Gain dan K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 3, pp. 1179–1186, 2018, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1140.

[19] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4004

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

SENTIMENT ANALYSIS OF ONLINE GOJEK TRANSPORTATION SERVICES ON TWITTER USING THE NAÏVE BAYES METHOD

Muhammad Fahmi[1], Yuyun Yuningsih[2], Ari Puspita[3]*

Teknik Infomatika[1]
Sistem Informasi[2]
Universitas Nusa Mandiri
www.nusamandiri.ac.id
muhammad.mmf@nusamandiri.ac.id[1], yuyun.yyg@nusamandiri.ac.id[2]

Sistem Informasi[3*]
Universitas Bina Sarana Informatika
www.bsi.ac.id
ari.arp@bsi.ac.id[3]

(*) Corresponding Author

Abstract— *Social media is the most accessed internet content by internet users in Indonesia. This is not surprising, given the many benefits that social media provides, one of which is the benefit of self-expression. Self-expression can include many things, including emotional openness, which is the openness of a person in conveying the emotions he is feeling. Along with the development of social media, emotional disclosure is ubiquitous in social media, one of which is social media Twitter. With the development of information technology, means of transportation are also developing with the existence of online transportation services. Currently, the use of online transportation services has become a necessity, so it is necessary to conduct a sentiment analysis on online transportation services to find out how the public responds to these online transportation services. The purpose of this study is to analyze community responses by analyzing data in the form of tweets and then classifying them into positive, negative, and neutral classes using the Naïve Bayes method because the error rate obtained is lower when the dataset is large, besides that the accuracy of Naive Bayes and the speed is higher. high when applied to a larger dataset. The results of this study indicate that the neutral sentiment level of public tweets is greater than the level of positive sentiment and negative sentiment with an accuracy of 25.00%.*

Keywords: *Online Transportation, appraisal analysis, Twitter, Naïve Bayes.*

Intisari— *Media sosial merupakan konten internet yang paling banyak diakses pengguna internet di Indonesia. Hal ini tidak mengherankan, mengingat banyaknya manfaat yang diberikan media sosial, salah satunya adalah manfaat untuk mengekspresikan diri. Ekspresi diri dapat mencakup banyak hal, termasuk keterbukaan emosional, yaitu keterbukaan seseorang dalam menyampaikan emosi yang sedang dirasakannya. Seiring dengan perkembangan media sosial, keterbukaan emosional semakin banyak dijumpai di media sosial, salah satunya adalah media sosial Twitter. Berkembangnya teknologi informasi, alat transportasi juga berkembang dengan adanya jasa transportasi online. Saat ini penggunaan jasa transportasi online sudah seperti kebutuhan, maka perlu melakukan analisis sentimen terhadap jasa transportasi online untuk mengetahui bagaimana tanggapan masyarakat terhadap jasa tranportasi online tersebut. Tujuan dari penelitian ini adalah untuk menganalisa tanggapan masyarakan dengan analisis data yang berupa tweet kemudian diklasifikasikan menjadi kelas positif, negatif, dan netral menggunakan metode Naïve Bayes dikarenakan tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi naive bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar . Hasil penelitian ini menunjukkan bahwa dengan metode yang digunakan tingkat sentimen netral dari tweet masyarakat lebih besar dibandingkan dengan tingkat sentimen positif dan sentimen negatif dengan akurasi sebesar 25.00%.*

Kata Kunci: *Transportasi Online, Analisis sentimen, Twitter, Naïve Bayes.*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4004

INTRODUCTION

The development of technology and information is so fast. The breadth of internet service systems and the high influence of smart phones have made Indonesia one of the countries with the potential for developing online-based applications. One of them is in the field of public transportation. Users of technology and information systems that make transportation services more efficient to use, namely by ordering online via smart phone [1].

By registering as an online transportation driver, both private two-wheeled and four-wheeled vehicles, these private vehicles can be as useful as public transportation that can be ordered by the public, utilizing access to cellular telephone technology. In addition, this online transportation application is not only used as a means of transportation for the community but can also serve as a goods delivery service and food delivery. This online transportation phenomenon is becoming popular quickly because it offers the latest innovations regarding transportation combined with online communication technology so that it makes it easy for people to order transportation anywhere and anytime. [2].

With this phenomenon, many people express their opinions about online transportation in Indonesia through social media. Social media is a service that facilitates the exchange of information and topics on an ongoing basis with a broad scope [3]. One of the most popular social media in society is Twitter. Twitter is a social media that can connect many people with various topics from around the world [4]. Dengan menggunakan Twitter masyarakat dapat memberikan pendapat mereka tentang apapun yang terjadi secara langsung. Hal tersebut didukung dengan banyaknya masyarakat yang sudah memiliki telepon selular sehingga memudahkan untuk mengakses Internet untuk menggunakan media sosial. Populasi penduduk Indonesia saat ini mencapai 262 juta orang. Lebih dari 50 persen atau sekitar 143 juta orang telah terhubung jaringan Internet sepanjang 2017 [5].

Based on PeerReach research, Indonesia is the third most active Twitter user in the world, which means that Twitter users in Indonesia are among the most active in the world. [6] If we examine further the community's tweets, we will get a sentiment that can be analyzed.

In previous research[10]positive sentiment results of 88.60% and negative sentiment of 11.40% with an accuracy of 86.80%. The results show that the level of positive sentiment from public tweets is greater than the level of negative sentiment.

In this study, the authors analyzed public sentiment on Twitter to provide information about public satisfaction with online transportation services in Indonesia. The method to be used is the Naïve Bayes Classifier. One of the reasons this method was chosen is because the error rate obtained is lower when the dataset is large, besides that the accuracy of Naive Bayes and the speed is higher when applied to a dataset with a larger number, it has several advantages, including, simple, fast, and high accuracy. [7].

Data must go through the pre-processing stage before being classified. After being classified, you will get tweets that have positive or negative meaning.

Sentiment analysis is an analysis of an event from an opinion based on a person's attitude about an object. Sentiment analysis is usually done to collect and find out public opinion in Blog posts, Twitter, Facebook, and others. Sentiment analysis is needed with the aim of knowing public opinion on an object. These opinions can be negative or positive opinions depending on the public's view of the object. Therefore an analysis of these opinions is needed in this study so that it can be used as a benchmark for whether or not online transportation services are good according to customers.[8].

From the background of the problems above, the purpose of this study is to measure the accuracy, class precision, and class recall of the Naïve Bayes-based text classifier method to carry out sentiment analysis of online transportation services on Twitter.

MATERIALS AND METHODS

Data Collection Method, Population and Research Sample

A. Data Collection Method

In this chapter explains the steps undertaken by researchers included in the quantitative research methods, namely [9]:

1. Preparation Stage:

At this stage it is a stage that prepares material related to the selection of high achieving students and decision support systems, formulation of problems by gathering preliminary information to find out the background of the problem, identification of problems, goals and objectives scope and hypotheses, and compiling a study of literature relating to research.

2. Collecting data stage:

Data collection techniques used in this study is collecting data on Twitter to get data and conducting interviews and observations as a start to start research

3. Pre-processing:

Pre-processing is carried out in six stages, namely as follows following:

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4004

a. Cleansing: This stage is the character elimination stage non alphabetic to reduce noise. Deleted characters punctuation marks such as periods (.), commas (,), question marks (?),and exclamation points (!), as well as symbols such as the '@' sign for username, hashtag (#), emoticon, and website address.
b. Case Folding: Case folding is the stage for converting staged alphabetic characters cleanup to lowercase (lowercase).
c. Tokenizing: This stage functions as a breaker sentence based on each word that composes it, that is called terms or tokens. Tokenizing is broken down by space.
d. Slang word Normalization or Conversion:
This stage done so the words are shortened or extended into normal words according to the Big Dictionary Indonesian (KBBI). Cconversion is a process change non-standard words to standard words, this stage done with the help of a Tokenizing dictionary in words standard and check the word is in the dictionary slang or not. If the non-standard word is in dictionary, then the non-standard word is changed to the standard word is in the dictionary.
e. Filtering or Removing Stop words: This stage is processing so that words that are not important or not meaningful are deleted

4. Term weighting:

This research utilizes Term Frequency-Inverse Document Frequency (TF-IDF), which is implemented using Rapid Miner tools which done with the operator Process Document from Data (extension of Text Processing). Term Frequency (tf(w,d)) is considered to have a proportion of importance according to the total appearance in the text or document. Inverse Document Frequency (IDF) is a token weighting method that functions to monitor token occurrences in text sets. TF-IDF is a statistic to show the vitality of a word in a dataset or document [10]. Data that has gone through the pre-processing stage must be numeric form. TF-IDF is used to change data it becomes numeric. In the weight calculation using TF-IDF, the TF value per word is calculated first with the weight of each word is 1. IDF(word) is IDF value of each searched word, td is the number all existing documents, and df is the number the word appears in all documents.

Stages Study

The stages of research in making this research can be seen in Figure 1 below:

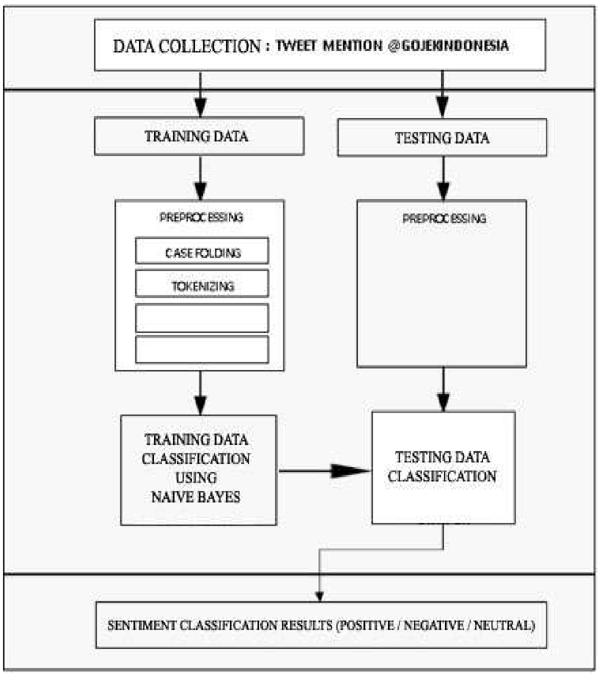


Figure 1. Research Stages

B. Population

The population can be interpreted as a generalization area consisting of: objects/subjects that have certain qualities and characteristics determined by the researcher to be studied and then conclusions drawn [11]

Population as a whole group, people, events or things that are interesting for researchers to study [11].

So the population is not only people, but also objects and other natural objects. The population is also not just the amount that exists in the object/subject being studied, but includes all the characteristics/values possessed by the subject or object.

C. Research Sample

The sample is a number of examples from a population that has the same characteristics as the population and is directly the target of research [11].

The sample is defined as part or subset of the population consisting of selected members of the population [12]

If the writer wants to do research on a large population, while the writer wants to research about this population and the writer has limited funds, manpower, time and other conditions, then the writer determines the sample using a simple random sampling technique (Simple Random Sampling) taken from several Twitter respondents by calculating the number of samples using the Slovin method with an error tolerance of 0.1 (10%).[13]

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4004

D. Research Instruments

The research instrument is a tool used to measure observed natural and social phenomena [14]. In this study there are research instruments to obtain the data needed to support the research. In this study, the authors obtained data with secondary data, namely online transportation data from this study obtained from the general public's Tweets using a simple random sampling technique (Simple Random Sampling).

E. Data Analysis Methods

The analytical method used is the Naïve Bayes Classifier (NBC) which is a simple probability classifier that applies Bayes' Theorem with the assumption of high independence. The advantage of using NBC is that it is small to determine parameter estimates as an independent variable, so only the variance of a variable in a class is needed to determine the classification, not the entire covariance matrix.[15]

RESULTS AND DISCUSSION

A. Research Results

The data collection process was taken from the crawling process of the RapidMiner application using the Twitter search operator with the query "@GojekIndonesia". Then the data is stored in excel format. The training data used when testing the data was taken from Twitter. Data testing was carried out using public opinion reviews about Sentiment Analysis for Online Transportation Services. Then testing and training is carried out so that accuracy is obtained. The following describes in more detail the research results obtained. The results of this study indicate that with the method used the level of neutral sentiment from public tweets is greater than the level of positive sentiment and negative sentiment with an accuracy of 25.00%.

B. Data Entry Process

In the process of entering data taken from excel data in the form of training data and test data which contains words from the results of word weighting using Microsoft Excel. An overview of word data from excel files before being entered into Rapid Miner as training and test data can be seen in table 1.

Table 1. Tweet Data From Twitter

No	Tweet	Label/Class
1	Why do I rarely get GO-FOOD vouchers now?	*Negative*
2	hello min, how come it's been 4 months since my go-jek account has never gotten any promo vouchers, has it been banned or what?	*Neutral*
3	hello, admin, how come I want to order Gofood, they say the server is busy all the time	*Neutral*
4	admin, my account has been blocked but emailed and called there is no solution, so what do you do? there is still gopay that can be used	*Negative*
5	admin, I'm asking for help, I ordered go food from half past 7 until now it hasn't arrived and the driver can't be contacted, I can't cancel.	*Neutral*
6	hi min! Today I ordered gofood with the Visa debit card payment method, when the order arrived "the server was full my order was canceled but my balance was drained according to the amount. Can you help min??	*Neutral*
7	already don't have this money Claim covid hurry up why disbursement account is turned off, money is not sent either	*Negative*
8	Just be fair, whoever claims first will be processed first. Don't get privileges, but many are given a long time. In the regulations it is clearly written that it applies to all drivers, there are no castes.	*Negative*
9	I want to log in, how come I can't?	*Neutral*
10	The voucher that was given to me can't be used. Please check your DM, min.	*Positive*
11	My account suddenly logs out on its own, I want to log in again but the verification link doesn't	*Positive*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4004

No	Tweet	Label/Class
	appear in the SMS	
12	hello gojek I have been called 3 times by gojek, but I answered but only answered hello hello. Until I'm confused, what error does my cellphone have? Btw, why are you calling, min?	*Positive*
13	down.. the application on my cellphone can't be used at all..	*Positive*
14	Hawo, etmin here, right? There's a discount of 20k, but how come when you order, you pay 70k?	*Positive*
15	Recently, why is the driver so hard to contact, since yesterday I often get drivers who can't be contacted, I've been waiting 30 minutes and then they just cancelled. we customers have no other option cuma just waiting, please how is this	*Negative*
16	why my account is blocked and my balance is red? Even though it doesn't violate the gojek community / and doesn't commit any fraud	*Negative*
17	I have registered for goride through the gopartner application and I have registered as a goride partner... Waiting for verification, until when should I wait... Thank you	*Neutral*
18	hi sis, I want to ask if the gofood restaurant sent the wrong food and I want to claim the food that was ordered, how about it how do you do it?	*Neutral*
19	Gojek has a habit like this every time it's about to overdue, what do you do on purpose to get fined? It goes on like this, where are your engineers whose salaries are tens of millions? since March there has always been a bug like this, fix it asap	*Neutral*
20	Before adding, it was difficult to find orders, let alone driver added	*Neutral*

C. Manufacture Process

The next step is to create each process in RapidMiner which consists of training data, data testing or tests, retrieve trainer methods, naïve Bayes, apply models and performance. Data imported into Rapid Miner as training data can later be seen in Figure 2.

Row No.	Sentimen	Text
1	Negative	kenapa skarang jrang dpet voucher GO-FOOD sih
2	Neutral	halo min, kenapa ya udah 4 bulan ini akun gojek saya gapernah ...
3	Neutral	halo min kok saya mau pesen gofood dibilangnya server busy te...
4	Negative	min akun saya kena block tapi diemail dan ditelfon gaada solusi...
5	Neutral	min minta tolong, saya pesen go food dari setengah 7 sampe s...
6	Neutral	hi min! Hari ini saya pesen gofood dengan metode pembayaran ...
7	Negative	daah ga punya duit ini.... Klaim covid cepetan napa cairinnya... A...
8	Negative	adil ajalah , siapa yang klaim duluan ya di proses duluan. Janga...
9	Neutral	saya mau login kok gakbisa bisa ya??
10	Positive	voucher yang dikasih ke saya tidak bisa digunakan nih. Tolong c...
11	Positive	ini akunku tiba-tiba ke logout sendiri, mau login lagi tapi link verif...
12	Positive	halo gojek saya udah 3 kali nih ditelepon sama gojek, tapi saya j...
13	Positive	down.. aplikasi di hape gw gabisa dipake sama sekali..
14	Positive	hawo etmin disini kan ket nya diskon 20k ya tp kok pas order ttp ...
15	Negative	ini belakangan ini drivernya kenapa susah bgt dihubungin ya, da...
16	Negative	kenapa akun saya teblokir dan saldo saya merah? Padahal tida...
17	Neutral	saya sudah daftar goride melalui aplikasi gopartner dan saya su...
18	Neutral	hai kak mau tanya kalau resto gofood salah kirim makanan dan ...
19	Neutral	Gojek kebiasaan bgt begini tiap udah mau overdue, sengaja bia...
20	Neutral	Sebelum ditambah aja susah nyari orderan ka, apalagi driverny...

Figure 2. Upload training data in .CSV format

Data imported into Rapid Miner as testing data can be seen in Figure 3.

Row No.	Sentimen Manual	Text
1	Negative	tolong tanggapi email saya dengan nomor tiket 72654451. S...
2	Neutral	kak saya mau hapus akun gojek yang ternyata nomor telpon ...
3	Positive	kapan keluarin voucher 75% lg? i need that
4	Negative	ini knp abangnya marah marah pls aku gaterima ya digininin
5	Positive	hai min mau tanya tentang pergantian rekening buat gofood ...
6	Positive	ada lowongan driver dgn sepeda?
7	Positive	voucher gofood nya dibanyakin dong min, masa punyaku vou...
8	Positive	bagaimana cara menghapus akun gojek yang sudah tidak di...
9	Negative	haloo tadi aku pesen gofood tapi pesenannya gak lengkap u...
10	Negative	min tolong dong ini saya kirim pengaduan saya tentang Go F...
11	Positive	min mau tanya, kalo pesen gobox tuh sama kaya kalau pese...
12	Neutral	minn knp yaa top up GO-PAY di alfamart katanya akun aku ga...
13	Negative	halo selain via twitter kalo mau nanya" soal gojek kemana la...
14	Negative	top up aku kok ga masuk ya dari m bangking aku sendiri? pa...
15	Positive	Tolong ni ya buat gojek masa yang jauh2 bisa kecantol gini? ...
16	Negative	kenapa ga bisa di pake ya promo nya?? padahal udah menc...
17	Neutral	Halo, saya mau tanya ini kenapa gini terus ya kl lagi order? J...
18	Negative	saya sdh bayar bpjs tgl 24 Juni selama 3 bulan . Saldo terpot...
19	Neutral	kok promo pengguna baru gojek tokopedia gabisa sih
20	Neutral	kok akunku jarang dapet promo sih min?padahal aku sering ...

Figure 3 Upload Testing Data in .CSV Format

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4004

After the data import process is complete, then carry out the process flow design process and input other processes such as the retrieve trainer method, naïve Bayes, apply model and performance. The following method can be seen in Figure 4.

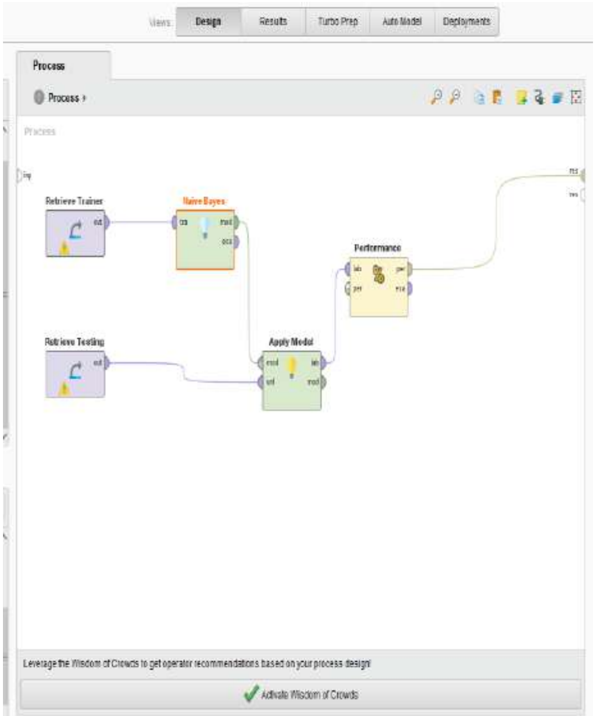


Figure 4 Adding Naïve Bayes Design, Apply Model and Performance

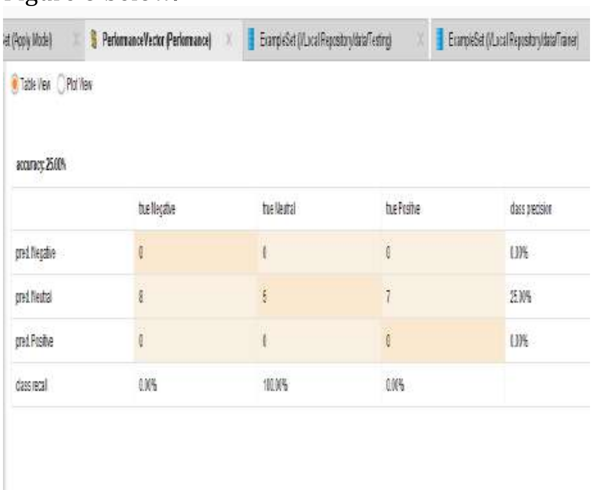
In figure 4 explain the training data used is 20 After pre-processing the training data remains at 20, the next stage is the classification of sentiment analysis. This stage is the stage to provide training and implement various machine learning algorithms. Figure 4 shows the contents of the "Cross Validation" operator in a RapidMiner application. In this process, two classification operators "Naive Bayes" are used After the experiment and the proposed model are created, then experiments are carried out by testing existing models with grouped datasets that become training data and testing data. The "Performance" operator is an operator for displaying accuracy, precision and recall results.

D. Rapid Miner Analysis Results

After processing the Rapid Miner, we get data that has been tested by the methods in Rapid Miner. The display of the analysis results can be seen in Figure 5 below.

Figure 5. Display of Data Analysis Results

The Rapid Miner can also see the performance results from the naïve Bayes calculations, so the performance results obtained from the Naïve Bayes analysis results are 100% Neutral, as can be seen in Figure 6 below.



accuracy: 25.00%

	true Negative	true Neutral	true Positive	class precision
pred. Negative	0	0	0	0.00%
pred. Neutral	8	5	7	25.00%
pred. Positive	0	0	0	0.00%
class recall	0.00%	100.00%	0.00%	

Figure 6. Accuracy Result Display

CONCLUSION

Based on the results of research and discussion carried out starting from the design stage to testing, it can be concluded that the results of the analysis using Rapid Miner for the Pred.Negative class get True Negative results, namely 0, True Neutral, namely 0, True Positive, namely 0 and results from class precision, namely 0.00%. the results of the analysis using the Rapid Miner for the Pred.Neutral class get True Negative results, namely 0, True Neutral, namely 5, True Positive, namely 7 and the results of the precision class, namely 25.00%, the results of the analysis using Rapid Miner for the Pred.Positive class get True Negative results,

CC BY NC

namely 0, True Neutral, namely 0, True Positive, namely 0 and the results of class precision, namely 0.00%, and the results of analysis using Rapid Miner for class recall get True Negative results, namely 0.00%, True Neutral, namely 100.00%, True Positive, namely 0.00%. So sentiment analysis on Gojek online transportation is more neutral by Twitter users than positive or negative values.

REFERENCE

[1] B. Abdullah and B. A. Saebani, *Metode penelitian ekonomi islam : Muamalah*. Bandung: Pustaka setia, 2014.

[2] I. Verawati and B. S. Audit, “Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022.

[3] Hartanto, “Hartanto 2017 text mining dan sentimen analisis twitter,” *J. Psikol. Ilm.*, vol. 9, no. 1, pp. 18–25, 2017.

[4] A. Wandani, “Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes,” *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 651–665, 2021.

[5] S. Fransiska and Yolanda, “Analisis Sentimen Twitter Untuk Review Film Menggunakan Algoritma Naive Bayes Classifier (Nbc) Pada Sentimen R Programming,” *J. Siliwangi*, vol. 5, no. 2, pp. 68–71, 2019.

[6] S. Sumanto and L. S. Marita, “Fuzzy Simple Additive Weighting (Fsaw) Untuk Analisa Kelayakan Pemberian Kredit Mobil,” *J I M P - J. Inform. Merdeka Pasuruan*, vol. 2, no. 2, pp. 1–13, 2017.

[7] E. Dwianto and M. Sadikin, “Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine,” *Format J. Ilm. Tek. Inform.*, vol. 10, no. 1, p. 94, 2021.

[8] W. Supriyanti, Kusrini, and A. Amborowati, “Perbandingan Kinerja Algoritma C4.5 Dan Naive Bayes untuk Ketepatan Pemilihan Konsentrasi Mahasiswa,” *J. Inf. Politek. Indonusa*, vol. 1, no. 3, pp. 61–67, 2016.

[9] T. T. Widowati and M. Sadikin, “Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021.

[10] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, “Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 108–118, 2021.

[11] M. W. A. Putra, Susanti, Erlin, and Herwin, “Analisis Sentimen Dompet Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020.

[12] E. Sumartono and Y. Astria, “Strategi Nafkah Petani Sawit Di Desa Penarik Kecamatan Penarik Kabupaten Mukomuko,” *MAHATANI J. Agribisnis (agribus. Agric. Econ. Journal)*, vol. 1, no. 2, pp. 80–95, 2019.

[13] V. Kevin, S. Que, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization),” vol. 9, no. 2, pp. 162–170, 2020.

[14] Bustami, “Penerapan Algoritma Naive Bayes,” *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.

[15] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, “Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5068–5074, 2021.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.2463

COMPARISON OF DIFFERENT KERNEL FUNCTIONS OF SVM CLASSIFICATION METHOD FOR SPAM DETECTION

AAIN Eka Karyawati[1*)]; Komang Dhiyo Yonatha Wijaya[2]; I Wayan Supriana[3]

Informatics Study Program
Universitas Udayana
https://if.unud.ac.id
eka.karyawati@unud.ac.id[1]; komangdhiyo66@gmail.com[2]; wayan.supriana@unud.ac.id[3]

(*) Corresponding Author

Abstract—*Today, the use of e-mail, especially for formal online communication, is still often done. There is one common problem faced by e-mail users, which is the frequent receiving of spam messages. Spam messages are generally in the form of advertising or promotional messages in bulk to everyone. Of course this will cause inconvenience for people who receive the SPAM message. SPAM e-mails can be interpreted as junk messages or junk mail. So that spam has the nature of sending electronic messages repeatedly to the owner of the e-mail. This is abuse of the messaging system. One way to solve the spam problem is to identify spam messages for automatic message filtering. Several machine learning based methods are used to classify spam messages. In this study, a comparison was made between several kernel functions (i.e., linear, degree 1 polynomial, degree 2 polynomial, degree 3 polynomial, and RBF) of the SVM method to get the best SVM model in identifying spam messages. The evaluation results based on the Kaggle 1100 dataset showed that the best model were the SVM model with a linear kernel function and a degree 1 polynomial, where both models returned Precision = 0.99, Recall = 0.99, and F1-Score = 0.98. On the other hand, the RBF kernel produced lower performance in terms of Precision, Recall, and F1-Score of 0.95, 0.95, and 0.94, respectively.*

Keywords: *Spam, SVM, Kernel Function, Classification.*

Intisari—*Dewasa ini, penggunaan e-mail, khususnya untuk komunikasi formal secara online, masih sering dilakukan. Ada satu masalah umum yang dihadapi oleh pengguna e-mail, yaitu seringnya menerima pesan spam. Pesan spam umumnya berupa pesan iklan atau promosi secara massal kepada semua orang. Tentu hal ini akan menimbulkan ketidaknyamanan bagi orang yang menerima pesan SPAM tersebut. e-mail SPAM dapat diartikan sebagai pesan sampah atau junk mail. Sehingga spam memiliki sifat mengirimkan pesan elektronik secara berulang-ulang kepada pemilik e-mail tersebut. Ini adalah penyalahgunaan sistem pesan. Salah satu cara untuk mengatasi masalah spam adalah dengan mengidentifikasi pesan spam untuk pemfilteran pesan otomatis. Beberapa metode berbasis pembelajaran mesin digunakan untuk mengklasifikasikan pesan spam. Pada penelitian ini dilakukan perbandingan antara beberapa fungsi kernel (yaitu linear, polinomial derajat 1, polinomial derajat 2, polinomial derajat 3, dan RBF) dari metode SVM untuk mendapatkan model SVM terbaik dalam mengidentifikasi pesan spam. Hasil evaluasi berdasarkan dataset Kaggle 1100 menunjukkan bahwa model terbaik adalah model SVM dengan fungsi kernel linier dan polinomial derajat 1, dimana kedua model mengembalikan Precision = 0.99, Recall = 0.99, dan F1-Score = 0.98. Di sisi lain, kernel RBF menghasilkan kinerja yang lebih rendah dalam hal Precision, Recall, dan F1-Score masing-masing sebesar 0,95, 0,95, dan 0,94.*

Kata Kunci: *Spam, SVM, Fungsi Kernel, Klasifikasi..*

INTRODUCTION

Today, the use of e-mail, especially for formal online communication, is still often done. There is one common problem faced by e-mail users, which is the frequent receiving of spam messages. Spam messages are generally in the form of advertising or promotional messages in bulk to everyone.According to J. Clement as of December 2019 the number of spam e-mails covered 57.26% of the total number of e-mails. Spam is often done for advertising, to get people who are spammed to reply to the message, or to annoy people who are spamming. For this reason, an identification of spam is needed to filter out Spam.

Identification is a specifik task of classification. One approach that usually used to do classification is the machine learning-based

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

method. There have been many studies used machine learning-based spam classification [1]–[6].

A machine learning has the advantage that it is easy to implement and good for high-dimensional data. However, it has the disadvantage of requiring unbiased and large amounts of data. In addition, adjusting parameters and complexity of the model is needed to select the best model.

In this research, a machine-learning based methode namely SVM is used to classify e-mail into two classes (i.e., Spam and not Spam). The aim of this reasearh is to select the best SVM model by comparing some kernel function of the SVM method, besides the parameters.

Several studies [7]–[10] used SVM for spam classification. [7] compared KNN, linear kernel SVM and RBF kernel SVM method. In this study, it was found that the KNN method at k=3 produced the best accuracy of 92.28% while the best accuracy in the SVM method was obtained using the SVM linear kernel with an accuracy of 96.6%. It can be concluded that the SVM method is better than KNN. [8] compared the Naive Bayes method and the SVM method with the RBF kernel to identify Instagram comment spam. The results showed that the SVM method produced an accuracy of 78.49%, which is better than the Naive Bayes method which produced an accuracy of 77.25%.

Other study [9] proposed a combination of KNN and SVM method. It used KNN-based sampling strategy to find close neighbors to improve the performance of the SVM method. The results of the study based on publicly available dataset (Dredze) showed the accuracy increased to about 98%. [10] proposed a new spam detection method that effective in distinguishing spam from its content. During classifying the dataset, the proposed classifier obtained a classification accuracy of 95.32 percent.

In this study several kernel functions (i.e., Linear, Polynomial, and RBF) were investigated to obtain the best SVM model for classifying Spam e-mails. Some experiments were conducted to determine the effect of parameter changes for each kernel function. The SVM performance was measured using the Precision, Recall, and F-Measure metrics.

MATERIALS AND METHODS

Research Steps

The general flow of this research starts from collecting raw data, then data preprocessing (i.e., tokenization, case folding, stop word removal, and stemming), then TF-IDF weighting, training each SVM with various kernels, and finally, evaluating the best SVM. model for each kernel.

Data Collection

The e-mail dataset was collected from https://www.kaggle.com/datasets/venky73/spam-mails-dataset which contains both spam and non-spam e-mails extracted from the e-mail body. The total number of 1100 data was 550 Spam data and 550 non-Spam data, of which 1000 data were used as training data (i.e., to select the best model of each kernel), and 100 data are used for testing (i.e., to evaluate the best model). The feature data used was word frequency. The data was in the form of text in .csv format. An example of a non-Spam e-mail and a Spam e-mail can be seen in Figure 1 and Figure 2, respectively.

0 Subject: research mike , vince and i are eager to see if our group can play a role in helping you in your development work using some combination of the or experts in our group and the resources to which we have access at stanford . can we get together for a short planning session when you are next in houston ? please let me know your schedule , or have your assistant coordinate a time with vince ' s assistant , shirley crenshaw (x 35290) . thanks , stinson

Figure 1. An Example of Non-Spam e-mail

1 Subject: having problems in bed ? we can help ! cialis allows men to enjoy a fully normal sex life without having to plan the sexual act . if we let things terrify us , life will not be worth living . brevity is the soul of lingerie . suspicion always haunts the guilty mind .

Figure 2. An Example of Spam e-mail

TF-IDF Weighting

The term weighting with TF-IDF starts by calculating the term frequency (tf). After that, a temporary weight calculation is carried out with equation (1) and Wf is obtained. To reduce the value of terms that occur frequently, the document frequency (df) of the term is calculated, followed by the calculation of the inverse-document frequency (idf) by equation (2). To get the tf-idf weight, the calculation is carried out according to equation (3) and the result is a term weight matrix. The results of this weighting will be used in the classification process with the SVM method.

Calculate term frequency (tf) by calculating the frequency of terms in the document and term weight (Wf) .

$$Wf_{t,d} = \begin{cases} 1 + log tf t, d, jika f_{t,d} > 0 \\ 0, jika f_{t,d} \leq 0 \end{cases} \quad \text{......................} (1)$$

With Wft,d is for weigh term t in document d, tf is for term frequency.

Calculate the document frequency (df) by calculating the frequency of the document where the term is located. Calculate the inverse document frequency (idf) [11].

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.2463

$idf_t = log \frac{N}{dft}$ .... (2)

With idf_t is the inverse document frequency for term t, df_t is document frequency of term t, and N is total frequency of documents.
Calculated the value of TF-IDF weighting [11].

$Wtf\text{-}idf_{t,d} = Wf_{t,d} * idf_t$ .... (3)

With Wtf-idf is the value of tf-idf weighting, Wf is the weight of term, and idf is inverse document frequency.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) algorithm is a supervised learning method that produces an input-output mapping function from a series of training data that already has a label [12], [13]. Nonlinear kernel functions are often used to convert the input data to a high-dimensional feature space where the input data becomes more separable than the original input space. The algorithm of SVM used in this research is shown in Figure 3.

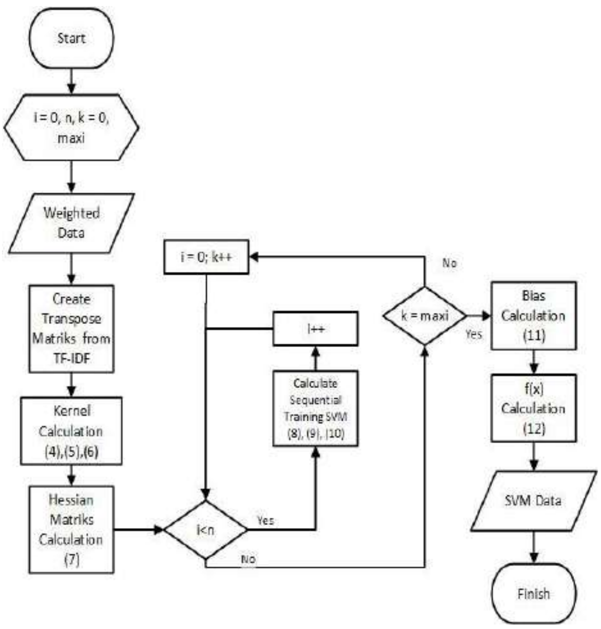


Figure 3. The flowchart of SVM

The SVM kernel used in this study were a polynomial, linear and RBF kernel using equation (4), (5), and (6), respectively [14].

Linear Kernel

$K(x_i, x_j) = x_i^T x_j$ .... (4)

Polynomial Kernel

$K(xi, xj) = x_i^T x_j + C^d$ .... (5)

RBF Kernel

$K(x_i, x_j) = exp\left(-\gamma |x_i - x_j|^2\right), \gamma > 0$ .... (6)

With $K(x_i, x_j)$ is kernel fuction, x_i is i-th data, x_j is j-th data, C for slack variable, d for degree, and γ for learning rate.

The steps in using the SVM method are as follows [6]:

a. Initiation of parameters used such as λ and γ (error rate).
b. Calculate the Hessian matrix.

$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2)$ .... (7)

With D_{ij} is Hessian matrix value, y_i is i-th class, y_j is j-th class, and λ for error control.

c. Starting from the 1st data to the nth data, do the calculation iterations.

$\varepsilon_i = \sum_{j=1}^{n} \alpha_j D_{ij}$ .... (8)
$\delta\alpha_i = min\{max[\gamma(1 - \varepsilon_i), -\alpha_i], C - \alpha_i\}$ .... (9)
$\alpha_i = \alpha_i + \delta\alpha_i$ .... (10)

With ε is error value and α_i is support vector.

$b = -\frac{1}{2} [(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x^-)) + (\sum_{i=1}^{n} \alpha_i y_i K(x_i, x^+))]$

d. From the previous calculation, the largest value of α_i is sought and calculations are carried out to determine the bias.

.... (11)

With b is for bias value.

e. To find out the results of the sentiment analysis, the f(x) function is calculated.

$f(x) = \sum_{i=0}^{n} \alpha_i y_i K(x_i, x) + b$ .... (12)

With f(x) is for classification function.

Selecting the Best Model

Model selection was done by adjusting the SVM parameters, namely learning rate (γ), error control (λ), and d (polynomial degree), for each kernel. The best model was determined by measuring the model's performance (i.e., F-Measure). The highest F-Measure in each particular combination of kernel parameters was chosen as the best model for that kernel.

Evaluation

The best models of each kernel were evaluated based on 100 testing data. The evaluation metrics used were Precision, Recall, and F-Measure [15].

RESULTS AND DISCUSSION

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

The SVM Model Selection for Liner Kernel

The Experiments of λ Changes

Table 1 showed the results of the effect of λ changes to the results of the SVM classifier performance with linear kernel. The experiment was carried out with γ = 0.1. The B value was started from 0.01 and progressed to a value of 2. As can be seen, there were no changes in term of Precision, Recall, and F-Measure or the value of B has no effect on the linear kernel.

Table 1. Effect of λ Changes in Linear Kernel

λ	Precision	Recall	F-Measure
0,01	0.9913	0.991	0.9909
0,1	**0.9913**	**0.991**	**0.9909**
1	0.9913	0.991	0.9909
2	0.9913	0.991	0.9909

The Experiments of γ Changes

Table 2 showed the results of testing the effect of changing parameter γ on the results of SVM classifier performance with a linear kernel. The test was carried out with λ = 0.1 and γ started from 0.0001 which was continued until γ = 0.1. As can be seen in Table 2, there was no change in the value of the performance measure (ie, Precision, Recall or N-Measure). So, similar to change B, change γ has no effect on the linear kernel.

Table 2. The Effect of γ Changes in Linear Kernel

γ	Precision	Recall	F-Measure
0,0001	0.9913	0.991	0.9909
0,001	0.9913	0.991	0.9909
0,01	0.9913	0.991	0.9909
0,1	**0.9913**	**0.991**	**0.9909**

The Classification Result of SVM with Polynomial Kernel

The following were the results obtained from the classification using the SVM with polynomial kernel. The experiments was conducted using the 10-fold cross validation method.

The SVM Model Selection for Degree 1 Polynomial Kernel

The Experiments of λ Changes

Table 3 showed the results of testing the effect of λ canges to the results of the SVM classifier performance with a degree 1 polynomial kernel. The testing was carried out with γ = 0.1 and λ was started from 0.01 which continued until λ = 2. From Table 3, it can be seen that there were no change in the value of Precision, Recall and F-measure. The λ values has no effect on polynomial kernel of degree 1.

Table 3. Effect of λ Changes in degree 1 polynomial kernel

λ	Precision	Recall	F-Measure
0,01	0.9923	0.992	0.9919
0,1	**0.9923**	**0.992**	**0.9919**
1	0.9923	0.992	0.9919
2	0.9923	0.992	0.9919

The Experiments of γ Changes

Table 4 showed the results of testing, observing the effect of γ canges to the results of the SVM classifier with a degree 1 polynomial kernel. The test was carried out with λ = 0.1, and γ was started from 0.0001 which continued until 0.1. The best results was found at γ = 0.001 with an average f-measure value of all folds of 0.9919 or 99.19%. As can be seen in Figure 4, the greater the value of γ (learning rate), the lower the value of SVM performance, the evaluation results increase to a peak at 0.001, where after that it decreases. In addition, experiments also showed that a learning rate that was too small gived poor results.

Table 4. Effect of γ Changes in degree 1 polynomial kernel

γ	Precision	Recall	F-Measure
0,0001	0.942	0.933	0.9324
0,001	**0.9931**	**0.993**	**0.992**
0,01	0.9923	0.992	0.9919
0,1	0.9913	0.991	0.9909

The SVM Model Selection for Degree 2 Polynomial Kernel

The Experiments of λ Changes

Table 5 showed the results of the research on the effect of the value of γ changes to SVM classifier performance with a degree 2 polynomial kernel. The experiments was conducted with γ = 0.1 and λ was started from 0.01 which continued to 2. The best result was found when l λ = 2 with an average F-Measure of all folds of 0.8568. Figure 5 showed that the classification performance increase to a peak at λ = 2. Because the value of λ is a value that indicates the degree of importance of the occurrence of misclassification and the greater the value of λ, the smaller the error of classification that can be allowed.

Table 5. Effect of Parameter λ in degree 2 polynomial kernel

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.2463

λ	Precision	Recall	F-Measure
0,01	0.8592	0.8379	0.8347
0,1	0.8592	0.8379	0.8347
1	0.8548	0.845	0.8422
2	**0.8766**	**0.859**	**0.8568**

The Experiments of γ Changes

Table 6 showed the results of γ changes to the SVM classifier performance using a degree 2 polynomial kernel. The experiments were carried out with λ = 2 and γ was started from 0.0001 which continues to 0.1. The best result was found at γ = 0.01 with an average F-Measure value of all folds of 0.8568. Figure 6 showed that the highest performance was at γ = 0.1. Experiments also showed that a learning rate that was too small gave poor results.

Table 6. Effect of γ Changes in Degree 2 Polynomial Kernel

γ	Precision	Recall	F-Measure
0,0001	0.6616	0.544	0.4223
0,001	0.8026	0.6839	0.6432
0,01	0.8796	0.845	0.8384
0,1	**0.8766**	**0.859**	**0.8568**

The SVM Model Selection for Degree 3 Polynomial Kernel

The Experiments of λ Changes

Table 7 showed the results λ changes to SVM classifier performance with a degree 3 polynomial kernel. The experiments was conducted with γ = 0.1 and λ was started from 0.01 which continued to 2. The best results were found when λ = 0.01 and λ = 0.1 with an average F-Measure value of all folds of 0.9919. As can be seen from Figure 7, the classification performance results were at their peak at = 0.01 and 0.01, after which they decreased. Because λ is the degree of importance of the occurrence of misclassification, then the greater the value of λ, the smaller the error of classification that can be allowed. In the classification model using a degree 3 polynomial kernel, overfitting started occured when λ = 1 so that the classification results drop drastically.

Table 7. Effect of λ Changes in Degree 3 Polynomial Kernel

λ	Precision	Recall	F-Measure
0,01	0.7831	0.715	0.6939
0,1	**0.7831**	**0.715**	**0.6939**
1	0.772	0.692	0.6673
2	0.752	0.693	0.6678

The Experiments of γ Changes

Table 8 showed the results of the effect of γ changes to the SVM classifier performance using a degree 3 polynomial kernel. The experiments were carried out with λ = 0.1 and γ was started from 0.0001 which continued to 0.1. The best result was shown when γ = 0.01 with the average F-Measure value of all folds of 0.75. As can be seen from Figure 7, that the classification performance result was at its peak at γ = 0.01 where after it has decreased. Experiments showed that a learning rate that is too small gives poor results.

Table 8. Effect of γ Changes in Degree 3 Polynomial Kernel

γ	Precision	Recall	F-Measure
0,0001	0.7127	0.548	0.4302
0,001	0.7571	0.6	0.5249
0,01	**0.8252**	**0.763**	**0.75**
0,1	0.7831	0.715	0.6939

The SVM Model Selection for RBF Kernel

The following are the results obtained from the experiments using the SVM with RBF kernel. Experiments were conducted via the k-fold cross validation method with k = 10.

The Experiments of λ Changes

Table 9 shows the results of the study on the effects of the λ changes to the SVM classifer performance using a RBF kernel. The experiments were carried out with γ = 0.1 and λ was started from 0.01 which continued to 2. The best result was found λ = 0.01 and 0.1 with an average F-measure value of all folds of 0.4494 (low value). From that experiments showed poor performance at learning rate (γ) = 0.1 .

Table 9. Effect of λ Changes in RBF Kernel

λ	Precision	Recall	F-Measure
0,01	0.716	0.5589	0.4494
0,1	**0.716**	**0.5589**	**0.4494**
1	0.7131	0.549	0.4316
2	0.6418	0.517	0.3711

The Experiments of γ Changes

Table 10 shows the results of the γ changes with to the SVM classifier performance with RBF kernel. The experiments were conducted with λ = 0.1 and γ was started from 0.0001 which continued to 0.1. The best result was found γ = 0.0001 and 0.001 with an average F-Measure value of all folds

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

of 0.9676. From Figure 10 showed that the classification result was at its peak at γ = 0.001 where after it has decreased. The experiments showed that the greater the value of γ (learning rate), the worse the classification results.

Table 10. Effect of γ Changes in RBF Kernel

γ	Precision	Recall	F-Measure
0,0001	0.9702	0.968	0.9679
0,001	**0.9702**	**0.968**	**0.9679**
0,01	0.8547	0.794	0.7842
0,1	0.7160	0.5589	0.4494

Evaluation the Best Models

After conducting experiments to select the best model for each kernel (i.e., 5 models), the evaluation for measuring performance for those model using the testing data conducted. The best models obtained were the linear kernel model with λ = 0.1 and γ =0.1, the degree 1 polynomial kernel model with, λ = 0.1 and γ = 0.01, the degree 2 polynomial kenel model with λ = 2 and γ = 0.1, the degree 3 polynomial kernel model with λ = 0.1 and γ = 0.01, and RBF kernel model with λ = 0.1 and γ = 0.001

To find out unbiased results, it is necessary to use some testing data outside of the training data. The evaluation using the testing data showed that the linear kernel and the degree 1 polynomial kernel resulted the best performance of Precision, Recall, and F-Measure which were 0.99, 0.99, and 98%, respectively. It was compared to RBF which resulted Precision, Recall, and F-Measure which were 0.95, 0.95, and 0.94, respectively. The worst model was showed by the degree 3 polynomial kernel which returned Precision, Recall, and F1-Measure of 0.85, 0.79, and 0.78, respectively.

Table 11. Comparison of Results on SVM Kernel

Method	Precision	Recall	F-Measure
Linear	**0.9901**	**0.99**	**0.98**
Degree 1 Polynomial	**0.9901**	**0.99**	**0.98**
Degree 2 Polynomial	0.903	0.89	0.8891
Degree 3 Polynomial	0.8521	0.79	0.7803
RBF	0.9545	0.95	0.9487

Conclusion

Evaluation of the best model of each kernel model (i.e., 5 models: linear, degree 1, 2, 3 poynomial, and RBF kernel) using the testing data of identification Spam e-mail showed that the linear kernel and the kernel of degree 1 polynomial produced the best Precision, Recall, and F-Measure performances of 0.99, 0.99, and 98%, respectively. Compared to RBF kernel which produced Precision, Recall, and F-Measure of 0.95, 0.95, and 0.94, respectively. The worst model was shown by a degree 3 polynomial kernel which produces Precision, Recall, and F1-Measure of 0.85, 0.79, and 0.78, respectively.

Parameter λ in the SVM classification serves as the degree of importance of the occurrence of misclassification. The greater the value of λ, the greater the chance of overfitting and the smaller the value of λ, the greater the occurrence of underfitting. In this study, overfitting generally begins to occur at a value of λ = 1. Parameter γ in the SVM classification functions as a determinant of learning rate, where the greater the value, the F-Measure value of the SVM classification results will decrease. Learning rate that is too small gives poor results. In this study the results were low because the parameter value was too small, occurring at γ = 0.0001.

REFERENCE

[1] N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. F. Othman, and W. I. S. Wan Din, “Youtube spam detection framework using naïve bayes and logistic regression,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.

[2] K. Agarwal and T. Kumar, “Email Spam Detection Using Integrated Approach of Naïve Bayes and Particle Swarm Optimization,” *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018*, no. November, pp. 685–690, 2019, doi: 10.1109/ICCONS.2018.8662957.

[3] Y. Vernanda, S. Hansun, and M. B. Kristanda, “Indonesian language email spam detection using n-gram and naïve bayes algorithm,” *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 2012–2019, 2020, doi: 10.11591/eei.v9i5.2444.

[4] N. Sutta, Z. Liu, and X. Zhang, “A study of machine learning algorithms on email spam classification,” *EPiC Series in Computing*, vol. 69, pp. 170–179, 2020, doi: 10.29007/qshd.

[5] Z. Ge, “A fusion algorithm model based on KNN-SVM to classify and recognize spam,” *Journal of Physics: Conference Series*, vol. 1982, no. 1, 2021, doi: 10.1088/1742-6596/1982/1/012069.

[6] N. Sun, G. Lin, J. Qiu, and P. Rimba, “Near real-time twitter spam detection with machine learning techniques,” *International Journal*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.2463

of Computers and Applications, vol. 44, no. 4, pp. 338–348, 2022, doi: 10.1080/1206212X.2020.1751387.

[7] S. N. D. Pratiwi and B. S. S. Ulama, "Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor," *Jurnal Sains dan Seni ITS*, vol. 5, no. 2, pp. 344–349, 2016.

[8] A. R. Chrismanto and Y. Lukito, "Identifikasi Komentar Spam Pada Instagram," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 8, no. 3, p. 219, 2017, doi: 10.24843/lkjiti.2017.v08.i03.p08.

[9] Y. K. Zamil, S. A. Ali, and M. A. Naser, "Spam image email filtering using K-NN and SVM," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, p. 245, 2019, doi: 10.11591/ijece.v9i1.pp245-254.

[10] G. A. Reddy and B. I. Reddy, "Classification of Spam Text using SVM," *Journal of University of Shanghai for Science and Technology*, vol. 23, no. 08, pp. 616–624, 2021, doi: 10.51201/jusst/21/08437.

[11] C. D. Manning, P. Raghavan, and S. Hinrich, *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2009. [Online]. Available: http://www.informationretrieval.org/

[12] R. J. Roiger, *Data Mining: A Tutorial-Based Primer*, 2nd ed. Boca Raton: CRC Press Taylor & Francis Group, 2017.

[13] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston: Pearson Addison-Wesley, 2006.

[14] C. C. Aggarwal, *Data Mining: The Textbook*. Switzerland: Springer, 2015. doi: 10.1007/978-3-319-14142-8.

[15] V. Jayaswal, "Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score," *Towards Data Science*, 2020. https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262 (accessed Sep. 08, 2021).

BY NC

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4077

EFFECTIVE BREAST CANCER DETECTION USING NOVEL DEEP LEARNING ALGORITHM

Irawadi Buyung[1]; Agus Qomaruddin Munir[2]; Putra Wanda[3]*

Dept. Electrical Engineering[1]
Dept. Information Technology[2]
Dept. Informatics[3]
Universitas Respati Yogyakarta
https://www.respati.ac.id
buyungirawadi@gmail.com, agusqmnr@respati.ac.id, putra.wanda@respati.ac.id

(*) Corresponding Author

Abstract— *Ultrasound is one of the most popular screening methods for finding breast cancer. However, the diagnosing process becomes complex due to the scarcity of experienced radiologists. We are motivated to use deep learning to address issues with medical image recognition because of its promising performance in various computer vision challenges. We present a rapid and precise breast cancer detection approach based on the Rapid-CNN. To undertake this experiment, we gather datasets related to breast cancer, pre-process them, train models, and assess the performance of the trained models. This model's bounding box detection of breast cancer has a training accuracy of up to 98.03% and a minimal loss of 0.78%. This model can detect the bounding box that is more than what it should be, so we applied NMS to eliminate the prediction of the bounding box that is less precise to increase accuracy.*

Keywords: *Breast Cancer Detection, Deep Learning, Rapid-CNN.*

Intisari— *Salah satu alat skrining yang paling umum untuk deteksi kanker payudara adalah ultrasound. Namun, kurangnya ahli radiologi yang mumpuni menyebabkan proses diagnosis menjadi tugas yang sulit. Pencapaian Deep Learning yang sangat bagus dalam berbagai masalah aplikasi komputer menginspirasi kami untuk menerapkan teknologi tersebut pada masalah pengenalan citra medis. Pada artikel ini, kami mengusulkan model deteksi Rapid-CNN untuk mendeteksi kanker payudara dengan cepat dan akurat. Eksperimen ini mengumpulkan dataset kanker payudara, melakukan pra-pemrosesan, melatih model, dan mengevaluasi kinerja model. Berdasarkan hasil percobaan diperoleh bahwa model ini dapat mendeteksi kanker payudara menggunakan bounding boxes dengan akurasi mencapai 98,03% dengan nilai loss hingga 0,78% pada proses training. Dalam model ini dimungkinkan untuk mendeteksi bounding box dengan lebih akurat, sehingga kami menerapkan NMS untuk menghilangkan prediksi bounding box yang kurang tepat untuk meningkatkan akurasi.*

Kata Kunci: *Deteksi, Kanker, Payudara,, Deep Learning, Rapid-CNN.*

INTRODUCTION

Breast cancer is one of the most common cancers in women for 15% of yearly fatalities [1]. Early identification of breast cancer can increase lifespan, reduce mortality risk, and improve quality of life [2]. Moreover, breast cancer screening can achieve early detection of ambiguous breast lesions. A systematic approach to breast screening is breast imaging diagnosis, which comprises breast MRI, mammography, and breast ultrasound [3].

Ultrasound is one of the most often utilized screening tools for finding breast cancer because of its pain-free, pleasant operation and outstanding real-time performance. Unfortunately, the ultrasonic instrument's high sensitivity makes it susceptible to the effects of various body tissues and their surroundings, resulting in many speckling noises and making diagnosis challenging. Moreover, the diagnostic effectiveness may suffer from a shortage of experienced radiologists, and 10–30% of diagnoses are missed [1] [2].

Conventional machine learning (ML) methods require large amounts of manual segmentation annotation data to train and test models for the classification or segmentation of ultrasound images. On the other hand, manual labeling is costly, time-consuming, and labor-intensive, significantly

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

raising the cost of system development [4]. Several papers have proposed methods for breast cancer detection. An article offered K-Nearest Neighbor (KNN) and Decision Tree to classify breast cancer. After selecting the Principal Component Analysis (PCA) technique, Wisconsin Diagnostic Breast Cancer (WDBC) dataset verified these two machine learning algorithms. Based on the findings of the experiments, the KNN classifier outperformed the decision tree classifier in the classification of breast cancer [5].

A report proposes ANN for breast cancer classification to increase classification accuracy. The Taguchi method initially counts the quantity of matched neurons in one of the hidden layers of the ANN. In accordance with the outcomes of the Taguchi approach, the model then goes through the training process, choosing the appropriate number of hidden neurons for the hidden layer. According to the findings of the experiments, this technique can give a classification accuracy for breast cancer that is 98.8% [6]. The accuracy, sensitivity, and specificity of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers were examined in this study. The Mammographic Image Analysis Society uses the model to conduct its training procedure (MIAS). According to the experiment's outcomes, the SVM classifier performed better than the KNN classifier [7].

Deep learning technology has allowed image recognition to discover target areas in medical images and classify detected target features. Deep learning's detection and classification technique are comparable to doctors' operating procedures to determine diagnoses based on ultrasound results. Thus, the approach becomes a new solution to the earlier issues [8]. The current paper proposed CNN and Uniform Experimental Design (UED) to classify breast cancer. UED uses regression analysis to optimize CNN parameters [9][10]. Another study explored a comparative classification of breast MRI tumors using human-engineered radionics, Transfer Learning from Deep Convolutional Neural Network (DCNN), and the Fusion Method [11]. DCNN shows excellent potential for classifying several very various fine-grained objects. Therefore, further study proposed a deep learning method based on Bilinear Convolutional Neural Networks (BCNNs) for the acceptable category of breast cancer histopathology images [12].

Therefore, we propose Rapid-CNN to detect breast cancer by utilizing breast images. We establish an effective model to solve the breast cancer detection issue. In breast cancer detection, we present several crucial contributions to this study, particularly in the categorization of breast cancer using learning methods as follows:

1. We introduce a novel technique for detecting breast cancer utilizing the Rapid-CNN algorithm to train the dataset to develop a viable model. We use a large dataset of breast ultrasound images to build our model.
2. We build a model to detect breast cancer as a solution to find the location of breast cancer more effectively than traditional machine learning techniques.
3. We test a model to achieve high accuracy in detecting breast cancer effectively using unseen image features. We modify several parameters to achieve the best accuracy value to create the best training model.

The following is a breakdown of the journal's organizational structure: Part II expands upon earlier findings. Part III discusses the issue description of the study. The experimental design, comprising a feature learning method, a dataset, and preprocessing, is described in Section IV. At the same time, the study's findings and in-depth analysis are presented in Section V. Section VI concludes by discussing the conclusion.

MATERIALS AND METHODS

Based on annotated and actual breast images, Rapid-CNN is used to detect breast cancer in this study. By inserting an RPN layer, this method addresses the sluggish R-CNN performance issue. The obtained feature mapping is entered into the RPN network to identify potential target regions, and then the mapping features and potential target regions are entered into the ROI network. Rapid CNN structural properties have a loss function comparable to multitasking [24]. This study uses a deep learning architecture to develop the detection model in multiple steps. Fig 1. depicts several steps to conduct the study.

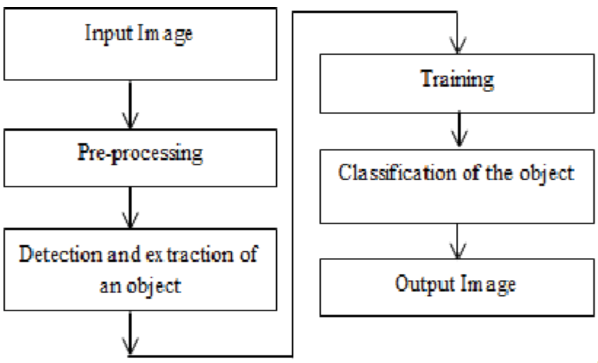


Fig.1: Experiment process Breast Cancer Detection using Rapid-CNN

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

In this study, classification assigns classes to images by finding similar features in images belonging to different categories and using them to identify and label images. In this part, we present a detailed explanation of the steps involved in detection using Rapid-CNN:

1. Input Image: The first step in detection using deep learning is to input an image. This can be a single image, a sequence of images, or a video feed.

2. Preprocessing: Before the image can be used for detection, it must be preprocessed to improve its quality and reduce noise. This can involve resizing the image, normalizing the pixel values, and performing other transformations such as color space conversions, image filtering, or data augmentation techniques.

3. Detection: The next step is to use a detection algorithm to identify objects within the image. This involves dividing the image into regions and analyzing each region to determine whether it contains an object. In this study, we utilize Rapid-CNN as a detection algorithm to construct a breast cancer detection model

4. Training: Before the detection algorithm can be used, it must be trained on a large dataset of images containing the objects it is meant to detect. During training, the Rapid-CNN algorithm learns to recognize the features associated with each object class, such as shape, color, and texture.

5. Classification: Once the objects have been detected within the breast image, the next step is to classify them. This typically involves using a separate classification algorithm, such as a deep neural network, to identify the specific class of each object. For example, if the thing is a breast cancer feature, the classification algorithm would label it as such.

6. Output: The output of the detection process is typically a bounding box that surrounds each detected object, along with the label for each object. This output can be used for various purposes, such as tracking objects in a video feed or identifying objects within a scene. Additionally, the accuracy and precision of the detection algorithm can be evaluated by comparing its output with ground-truth data.

In this study, Rapid-CNN detection involves a series of steps, from inputting an image to preprocessing it, detecting objects, training the algorithm, classifying the breast cancer class, and generating output. These steps can be iterated upon and refined to improve the accuracy and precision of the detection algorithm.

A. Proposed Method

This study presents a new CNN architecture, dubbed Rapid-CNN, in which area proposal generation occurs before the convolution layer. This step is rumored to decrease performance when working with large images. This NN resolves the performance issue by implementing the RPN layer and eliminating the current production of region proposals. Rapid-CNN suggested a solution to the performance problem. After feature extraction is performed, the model estimates RPN [25].

Rapid-first CNN's component is the region suggestion method, which gives bounding boxes or locations for likely picture objects. Commonly, a CNN is used to extract features from these objects during the second stage, which is feature creation. The third layer is a classification layer that predicts the object's class membership. The fourth layer is a regression layer that establishes the object's bounding box coordinates. This issue is addressed by the Rapid-CNN study, which generates regional suggestions using the RPN, reduces region proposal time, and permits the region proposal stage to share layers with future detection stages [26].

In the proposed network, an image's objective function is defined as:

$$L(\{P_i\}),\{t_i\} = \frac{1}{N_{cls}} \sum_i L_{cls} \ (P_i, P_i^*) + \lambda \sum_i P_i^* \ L_{reg}\,(t_{i,}t_i^*)$$

P_i is the probability that the anchor will be predicted as a target.

$$P_i^* = \begin{Bmatrix} 0 & negative\_lable \\ 1 & positive\_lable \end{Bmatrix}$$

$t_i = \{t_x, t_y, t_w, t_h\}$ is a vector representing the predicted bounding box's four-parameter coordinates. t_i^* is the coordinate vector of the positive anchor in the ground truth bounding box.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

Table 1. Mathematic notation of the Rapid-CNN

Notation	Description
P_i^*	The probability that the anchor is predicted to be the target.
t_i^*	The coordinate vector of the ground truth bounding box corresponds to the positive anchor.
t_i	The vector, represents the four-parameter coordinates of the predicted bounding box.
L_{cls}	The cross-entropy loss of binary classification (target & non-target)
L_{reg}	the regression loss
R	Smooth L1 function.
$L1$	Function.

In the Rapid-CNN process, each suggested region for each image requires feed-forward from CNN, even though the regions may overlap. Second, Rapid-CNN runs three separate models: a feature extraction model, a classification model, and a regression model. Rapid-CNN is a new solution to deal with both problems. The overall picture, as well as the extraction stage, have both been improved. Unlike R-CNN, it integrates all models into a single network, including feature extraction, classification, and detection. Second, the number of times a regional CNN has to be run has been reduced to one per image [24].

B. Main Idea

The main goal of this paper is to use the Rapid-CNN algorithm to develop a model based on breast ultrasound images to detect breast cancer. Rapid-CNN comprises two key processes for detecting and categorizing breast lesions [27]. Rapid-CNN starts with an ultrasonic image as input and outputs a rectangle box around the desired item. Second, the RPN, trained with ground truth data to give Regional Proposals, passes the convolution feature map through it. Therefore, the feature map is sent into the RPN, which generates a set of predictive-score areas [25] [28].

C. Dataset

In this experiment, breast ultrasound photos are collected from Kaggle.com to build our detection model. The data is then separated into training and testing datasets. Learning models are constructed using training datasets, whereas testing datasets are used to evaluate the performance or accuracy of the models. Here, the researcher divides the dataset into 80 percent for training and 20 percent for testing. Table 2 shows the details in terms of the distribution of the dataset used in the study, as follows:

Table 2. Distribution of the dataset

Dataset	**Sample**
Data Training (80%)	840
Data Testing (20%)	210
Total	**1.050**

D. Data Pre-Processing

Pre-processing is a stage to process high-resolution photographs; because high-resolution photo processing takes a long time, it must reduce the image size. Then, we convert the image to grayscale. After that, we perform noise removal to find and remove unwanted noise from the digital image. We sort each sample price by magnitude. The sample median in the window is the middlemost value, which can be a filter output. Grouping photos into segments is necessary for recording changes to image attributes. Image analysis is performed pixel by pixel after segmentation, and each pixel is labeled based on whether the gray level pixel is greater or less than the threshold value. As a result, segmented image analysis becomes easier [29].

E. Detection Method

To conduct our study, we used Rapid-CNN to detect breast cancer. We collected a dataset containing original and annotated breast images. This study builds a detection model of breast cancer with a training and testing process. Before the pre-processing stage, we split the dataset into two classes: Annotate and image, with 80% as the training dataset and 20% as the testing dataset.

During pre-processing, a detection model is simplified by resizing the image, removing undesired noise from the digital image, creating a bounding box for the position of the breast cancer target, and then labeling it based on the image. We extract vector values from features before putting them into the training and testing procedure.

After preprocessing, the training dataset is used to train the model. At the testing phase of the procedure, we utilize the testing dataset to evaluate model performance via data validation. In addition, a viable model was constructed and then validated using vector test data to assess the model's ability to detect breast cancer spots. Fig 1. depicts several steps to conduct the study.

RESULTS AND DISCUSSION

This work incorporates original and annotated breast photos in the training process. A low error rate proves that the model's performance is

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4077

satisfactory. In this procedure, we set epochs = 50 to train the model by modifying various hyperparameters for optimal performance. Using the hyperparameter setting, the loss result is 0.2743, while at epoch 49, the loss result is 0.0780. These data suggest that more epochs can result in a small score drop. Our proposed model is adequate for detecting breast cancer with robust detection outcomes based on the training data.

In the testing process, we analyze 88 annotated breast cancer data to test the model and get a red bounding box to mark the location of breast cancer. The testing process produces expected output and model output detection. Fig. 2 shows desired output and model output detection.

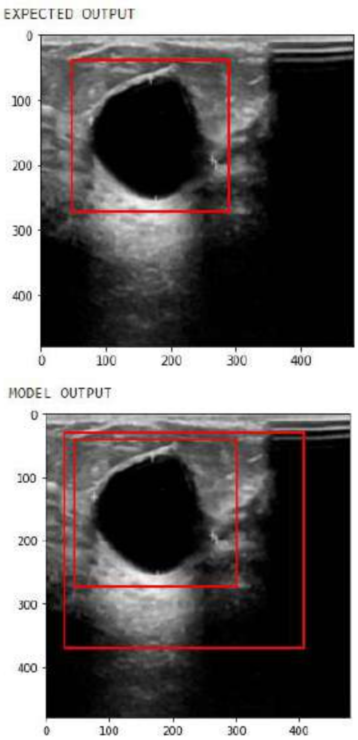


Fig. 2 Detection results

In Fig. 2, the expected outcome demonstrates accurate breast image recognition. In contrast, the model output implies that our model can detect the target. The testing results show that the intended output contains one bounding box, whereas the model output may contain many bounding boxes. In the subsequent testing phase, we determine actual and model detection based on the number of breast cancer spots denoted by a bounding box. The actual detection and model detection outcomes are graphically represented. Fig. 3 shows the actual detection results and the detection results of our model.

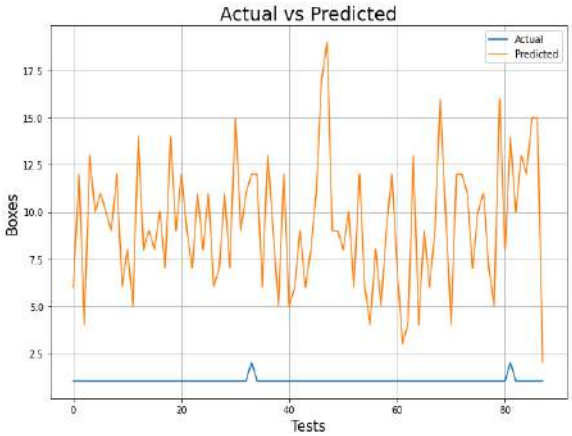


Fig. 3 Actual vs Predicted testing process

Fig. 3 shows the blue line that indicates the actual test result of accurate breast image detection, while the orange line indicates the effect of model detection. Based on the testing result, our model can detect cancer more accurately than the real detection image approach.

In the following process, we calculate the ratio value of actual and model detection. Fig. 4 shows the actual vs. Predicted ratio

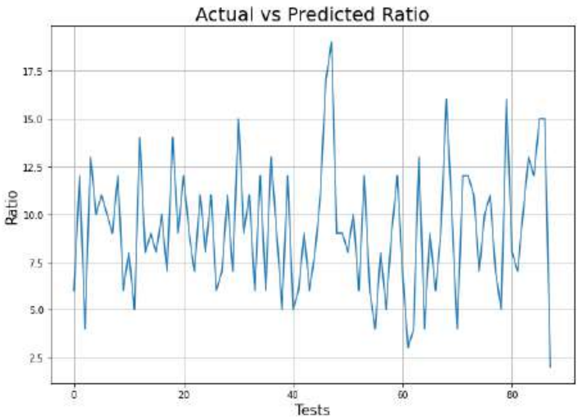


Fig. 4 Actual vs predicted ratio of the testing process

Fig. 4 displays a graph of the actual detection ratio with our model's detection. The ratio graph shows the number of detection results from our model divided by actual detection results. Our model can harvest better detection than the actual detection in breast image detection issues. In the previous dataset, there were redundant and overlapping bounding boxes. Therefore, we can conclude that our proposed model can detect more objects. Finally, after several stages, the Rapid-CNN can detect areas of breast cancer with several settings to obtain the best performance detection and accuracy value.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4077

Table 3. Training and testing results with different optimizer function with Rapid-CNN.

Hyperparameter	Training Accuracy	Testing Accuracy
Epoch = 50 Batch size = 64 *Learning rate = 0.09*	98.03%	96.23%

Hyperparameter	Training Loss	Testing Loss
Epoch = 50 Batch size = 64 *Learning rate = 0.09*	0.78%.	2,13%

CONCLUSION

Detecting breast ultrasound pictures is a key obstacle in detecting breast cancer. The existing literature suggests employing conventional machine-learning techniques to address this difficulty. Yet, manual feature engineering is costly and time-consuming. To improve the performance of the detection model through autonomous feature engineering, Rapid-CNN is recommended for the development of a breast detection model.

Based on the experimental result, the Rapid-CNN can produce higher accuracy and tiny loss during the training process. In the training process, the study set some hyperparameters Epoch = 50. The training process has a loss of 0.0780, region box loss of 0.0471, abjectness loss of 0.0005, and RPN box loss of 0.0014. In the testing process, the study produces more bounding boxes with a testing accuracy of up to 96.23% and a minimum loss of 2,13%. Therefore, the model can be a promising solution to deal with breast cancer detection challenges accurately and in real-time.

Future studies can adopt another method, such as GAN architecture, to enhance this model. GAN can produce high-quality images and provide an accurate medical image analysis solution. With the development of additional features, the usage of a neural network that is dynamic should produce greater precision.

REFERENCE

[1] Pang, Ting, et al. "Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification". Elsevier Inc. 2021.
[2] Zhuang, Zhemin, et al. "Breast Ultrasound Tumor Image Classification Using Image Decomposition and Fusion Based on Adaptive Multi-Model Spatial Feature Fusion". Elsevier Inc. 2021.
[3] Chiao, Jui Ying., et al. "Detection and Classification the Breast Tumors using Mask R-CNN on Sonograms". Medicine. 2019.
[4] Shu, Xin, et al. "Deep Neural Networks with Region-based Pooling Structures for Mammographic Image Classification". IEEE Transaction on Medical Imaging. 2020
[5] Rajaguru, Harikumar, et al. "Analysis of Decision Tree and K-Nearest Neighbor Algorithm in the Classification of Breast Cancer". Asian Pasific Journal of Cancer Prevention, Vol 20.2019.
[6] Rahman, Md Akizur., et al. "Artificial Neural Networks with Taguchi Method for Robust Classification Model to Improve Classification Accuracy of Breast Cancer". PeerJ Computer Science.2021.
[7] Osman, Yahia and Umar Alqasemi. "Breast Cancer Computer-Aided Detection System based on Simple Statistical Features and SVM Classification". International Journal of Advanced Computer Science and Applications (IJACSA).2020
[8] Lin, Bor-Shing, et al. "Using Deep Learning in Ultrasound Imaging of Bicipital Peritendinous Effusion to Grade Inflamation Severity". Journal of Biomedical and Health Informatics.2020.
[9] Mewada, Hiren K., et al. "Spectral-Spatial Features Integrated Convolutional Neural Network for Breast Cancer Classification". Sensors.2020.
[10] Lin, Cheng-Jian and Shiou-Yun Jeng. "Optimization of Deep Learning Network Parameters Using Uniform Experimental Design for Breast Cancer Histopathological Image Classification". Diagnostics.2020.
[11] Whitney, Heather M., et al. "Comparison of Breast MRI Tumor Classification Using Human-Engineered Radiomics, Transfer Learning From Deep Convolutional Neural Networks, and Fusion Methods". Proceedings of The IEEE.2019.
[12] Liu, Weihuang., et al. "Fine-Grained Breast Cancer Classification with Bilinear Convolutional Neural Networks". Frontiers in Genetics.2020.
[13] GK, Rajini, et al. "Statistical Detection of Breast Cancer by Mammogram Image". Asian Journal of Pharmaceutical and Clinical Research. 2017
[14] Nithya, et al. "Robust Minimal Spanning Tree Using Intuitionistic Fuzzy C-means Clustering Algorithm for Breast Cancer Detection". American Journal of Neural Networks and Applications.2019
[15] Nascimento, Carmina Dessana Lima, et al. "Breast Tumor Classification in Ultrasound Images using Support Vector Machine and

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

Neural Networks". Research on Biomedical Engineering.2016

[16] Chtihrakkannan,R , et al. "Breast Cancer Detection using Machine Learning". International Journal of Innovative Technology and Exploring Engineering (IJITEE).2019

[17] Li, Mochen, et al. "Machine Learning-Based Decision Support System for Early Detection of Breast Cancer". Indian Journal of Pharmaceutical Education and Research.2020

[18] Ibeni, Wan Nor LWH, et al. "Comparative Analysis on Bayesian Classification for Breast Cancer Problem". Bulletin of Electrical Engineering and Informatics.2019

[19] Hamzah, Mohammad Diqi, and Antomy David Ronaldo. "Effective Soil Type Classification Using Convolutional Neural Network". International Journal of Informatics and Computation (IJICOM) Vol. 3, No.1. 2021

[20] Paryadi, Catur, M.Diqi, and Sri Hasta Mulyani. "Implementation of CNN for Plant Leaf Classification". International Journal of Informatics and Computation (IJICOM) Vol. 2, No.2. 2020

[21] Abidin, Adnan, Hamzah, and Marselina Endah H. "Efficient Fruits Classification Using Convolutional Neural Network". International Journal of Informatics and Computation (IJICOM) Vol. 3, No.1. 2021

[22] Shamy, S., and J.Dheeba. "A Research on Detection and Classification of Breast Cancer using K-Means GMM and CNN Algorithms". International Journal of Engineering and Advanced Technology (IJEAT).2019.

[23] Hameed, Zabit, et al. "Breast Cancer Hispathology Image Classification Using an Ensemble of Deep Learning Models". Sensors.2020.

[24] Zhao, Yuntau, et al. "A Malware Detection Method of Code Texture Visualization Based on an Improved Faster RCNN Combining Transfer Learning". IEEE.2020.

[25] Rahmat, Taufik, et al. "Chest X-Ray Image Classification using Faster R-CNN". Malaysian Journal of Computing.2019.

[26] Singh, Sunil, et al. "Face Mask Detection using YOLOv3 and Faster R-CNN Models: Covid -19 Environment". Multimedia Tools and Applications. 2021

[27] Wei, Kaizhen, et al. "Faster Region Convolutional Neural Networks Applied to Ultrasonic Images for Breast Lesion Detection and Classification". IEEE. 2020

[28] Abd-Ellah, Mahmoud Khaled, et al. "Automatic Diagnosis of Common Carotid Artery Disease using Different Learning Techniques". Journal of Ambient Intelligence and Humanized Computing. 2021

[29] AG, Neela, et al. "A Breast Cancer Detection using Image Processing and Machine Learning Techniques". International Journal of Recent Technology and Engineering (IJRTE).2019

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3892

NON-CASH FOOD ASSISTANCE PROGRAM BENEFICIARIES BASED ON COPRAS AND CODAS

Dwi Marisa Midyanti[1*]; Syamsul Bahri[2]; Hafizhah Insani Midyanti[3]

Rekayasa Sistem Komputer[1,2]
Universitas Tanjungpura
www.untan.ac.id
dwi.marisa@siskom.untan.ac.id[1*]; syamsul.bahri@siskom.untan.ac.id[2]

Program Studi Musik[3]
Universitas Pendidikan Indonesia
fpsd.upi.edu
dicemidyanti@upi.edu[3]

(*) Corresponding Author

Abstract—*Determination of recipients of the Non-Cash Food Assistance Program (BPNT) is a matter that causes problems if it is not carried out in an objective, transparent, and targeted manner. Previous studies on BPNT were based on a specific method, which did not use a negative trend in the criteria. In this study, the Multi-Criteria Decision Making (MCDM) approach was used to recommend the recipients of the BPNT program in Tambelan Sampit Sub-district, Pontianak. MCDM is a technique of a Decision Support System that functions to support policymakers in making more objective decisions. Two MCDM models were used in this study, namely COPRAS and CODAS. This study aimed to determine the best model and measure the degree of similarity between the results obtained from different methods based on the Spearman rank correlation method Spearman's rank correlation method was used to determine the best model and measure the degree of similarity between the results obtained from different models. Spearman rank correlation shows that COPRAS and CODAS have a strong positive correlation of 0.89899. The combined COPRAS-CODAS ranking model produces a very strong positive correlation value of 0.9744 for both methods, so the model is used for recommendations for BPNT program recipients.*

Keywords: *BPNT, MCDM, COPRAS, CODAS, Spearman Rank Correlation*

Intisari— Penetapan penerima Program Bantuan Pangan Non Tunai (BPNT) merupakan hal yang menimbulkan permasalahan jika tidak dilaksanakan secara objektif, transparan, dan tepat sasaran. Studi sebelumnya tentang BPNT didasarkan pada metode tertentu, yang tidak menggunakan tren negatife pada kriteria. Pada penelitian ini, pendekatan *Multi-Criteria Decision Making* (MCDM) digunakan untuk rekomendasi penerima program BPNT pada Kelurahan Tambelan Sampit, Pontianak. MCDM merupakan bagian dari Sistem Pendukung Keputusan yang berfungsi untuk mendukung pemangku kebijakan dalam pengambilan keputusan yang lebih objektif. Dua model MCDM digunakan dalam penelitian ini, yaitu COPRAS dan CODAS. Tujuan penelitian ini adalah menentukan model terbaik dan mengukur tingkat kesamaan antara hasil yang diperoleh dari metode yang berbeda berdasarkan metode korelasi rank Spearman. Metode korelasi rank Spearman digunakan untuk menentukan model terbaik dan mengukur tingkat kesamaan antara hasil yang diperoleh dari model yang berbeda. Korelasi rank Spearman menunjukkan bahwa COPRAS dan CODAS memiliki korelasi positif kuat sebesar 0.89899. Model gabungan ranking COPRAS-CODAS menghasilkan nilai korelasi positif sangat kuat sebesar 0.9744 terhadap kedua metode, sehingga model tersebut digunakan untuk rekomendasi penerima program BPNT.

Kata Kunci*: BPNT, MCDM, COPRAS, CODAS, Spearman Rank Correlation.*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

INTRODUCTION

The Non-Cash Food Assistance Program, also known as BPNT, is a government program to help poor people who lack food so that they can get food for household needs [1]. BPNT is a food-specific social program distributed non-cash from the government to beneficiary families every month through an electronic money mechanism that is used only to buy food at food vendors called E-warong, in collaboration with channeling banks [2]. Candidates for BPNT recipients are people who are proposed by the head of the Neighbourhood to the Urban village, and then the community data is managed by the Social Service [3]. The success of the BPNT program is based on the achievement of the 6T indicators, namely Right on target, Right quantity, Right price, Right time, Right quality, and Right administration[1].

To achieve targeted indicators for prospective BPNT recipients to be more objective, researchers [3] used the Composite Performance Index (CPI) to determine the priority of BPNT recipients in Sampit Pontianak Village. In this study, 14 assessment criteria were used to obtain the priority ranking of BPNT recipients. However, negative trends were not used in the criteria.

CPI is one method to solve ranking problems. Other methods that can overcome the ranking are COmplex proportional assessment (COPRAS) and COMBinative Distance-based Assessment (CODAS). According to Zavadskas, et al. in [4] the COPRAS method is a Multi-Criteria Decision Making (MCDM) method that is used to evaluate alternatives where the ratio based on two measures, the sum of the performance of the favorable criteria and the sum of the unfavorable criteria. The COPRAS method is used for the assessment of ICT development in G7 countries [5], vulnerability mapping of sub-watershed erosion [6], green logistics and green supply chain management [7], Decision Making for New Student Admissions at MTsN Bangkalan [8], motorcycle selection Electricity [9], multi-criteria decision making for hybrid wind power plants [10], Supplier Selection at ABC Mining Companies in Indonesia [11], and Determination of Potential Zones for the Pasir Batu Mine [12].

Ghorabaee et al. developed CODAS in 2016 to address the ranking issue. This method uses Euclidean distance as the primary measure and Taksicab distance as a secondary measure, and this distance is calculated based on the negative-ideal point. The CODAS method's alternatives that have a long distance are more desirable [13]. CODAS method was used in research on the weighting parameters of the CODAS method [14], Material Selection [15], Agricultural Supplier Selection [16], studies on heat transfer optimization [17], and Evaluation of the Usefulness of Multi-Criteria Health Applications in Type 2 Diabetes Mellitus [18].

In contrast to research [3] , this study will use the criteria of benefits and costs as part of the algorithm contained in COPRAS and CODAS.

This study aims to determine the best model and measure the similarity between the results obtained from different methods based on the Spearman rank correlation method on priority BPNT recipients in Tambelan Sampit Village, Pontianak. The method used in this research is COPRAS and CODAS. In addition, we propose a combined model of COPRAS and CODAS means and quantify the results of Spearman's correlation of the model against the COPRAS and CODAS methods.

MATERIALS AND METHODS

Materials

At this stage, the value of the sub-criteria data in the form of text is converted into numerical data so that it can be calculated using the COPAS and CODAS methods. There are 11 Benefits Criteria and 3 cost criteria in this research.

Table 1. Criteria and Value

Criteria	Benefit / Cost	Value (%)
Building ownership	Benefit	18
The widest type of floor	Benefit	8
The widest type of wall	Benefit	8
The widest type of roof	Benefit	5
Source of drinking water	Benefit	5
How to get drinking water	Benefit	5
Main source of light	Benefit	7
Main fuel/energy for cooking	Benefit	5
Toilet facility users	Benefit	6
Toilet type	Benefit	5
Final disposal of feces	Benefit	4
Surface area	Cost	11
Number of Bedrooms	Cost	8
Installed PLN power	Cost	5

This study uses the categories of benefits and costs, as in table 1, which are not included in the study [3]. Another difference is the criteria for Land Area, Number of Bedrooms, and If it is PLN electricity, then the installed power uses numerical data filled in by the BPNT recipient.

Methods

This study will use the COPRAS and CODAS methods for recommendations for BPNT Program beneficiaries. The data used in this study comes from research [3]. Calculations were performed using COPRAS based on the method in the study [9]

CC BY NC
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3892

and CODAS based on the method in the study [16]. The research stage of this research can see in Figure 1.

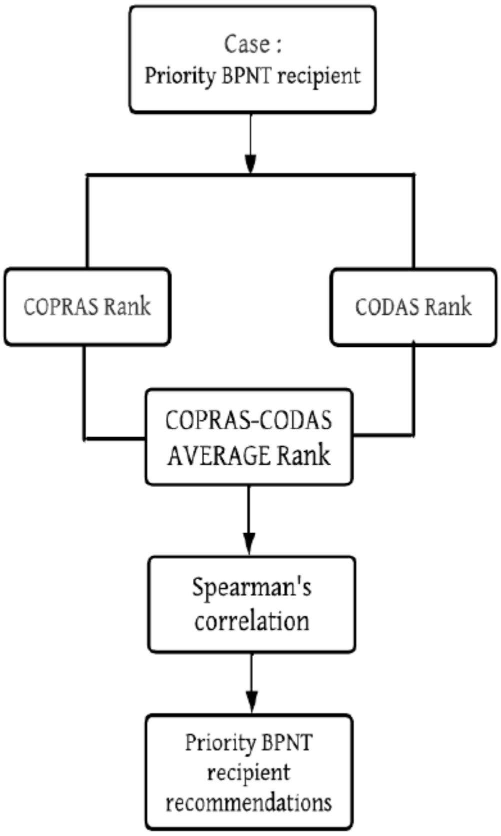


Figure 1. Research Stage

Based on figure 1, there are several stages of research conducted, including:

1. Anasila case. In this study, a case analysis was carried out related to the problems that occurred in the Tambelan Sampit Sub-district Pontianak to determine the priority of BPNT recipients so that they were right on target.
2. The calculation process uses the COPRAS method to obtain BPTN beneficiary rankings based on the quantitative utility value (Ui) from the order of the largest Ui to the smallest Ui value. Calculations also use the CODAS method to obtain the ranking of BPNT recipients based on the assessment score (Hi) value by sorting the highest value to the smallest Hi value.
3. Ranking using the average COPRAS-CODAS ranking. This ranking uses as a research [6], which carried out the final ranking based on the average ranking of the 4 MCDM methods.
4. Perform Spearman correlation calculations to measure the similarity between the ratings of the two MCDM models. If the average correlation value of COPRAS-CODAS to COPRAS or CODAS is better than the correlation between CODAS and COPRAS, then a recommendation for beneficiaries of the BPNT Program will be given based on the average rating of COPRAS-CODAS.
5. Provide priority recommendations for BPNT recipients based on the average COPRAS-CODAS ranking

RESULTS AND DISCUSSION

We used 30 data alternative from the study [3]. Data is processed using the steps of the COPRAS and CODAS methods.

Step 1-6 from COPRAS methods generate value S_{+i} , S_{-i} , and Q_i, as in Table 2. *S_(+i)* and *S_(-i)* are the sum of the maximum and a sum of minimum weighted values. Q_i is the relative significance of the alternatives.

Table 2. The results from steps 1 to 6 use the COPRAS method

Alternative	Si+	Si-	Qi
A1	0.024	0.008	0.031
A2	0.025	0.008	0.033
A3	0.021	0.009	0.028
A4	0.024	0.006	0.034
A5	0.022	0.008	0.03
A6	0.028	0.008	0.036
A7	0.03	0.006	0.04
A8	0.026	0.012	0.031
A9	0.024	0.008	0.032
A10	0.024	0.011	0.03
A11	0.024	0.011	0.029
A12	0.023	0.008	0.031
A13	0.019	0.006	0.028
A14	0.033	0.009	0.039
A15	0.024	0.005	0.037
A16	0.025	0.008	0.032
A17	0.031	0.005	0.044
A18	0.023	0.005	0.036
A19	0.022	0.008	0.03
A20	0.027	0.006	0.037
A21	0.023	0.007	0.032
A22	0.032	0.009	0.038
A23	0.024	0.009	0.031
A24	0.023	0.009	0.03
A25	0.031	0.009	0.038
A26	0.034	0.005	0.047
A27	0.022	0.008	0.03
A28	0.022	0.009	0.029
A29	0.028	0.011	0.034
A30	0.021	0.009	0.028

The final results of the COPRAS method can see in Table 3. Ui value is the quantitative utility for the alternative. Rank the solution from top value to bottom value of quantitative utility (Ui).

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

Table 3. Results of COPRAS method

Alternative	Ui Value	Rank
A1	65.56	17
A2	69.31	13
A3	58.30	30
A4	71.14	12
A5	63.83	21
A6	76.05	9
A7	84.22	3
A8	65.36	18
A9	67.28	15
A10	62.87	24
A11	61.91	26
A12	65.32	19
A13	59.06	29
A14	82.16	4
A15	78.28	7
A16	68.04	14
A17	93.11	2
A18	75.46	10
A19	63.13	22
A20	77.79	8
A21	66.81	16
A22	80.06	5
A23	64.88	20
A24	62.93	23
A25	80.00	6
A26	100.00	1
A27	62.51	25
A28	60.93	27
A29	71.46	11
A30	59.08	28

The ranking generated by COPRAS method is as follows:

A26>A17>A7>A14>A22>A25>A15>A20>A6>A18>A29>A4>A2>A16>A9>A21>A1>A8>A12>A23>A5>A19> A24>A10>A27>A11>A28>A30>A13>3.

Tabel 4 is final results of the CODAS method. Hi value is an assessment score for each possibility. Sort the possibilities by decreasing the assessment score value Hi. The option with the highest is the best.

Table 4. Results of CODAS method

Alternative	Hi Value	Rank
A1	-0.834	11
A2	-1.351	16
A3	-1.629	26
A4	-1.366	17
A5	-1.561	23
A6	2.320	7
A7	3.948	3
A8	-1.491	19
A9	-0.309	10
A10	-1.795	29
A11	-1.658	27
A12	-1.435	18
A13	-1.568	25
A14	2.761	6
A15	-1.063	14
A16	-0.917	12
A17	4.750	2
A18	-0.955	13
A19	-2.174	30
A20	0.507	9
A21	-1.549	22
A22	3.480	4
A23	-1.500	20
A24	-1.518	21
A25	3.302	5
A26	6.572	1
A27	-1.298	15
A28	-1.567	24
A29	1.641	8
A30	-1.742	28

The ranking generated by CODAS method is as follows:

A26>A17>A7>A22>A25>A14>A6>A9>A20>A9>A1>A16>A18>A15>A27>A2>A4>A12>A8>A23>A24>A21>A5>A28>A13>A3>A11>A30>A10>A19

With Spearman rank correlation, the correlation value between COPRAS and CODAS methods is 0.89899. It means that the ranking results using the COPRAS and CODAS methods have a strong positive correlation. Due to the strong correlation of methods, the ranking results of the two methods can average to obtain recommendations for prospective recipients of BPNT in Sampit Pontianak Village. The final results of the Average Rank can see in Table 5.

Table 5. Result of Average Rank

Alternative	COPRAS Rank	CODAS Rank	AVERAGE Rank
A1	17	11	14
A2	13	16	15
A3	30	26	28
A4	12	17	15
A5	21	23	22
A6	9	7	8
A7	3	3	3
A8	18	19	19
A9	15	10	13
A10	24	29	27
A11	26	27	27
A12	19	18	19
A13	29	25	27
A14	4	6	5
A15	7	14	11
A16	14	12	13
A17	2	2	2
A18	10	13	12
A19	22	30	26
A20	8	9	9
A21	16	22	19
A22	5	4	5
A23	20	20	20
A24	23	21	22
A25	6	5	6
A26	1	1	1
A27	25	15	20
A28	27	24	26
A29	11	8	10
A30	28	28	28

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2022
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.3892

The correlation between COPRAS, CODAS, and the average ranking using COPRAS and CODAS can see in Table 5.

Table 6. Spearman Rank Correlation obtained by different MCDM

	COPRAS	CODAS	COPRAS -CODAS AVERAGE
COPRAS	1	0.89899	0.9744
CODAS	0.89899	1	0.9744
	COPRAS	CODAS	COPRAS -CODAS AVERAGE
COPRAS - CODAS AVERAGE	0.9744	0.9744	1

From Table 6, it can see that the correlation value between COPRAS and the average rank of COPRAS and CODAS, with the correlation value between CODAS and the average rank of COPRAS and CODAS, is equal to 0.9744. This means that the nature of the correlation is very strong and positive, so it can be recommended for ranking.

Recommendations for recipients of BPNT funds from Sampit Pontianak Village based on the average ranking using the COPRAS and CODAS methods are as follows:

COPRAS-CODAS average rank:

A26>A17>A7>A22>A14>A25>A6>A20>A29>A15> A18>A9>A16>A1>A2>A4>A8>A12>A21>A23>A27 >A5>A24>A28>A19>A10>A11>A13>A3>A30.

CONCLUSION

In this study, recommendations for prospective recipients of BPNT in Sampit Pontianak Village were obtained from ranking using the COPRAS, CODAS, and average rankings from COPRAS and CODAS, to overcome the achievement of the right target indicator, which is one indicator of the success of the BPNT program. The ranking generated by the COPRAS method differs from the CODAS method ranking. Still, it has a Spearman correlation value of 0.89899, which means it has a strong positive correlation. The average ranking of the COPRAS and CODAS methods has a very strong positive trait with a value of 0.9744 for both methods, so the rankings generated from the average COPRAS and CODAS rankings can use to determine the prospective BPNT recipients in Sampit Pontianak Village.

REFERENCE

[1] E. Y. Yunus, “Implementasi Program Bantuan Pangan Non Tunai (Bpnt) Di Kecamatan Kanigaran Kota Probolinggo,” *Reformasi*, vol. 9, no. 2, pp. 138–152, 2019, doi: 10.33366/rfr.v9i2.1454.

[2] I. Fadlurrohim, S. A. Nulhaqim, and S. Sulastri, “Implementasi Program Bantuan Pangan Non Tunai (Studi Kasus Di Kota Cimahi),” *Share Soc. Work J.*, vol. 9, no. 2, pp. 122–129, 2020, doi: 10.24198/share.v9i2.20326.

[3] S. F. Azzahra, *Implementasi Metode Composite Performance Index (CPI) Dalam Penentuan Calon Penerima Bantuan Pangan Non Tunai (BPNT) (Studi Kasus : Kelurahan Tambelan Sampit)*, Skripsi, Pontianak : Universitas Tanjungpura, 2022.

[4] M. Keshavarz-Ghorabaee, M. Amiri, E. K. Zavadskas, Z. Turskis, and J. Antucheviciene, “Simultaneous Evaluation of Criteria and Alternatives (SECA) for Multi-Criteria Decision-Making,” *Informatica*, vol. 29, no. 2, pp. 265–280, 2018, doi: 10.15388/Informatica.2018.167.

[5] N. Ersoy, “Comparative Analysis of MCDM Methods for The Assessment of ICT Development in G7 Contries,” *KAÜİİBFD*, vol. 13, no. 25, pp. 55–73, 2022, doi: 10.1016/j.omega.2015.05.013.

[6] B. Ghosh and S. Mukhopadhyay, “Erosion Susceptibility Mapping of Sub-Watersheds for Management Prioritization using MCDM-based Ensemble Approach,” *Arab. J. Geosci.*, vol. 14, no. 36, pp. 1–18, 2021, doi: 10.1007/s12517-020-06297-4.

[7] N. A. Osintsev, “Multi-Criteria Decision-Making Methods in Green Logistics,” *World Transp. Transp.*, vol. 19, no. 5(96), pp. 231–240, 2021, doi: 10.30932/1992-3252-2021-19-5-13.

[8] Y. Kustiyahningsih, Husni, and I. Q. Aini, “Integration of FAHP and COPRAS Method for New Student Admission Decision Making,” in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 2020, no. November, pp. 1–6, doi: 10.1109/ICVEE50212.2020.9243260.

[9] A. Patel, S. Jha, R. Soni, and K. Fuse, “Notice of Removal: Comparative study of MCDM Techniques COPRAS and TOPSIS for selection of Electric Motorcycles,” in *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, 2020, pp. 54–59, doi: 10.1109/ICIEA49774.2020.9101932.

[10] H. S. Dhiman and D. Deb, “Fuzzy TOPSIS and fuzzy COPRAS based Multi-Criteria Decision

Making for Hybrid Wind Farms," *Energy*, vol. 202, pp. 1–10, 2020, doi: 10.1016/j.energy.2020.117755.

[11] R. P. Setyono and R. Sarno, "Comparative Method of MOORA and COPRAS Based on Weighting of the Best Worst Method in Supplier Selection at ABC Mining Companies in Indonesia," in *International Conference on Information and Communications Technology (ICOIACT)*, 2019, pp. 354–359, doi: 10.1109/ICOIACT46704.2019.8938520.

[12] A. S. Ajrina, R. Sarno, and R. V. H. Ginardi, "Comparison of moora and copras methods based on geographic information system for determining potential zone of pasir batu mining," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, 2019, pp. 360–365, doi: 10.1109/ICOIACT46704.2019.8938465.

[13] M. Keshavarz Ghorabaee, E. K. Zavadskas, Z. Turskis, and J. Antucheviciene, "A new combinative distance-based assessment (CODAS) method for multi-criteria decision-making," Econ. Comput. Econ. Cybern. Stud. Res., vol. 50, no. 3, pp. 25–44, 2016.

[14] M. Regaieg and H. M. Frikha, "Inferring Criteria Weight Parameters in CODAS method," *Int. J. Multicriteria Decis. Mak.*, vol. X, no. Y, pp. 1–19, 2021, doi: 10.1504/ijmcdm.2021.10044853.

[15] E. M. Abdelkader, A. Al-Sakkaf, and G. Alfalah, "Optimizing Material Selection using a Hybridized Multi-Attribute Decision Making Model," *WSEAS Trans. Syst. Control*, vol. 16, pp. 404–421, 2021, doi: 10.37394/23203.2021.16.36.

[16] L. Thi Diem My, C.-N. Wang, and N. Van Thanh, "Fuzzy MCDM for Improving the Performance of Agricultural Supply Chain," *Comput. Mater. Contin.*, vol. 73, no. 2, pp. 4003–4015, 2022, doi: 10.32604/cmc.2022.030209.

[17] V. Sivalingam, P. G. Kumar, R. Prabakaran, J. Sun, R. Velraj, and S. C. Kim, "An Automotive Radiator With Multi-Walled Carbon-based Nanofluids: A study on Heat Transfer Optimization using MCDM Techniques," *Case Stud. Therm. Eng.*, vol. 29, pp. 1–17, 2022, doi: 10.1016/j.csite.2021.101724.

[18] K. Gupta *et al.*, "Multi-Criteria Usability Evaluation of mHealth Applications on Type 2 Diabetes Mellitus Using Two Hybrid MCDM Models: CODAS-FAHP and MOORA-FAHP," *Appl. Sci.*, vol. 12, no. 4156, pp. 1–26, 2022, doi: 10.3390/app12094156.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4060

ERFORMANCE ANALYSIS OF ALEXNET CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURE WITH IMAGE OBJECTS OF RICE PLANT LEAVES

Adi Fajaryanto Cobantoro[1*]; Fauzan Masykur[2]; Kelik Sussolaikah[3]

Information Engineering[1,2]
Universitas Muhammadiyah Ponorogo
umpo.ac.id
adifajaryanto@umpo.ac.id[1*], fauzan@umpo.ac.id[2]

Information Engineering[3]
Universitas PGRI Madiun
unipma.ac.id
kelik@unipma.ac.id

(*) Corresponding Author

Abstract— *Rice is a staple food consumed by Indonesian people, even 75% of the world's population consumes rice and it is mostly found in Asia. Rice derived from pounded rice is a staple food so it can be consumed. In the process of planting rice, pests and diseases are not spared so that it can affect crop yields. Pest and disease attacks need fast, accurate and precise handling so that crop failures. In this paper, we will discuss the classification of leaf diseases of rice plants using the Convolutional Neural Network (CNNN) algorithm, especially the Alexnet architecture. There are 4 types of disease, namely Brown spot, Leafblast, Hispa and Healthy. Models built based on the Alexnet architecture may have differences in the level of accuracy and loss compared to other architectures due to the different stages in the sequential model formation. The dataset used is public data from Kaggle consisting of 4 classes with a total of 1,600 images. In each class the dataset is divided for training, testing and validation datasets with a ratio of 70:20:10. As for tools in the process of training datasets using Google Colab from Google. After going through the stages of the research, the research results obtained are accuracy worth 99,22%, mean average precision worth 0,24 and loss worth 0,05.*

Keywords: *Convolutional Neural Network, Alexnet, Rice leave image, CNN Classification.*

Intisari— *Beras merupakan bahan makanan pokok yang dikonsumsi masyarakat Indonesia bahkan 75 % penduduk dunia mengkomsumsi beras dan sebagai besar terdapat di Asia. Beras yang berasal dari padi yang ditumbuk menjadi bahan makan pokok supaya bisa dikonsumsi. Pada proses penanaman padi tidak luput dari serangan hama dan penyakit sehingga bisa mempengaruhi terhadap hasil panen. Serangan hama dan penyakit perlu penanganan yang cepat, akurat dan tepat supaya kegagalan panen. Pada penelitian kali ini akan membahas klasifikasi jenis penyakit daun tanaman padi dengan algoritma Convolusional Neural Network (CNNN) khususnya arsitektur Alexnet. Jenis penyakit pada daun tanaman padi ada 4 kelas yakni Brownspot, Leafblast, Hispa dan Healthy. Model yang dibangun berdasarkan arsitektur Alexnet dimungkinkan terjadinya perbedaan pada tingkat akurasi dan loss daripada arsitektur yang lain dikarenakan adanya perbedaan tahapan pada pembentukan sequential model. Dataset yang digunakan merupakan data publik dari kaggle terdiri dari 4 kelas dengan total citra sebanyak 1.600 citra. Pada setiap kelas dataset dibagi untuk dataset training, testing dan validasi dengan perbandingan 70:20:10. Adapun sebagai tools pada proses men-training dataset menggunakan google colab dari google. Setelah melewati tahapan-tahapan penelitian didapatkan hasil penelitian yakni akurasi senilai 99,22%, presisi rata-rata senilai 0,24 dan loss senilai 0,05.*

Kata Kunci: *Convolutional Neural Network, Alexnet, Citra Daun Padi, Klasifikasi CNN.*

INTRODUCTION

Indonesia is an agricultural country with the majority of the population working as farmers [1] either as landowners or as farm laborers. So far, the main crops grown by farmers are rice and corn [2] because they are staple foods. In addition, the staple foods grown are sweet potatoes, cassava, vegetables and tubers. Various crops grown by farmers show that Indonesia is dependent on the agricultural

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

sector and plays an important role in the national economy. In the rice planting process, pests and diseases are not spared which can interfere with rice growth and can even cause crop failure. Regular monitoring must be carried out by farmers to find out the progress of growth from the planting season to entering the harvest period. There are various types of pests that attack rice plants including insects, leafhoppers, mammals, and invertebrate animals [3].

Detection of rice plant diseases requires a fast, accurate and precise time because to prevent rice damage which results in a decrease in rice yields. The speed of rice disease detection using the deep learning convolutional neural network (CNN) method. CNN can achieve a high degree of accuracy by using leaf images with controlled lighting and background conditions [4]. Even though the presence of AI helps in the rapid detection of rice diseases, the presence of experts is still needed as a comparison in the process of determining the results of the disease detection process.

Pest and disease attacks on rice plants can be seen, one of which can be identified through the color of the leaves because the leaves will have certain characteristics according to the type of pest and disease. According to Ramesh [5] in 2020 there are 4 types of leaves based on the disease (Brownspot, Leafblast, Hispa and Healthy). These types of infected leaves can be identified and classified using deep learning algorithms to look for certain characteristics to distinguish one from another. The algorithm used is a Convolutional Neural Network (CNN). This paper specifically uses the CNN Alexnet architecture because Alexnet was the winner of the Imagenet Competition in 2012 [6]. Alexnet itself has 8 convolutional layers and 25 layers and 60 million parameters [7].

CNN is a development method for Multi-Layer Perception (MLP) which has few parameters because it does not require pre-processing, segmentation and feature extraction values [8]–[12]. The Alexnet architecture itself has 5 conv layers, 3 pooling layers, 2 dropout layers, and 3 fully connected layers. CNN can classify an image based on a predetermined class. In addition, CNN has high accuracy because it has the number of feature extractions from the results of the convolution process and the number of neurons used. The iteration weight also influences the image classification results. The CNN architecture develops with various methods and certain features, these architectures include GoogleNet, Resnet, VGG-16, VGG-19, SqueezeNet, MobileNet [13], [14].

Research conducted by Sibiya et al in 2019, they conducted research using 3 classes of classification of diseases in corn plants with 3 categories and using the CNN 50 hidden layer architecture consisting of a convolution layer with a kernel filter having a median of 24, ReLU and a pooling layer. Research conducted using 100 images per class with a ratio of 70% for training and 30% for testing with research results showing an accuracy rate of 92.85% [15].

Another research is research on corn leaf diseases with 8 classes of classification of leaf types.

This study uses the CNN architecture, namely googleNet or Inception-V1 with a total of 3,672 images used, 80% for training data and 20% for test data [16]. The results of the research conducted by Zhang et al resulted in an accuracy of 98.9%.

The final goal of this paper is to classify the image of rice plant leaves with 4 classes. Image classification uses the Alexnet architecture and the training process uses google colab. Analysis of the classification training process using a confution matrix consisting of accuracy, precision, recall, F1-score and loss.

MATERIALS AND METHODS

The research stages applied are using the CNN architecture, especially the Alexnet architecture. The Alexnet architecture chart can be seen in Figure 1. The Alexnet architecture is used to obtain feature extraction by training datasets which are first grouped into training and testing datasets as well as validation. The comparison of the three dataset groups is 70:20:10.

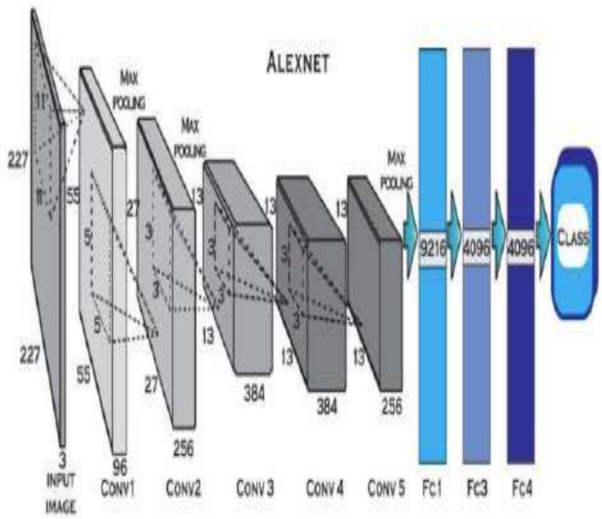


Figure 1 Architecture of Alexnet [17].

The alexnet architecture has 5 stages of layer convolution as shown in Figure 1. This architecture applies max pooling as an approach in determining feature extraction. After passing through the

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4060

convolution layer, the next stage is fully connected for 3 stages before heading to the classification process according to the class on the input. The kernel used in the alexnet architecture consists of 11x11, 5x5, 3x3 contained in the covolution layer.

The dataset used is images of rice plant leaves obtained from public data on kaggle.com as many as 2,150 images of rice plant leaves. The image of rice plant leaves is divided into 4 classes namely Brownspot, Leafblast, Hispa and Healthy.

Table 1 Image of Rice Plant Leaves

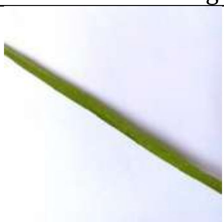
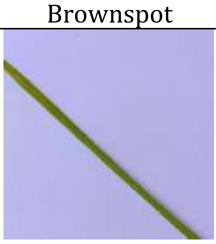
 Brownspot	 Leafblast
 Hispa	 Healthy

Table 1 explain the image of a rice plant leaf. The 4 dataset classes are based on the diseases suffered on rice leaves including healthy leaves. Differences in leaf patterns in each class indicate differences in pixel colors in each leaf so that feature extraction on each leaf can be identified.

Research stage starts with collecting datasets in the form of images of rice plant leaves then the second stage is dividing the image dataset into 4 classes. Data acquisition was carried out by the publisher on Kaggle by listing classes according to the characteristics of the rice leaves. Class division consists of brownspot, hispa, leafblast and healthy. Third stage is training process on the Alexnet architecture with google colab. The dataset training process for building models is carried out using Google Colab because it is open source, but if additional features are needed you can use Google Colab Pro or Google Colab Pro+.

Dataset training process involves various epoch combinations that function to determine the best model. Evaluation of the model uses the dataset testing then as a determination of the best model using the validation dataset. The next stage is the stage of calculating the training results using the convolution matrix to calculate accuracy, recall, precision and F1-Score. The stages of conducting the research are presented in Figure 2.

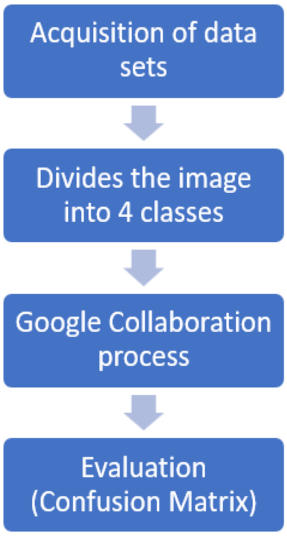


Figure 2. Research Stages

Evaluation of the performance of the CNN model built using a confusion matrix involving TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative). Several parameters are calculated using the confusion matrix, namely accuracy, precision, recall, F1-Score. The existence of a dataset in the CNN learning process greatly determines the results of the classification, therefore the success of the classification of the model built depends on the comparison between the actual value and the predicted value. There are several parameters whose values need to be known to determine model performance. The size of the confusion matrix depends on the number of classes used because the comparison between actual and predicted must be balanced.

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN} \text{.........................} (1)$$

$$Presicion = \frac{TP}{TP+FP} \text{...} (2)$$

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

$$Specification = \frac{TN}{TN+FP} \text{.................................} (4)$$

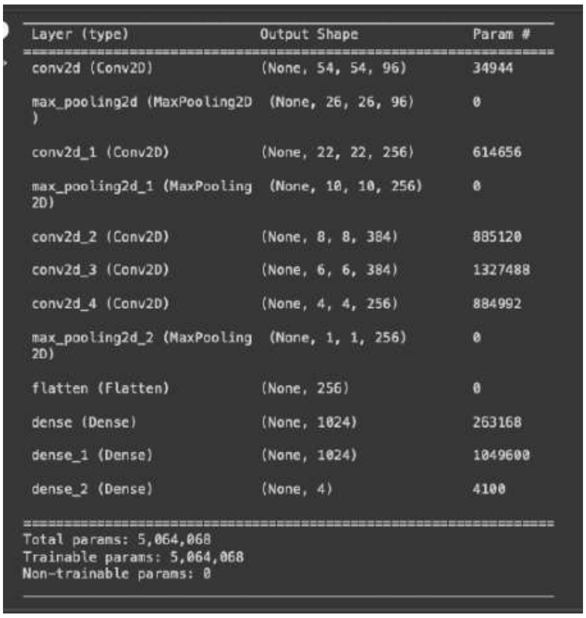
$$F_1 = 2\,x\,\frac{Precision\ x\ Recall}{Presicion+Recall} \text{..............................} (5)$$

In the CNN algorithm, to overcome the lack of datasets in the training data or dataset imbalances that are not evenly distributed with each other, the analysis uses the traditional transformation data augmentation analysis method which is commonly applied to the CNN algorithm. The performance or performance of CNN is also influenced by the number of datasets, meaning that the more datasets used, the better the classification results.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

RESULTS AND DISCUSSION

The Alexnet architecture used in this study consists of several stages starting from the convolution process, max pooling to flaten. Complete Alexnet architecture can be seen in Figure 4. The CNN network is used to extract features from the input image because the CNN network has a strong ability to search for image features. The input image dataset will be subjected to a convolution operation on each layer to look for the most important features of the image. The earliest layer has low spatial information or no image features can be found, while the high layer will have more spatial information so that features will be found in the image. Finally, the pooling layer is used as the most important feature and is defined as an image feature. The results of image feature extraction with CNN are used as input in making a graph of classification result (Figure 3).



```
Layer (type)                  Output Shape          Param #
=================================================================
conv2d (Conv2D)               (None, 54, 54, 96)    34944

max_pooling2d (MaxPooling2D   (None, 26, 26, 96)    0
)

conv2d_1 (Conv2D)             (None, 22, 22, 256)   614656

max_pooling2d_1 (MaxPooling   (None, 10, 10, 256)   0
2D)

conv2d_2 (Conv2D)             (None, 8, 8, 384)     885120

conv2d_3 (Conv2D)             (None, 6, 6, 384)     1327488

conv2d_4 (Conv2D)             (None, 4, 4, 256)     884992

max_pooling2d_2 (MaxPooling   (None, 1, 1, 256)     0
2D)

flatten (Flatten)             (None, 256)           0

dense (Dense)                 (None, 1024)          263168

dense_1 (Dense)               (None, 1024)          1049600

dense_2 (Dense)               (None, 4)             4100

=================================================================
Total params: 5,064,068
Trainable params: 5,064,068
Non-trainable params: 0
```

Figure 3. Architecture of Alexnet

Some of the parameters used in the Alexnet training model can be seen in Table 2. The tools used are Google Colab with GPU runtime type. It takes 2 hours for the training process to produce a classification model for the image of rice plant leaves.

Table 2. Parameter Training Model Alexnet

Parameter	Value
Epoch	50
Step Per Epoch	8
Val_Steps	8
Verbose	1

The accuracy of the classification model using Alexnet on leaf images of rice plants can be seen after a 2-hour dataset training process with 100 epochs and 8 steps per epoch resulting in an accuracy of 99.2%. While the loss in the training process is 0.05. Figure 4 shows a graph of model accuracy and loss. From the graph shown, it is known that loss decreases throughout the epoch and the accuracy of the longer the training process shows an increasing trend up to close to 1 or 99.2% to be exact.

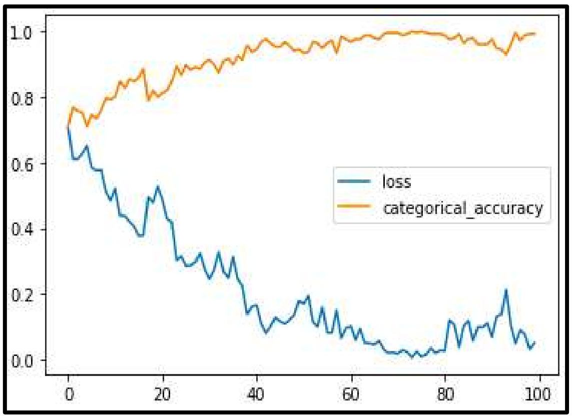


Figure 4. Graph of Accuracy and Loss

The performance measurement of the alexnet model is calculated based on the confusion matrix which shows the comparison between *true labels* and *predicted labels.* The confusion matrix from the results of the Alexnet training model is presented in Figure 6. The measurement results of accuracy, precision, recall, f1-score are known from the confusion matrix and the results are presented in Table 3.

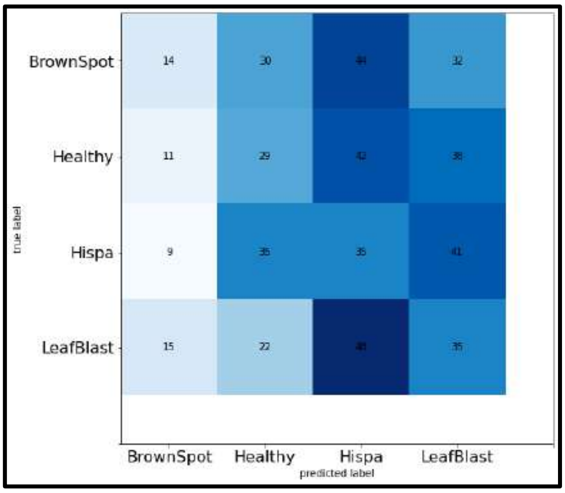


Figure 6. Matrix of Alexnet Performance Confusion

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4060

The Score of accuracy, precision, recall and F1-Score are calculated based on the confusion matrix and using equations (1), (2), (3), (4), (5). The score of several calculated parameters are used as evaluation material for the performance of the mode being built.

Table 3. Evaluation Material Table

No Label	Presisi	recall	f1-score	support
0	0,29	0,12	0,17	120
1	0,25	0,24	0,25	120
2	0,21	0,29	0,24	120
3	0,24	0,29	0,26	120

While the accuracy score of the model in the classification of rice leaf imagery shows a value of 99.22% and loss in the Alexnet model is 0.05.

CONCLUSION

The Alexnet architecture used in this study consists of several stages starting from the convolution process, max pooling to flatten. Complete Alexnet architecture can be seen in Figure 4. The CNN network is used to extract features from the input image because the CNN network has a strong ability to search for image features. The input image dataset will be subjected to a convolution operation on each layer to look for the most important features of the image. The earliest layer has low spatial information, or no image features can be found, while the highest layer will have more spatial information so that features will be found in the image. Finally, the pooling layer is used as the most important feature and is defined as an image feature. The results of image feature extraction with CNN are used as input in making a graph of classification results.

REFERENCE

[1] Taufan, Rafael, and Sebastiana, "Analisis Kelembagaan Hulu Industri Tape di Desa Sumber Tengah Kecamatan Binakal Kabupaten Bondowoso," *Jurnal Ekonomi Ekuilibrium*, vol. 2, no. 2, pp. 40–51, 2018.

[2] A. Amzeri, "Tinjauan Perkembangan Pertanian Jagung Di Madura Dan Alternatif Pengolahan Menjadi Biomaterial," *Rekayasa*, vol. 11, no. 1, p. 74, 2018, doi: 10.21107/rekayasa.v11i1.4127.

[3] M. Sayuthi, A. Hanan, P. Satriyo, and Muklis, "Distribusi hama tanaman padi (Oryza sativa L .) pada fase vegetatif dan generatif di Provinsi Aceh," *Jurnal Agroecotenia*, vol. 3, no. 1, pp. 1–10, 2020.

[4] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Tomato leaf segmentation algorithms for mobile phone applications using deep learning," *Comput Electron Agric*, vol. 178, Nov. 2020, doi: 10.1016/j.compag.2020.105788.

[5] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 249–260, 2020, doi: 10.1016/j.inpa.2019.09.002.

[6] W. Setiawan, "Perbandingan Arsitektur Convolutional Neural Network Untuk Klasifikasi Fundus," *Jurnal Simantec*, vol. 7, no. 2, pp. 48–53, 2020, doi: 10.21107/simantec.v7i2.6551.

[7] Alex Krizhevsky, "Handbook of approximation algorithms and metaheuristics," *Handbook of Approximation Algorithms and Metaheuristics*, pp. 1–1432, 2007, doi: 10.1201/9781420010749.

[8] D. Irfansyah, M. Mustikasari, and A. Suroso, "Arsitektur Convolutional Neural Network (CNN) Alexnet Untuk Klasifikasi Hama Pada Citra Daun Tanaman Kopi," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 2, pp. 87–92, 2021.

[9] N. Purnama, "HERBAL PLANT DETECTION BASED ON LEAVES IMAGE USING CONVOLUTIONAL NEURAL NETWORK WITH MOBILE NET ARCHITECTURE," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 6, no. 1, 2020, doi: 10.33480/jitk.v6i1.1400.

[10] M. Defriani and I. J. Jaelani, "Recognition of Regional Traditional House in Indonesia Using Convolutional Neural Network (CNN) Method," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 4, no. 2, pp. 104–115, Jul. 2022, doi: 10.47709/cnahpc.v4i2.1562.

[11] A. F. Cobantoro, F. Masykur, and M. Rizqi Rosyadi, "Implementation of Bot Telegram as Broadcasting Media Classification Results of Convolutional Neural Network (CNN) Images of Rice Plant Leaves," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 1, pp. 1–9, Jan. 2023, doi: 10.47709/cnahpc.v5i1.1976.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

[12] F. Masykur, K. Adi, and O. D. Nurhayati, "Classification of Paddy Leaf Disease Using MobileNet Model," in *2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED)*, 2022, pp. 1–4. doi: 10.1109/ICCED56140.2022.10010535.

[13] K. A. and O. D. N. F. Masykur, "Classification of Paddy Leaf Disease Using MobileNet Model," in *2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED)*, 2022. doi: 10.1109/ICCED56140.2022.10010535.

[14] N. A. Prasetyo, "MOBILE WEB APPLICATION PURWOKERTO TRADITIONAL FOOD GAME CLASIFICATION USING MOBILENET V2," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 6, no. 1, pp. 33–40, 2020, doi: 10.33480/jitk.v6i1.1418.

[15] M. Sibiya and M. Sumbwanyambe, "A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks," *AgriEngineering*, vol. 1, no. 1, pp. 119–131, Mar. 2019, doi: 10.3390/agriengineering1010000.

[16] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol. 6, pp. 30370–30377, 2018, doi: 10.1109/ACCESS.2018.2844405.

[17] I. A. Sabilla, "Arsitektur Convolutional Neural Network (Cnn) Untuk Klasifikasi Jenis Dan Kesegaran Buah Pada Neraca Buah," *Tesis*, no. 201510370311144, pp. 1–119, 2020.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4112

IMPLEMENTATION OF THE SIMPLE MULTI-ATTRIBUTE RATING TECHNIQUE METHOD IN DSS SELECTION OF EXTRACURRICULAR ACTIVITIES

Suratun[1]; Novita Br Ginting [2*];Fitria Rachmawati [4]; Wisnu Indra Giri[5].

[1]Electrical Engineering Study Program, [2, 4]Information System Study Program, [5]Informatic Egineering Study Program
Ibn Khaldun Bogor University
https://www.uika-bogor.ac.id
[1]suratun@uika-bogor.ac.id, [*2)] novitawahab@uika-bogor.ac.id, , [4] fitria@uika-bogor.ac.id,
[5]wisnuindragiri673@gmail.com

Frieyadie [3]
[3]Infromation Technology
Universiti Kuala Lumpur Malaysia
https://www.unikl.edu.my
[3]frieyadie@s.unikl.edu.my

(*) Corresponding Author

Abstract—Extracurricular activities at SMAN 1 Cibungbulang are a means of exploring student competencies. Currently, students can choose extracurriculars without considering their interests and talents. It creates problems for extracurricular coaches because it is challenging to explore student competencies. In this study, the Simple Multi-Attribute Rating Technique (SMART) method was implemented and implemented in a decision support system to overcome this problem. This study used four alternatives: Basketball, Volly, Aikido, and Futsal. At the same time, the criteria are interests, talents, schedules, and achievements. The type of criteria used is benefit criteria. The decision support system is designed using object-oriented design and built on a web-based basis. The decision support system is access by every user registered on the system.

Keywords*:* Extracurricular*,* SMART Method*,* Decision Support System

Intisari— Kegiatan extrakurikuler di SMAN 1 Cibungbulang menjadi sarana dalam mengexplore kompetensi siswa. Saat ini siswa dapat memilih extrakurikuler tanpa mempertimbangkan minat dan bakatnya. Hal ini menimbulkan masalah bagi Pembina ektrakurikuler karena sulit mengeksplore kompetensi siswa. Untuk mengatasi masalah tersebut, maka dalam penelitian ini di implementasikan metode *Simple Multy Atribute Rating Technique* (SMART) dan di implementasikan dalam system pendukung keputusan. Pada penelitian ini menggunakan empat alternatif yaitu Basket, Volly, Aikido, dan Futsal. Sedangkan kriteria adalah minat, bakat, jadwal, dan prestasi. Jenis kriteria yang digunakan adalah benefit criteria. System penunjang keputusan dirancang menggunakan object oriented design, dan dibangun berbasis web. system penunjang keputusan dapat diakses oleh setiap user yang telah di daftrkan pada system.

Kata Kunci*:* Ekstrakurikuler, Metode SMART, Sistem Penunjang Keputusan*.*

INTRODUCTION

Developing students' potential according to students' talents, interests, creativity, and passion is one of the goals of national education. Developing students' potential requires systematic and continuous student development from the school. Student development through extracurricular and co-curricular activities.

"Student development materials consist of 1) faith and devotion to God Almighty. 2) Sublime ethics or noble morals. 3) Superior personality, national insight, and defending the country. 4) Academic, artistic, and sporting achievements according to talents and interests. 5) Democracy, human rights, and 6) Political education, environment, social sensitivity, and tolerance in a plural society. 7) Creativity, skills, and entrepreneurship. 8) Physical

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

quality, health, and nutrition based on diversified sources of nutrition. 9) Literature and culture. 10) Information and Communication Technology. Communication in English"[1].

SMA Negeri 1 Cibungbulang also did not escape student development activities. There are fourteen (14) extracurricular activities carried out, namely: 1) sports (badminton and volleyball), 2) sports (basketball and futsal), 3) sports (karate, silat, taekwondo), 4) women's scouts, 5) men's scouts, 6) rohis akhwat, 7) rohis Ikhwan, 8) PMR, 9) English Club, 10) KIR, 11) C. Move@Art, 12) Traditional Arts, 13) Band and 14) Paskibra[2].

Each student is free to choose the extracurricular they are interested in. Still, many students do not know what extracurricular suits their interests and talents, so many jump on the bandwagon. The extracurricular organization's performance makes it difficult for the coach to explore the potential of each student. Extracurricular activities are student activities outside of teaching and learning; extracurricular activities influence students' interests, talents, and achievements. Through extracurricular activities in the school environment, in the future, students can gain knowledge and experience that can be useful for them in everyday life. A decision support system is needed in recommending extracurricular activities that match the interests and talents of students.

Several literature reviews have been carried out that various methods can carry out extracurricular selection activities, one of which is using SMART (Simple Multi-Attribute Rating Technique which can be applied to decision support systems. Based on the results of research by Alif Catur et al., who tested the performance of the SMART method and SAW, the results showed that the SMART way could be used in providing choice recommendations even though the lowest value of a criterion reaches 0[3].

The SMART (Simple Multi-Attribute Rating Technique) method is a multi-attribute decision-making method. This multi-attribute decision-making technique supports decision-makers in choosing multiple alternatives[4]. Several studies have successfully applied the SMART method in solving the problem of selecting an option with many attributes through a decision support system for extracurricular selection carried out by students[5].

The application of SMART method is also applied in the SPK performance assessment, which includes 8 (eight) criteria, namely 1) social networking, 2) leadership, 3) communication, 4) integrity, 5) emotional control, 6) administrative management, 7) creativity and 8) independence[6]. The SMART method is also used in determining the provision of annual bonuses to employees, using 13 criteria, namely, work discipline, responsibility, initiative, creativity, ability to consider and make decisions, adaptability, attitude towards superiors, impressions of behavior, motor skills, physical condition, the amount of work produced and the quality of work results produced[7]. In addition, the SMART Method is also used in facilitating the determination of recipients of the Quran hafiz scholarship, based on four criteria, namely memorization of the Quran, high school academic scores (Senior High School), parents' work, and the number of dependents of parents[8]. In assisting fishermen with Smart Fishing tools to realize the Champion Fishermen and Champion Fish Warehouse programs in West Java, they also apply the SMART Method with the following criteria: 1. Photocopy of KTP, 2. Photocopy of KK Family Card, 3. Photocopy of DKK Ship Ownership Document, 4. Photo of FK Ship, 5. Statement of Ability to Operate, and 6. Maintaining SMART FISHING Tools[9]. SMART is also used in completing the selection of financial assistance in developing food businesses for the community, with legality criteria, the experience of trade activities, having AD / ART, grinding machines, and a storage warehouse area[10].

The SMART method is also used to select participants for basketball tournaments, making it easier for coaches to select participants[11]. The SMART method is also used to complete the coffee suppliers' selection at the Coffee Shop. The criteria are price, quality, delivery time, shipping cost, color, and post-harvest duration. Able to help Coffee Shop business people determine coffee suppliers[12]. The SMART method is also used to determine the beneficiary communities affected by covid by using the criteria of Integrated Community Welfare Data (DTKS) consisting of 10 standards, namely: home conditions, income, electricity voltage, education, employment, water sources, cooking fuel, age, and dependents. It makes it easier for village officials to decide on recipients of community social assistance[13]. The SMART method is also used to solve the problem of providing scholarships to lecturers for further study of doctoral education. The criteria used are NIDN, Period of Service, Teaching, Educational Qualifications, Research, Service, Functional Position, and Journal Publication[14]. The SMART method also solves the problem of selecting bamboo suppliers. The criteria used are Price, Quality, Delivery Time, and Service Provided[15]. They refer to several studies that the SMART method can solve problems in choosing alternatives with many criteria. Based on this, this study tries to apply the SMART method in solving extracurricular selection problems at SMAN I Cibungbulang.

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4112

MATERIALS AND METHODS

The research methodology uses the Waterfall method, as shown in figure 1. Calculation analysis using the SMART (Simple Multi-Attribute Rating Technique) method, shown in figure 2.

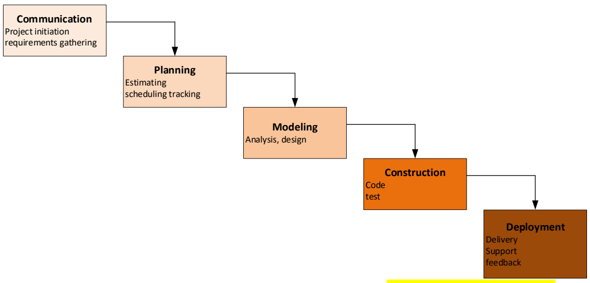


Figure 2. Waterfall method[16][roger presman]

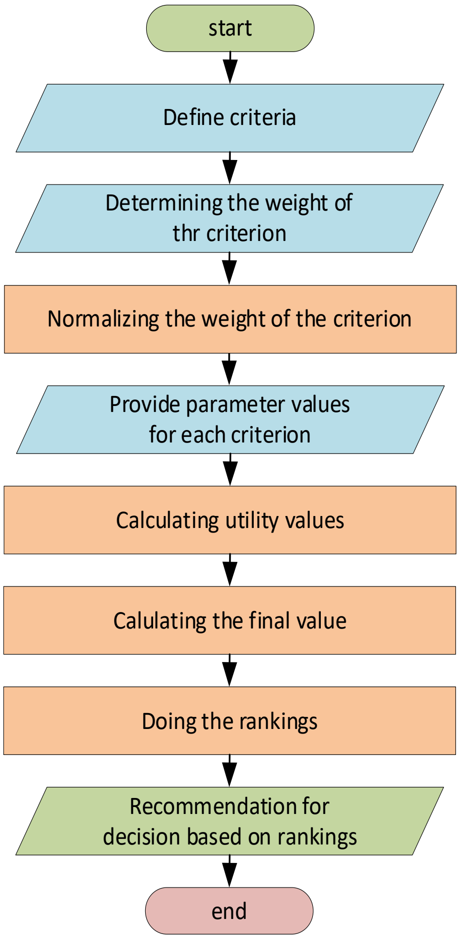


Figure 2. Stages of analysis with the SMART method

Step communication through interviews, observation, and filling out questionnaires by students. This process is carried out to determine how the current business system is running, what is the extracurricular data, the criteria data, parameter value data for each criterion, and Student sample data is used in the analysis stage.

Step Planning to develop a work schedule for the system.

Step Modeling is carried out by analyzing problems with the SMART method and designing the needs of a decision support system.

Step construction to build a decision support system application. System development using the PHP programming language, a web server using Apache, a Google Chrome browser, and a MySQL database server. They are testing the system using a black box. The data sources for testing used are: 1) alternative data of Basketball, Volly, Aikido, and Futsal, 2) data on interest criteria, aptitude, training schedule, and student achievement.

The SMART (Simple Multi-Attribute Rating Technique) method uses the analysis data. The steps taken in the formulation of the SMART strategy are as follows:

1. Define criteria. The decision maker sets the data of the requirements. In the study, table 1 displays alternative data.

Table 1. List of Criteria

No	Code of Criteria	Name of Criteria
1	C1	Minat
2	C2	Bakat
3	C3	Jadwal
4	C4	Prestasi

2. Determine alternatives. Decision makers fix alternative data. Table 2 displays alternative data.

Table 2. List of Alternatives

No	Code of Alternative	Name of Alternative
1	A1	Basket
2	A2	Volly
3	A3	Aikido
4	A4	Futsal

3. Determine the weight of the criteria. They are determined on each criterion using an interval of 1-100 for each criterion with the most critical priority. The criteria used are the Benefit Criteria, shown in table 3.

Table 3. Bobot Criteria

No	Code of Criteria	Name of Criteria	Wight for Criteria
1	C1	Minat	40
2	C2	Bakat	15
3	C3	Jadwal	35
4	C4	Prestasi	10
		Total	**100**

Normalize the criterion's weight by comparing the value of the criterion weight with the sum of the weights of the criterion in the following equation 1.

$$W_i = \frac{w_i'}{\sum_{J=1}^{m} w_j} \quad \text{.......... (1)}$$

W_i: Normalized criterion weights for i-th criteria.

W_i' : Weight of the i-th criterion.
W_j: Weight of criteria
j: 1,2,3,...,m Number of criteria

Table 4. Normalization Results

No	Code of Criteria	Name of Criteria	Wight for Criteria	normalization
1	C1	Minat	40	40/100= 0,4
2	C2	Bakat	15	15/100=0,15
3	C3	Jadwal	35	35/100=0,35
4	C4	Prestasi	10	10/100 =0,1
Total				1

4. Provide parameter values for each criterion. The criteria values for each alternative are both quantitative and quantitative. This research uses qualitative and is converted to quantitative values. Grades are obtained through student representative interviews. The parameter values for each criterion are shown in Table 5.

Table 5. Parameter values for each criterion

	Minat	Bakat	Jadwal	Prestasi
Sangat	4	4	4	4
Cukup	3	3	3	3
Kurang	2	2	2	2
Tidak	1	1	1	1

5. Calculates the utility for each criterion. Assignment of utility values. Assignment of utility values. Data were obtained from questionnaires that students filled out. The data is shown in table 6.

Table 6. Test data

Code of Alternative	Criteria			
	Minat	Bakat	Jadwal	Prestasi
A1	4	3	1	1
A2	3	2	4	2
A3	2	1	4	2
A4	3	1	4	3
Minimum	**2**	**1**	**1**	**1**
Maximum	**4**	**3**	**4**	**3**

Because the criterion type is benefit criteria, equation 2 calculates utility values. The calculation results are shown in Table 7.

$$Uj_{(ai)} = \frac{(Cout - Cmin)}{(Cmax - Cmin)} \quad \text{.......... (2)}$$

$Uj_{(ai)}$: j-th criterion utility value for i-th alternative

$Cout$: I-th alternative criterion value
$Cmax$: Maximum criterion value
$Cmin$: Minimum criterion value
Sample calculation

$$Uminat_{(A1)} = \frac{(4-2)}{(4-2)} = \frac{2}{2} = 1$$

Table 7. Utility calculation results

Code of Alternative	Nilai Utility			
	Minat	Bakat	Jadwal	Prestasi
A1	1	1	0	0
A2	0.5	0.5	1	0.5
A3	0	0	1	0.5
A4	0.5	0	1	1

6. They calculate the final value, Determining the absolute value of each by multiplying the value obtained from the normalization of the common data criterion value by the normalization value of the weight of the criteria. Then add up the multiplication value, which is used in equation 3. Moreover, the final calculation results are shown in Table 8.

$$U_{(ai)} = \sum_{j=1}^{m} wj * uj(ai) \quad \text{.......... (3)}$$

$U_{(ai)}$: The total value for the i-th alternative.
wj: The value of the weight of the j-th criterion, which is already normalized.
uj: The value of the j-th criterion utility for the i-th alternative

Example calculation
Alternative final value A1 [Basket]
$=(w_{A1} * u_{A1(a1)}) + (w_{A2} * u_{A2(a1)}) + (w_{A3} * u_{A3(a1)}) + (w_{A4} * u_{A4(a1)})$
$= (1 * 0.4) + (1 * 0.15) + (0 * 0.35) + (0 * 0.1)$
$= (0.4) + (0.15) + (0) + (0)$
$= 0.55$

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4112

Table 8. The result of the calculation of the final value

Code of Alternative	Criteria				The final result
	Minat	Bakat	Jadwal	Prestasi	
A1	0.4	0.15	0	0	0.55
A2	0.2	0.075	0.35	0.05	0.68
A3	0	0	0.35	0.05	0.40
A4	0.2	0	0.35	0.1	0.65

Squealing. The final results are sorted from the largest to the smallest, then obtained in table 9.

Table 9. Ranking Results

Ranking	Alternative		Final
	code	name	grade
1	A2	Volly	0.68
2	A4	Futsal	0.65
3	A1	Basket	0.55
4	A3	Aikido	0.40

Based on the results of the ranking can be recommended to the student that Volly's extracurriculars are the best choice first, second, Futsal, third, Basketball, and fourth, Aikido.

RESULTS AND DISCUSSION

It is implementing the results of data analysis in a decision support system. So an analysis of system needs is carried out by studying business processes that run, shown in figure 4. Next, describe the business processes developed as shown in Figure 5, and also depict the use case diagram shown in figure 6.

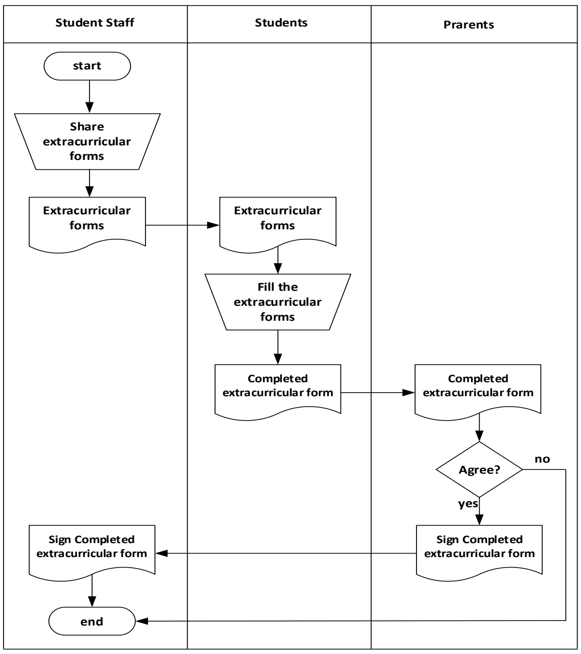


Figure 4. Existing business processes

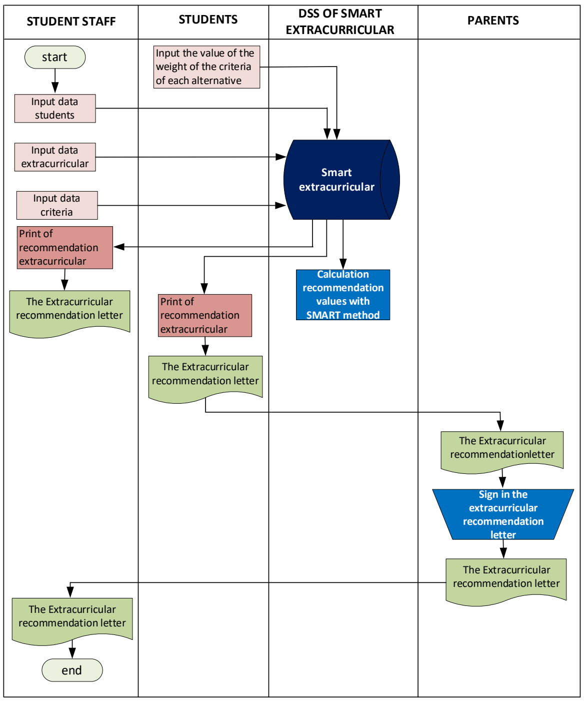


Figure 5. Developed business processes

In business processes that develop like figure 5, student staff will input student data, extracurricular data, and data criteria. Students will fill in the parameter values of each criterion, and then the system will process them with a SMART model. Furthermore, the system will display recommendations based on the ranking of values.

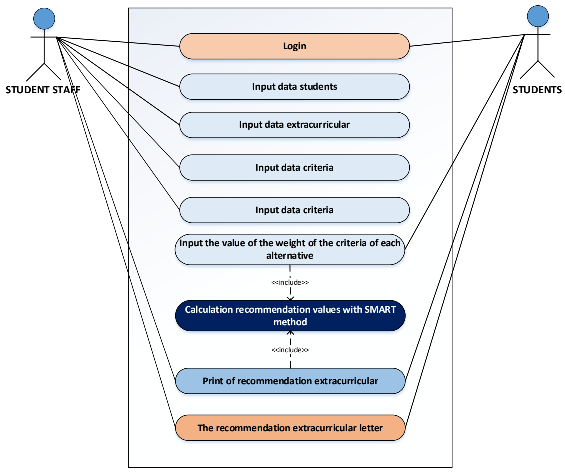


Figure 6. Use case diagram

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4112

The use case diagram, it is shown how the relationship between the actor and the system. In the use case diagram, two actors are student staff and students whose authorization is each. There are nine use cases.

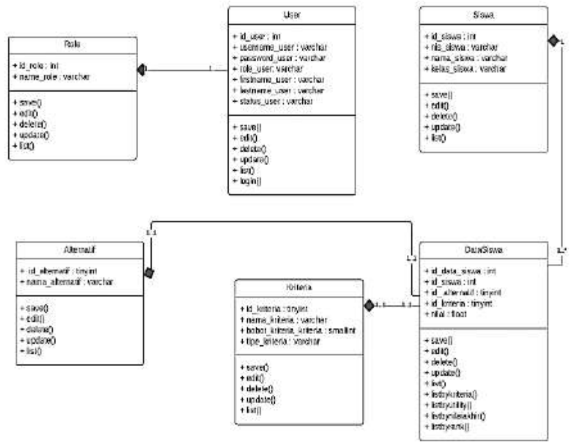


Figure 7. Class diagram

The diagram class illustrates the object's structure; six objects are in the extracurricular selection DSS.

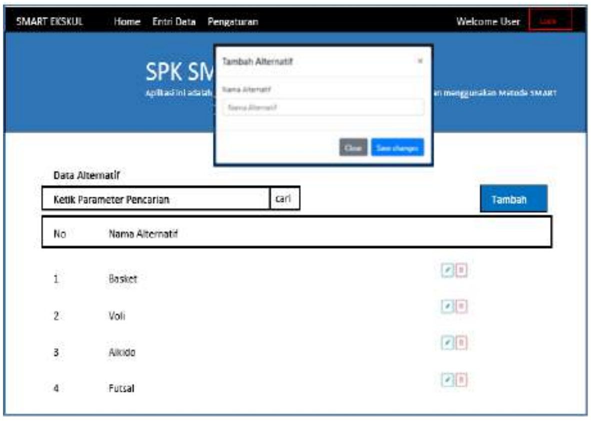


Figure 9. Input data alternative

After the system design stage is completed, the decision support system is built. Figure 9 shows how alternatives are inputted into the system. There are four alternatives Basket, Volly, Aikido, and Futsal

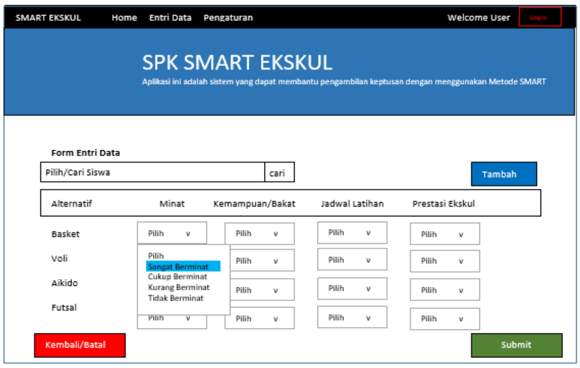


Figure 10. Input parameter values of each criterion

Students collect input data on each criterion's parameters to find extracurricular recommendations based on their interests and talents. The process of inputting parameter data is shown in figure 10.

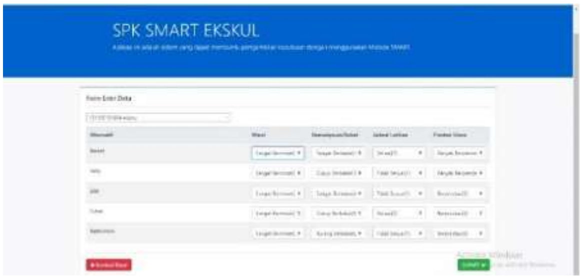


Figure 11. Calculating the final value with SMART

After the input data parameters of each criterion are carried out, the system will calculate the final value of each alternative using the SMART method. Inputting the value of each parameter is filled with qualitative, and the system will convert to quantitative value. This process is shown in figure 11.

CONCLUSION

The problem of choosing one alternative from many criteria that are often experienced daily in organizations to make decisions can be solved by the SMART method. The SMART method has two types of criteria: benefit and cost. Parameter values can be given qualitatively and quantitatively. Parameter values can be given qualitatively and quantitatively. This research uses the SMART method to solve the problem of extracurricular selection. Four alternatives are used: Basketball, Volly, Aikido, and Futsal. The criteria used are interests, talents, training schedules, and achievements. The resolution of this problem is also supported by the decision support system (DSS) application. The DSS application makes it easier for student guardians, staff, students, and parents to

CC BY NC

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4112

manage data and obtain information. The study succeeded in applying the SMART method in DSS to solve the problem of extracurricular selection recommendations at SMAN Cibungbulang, Bogor regency.

REFERENCE

[1] M. P. N. R. Indonesia, *Pembinaan Kesiswaan*. Indonesia, 2008.

[2] S. N. 1 Cibungbulang, *Keputusan Kepala SMA Negeri 1 Cibungbulang Kab. Bogor*. Indonesia.

[3] A. C. Murti and W. A. Triyanto, "Analisa Metode Simple Multi Attribute Rating Technique (Smart) Dan Simple Additive Weighting (Saw) Dalam Mendukung Gerakan Masyarakat Hidup Sehat," *Indones. J. Technol. Informatics Sci.*, vol. 2, no. 2, pp. 81–84, 2021, doi: 10.24176/ijtis.v2i2.6310.

[4] F. Nasher and M. W. Fratama, "Sistem Pendukung Keputusan Kemajuan Yayasan Menggunakan Metode SMART (Simple Multi Attribute Rating Technique) (Studi Kasus : Yayasan Pendidikan Daarut Taqwa Asyamsuriyyah Sabandar)," *Media J. Inform.*, vol. 12, no. 1, p. 29, 2020, doi: 10.35194/mji.v12i1.1190.

[5] F. Syakti and B. Tujni, "Implementasi Simple Multi Attribute Rating Technique Dalam Pemilihan Minat Ekstrakurikuler Siswa," *J. Inform.*, vol. 20, no. 1, pp. 1–12, 2020, doi: 10.30873/ji.v20i1.1738.

[6] A. Prayoga and dkk, "Evaluasi Kinerja Kepolisian Berdasarkan Kriteria Pengguna Menggunakan Metode Smart (Studi Kasus Polsek Makasar Jakarta Timur)," *J. Informatics Adv. Comput.*, vol. 1, no. 1, 2020, [Online]. Available: https://journal.univpancasila.ac.id/index.php/jiac/article/view/1398/891.

[7] K. Sitorus, D. Dedih, and A. B. Purba, "Penentuan Bonus Tahunan Menggunakan Metode Simple Multi Attribute Rating Technique," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 20–27, 2022, doi: 10.35969/interkom.v16i4.193.

[8] S. Hamidani and V. Amalia, "Penentuan calon mahasiswa penerima beasiswa hafiz Alquran pada tingkat perguruan tinggi menggunakan metode SMART," *Teknologi*, vol. 10, no. 2, pp. 85–92, 2020, doi: 10.26594/teknologi.v10i2.2036.

[9] N. D. Nuris, "Penentuan Pemberian Bantuan Smart Fishing Kepada Nelayan Menggunakan Metode SMART," *J. ICT Inf. Commun. Technol.*, vol. 20, no. 2, pp. 406–411, 2021, doi: 10.36054/jict-ikmi.v20i2.371.

[10] M. Guntur and R. Yanto, "Penerapan Metode SMART untuk Seleksi Kelayakan Penerima Bantuan Pengembangan Usaha Pangan Masyarakat," *Telematika*, vol. 12, no. 2, pp. 149–159, 2019, doi: 10.35671/telematika.v12i2.826.

[11] S. S. Putro, E. Rahmanita, and F. Khumairoh, "Penerapan Metode Smart Untuk Seleksi Peserta Turnamen Pada Cabang Olahraga Bola Basket," *J. Simantec*, vol. 7, no. 2, pp. 60–71, 2019, doi: 10.21107/simantec.v7i2.6691.

[12] R. A. S. PRAYOGA and S. W. PUTRI, "Sistem Pendukung Keputusan Memilih Pemasok Kopi Dengan Metode Smart Pada Coffee Shop Abc Ponorogo," *... Inform. Sains dan ...*, vol. 7, no. April, pp. 69–78, 2022, [Online]. Available: http://journal.uin-alauddin.ac.id/index.php/instek/index.

[13] A. Nurhidayat, A. Supriyanto, and E. Nurraharjo, "Sistem Pendukung Keputusan Metode SMART Dalam Penentuan Pemberian Bantuan Sosial Berdasarkan DTKS Di Desa Bebengan," *J. Sains Komput. Inform. (J-SAKTI*, vol. 6, no. 2, pp. 1022–1031, 2022, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jsakti%0ASistem.

[14] A. Ardi and I. Fadhli, "Sistem Pendukung Keputusan Pemilihan Beasiswa Doktor Untuk Dosen Potensial Dengan Metode Smart," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 7, no. 1, pp. 39–46, 2020, doi: 10.33330/jurteksi.v7i1.911.

[15] M. N. Amalia and M. Ary, "Sistem Pendukung Keputusan Pemilihan Supplier Dengan Menggunakan SMART Pada CV. Hamuas Mandiri," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 127–134, 2021, doi: 10.34128/jsi.v7i2.322.

[16] B. R. M. Roger S. Pressman, Ph.D., *Software Engineering A Practitioner's Approach*, Eighth edition. New York: McGraw-Hill Education, 2021.

CC BY NC

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi
No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

SENTIMENT ANALYSIS OF CONTENT PERMENKOMINFO NO.5 OF 2020 USING A CLASSIFICATION ALGORITHM

Mohammad Amada[1]; Munawar[2*]; Marzuki Pilliang[3]

Master of Computer Science[1,2,3]
Esa Unggul University
https://www.esaunggul.ac.id/
arabamada@gmail.com, moenawar@gmail.com, marzuki.pilliang@ieee.org

Abstract*—This study aims to evaluate the impact of the policy issued by the Minister of Communication and Information Technology (PERMENKOMINFO No.5 of 2020) on the public's ability to access content through Private Scope Electronic System Providers (PSE). The study uses sentiment analysis and data classification methods to analyze the content of PERMENKOMINFO No.5 of 2020 and provides results on the accuracy of sentiment prediction. The results of the study show that the data classification method in sentiment analysis can provide accurate results in predicting the sentiment towards the content of PERMENKOMINFO No.5 of 2020. The study also highlights the need for improvement and better policy to ensure the interests of the public in accessing online information. The negative sentiment of 80.34% obtained through sentiment analysis provides important contributions for policy evaluation and feedback for improvement. This study provides valuable insights into the public's sentiment towards the PERMENKOMINFO No.5 of 2020 policy and its impact on their ability to access content. It also contributes to understanding the legal uncertainty in accessing content and reinforces the case for better policy to ensure the interests of the public.*

Keywords*: Sentiment analysis, PERMENKOMINFO No.5 of 2020, Content Regulation.*

Intisari— Penelitian ini bertujuan untuk mengevaluasi dampak atas kebijakan yang diterbitkan oleh Menteri Komunikasi dan Teknologi Informasi (PERMENKOMINFO No.5 Tahun 2020) pada kemampuan masyarakat untuk mengakses konten melalui Penyelenggara Sistem Elektronik (PSE) Lingkup Pribadi. Studi ini menggunakan metode analisis sentimen dan klasifikasi data untuk menganalisis isi PERMENKOMINFO No.5 Tahun 2020 dan memberikan hasil tentang akurasi prediksi sentimen. Hasil studi menunjukkan bahwa metode klasifikasi data dalam analisis sentimen mampu memberikan hasil yang akurat dalam memprediksi sentimen terhadap isi PERMENKOMINFO No.5 Tahun 2020. Studi ini juga menyoroti kebutuhan untuk peningkatan dan kebijakan yang lebih baik untuk memastikan kepentingan masyarakat dalam mengakses informasi online. Sentimen yang negatif sebesar 80.34% yang didapatkan melalui analisis sentimen memberikan kontribusi penting bagi evaluasi kebijakan dan memberikan masukan untuk perbaikan. Studi ini memberikan wawasan penting tentang sentimen masyarakat terhadap kebijakan PERMENKOMINFO No.5 Tahun 2020 dan dampaknya pada kemampuan masyarakat untuk mengakses konten. Studi ini juga berkontribusi pada pemahaman tentang ketidakpastian hukum dalam mengakses konten dan memperkuat case untuk kebijakan yang lebih baik untuk memastikan kepentingan masyarakat.

Kata Kunci*: Analisis sentimen, PERMENKOMINFO No.5 Tahun 2020, Pengaturan Konten .*

INTRODUCTION

The Minister of Communication and Information Technology issued a regulation, Ministerial Regulation Number 5 of 2020, which limits the activities of Private Scope Electronic System Providers (PSE Lingkup Privat). Through Press Release No. 296 in 2022, the Ministry of Communication and Information Technology set a deadline until July 27, 2022, for unregistered PSEs to process their registration, or they will be subject to temporary access termination sanctions. On July 30, 2022, unregistered PSEs were subjected to temporary service termination sanctions by the Ministry of Communication and Information Technology, making their services unavailable to users. This access closure triggered public responses on various social media platforms, including Twitter, and created the hashtag #BlokirKominfo, which shows various responses to the legal uncertainty for the public in accessing content after the imposition of sanctions on PSE by the Ministry of Communication and Information Technology.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4091

Gap that creates legal uncertainty for the public in accessing content can be found in the administrative sanctions imposed by the Minister on PSEs that do not comply with the registration provisions as stated in Article 2 and 6. Administrative sanctions such as access termination to electronic systems and even the revocation of electronic system provider registration can create legal uncertainty for the public in accessing content through the electronic system.

To understand the rejection that arose from the policy outlined in PERMENKOMINFO No. 5 of 2020 related to legal uncertainty in accessing content, it is necessary to analyze the content of the regulation. Sentiment analysis is the process of processing unstructured words found in a word or sentence to predict expressive value in the form of positive, negative, or neutral responses [1]. Sentiment analysis seeks to determine the polarity of text by measuring its level of subjectivity, which can result in a positive, neutral, or negative value [2]. The sentiment analysis research on the content of PERMENKOMINFO No.5 Year 2020 aims to identify public sentiment towards the policies outlined in it.

In sentiment analysis, several classification methods can be used. One is Naive Bayes, which operates based on probability and Bayesian theorem [3]. In this method, the assumption is that the presence or absence of an attribute is not related to other attributes. It calculates the probability of each previously determined data feature, calculates the likelihood value by multiplying the value of each probability, and then predicts the data based on the previously determined label [4]. In addition, Support Vector Machine (SVM) can also be used as a classification method. SVM is a hypothetical spatial learning system in the form of a linear function in multidimensional features trained using optimization theory [5]. Finally, the K-Nearest Neighbor (K-NN) method divides new data based on the distance between new data and the nearest neighbors [6]. Several studies have utilized sentiment analysis and data classification as their research methods to analyze and understand public sentiment towards various issues.

Yoga Vikriansyah Wijaya, et al [5] performed sentiment analysis on Twitter social media using the Support Vector Machine classification algorithm on policies related to the ITE Law. The test data accuracy was 84%, the recall value was 65%, and the f1-score was 71% for each sentiment class, with the majority indicating that 74.10% of the public is against the ITE Law.

Ahmad Syaifuddin and Mohammad Muslimin [7] analyzed opinion sentiment on Twitter using the Lexicon Based method on the Kominfo PSE policy. The sentiment analysis results were 80.1% negative towards the opinions expressed by the public on social media Twitter in a dataset of 3,300 tweets, leading to the conclusion that the public opposes the PSE policy.

The research conducted by Yoga Vikriansyah Wijaya, et al, as well as Ahmad Syaifuddin and Mohammad Muslimin, focused on public sentiment analysis regarding the PSE policy and ITE Law through the social media platform Twitter. However, this research has not demonstrated an in-depth understanding of the legal uncertainty for the public in accessing content. Therefore, sentiment analysis of the content of the PSE policy and ITE Law is necessary to understand the gap in legal uncertainty for the public in accessing content. This is an important issue to understand the implications of the PSE policy and ITE Law for the public.

The next research will focus on sentiment analysis of the content of Ministerial Regulation Number 5 of 2020. Sentiment analysis will help to understand how the public perceives the content of the policy and whether there are legal issues for the public in accessing content. This research will help researchers obtain objective and detailed information about the gap.

MATERIALS AND METHODS

The research that will be conducted combines text analysis approach, namely Text Mining, and sentiment analysis technique. Text Mining is the process of searching for patterns in text to extract useful information, while sentiment analysis is the process of understanding the polarity of a text and categorizing it as positive, negative, or neutral [8], [9]. In this research, the analysis process will go through several stages, such as dataset retrieval, Text Preprocessing, sentiment labeling with InSet Lexicon, word weighting with TF-IDF, and classification using Support Vector Machine, Naïve Bayes Classifier, and K-Nearest Neighbors. The research steps will be explained in Figure 1.

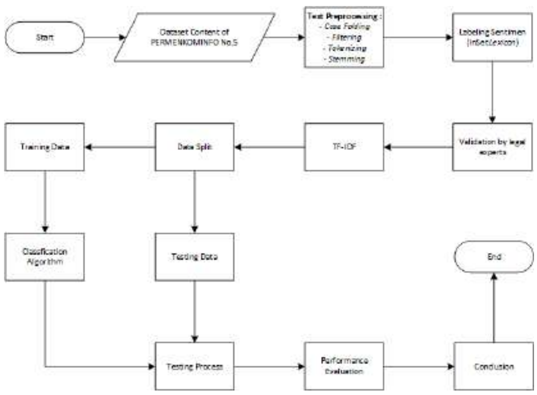


Figure 1. Research Procedure

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

2.1. Data Collection

In this research, the dataset used is derived from the content of Regulation of the Minister of Communication and Information Technology No. 5 of 2020. The dataset can be downloaded through the official website of the Legal Documentation and Information Network of the Ministry of Communication and Information Technology (JDIH Kominfo) at https://jdih.kominfo.go.id/produk_hukum/view/id/759/t/peraturan+menteri+komunikasi+dan+informatika+nomor+5+tahun+2020. The dataset consists of 178 section, starting from the opening until the end. In this study, sentiment analysis will be focused on articles 7 and 8, which contain Administrative Sanctions and Normalization and consist of 8 sections.

2.2. *Text Preprocessing*

Text Preprocessing is a process that is applied to change and clean text to be structured [8], [10] so that the text becomes quality when analyzed. At this stage, the dataset is processed in order to clean up words, characters or symbols that have no value or influence on the analysis process and normalize words to become standard words/language. This process is divided into several stages, and the following steps are carried out :

1. Case Folding

Case Folding is a procedure used to equate the shape of letters in data into a lowercase form [11]. The output of the Case Folding process is described in table 1.

Table 1. Application of Case Folding in Datasets

Before	After
Dalam hal sanksi administratif yang diberikan kepada PSE Lingkup Privat sebagaimana dimaksud pada ayat (1) adalah Pemutusan Akses terhadap Sistem Elektronik (access blocking), Menteri melakukan Normalisasi berdasarkan pengajuan rekomendasi oleh Kementerian atau Lembaga atas dasar layanan PSE lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.	dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat (1) adalah pemutusan akses terhadap sistem elektronik (access blocking), menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.

2. Filtering

Filtering is the process of removing words/symbols which are considered to have no meaning [12]. In the Filtering stage, the text that the Case Folding process has carried out will be cleaned of punctuation, numbers, hashtags, mentions (user embeds), slang conversions and correction of wrong/typo words (slang words) from the dataset used. The output of Filtering is described in table 2.

Table 2. Application of Filtering in Datasets

Before	After
dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat (1) adalah pemutusan akses terhadap sistem elektronik (access blocking), menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan.	dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat adalah pemutusan akses terhadap sistem elektronik access blocking menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan

3. Tokenizing

Tokenizing is splitting each existing sentence into words or tokens [13]. In the Tokenizing process, the entire dataset will be broken down per word so that each word is separate and does not become a complete sentence. This process makes it easier for the next process, namely the Stemming process. The output of Tokenizing is described in table 3.

Table 3. Application of Tokenizing in Datasets.

Before	After
dalam hal sanksi administratif yang diberikan kepada pse lingkup privat sebagaimana dimaksud pada ayat adalah pemutusan	['dalam', 'hal', 'sanksi', 'administratif', 'yang', 'diberikan', 'kepada', 'pse', 'lingkup', 'privat', 'sebagaimana', 'dimaksud', 'pada', 'ayat', 'adalah',

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4091

akses terhadap sistem elektronik access blocking menteri melakukan normalisasi berdasarkan pengajuan rekomendasi oleh kementerian atau lembaga atas dasar layanan pse lingkup privat yang telah memenuhi ketentuan peraturan perundangundangan	'pemutusan', 'akses', 'terhadap', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'melakukan', 'normalisasi', 'berdasarkan', 'pengajuan', 'rekomendasi', 'oleh', 'kementerian', 'atau', 'lembaga', 'atas', 'dasar', 'layanan', 'pse', 'lingkup', 'privat', 'yang', 'telah', 'memenuhi', 'ketentuan', 'peraturan', 'perundangundangan']

4. Stemming

In the Stemming process, datasets containing affixes to each word will be converted to base words [14]. In this process, each word fraction of the tokenizing process will be normalized into a root word. In this Step, the conjunctions or conjunctions will be removed. The output of stemming is described in table 4.

Table 4. Application of Stemming in Datasets.

Before	After
['dalam', 'hal', 'sanksi', 'administratif', 'yang', 'diberikan', 'kepada', 'pse', 'lingkup', 'privat', 'sebagaimana', 'dimaksud', 'pada', 'ayat', 'adalah', 'pemutusan', 'akses', 'terhadap', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'melakukan', 'normalisasi', 'berdasarkan', 'pengajuan', 'rekomendasi', 'oleh', 'kementerian', 'atau', 'lembaga', 'atas', 'dasar', 'layanan', 'pse', 'lingkup', 'privat', 'yang', 'telah', 'memenuhi', 'ketentuan', 'peraturan', 'perundangundangan']	['sanksi', 'administratif', 'pse', 'lingkup', 'privat', 'ayat', 'putus', 'akses', 'sistem', 'elektronik', 'access', 'blocking', 'menteri', 'normalisasi', 'dasar', 'aju', 'rekomendasi', 'menteri', 'lembaga', 'dasar', 'layan', 'pse', 'lingkup', 'privat', 'penuh', 'tentu', 'atur', 'perundangundangan']

2.3. Sentiment Labeling

Sentiment labelling is done using a lexicon-based approach to datasets that have been preprocessed by Text Preprocessing using the InSet Lexicon [15]. Assessment is done by calculating the value of each word by matching the existing words from each sentence in the word dictionary dataset that has been given a weight. After each word gets a weight value, a calculation is performed to get the total value (polarity score) of each sentence to be given a sentiment label. In general, the sentiment label is stated as "Negative" if the polarity score obtained is below 0, the "Positive" label is given if the polarity score obtained is above 0, and the "Neutral" label is given if the polarity score has a value of 0. The output of the sentiment prediction labelling process with the InSet Lexicon is described in table 5.

Table 5. Sentiment Analysis Results Based on Compound Score

Text	Polarity Score	Sentiment Prediction
selenggara jasa akses internet internet service provider singkat isp selenggara jasa multimedia selenggara jasa layan akses internet hubung jaring internet publik	28	Positive
menteri kena sanksi administratif pse lingkup privat daftar pasal pasal tanda daftar lapor ubah informasi daftar pasal informasi daftar pasal ayat pasal ayat pasal ayat	-36	Negative
selenggara sistem elektronik orang selenggara negara badan usaha masyarakat sedia kelola operasi sistem elektronik sendirisendiri bersamasama guna sistem elektronik perlu perlu	0	Neutral

Sentiment analysis was performed on a dataset consisting of 8 article lines. The sentiment prediction graph for the dataset is displayed in Figure 2, which shows the sentiment distribution for the articles in the dataset based on the InSet Lexicon dictionary. This dictionary commonly categorizes words in a text into positive, negative, or neutral sentiment categories. The graph indicates that the majority of the articles in the dataset had a negative sentiment, suggesting that the articles tended to convey messages or information that were less joyful or poignant. Articles with positive

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

and neutral sentiments were present in smaller proportions. The results showed that the majority of the articles, 80.34%, had a predicted negative sentiment, 17.42% of the articles had a positive sentiment, and 2.25% of the articles had a neutral sentiment.

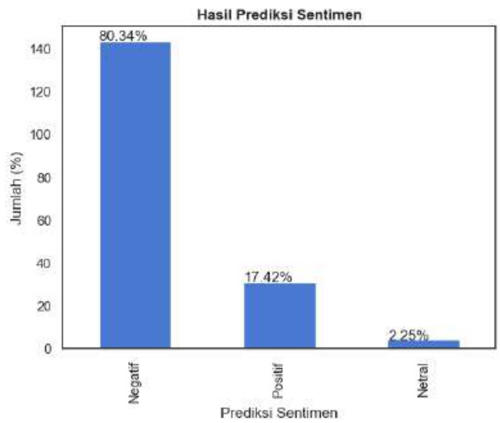


Figure 2. Sentiment Predicted Distribution of Datasets with InSet Lexicon

The results of the sentiment analysis implicitly indicate that articles 7 and 8 of Ministerial Regulation Number 5 of 2020 have the potential to create legal uncertainty for the public in accessing content, which is consistent with the research findings showing negative results. The following is the result of sentiment prediction analysis on the contents of articles 7 and 8 of Ministerial Regulation Number 5 of 2020 on Communication and Information Technology, as described in Table 6.

Table 6. Results of Sentiment Prediction in Articles 7 and 8

Articles	Section	Text Preprocessing	Score
7	1	pse lingkup privat daftar ayat huruf menteri sanksi administratif putus akses sistem elektronik access blocking	-7
7	2	pse lingkup privat tanda daftar lapor ubah informasi daftar ayat huruf informasi daftar ayat huruf menteri sanksi administratif tegur tulis surat elektronik electronic mail media elektronik henti pse lingkup privat indah tegur tulis ayat huruf putus akses sistem elektronik access blocking cabut tanda daftar selenggara sistem elektronik pse lingkup privat konfirmasi jangka tujuh henti ayat huruf	-30
7	3	pse lingkup privat penuh tentu daftar pasal pasal menteri normalisasi sistem elektronik putus akses access blocking ayat	-18
7	4	pse lingkup privat baru informasi daftar menteri normalisasi sistem elektronik henti ayat huruf	-14
7	5	pse lingkup privat daftar ulang informasi daftar menteri normalisasi sistem elektronik putus akses sistem elektronik cabut tanda daftar selenggara sistem elektronik ayat huruf	-15
7	6	menteri kena sanksi administratif pse lingkup privat dasar mohon menteri lembaga dasar langgar atur perundangundangan bidang menteri lembaga milik wenang sesuai tentu atur perundangundangan	-3
8	1	sanksi administratif pse lingkup privat ayat putus akses sistem elektronik access blocking menteri normalisasi dasar aju rekomendasi menteri lembaga dasar layan pse lingkup privat penuh tentu atur perundangundangan	-16
8	2	pse lingkup privat tanggung selenggara sistem elektronik kelola informasi elektronik dokumen elektronik sistem elektronik andal aman tanggung	-13

The sentiment analysis on Articles 7 and 8 shows negative outcomes and implicitly indicates the potential for legal uncertainty for the public in accessing content, which is consistent with the

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4091

research findings. These results are supported by validation from two legal experts who deemed the analysis results as valid and in accordance with prevailing conditions. Legal expert that confirmed it was conducted accurately and in line with legal principles. The experts considered the analysis results reliable and accurate.

The results of the sentiment prediction are visualized in graphical form. WordCloud is a method for visualizing words based on the number of occurrences or their frequency in a document [16], The more often the word is used, the greater the visualization of the word. The following is a word visualization based on the frequency of word use in the content of PERMENKOMINFO No. 5 of 2020, which is shown in figure 3.

Figure 3. WordCloud for PERMENKOMINFO Content No.5 of 2020

The results of the word visualization applied in the WordCloud method show that the words "akses" and "sistem" are the most frequently used words in the content of PERMENKOMINFO No. 5 of 2020. This indicates how important both words are in the context of the regulation and suggests that access and system are the main focus of the regulation. This visualization also provides useful information to understand the context and main focus of the content of the regulation. The frequency of word occurrences is also explained in Figure 4 based on the 15 words with the most usage in the dataset for the article on PERMENKOMINFO No.5 of 2020.

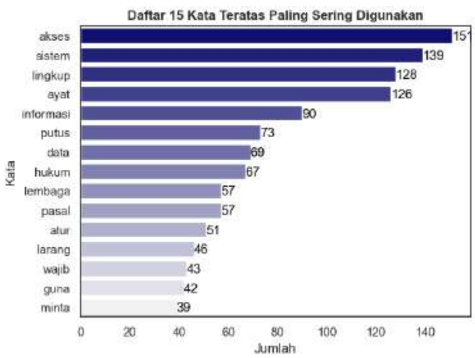


Figure 4. List of Popular Words that Often Appear

After the sentiment analysis has been conducted and approved by legal experts, the next step is to assign weights to the analyzed words that have been given sentiment values using the TF-IDF method. Term Frequency -Inverse Document Frequency is a statistical analysis technique that analyses the frequency of occurrence of words and is combined with inverse [17]. The purpose of weighting words using the TF-IDF method on the sentiment analysis results of the content of Ministerial Regulation No. 5 of 2020 is to provide different weights to each word in the document based on its meaning in the text. This weighting helps to identify the most important words in the document and to identify the sentiment polarity. The results of the TF-IDF calculations on the dataset are shown in Figure 5.

	TF-IDF
access	3.142863
acu	5.494239
adil	3.884801
administratif	3.548328
adu	4.801091
...	...
web	5.494239
website	4.801091
wenang	3.622436
wilayah	3.990161
yurisdiksi	5.494239

439 rows × 1 columns

Figure 5. TF-IDF Calculation Results on the PERMENKOMINFO Content Dataset No. 5 of 2020

After performing TF-IDF weighting, the next step in sentiment analysis research on the content of Ministerial Regulation No. 5 of 2020 is to perform data split. Data split is done to divide the data into two parts, which are training data and testing data. The training data is used to train the data classification model, while the testing data is used to test the performance of the trained model. After the data is split, the next step is to use data classification methods, namely Naive Bayes Classifier, Support Vector Machine, and K-Nearest Neighbors. These three methods are used to predict the sentiment of the given text, whether the previously predicted sentiment results are tested to see if they are positive, negative, or neutral. The NBC method works based on calculating the probability of each sentiment class based on the occurrence of words in the text, while the SVM method uses optimization theory-based learning algorithms and linear

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4091

functions in multidimensional features. The KNN method works by grouping new data based on the distance between the new data and some of the nearest data. In this study, all three data classification methods were used to test the sentiment of the content analysis of Ministerial Regulation No. 5 of 2020 and to test the performance of the trained model using testing data. The following are the results of the model testing for the sentiment analysis on the content of Ministerial Regulation No. 5 of 2020.

2.4. Support Vector Machine

The test results with the SVM algorithm, with a test ratio of 80:20, obtained a test accuracy of 83%, a precision value of 82%, a recall value of 100%, and an f1-score value of 90%, which is explained in Figure 6.

	precision	recall	f1-score	support
Negatif	0.82	1.00	0.90	28
Netral	0.00	0.00	0.00	2
Positif	1.00	0.33	0.50	6
accuracy			0.83	36
macro avg	0.61	0.44	0.47	36
weighted avg	0.81	0.83	0.79	36

Figure 6. Testing Result of Support Vector Machine

2.5. Naïve Bayes Classifier

Tests carried out with the NBC algorithm with a test ratio of 80:20 obtained 78% accuracy, 78% precision, 100% recall, and 88% f1-score, which is explained in Figure 7.

	precision	recall	f1-score	support
Negatif	0.78	1.00	0.88	28
Netral	0.00	0.00	0.00	2
Positif	0.00	0.00	0.00	6
accuracy			0.78	36
macro avg	0.26	0.33	0.29	36
weighted avg	0.60	0.78	0.68	36

Figure 7. Testing Result of Naive Bayes Classifier

2.6. K-Nearest Neighbors

The test results with the KNN model, with a test ratio of 80:20, obtained an accuracy value of 83%, a precision value of 87%, a recall value of 93%, and an f1-score value of 90%, which can be seen in Figure 8.

	precision	recall	f1-score	support
Negatif	0.87	0.93	0.90	28
Netral	0.00	0.00	0.00	2
Positif	0.67	0.67	0.67	6
accuracy			0.83	36
macro avg	0.51	0.53	0.52	36
weighted avg	0.79	0.83	0.81	36

Figure 8. Testing Result of K-Nearest Neighbors

RESULTS AND DISCUSSION

The results of the data classification testing showed that the use of Naive Bayes method was able to provide testing accuracy of 78%, while the SVM and KNN methods were able to provide testing accuracy of 83%. This testing was conducted on 178 articles, including a focus on articles 7 and 8. The results of the comparison of method testing are described in Table 7.

Table 7. Test Results for Each Classification Method

Method / Rated	Support Vector Machine	Naïve Bayes	K-Nearest Neighbors
accuracy	83%	78%	83%
precision	82%	78%	87%
recall	100%	100%	93%
f1-score	90%	88%	90%

This indicates that the use of data classification methods in sentiment analysis can provide accurate results in predicting the sentiment towards the content of PERMENKOMINFO No.5 Tahun 2020, especially in articles 7 and 8 that are the main focus of this research. With good accuracy results, this study can contribute significantly to understanding public sentiment towards the policy. These results indicate that the research was conducted correctly and validly. The fact that the results were also supported by validation from two legal experts strengthens the statement that the sentiment analysis conducted was in accordance with legal standards and produced correct results. Both the results of data classification and legal expert validation can be concluded that the policy in PERMENKOMINFO No.5 Tahun 2020 related to legal uncertainty in accessing content is considered negative by the public. Therefore, there is a need for improvement and better policy to ensure the interests of the public in accessing online information.

CONCLUSION

The conclusion of this study is that the data classification method in sentiment analysis can provide accurate results in predicting the sentiment towards the content of PERMENKOMINFO No.5 of 2020. The study also shows the need for improvement and better policy to ensure the interests of the public in accessing online information. The negative sentiment result of 80.34% obtained in the study contributes to being an input for evaluation. This study provides valuable insights into the public's sentiment towards the policy in PERMENKOMINFO No.5 of

2020 and its impact on their ability to access content. The study contributes to the understanding of legal uncertainty in accessing content and the need for better policy to ensure the interests of the public.

REFERENCE

[1] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.

[2] M. Pilliang, H. Akbar, and G. Firmansyah, "Sentiment Analysis for Super Applications in Indonesia: A Case Study of Gov2Go App," in *2022 3rd International Conference on Electrical Engineering and Informatics (ICon EEI)*, 2022, pp. 80–85. doi: 10.1109/IConEEI55709.2022.9972291.

[3] I. Gusti Naufhal Daffa Adnyana, F. Adams, and A. Windari Oktavia, *Analisis Sentimen Terhadap Undang-Undang Cipta Kerja Menggunakan Metode Naïve Bayes*. 2021.

[KOMPARASI NAIVE BAYES, K-NEAREST NEIGHBOR, DAN RANDOM FOREST UNTUK PREDIKSI CALON MAHASISWA YANG DITERIMA ATAU MUNDUR," vol. 9, no. 7, pp. 1341–1348, 2022, doi: 10.25126/jtiik.202296737.

[5] …… Yoga, V. Wijaya, Y. Vikriansyah Wijaya, A. Erfina, and C. Warman, "Analisis Sentimen Seputar UU ITE menggunakan Algoritma Support Vector Machine", [Online]. Available: https://databoks.katadata.co.id/

[6] E. N. Hamdana, M. Balya, and I. Alfahmi, "JIP (Jurnal Informatika Polinema) Pengembangan Sistem Analisis Sentimen Berbasis Java Pada Data Twitter Terhadap Omnibus Law Menggunakan Algoritma Naïve Bayes dan K-Nearst Neighbor (K-NN)".

[7] A. Syaifuddin and M. Muslimin, "ANALISIS SENTIMEN PADA SOSIAL MEDIA TENTANG IMPLEMENTASI KEBIJAKAN PSE KOMINFO MENGGUNAKAN ALGORITME LEXICON BASED," *SMART CITY AND SUSTAINABLE DEVELOPMENT GOALS*, vol. 1, no. 1, 2022.

[8] A. Rahman, Wiranto, and A. Doewes, "Online News Classification Using Multinomial Naive Bayes." doi: 10.20961/itsmart.v6i1.11310.

[9] R. Rismayani, H. SY, T. Darwansyah, and I. Mansyur, "Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 169–176, Aug. 2022, doi: 10.34010/komputika.v11i2.6501.

[10] D. S. Suparno and M. Rosyda, "Penggunaan Text Modeling Untuk Identifikasi Kesalahan Penulisan Kata Pada Teks Pidato Bupati Banggai Sulawesi Tengah," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 779, Jul. 2021, doi: 10.30865/mib.v5i3.3051.

[11] Y. Afrillia, L. Rosnita, and D. Siska, "ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP ISU KESETARAAN GENDER DALAM PENERAPAN PERMENDIKBUDRISTEK NOMOR 30 TAHUN 2021 MENGGUNAKAN TEXTBLOB ANALYSIS OF TWITTER USER SENTIMENT TOWARDS TO ISSUE OF GENDER EQUALITY IN THE IMPLEMENTATION OF PERMENDIKBUDRISTEK NUMBER 30 OF 2021 USING TEXTBLOB," *Journal of Informatics and Computer Science*, vol. 8, no. 2, 2022.

[12] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, p. 1020, Aug. 2022, doi: 10.30865/jurikom.v9i4.4673.

[13] S. Mujahidin, B. Prasetio, and M. C. Cahyo Utomo, "Jurnal Vocational Teknik Elektronika dan Informatika", [Online]. Available: http://ejournal.unp.ac.id/index.php/voteknika/index

[14] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY*, vol. 2, no. 1, 2022.

[15] F. Koto and G. Y. Rahmaningtyas, "InSet Lexicon: Evaluation of a Word List forIndonesian Sentiment Analysis in Microblogs."

[16] M. Galih Pradana, "PENGGUNAAN FITUR WORDCLOUD DAN DOCUMENT TERM MATRIX DALAM TEXT MINING."

[17] A. M. Priyatno and L. Ningsih, "SISTEMASI: Jurnal Sistem Informasi Pembobotan TF-IDF untuk Mendeteksi Akun Spammer di Twitter berdasarkan Tweet dan Representasi Retweet dari Tweet TF-IDF Weighting to Detect Spammer Accounts on Twitter based on Tweets and Retweet Representation of Tweets." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4138

IMPLEMENTATION OF THE SMART METHOD IN THE DECISION SUPPORT SYSTEM FOR PENALTY RECOMMENDATIONS

Novita Br Ginting1*; Dahlia Widhyaestoeti 2; Jejen Jaenuddin 3; Fitria Rachmawati4; Yusup Maolani5

Information System Study Program 1234, Informatic Engineering Study Program5
Ibn Khaldun University Bogor
https://www.uika-bogor.ac.id
novitawahab@uika-bogor.ac.id 1*,dahlia@uika-bogor.ac.id-2, zen@uika-bogor.ac.id-3, fitria@uika-bogor.ac.id-4, yusup17maolani@gmail.com-5

(*) Corresponding Author

Abstract—Sanctions need to be given to students who do not obey school rules, and the purpose is to establish student discipline and ethics. SMK Gazza Wiguna 1 Megamendung Bogor district also imposes sanctions for students who violate the rules of conduct. The problem is that determining the violation point has not been done objectively. To overcome this, applying the Simple Multi-Attribute Rating Technique Method (SMART) method to the decision support system to determine recommendations for sanctions for students who commit violations. This system is designed to use UML tools and is built on a web basis. The system uses an alternative oral warning, first warning letter, second warning letter, third warning letter, and returned to parents (DO). The criteria include violation of school uniforms, tidiness, discipline, cleanliness, order, hospitality, commemorating national holidays, and additional provisions. The type of criteria used is the cost criteria because it is expected that students do not violate school rules; if students commit violations, it is hoped that students do not commit serious transgressions.

Keywords: decision support system; penalty; SMART method; Information System

Abstrak*—Pemberian sanksi perlu dilakukan bagi siswa yang tidak taat pada tata tertib sekolah, tujuannya adalah untuk membentuk disiplin dan budi pekerti siswa. SMK Gazza Wiguna 1 Megamendung kab. Bogor pun memberlakukan sanksi bagi siswa yang melakukan pelanggaran terhadap tata tertib. Permasalahannya adalah penetapan point pelanggaran masih belum dilakukan secara objektif. Untuk mengatasi hal tersebut maka penerapan metode Simple Multi Attribute Rating Technique Method (SMART) pada system penunjang keputusan untuk menentukan rekomendasi sanksi pada siswa yang melakukan pelanggaran. System ini dirancang menggunan tool UML, dan system dibangun berbasis web. System ini menggunakan alternatif peringatan lisan, surat peringatan pertama, surat peringatan kedua, surat peringatan ketiga, dan dikembalikan ke orang tua (DO), dan kriteria terdiri dari pelanggaran seragam sekolah, kerapihan, disiplin, kebersihan dan ketertiban, tata kerama, peringatan hari besar nasional, dan ketentuan tambahan. Jenis kriteria yang digunakan adalah cost kriteria, karena diharapkan siswa tidak melakukan pelanggaran terhadap tata tertib sekolah, bila siswa melakukan pelanggaran, diharapakan siswa tidak melakukan pelanggaran yang berat.*

Kata Kunci: : sistem penunjang keputusan; sanksi; SMART method; system informasi.

INTRODUCTION

Habituation of positive attitudes and behaviours in schools can be done through Ethics Growth activities such as 1) fostering the development of moral and spiritual values, 2) promoting the development of national importance and diversity, 3) developing positive interactions between students and teachers and parents, 4) taking care of themselves and the school environment, 5) developing the full potential of students, 6) involving parents and the community in schools [1]. For Ethics Growth habituation, schools need to enforce discipline to train students in self-control to be responsible and not affected by negative things [2][3]. The application of discipline in schools, especially for students, really needs to be taken seriously. The application of penalty is independent of punishment as a control for students who do not obey the established school rules. Discipline aims to educate students not to make the same mistakes so that the character of students with good ethics is formed [4]–[6].

SMK Gazza Wiguna 1 Megamendung Bogor regency is a private school that applies discipline and punishment in shaping the ethical character of

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

its students. Referring to the Decree of Principal Number: 421.7/026/SMKGW1/VII/2020 concerning Rules and Regulations for the Social Life of Schools of Students in SMK Gazza Wiguna 1 Megamendung regarding Violations and Sanctions, every violation committed by students will be given sanctions in the form of oral warnings, first, second, third written warnings, and dropouts known by parents [7]–[9].

The problem faced by picket teachers determines the point and punishment of violations in a feeling only so that there is a less objective determination of penalties. Besides, student violation data is recorded manually in the violation book. As a result, student coaching is challenging. Finding information is difficult because the data is recorded manually. Regarding this problem, SMK Gazza Wiguna 1 Megamendung wants a decision-supporting system to help school management manage student breach data to decide the suitable student penalty and the right coaching. This study aims to apply the SMART (Simple Multi-Attribute Rating Technique) method to the decision support system. The decision support system can record violations, give offenders weight, and produce output through sanctions recommendations. These recommendations will be material for school management in establishing sanctions that will be given to students who commit violations.

The SMART method is multi-criteria decision-making. In the SMART process, each alternative consists of several criteria and has a weight value that describes its importance compared to other measures. Weighting assesses each alternative to obtain the best alternative [10]. Another study used the SMART method to determine toddlers' growth and skinning criteria [11]. SMART method Determining the quality of clothing materials following consumer demands, consumers give the value professionally [12]. The selection of satellite system vendors applying twelve criteria for implementing the SMART method can be used to improve group decision-making [13]. The SMART method is also used for e-marketplace selection using Business View, Market Service View, Third-party B2B E-marketplace performance, Transaction View, and Infrastructure e View tools. Assessment is carried out by giving the value of the weight of the criteria quantitatively [14]. The SMART method is implemented in determining the level of drug addiction, with the criteria The frequency of Drug Use, History (Old) of Drug Use, Amount of Drug Use, Alcohol Screening Value, and Narcotics Screening Score. Alternative hatching using a qualitative manner is with light, medium, and heavy values [15].

Referring to literature that has applied the SMART method in solving multi-criteria problems, the problem of sanctioning students with multi-criteria can be solved by the application of the SMART method and to facilitate. The SMART process can be integrated into the decision support system.

MATERIALS AND METHODS

The research methodology uses the Waterfall method, as shown in figure 1. Calculation analysis using the SMART (Simple Multi-Attribute Rating Technique) method, shown in figure 2.

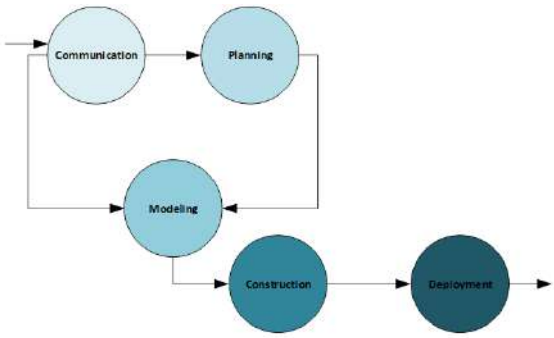


Figure 1. Research method [16]

Communication

Communication is carried out to understand the system that runs in the application of sanctions. Communication is carried out through interviews and observations of stakeholders, namely the principal, vice principal for student affairs, counselling teachers, picket officer teachers, and students. The need for data to analyze and build a system is also communicated at this stage.

Planning

At this stage, planning is carried out using methods to solve the problem of calculating the value of violations and planning the construction of the system. At this stage, SMK Gazza Wiguna 1, as the owner of the system, conducts elicitation, elaboration, negotiation, specifications, and validation, of the use of problem-solving models and strategies to be built

Modeling

At the modeling stage, a test of the application of the SMART method is carried out. The analysis step with the SMART process is shown in figure 2. The decision support system (DSS) design is also carried out at this stage. The system design method uses an object-oriented design with the Unified Modelling Language (UML) design tool.

Construction

DSS application development and application tests are carried out in the construction

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4138

stage. Applications are built with PHP programming language, database server MySQL, web service using Xampp v.3.3.0, and web browser using google chrome.

Deployment

At this stage, system installation is carried out at SMK Gazza Wiguna 1. Users test the system, delivery, support, and feedback.

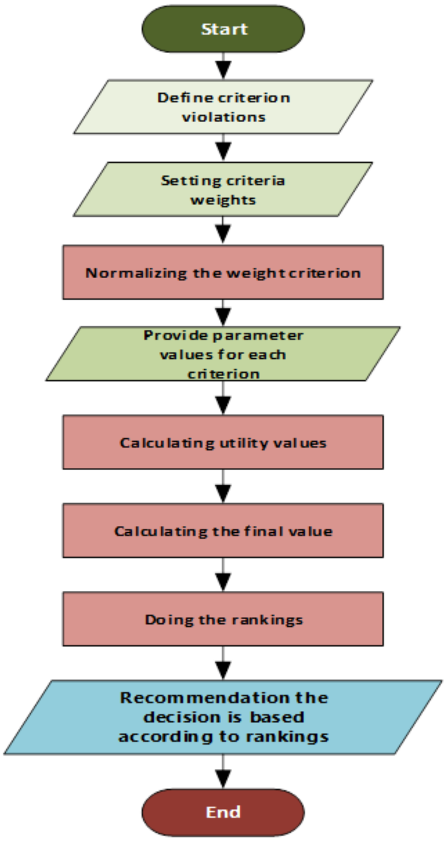


Figure 2. Stages of analysis with the SMART method

RESULTS AND DISCUSSION

This research was conducted based on the waterfall research method, as shown in figure 1. The data used in this study are 1) data on the regulation of the principal's decree number 421.7/026/SMKGW1/VII/2020, 2) picket teacher data, 3) student data, and 4) breach case book data. Then the alternative sanctions given to students who commit violations are shown in table 1.

Table 1. Data of Alternative

No	Penalty code	Penalty name	Range Point
1	PL	Verbal warnings	10-24
2	SP1	First warning letter	25-50
3	SP2	Second warning letter	51-80
4	SP3	Third warning letter	81-90
5	DO	Drop out	91->=100

The SMART method can be applied to obtain penalty recommendations. The steps performed refer to figure 2.

Step 1. Define criterion violations

Then the criteria for violation are shown in table 2.

Table 2. Data of criterion

No	Criterion Code	Criterion Name
1	P1	School Uniform
2	P2	Tidiness
3	P3	Discipline
4	P4	Cleanliness and Order
5	P5	Manners
6	P6	National holidays and memorials
7	P7	Additional Term

Step 2. Setting criteria weights

The principal has determined the value of the weight of the criteria based on decree number: 421.7/026/SMKGW1/VII/2020. The weight values are shown in table 3.

Table 3. the weight of the criterion

No	Criterion Code	Criterion Name	Weight
1	P1	School Uniform	5
2	P2	Tidiness	5
3	P3	Discipline	15
4	P4	Cleanliness and Order	5
5	P5	Manners	10
6	P6	National holidays and memorials	10
7	P7	Additional Term	50
		Amount	100

Step 3. Normalizing the weight of the criterion

After determining the weight of the criteria, the next step is normalized by following the formula equation 1. The result normalizing the importance of the measure is shown in table 4.

$$W_i = \frac{w_i'}{\sum_{j=1}^{m} w_j} \quad \text{...} (1)$$

W_i = The weight of the criteria is normalized for the i-th criterion.
W_i' = The weight of the i-th criterion.
W_j = The weight of the j-th criterion.
j = 1,2,3,...,m number of criteria.
Example

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v8i2.4138

Table 4. Result normalizing the weight criterion

No	Criterion Code	Criterion Name	Weight	W'_i
1	P1	School Uniform	5	0.05
2	P2	Tidiness	5	0.05
3	P3	Discipline	15	0.15
4	P4	Cleanliness and Order	5	0.05
5	P5	Manners	10	0.1
6	P6	National holidays and memorials	10	0.1
7	P7	Additional Term	50	0.5

Step 4. Provide parameter values for each criterion. This research uses quantitative values. Grades are obtained through a picket teacher representative and recorded in the book of violations. The parameter values for each measure are shown in Table 5.

Table 5. The parameter values for each criterion

Alternative Code	Criterion Code						
	P1	P2	P3	P4	P5	P6	P7
PL	5	5	10	5	20	30	50
SP1	10	10	20	15	30	50	75
SP2	15	30	30	30	50	50	80
SP3	20	30	40	30	75	50	90
DO	25	30	80	40	80	50	100

Step 5. Calculating utility values.

Calculate utility values by converting the criteria values on each measure into standard data criteria values. The utility value uses cost criteria because the requirements are "more desirable small values" than calculating utility value by following the formula equation 2.

$$\text{U}_i(ai) = \frac{(C_{max} - C_{out})}{C_{max} - C_{min}} \text{..}(2)$$

$\text{U}_i(ai)$ = i-th criteria utility value for an i-th alternative.
C_{max}= maximum criterion value.
C_{min}= minimum criterion value.
C_{out}= i-th criterion value.
Example
$\text{U}_i(ai)$=$\frac{(25-5\ )}{(\ 25-25)}$=$\frac{20}{20}$ =1.

The result calculating the utility value is shown in figure 6.

Table 6. Utility Values

Alternative code	P1	P2	P3	P4	P5	P6	P7
PL	1	1	1	1	1	1	1
SP1	0.75	0.8	0.85	0.75	0.83	0	0.5
SP2	0.5	0	0.71	0.28	0.5	0	0.4
SP3	0.25	0	0.57	0.28	0.083	0	0.2
DO	0	0	0	0	0	1	0

After getting the utility value, calculate the final value using the formula equation 3.

$$\text{U}(a_i)\sum_{j=1}^{m} w_j * u_{j(ai)} \text{..}(3)$$

$\text{U}(a_i)$ = The total value for the i-th alternative.
w_j = The already normalized value of the j-th criterion weight.
$u_{j(ai)}$= The value of the j-th criterion utility for the i-th alternative.
Example
=(0.5*1)+(0.05*1)+(0.15*1)+(0.05*1)+(0.1*1)+(0.1*1)+(0.5*1)
=0.5+0.05+0.15+0.05+0.1+0.1+0.5
=1.
The result of the calculation of the final value is sorted from the largest to the smallest value; the alternative with the most significant absolute value indicates the best choice. The result of the calculation of the final value is shown in table 7.

Table 7. Ranking table

Ranking	Alternative	
	Code	Final score
1	PL	1
2	SP1	0.5755
3	SP2	0.3955
4	SP3	0.2203
5	DO	0.1

Table 7 shows that the sanction recommendation given is "Oral Warning (PL)" because the best advice is the alternative with the highest ranking value. Alternative suggestions are shown in figure 3.

Figure 3. Alternative recommendations

After analyzing the data, the next step is to design and build a decision support system. System design starts by examining the running system. Analysis of the running system is shown in figure 4.

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4138

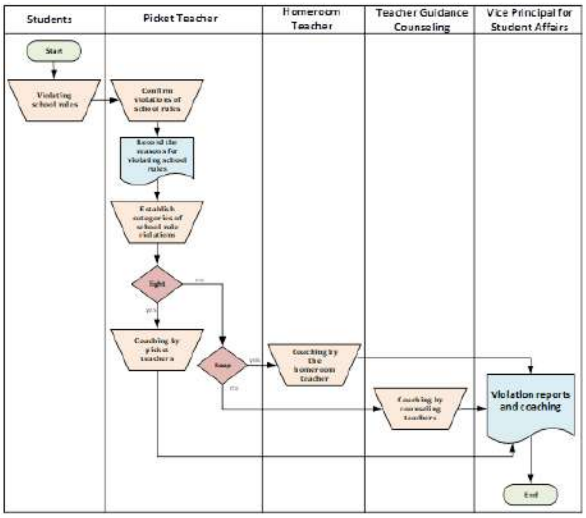


Figure 4. Existing process business

The picket teacher will record the data of the student who committed the violation and determine the point of a breach; if the end of a misdemeanour is, then the picket teacher will carry out the coaching. If the violation point is moderate, then the homeroom teacher will coach. Suppose the end of a breach is severe, then. After understanding the running process, design the system that will be developed. The system developed is a decision-supporting based system integrated with the SMART model to obtain recommendations that will be imposed on students who commit violations. The business process of the developed system is shown in figure 5.

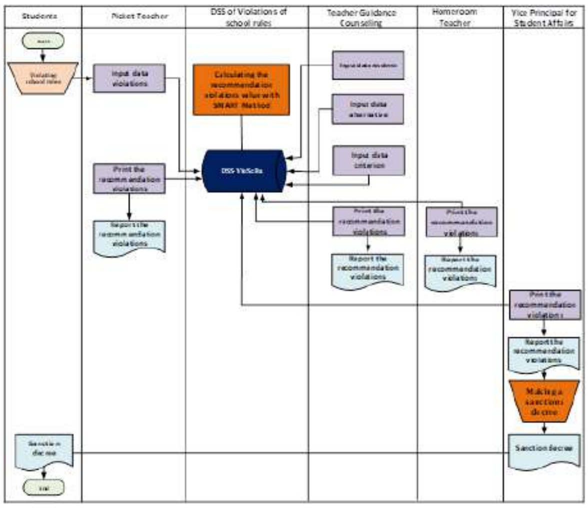


Figure 5. Developed business processes

After understanding the system model to be developed, the next step is to draw what processes actors can access, illustrated through the UML diagram. The use case diagram is shown in figure 6. There are six use cases on the system, and four actors can directly access the system.

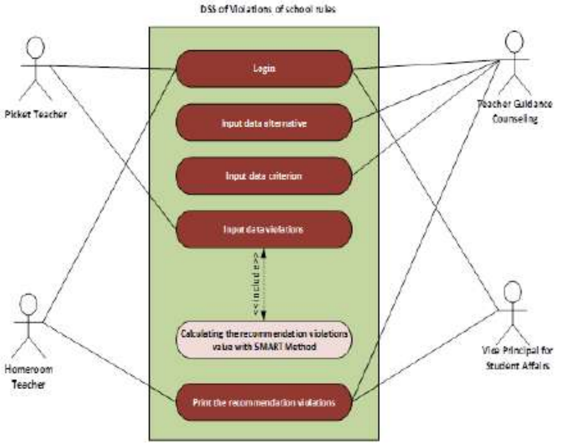


Figure 6. Use case diagram

After understanding the system model, the database needs are further designed through the UML diagram. The database is depicted through a class of graphs, shown in figure 7.

Figure 7. Class diagram

After the system design is completed, a decision support system application is built. Figure 8 shows the process of inputting data violations committed by picket teachers.

Figure 8. The process of inputting data violations

Figure 9 shows that the application of the SMART method was successfully implemented in the decision support system to inform the recommendations for appropriate sanctions for students who commit violations

Figure 9. Result of the recommendation penalty

CONCLUSION

This decision support system has succeeded in applying the SMART method. This system applies cost criteria because it is expected that students do not commit violations. This system uses four alternatives and seven groups of violation criteria. This system can shorten the discussion time between picket teachers, counselling teachers, vice class, and vice principals for student affairs in providing sanctions. The decision support system can record student violations, making it easier for stakeholders to decide the suitable sanctions.

The decision support system that has been built has not been able to provide treatment recommendations to students because this module has not yet been developed. The head decree on a

CC BY NC

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

penalty given to the student cannot be printed automatically because the decree's determination is still done manually after considering the results of the system recommendations.

REFERENCE

[1] Menteri Pendidikan Dan Kebudayaan Republik Indonesia, *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2015 Tentang Penumbuhan Budi Pekerti*. Jakarta, Indonesia, 2015, pp. 1–8.

[2] A. Novitasari, "Pemberian Reward and Punishment dalam Membentuk Karakter Disiplin Anak pada Sekolah Madrasah Ibtidaiyah," *Halaqa Islam. Educ. J.*, vol. 3, no. 1, pp. 27–33, Jun. 2019.

[3] N. Hernawati, "Pengaruh Penerapan Sanksi Berjenjang Terhadap Kedisiplinan Siswa Di Sdn Mekarwangi I Kecamatan Cihurip Kabupaten Garut," *J. Pendidik. UNIGA*, vol. 2, no. 1, pp. 10–16, 2017.

[4] Ridwan, "Penerapan Hukuman Untuk Meningkatkan Kedisplinan Siswa Di Smk Negeri 1 Pangkalan Bun Kab. Kotawaringin Barat," *Angew. Chemie Int. Ed. 6(11), 951–952.*, vol. 4, no. April, pp. 5–24, 2018.

[5] K. Rinaldi, "Penerapan Sanksi Terhadap Siswa/Siswi yang Melakukan Pelanggaran di Luar Sekolah," *JURPIKAT (Jurnal Pengabdi. Kpd. Masyarakat)*, vol. 3, no. 1, pp. 84–94, Mar. 2022.

[6] R. Choirunnisa, A. Yusuf, and V. Ameliasari, "Upaya Meningkatkan Karakter Peserta Didik Melalui Kegiatan Ekstrakurikuler Kepramukaan," *J. Pendidik. Dasar J. Tunas Nusant.*, vol. 4, no. 1, pp. 458–463, 2022.

[7] S. A. Cahyono, *Peraturan Dan Tata Tertib Kehidupan Sosial Sekolah Peserta Didik SMK Gazza Wiguna 1 Megamendung*. Bogor, Indonesia, 2020, pp. 1–9.

[8] R. A. S. Prayoga and S. W. Putri, "Sistem Pendukung Keputusan Memilih Pemasok Kopi Dengan Metode Smart Pada Coffee Shop Abc Ponorogo," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 7, no. 1, pp. 69–78, 2022.

[9] M. N. Amalia and M. Ary, "Sistem Pendukung Keputusan Pemilihan Supplier Dengan Menggunakan SMART Pada CV. Hamuas Mandiri," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 127–134, 2021.

[10] R. S. Hayati, R. N. Sari, Hardianto, E. Syahrin, F. P. Nasution, and Daifiria, "The Implementation of Smart Method for The Selection Of Concentration Computer Engineering Courses At The University Potensi Utama," in *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, 2019, pp. 1–5.

[11] A. Wahana, C. N. Alam, and S. N. Rohmah, "Implementation of the Simple Multi-Attribute Rating Technique Method (SMART) in Determining Toddler Growth," *J. Online Inform.*, vol. 5, no. 2, p. 169, Dec. 2020.

[12] N. Marpaung, A. Nata, and R. Yesputra, "Pemilihan Kain Berkualitas Dengan Metode Simple Multi Attribute Rating Technique Pada Sistem Pendukung Keputusan," *J. Sci. Soc. Res.*, vol. 5, no. 1, pp. 11–17, Feb. 2022.

[13] M. Sadly, Agustan, S. Yulianto, O. B. Bintoro, D. Sutrisno, and F. Alhasanah, "An Application of SMART Method in vendor selection of Satellite Systems Case study of Indonesia Remote Sensing Satellite Systems (InaRSSat)," in *2018 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES)*, 2018, vol. 5, pp. 1–6.

[14] L. Marlinda, Y. Durachman, E. Zuraidah, Tukino, Widivawati, and S. Hanadwiputra, "Selection of E-Marketplaces Factors Affecting Indonesian Women's Business Loyalty Using Simple Multi-Attribute Rating Technique (SMART) Method," in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, 2020, pp. 1–5.

[15] A. Tejawati, H. S. Pakpahan, and W. Susantini, "Drugs Diagnose Level using Simple Multi-Attribute Rating Technique (SMART)," in *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2018, no. Double L, pp. 357–362.

[16] R. S. Pressman and B. R. Maxim, *Software Engineering a Practitioner's Approach*, Eighth. New York: McGraw-Hill Education, 2015.

Accredited Rank 2 (Sinta 2) base on the Decree of the Dirjen Pendidikan Tinggi, Riset, dan Teknologi No. 225/E/KPT/2022, December 07, 2022. Published LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4086

COMPARISON OF SIFT AND ORB METHODS IN IDENTIFYING THE FACE OF BUDDHA STATUE

Linda Marlinda[1,2*] , Fikri Budiman[3], Ruri Suko Basuki[4], Ahmad Zainul Fanani[5]

Doctoral Program Of Computer Science Faculty[1,3,4,5]
Universitas Dian Nuswantoro, Semarang,
dinus.ac.id
marlinda@mhs.dinus.ac.id[1*], fikri.budiman@dsn.dinus.ac.id[3], ruri@dsn.dinus.ac.id[4],
a.zainul.fanani@dsn.dinus.ac.id[5]

Informatic Program of Faculty Of Information Technology[2]
Universitas Nusa Mandiri
https://nusamandiri.ac.id/
linda.ldm@nusamandiri.ac.id[2]

(*)Corresponding Author

Abstract—*The statue is part of the heritage facial recognition process which is immobile and artistically stylized. Identifying the similarities between the statues can help provide an important reference for tourism in recognizing the faces of the statues which are different and have almost the same characteristics in every country, especially in Indonesia, among the facial recognition of the statues based on the condition, color, and shape of the face. The purpose of this study is to apply the original images that have characteristics, partially done manually to various types of transformations and calculate matching evaluation parameters such as the number of key points in the image, the level of matching, and the required execution time for each algorithm. To confirm the efficiency of the proposed method, experiments were carried out on private data sets obtained from statues under low light conditions and in different poses. The data was taken based on the image of the Buddha's face and matched with the facial image of the Buddha statue available in the database using comparisons resulting from data processing using the Sift and ORB methods with various types of transformations. The result will be seen in the image that is matched with the best algorithm for each type of distortion. The faces tested are images of the faces of the Buddha statues that are recognized, and photos of some of the original statues that were not saved due to unclear lighting and camera distance factors. The results show that the number of key points generated is the number of key points, the ORB method gives fewer results compared to the SIFT method and the average SIFT recognition and processing time shows better performance for an average of 100% at a SIFT matching rate of 2% with time 0.400285 and the ORB method is 1% for the time 0.400961.*

Keywords: *images matching, face recognition, SIFT, ORB, statue, face buddha.*

Intisari— Arca merupakan bagian dari proses pengenalan wajah warisan yag tak bergerak serta bergaya artistik. Dalam mengidentifikasi kesamaan antara arca dapat membantu memberikan referensi penting untuk pariwisata dalam pengenalan wajah arca yang berbeda dan memiliki ciri khas yang hampir sama disetiap negara khususnya di Indonesia, diantara pengenalan wajah arca berdasarkan kondisi, warna, serta bentuk wajah. Tujuan penelitian ini untuk menerapkan gambar asli yang memiliki ciri khas, sebagian dilakukan secara manual ke berbagai jenis transformasi dan menghitung parameter evaluasi pencocokan seperti jumlah titik kunci dalam gambar, tingkat pencocokan, dan waktu eksekusi yang diperlukan untuk setiap algoritma. Untuk menguatkan efisiensi metode yang diusulkan, percobaan dilakukan pada kumpulan data private yang diperoleh dari arca dengan kondisi pencahayaan rendah dan pose berbeda. Data diambil berdasarkan citra wajah buddha dicocokkan dengan citra wajah arca budha yang tersedia dalam basis data menggunakan perbandingan yang dihasilkan dari pengolahan data dengan metode sift dan ORB dengan berbagai jenis transformasi. Hasilnya akan terlihat pada gambar yang dicocokan dengan algoritma terbaik untuk setiap jenis distorsi. wajah yang diuji adalah citra wajah arca buddha yang dikenali, dan foto beberapa arca asli yang tidak disimpan karena faktor pencahayaan yang kurang jelas dan faktor jarak kamera. Hasil menujukkan banyaknya key point yang dihasilkan jumlah poin kunci, metode ORB memberikan hasil yang lebih sedikit dibandingkan dengan metode SIFT dan rata-rata waktu pengenalan dan pemrosesan SIFT

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

menunjukkan kinerja yang lebih baik untuk rata rata 100 % pada tingkat pencocokan SIFT sebesar 2 % dengan waktu 0,400285 dan metode ORB sebesar 1 % untuk waktu 0,400961.

***Kata Kunci**: pencocokan gambar, pengenalan wajah, SIFT, ORB, arca, wajah buddha.*

INTRODUCTION

Technology assisting the face is currently growing, where aid research is still a big theme and involves many disciplines. Several face-matching applications will continue to develop and mastery of this technology will be indispensable for identifying the faces of intact or damaged statues in museums or Indonesian temples or various countries[1][2].

At present, the relief of Buddha statues with similar artistic models is still lacking in empirical assistance, and scientific quantitative relief methods are still lacking[3]. This research provides quantitative evidence for repairing virtual and physical statues using pictures or photographs, so that people can quickly judge the similarities between different Buddha statues, accurately find common feature points, and assist experts in repairing Buddha statues. Some examples of these Buddha statues are considered very similar and can be chosen as a reference for other Buddha statues[4][1][5].

Currently, facial recognition of statues or artifacts relies on traditional workers to repair or recognize statues due to the lack of available science and technology. In recent years, the protection of the temple has focused on keeping the building in good condition, but a quantitative assessment of the similarity of the Buddha statues inside the temple stupas has not received much attention from researchers[6][7]. The best evidence for cultural heritage is ideally original documentation, such as survey data and photographs. However, these documents have been largely lost to time. cultural heritage, especially temples, and statues, of the same period tend to have a uniform style. Therefore, searching for similar objects can provide an economical and scientific way to recover damaged relics[8][9].

This identification is necessary in order to distinguish the characteristics of the Buddha statues which have a high degree of facial resemblance in each country[10][7]. The process of numbering points is used to express the similarity index of matching feature points between two similar Buddha statues[11]. It is therefore necessary to identify and match the feature points of similar Buddha statues[12][13].

SIFT (Scale-Invariant Feature Transform) is a feature recognition technique that is often used for facial recognition. In this method, facial images are analyzed to look for unique features such as edges, points, and paths that distinguish one face from another[14]. Once these features are discovered, they can be used to compare new faces with a database of known faces to estimate the identity of those faces. SIFT has the ability to handle scaling and rotational changes in facial images, making it suitable for facial recognition under different conditions[8][15][16].
The SIFT method on facial images consists of several stages[17][18]:

a. Key Point Detection
   The SIFT algorithm uses a key point detection technique to find unique feature points in facial images.
b. Keypoint Descriptor
   Once the key points are found, SIFT generates a feature description of each key point which involves measuring the distribution of pixel intensity around that point.
c. Matching
   The feature description of the face image is compared with the feature description of the reference image to determine the degree of similarity.
d. Verification
   After the matching process is complete, the results are confirmed through a face verification algorithm to ensure that the recognized face matches the owner's face.
e. Identification
   If the verification is successful, the facial image is recognized and associated with the corresponding owner's identity.

SIFT has the ability to handle orientation and scale changes in facial images, making it an effective algorithm for face recognition[17].

The Scale-Invariant Feature Transform (SIFT) algorithm is a method that is widely used in face recognition. Experts consider SIFT to be an effective method because of its ability to handle differences in scale and orientation of facial images, as well as having a high degree of accuracy in recognizing faces[18]. However, SIFT has drawbacks in terms of execution time which is quite long and requires a lot of computational resources. Therefore, some experts are also looking for

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4086

alternative methods that are more efficient in terms of time and resources[19].

Researchers have opinions on improvements to the SIFT algorithm for facial recognition. These improvements include adding steps to the process of selecting facial features, thereby increasing accuracy in facial recognition. He also pointed out that the improvements yielded better results than the original SIFT method when applied to multiple facial datasets. Therefore, it can be said that SIFT improvements are very helpful in increasing facial recognition accuracy[20].

The SIFT algorithm is still effectively used in facial recognition even though the facial image has been manipulated. He proposed adding a validation step to the face recognition process with SIFT so that it can filter manipulated images and improve facial recognition accuracy. Therefore, the SIFT method is very helpful in facial recognition even though the image has been manipulated[21].

ORB (Oriented FAST and Rotated BRIEF) is a feature recognition technology that is often used in face recognition. In this method, facial images are analyzed to look for unique features such as edges, points, and paths that distinguish one face from another[22]. Once these features are discovered, they can be used to compare new faces with a database of known faces to estimate the identity of those faces. ORB has advantages in terms of research speed and efficiency compared to other feature recognition techniques such as SIFT. This makes ORB suitable for facial recognition applications on low-speed devices. ORB combines FAST key point detection techniques with BRIEF feature descriptions to create a unique representation of each feature in an image[23].

In general, ORB follows the same steps as SIFT in facial recognition, including key point detection, feature description, matching, verification, and identification. However, ORB uses different techniques for each of these stages[24].

Use of the ORB-PCA feature extraction technique for facial recognition. ORB-PCA is a combination of ORB and PCA (Principal Component Analysis) algorithms used to extract facial features. This technique has a faster execution time and requires fewer computational resources compared to the SIFT method. In addition, ORB-PCA also has fairly good facial recognition accuracy. Therefore, it can be said that the ORB-PCA technique can help in increasing the efficiency and accuracy of face recognition[22][25].

Use of LBP (Local Binary Pattern) and ORB features for facial expression recognition. He showed that the combination of LBP and ORB features has good accuracy in recognizing facial expressions. In addition, this method also has a fast execution time and requires little computational resources. Therefore, it can be said that the combination of LBP and ORB features can help in increasing the accuracy and efficiency of facial expression recognition[26][27].

A comparison of the SIFT and ORB methods is needed to provide information on which method can provide a level of matching of feature points, as well as the level of matching speed, accuracy, and robustness that must be considered comprehensively for algorithm selection. SIFT operator and ORB operator have high accuracy, and the results are relatively stable[24][22]. SIFT is more accurate than Surf but runs slower. Brisk operators and Orb operators are relatively fast, but their accuracy is poor. Again, given the characteristics of small cave sculptures, accuracy is the more important factor; Therefore, the SIFT operator was chosen as the feature point matching algorithm in this study.

MATERIALS AND METHODS

Taking images of Buddha statues from public data (pinters205facebuddha), with stacked Buddha images which aim to get the level of compatibility of the images in the identification process of the buddha's face using the Sift and ORB methods, with the following process:

a. The initial stage is the formation of datasets. The dataset comes from public data or private data that form the features of the Buddha statue, by preprocessing the data. Camera-based shooting system (2D) uses local features. For local features, key points are extracted to select the part of the image to be retained for the description part. Local key point detectors are used to detect regions of interest that are invariant to the transformation class (eg scaling, rotation, and translation) for each detected region.

b. Identification Stage. At the stage of dataset formation, there are six processes, namely preprocessing, segmentation of arcuate images, rescaling, feature extraction of points, feature extraction of distances between furcation points, and formation of feature vectors as datasets. In the second stage the identification stage, the image of the statue is processed as in the first stage to obtain the feature vector of the statue pattern. Then carried out the process of identification through the process of training and testing. Performance measurement results

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

of system identification are done by calculating the value of the confusion matrix.

c. The comparison stage uses two methods, SIFT and ORB.

RESULTS AND DISCUSSION

The research data obtained from the comparison method between the SIFT and ORB methods are as follows:

Image Acquisition

The images used in this study are red, green, and blue (RGB) images taken indoors. Relatively low light intensity. If the light intensity is low, the object will appear non-existent. Figure 1 shows the original image of buddha's face.

Figure 1. Original Image

The image is a representation of the figure 1, and the image is produced from the output of shooting in the form of a photo, the image itself has a lot of information in which there are many sets of pixels, and in image data, there is an influence on reducing image quality for image defects.

a. Elements in a Picture Image: Brightness, Contrast, Outline, and Color.
b. Pixels are the smallest part of an image, each pixel has a different value
c. Resolution is the level of image detail, high resolution is a good determinant of image quality.
d. Convolution Image manipulation process
e. Gaussian: Serves to smooth the image, in the Gaussian smoothing process it uses the normal distribution.

Preprocessing

The result of the pre-processing stage is an image that has a smaller pixel size than the original image and is reduced in size to 1/3 of the actual size. In addition, the processed image has a simple color space obtained from the gray scaling process. Figure 1 shows the preprocessing results of the original image.

Feature Extraction

This step is carried out to obtain the characteristics of each grayscale image method. The features extracted at this stage for both methods, SIFT and ORB, are in the form of numeric form with the term key point. Each key point value represents a characteristic according to the characteristics of the extracted object.

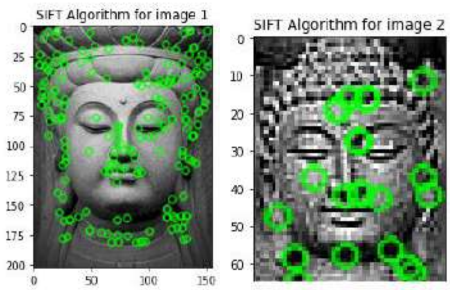


Figure 2. Detector, in this case, using SIFT

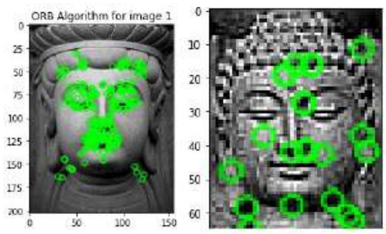


Figure 3. Detector, in this case using ORB

After using Brute Force, the results are shown below. The process of matching two images to mark crucial points uses FLANN Matches as a comparison of the two images

Feature Matching and Testing

This step was carried out to test the sensitivity of the SIFT and ORB methods to distortion effects such as rotation, scaling, and cropping. Tests are carried out by comparing pre-processed images with different images that have experienced various types of distortion

Rotation

The results of tests performed on distorted images with a 90-degree rotation are presented in Figure 3

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4086

and Table 1. Figure 2 shows the results of feature matching using the SIFT method. On the other hand, Figure 3 shows the results of feature matching using the ORB method.

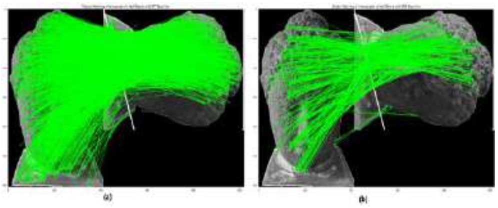


Figure 4 The results of feature matching at 90 degrees rotation

Table 1. Details of the test at an angle of 90 degrees rotation

Method	Key Point 1 (original image)	Key Point 2 (90 degrees)	Matches	Average Match Rate	Time (sec)
SIFT	1738	1735	1616	100%	0.557891
ORB	500	500	422	100%	0.616017

From the test results in Table 1, the results show the number of key points generated by each method. In terms of the number of key points, the ORB method gives less results than the SIFT method and the average SIFT recognition and processing time shows better performance.

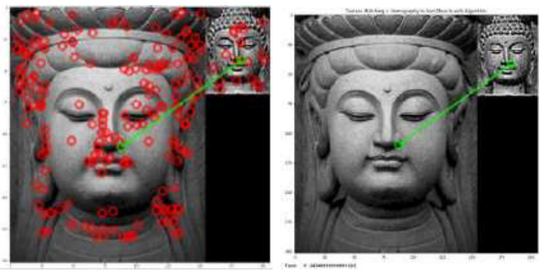


Figure 4 The results of feature matching at Flip Horizontal

Table 2. Details of the test at an angle of flip-horizontal

Method	Key Point 1 (original image)	Key Point 2 (90 degrees)	Matches	Average Match Rate	Time (sec)
SIFT	1738	1731	12	2%	0.400285
ORB	500	495	2	1%	0.400961

However, when the horizontal flip is performed, the two compact methods result in fast computation time. Unfortunately, both of them experienced a decrease in keypoint detection, as shown in Figure 4 and Table 2 where the SIFT method only has 7 key points and the ORB method is even smaller, namely 2 key points.

CONCLUSION

ORB differs from SIFT in that it uses a simpler and faster description of BRIEF, thereby making it more efficient in terms of computation time. ORB also has the ability to handle orientation and scale changes in facial images, making it an effective algorithm for face recognition. The results show that the number of key points generated is the number of key points, the ORB method gives fewer results compared to the SIFT method and the average SIFT recognition and processing time shows better performance for an average of 100% at a SIFT matching rate of 2% with time 0.400285 and the ORB method is 1% for the time 0.400961.

REFERENCE

[1] J. Ma, X. Jiang, J. Jiang, and Y. Gao, "Feature-guided Gaussian mixture model for image matching," *Pattern Recognit.*, vol. 92, pp. 231–245, 2019, doi: 10.1016/j.patcog.2019.04.001.

[2] J. Han, *China's Architecture in a Globalizing World: Between Socialism and the Market*. library.oapen.org, 2018.

[3] F. Di Paola, G. Milazzo, and F. Spatafora, "Computer aided restoration tools to assist the conservation of an ancient sculpture, The colossal statue of zeus enthroned," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 2W5, pp. 177–184, 2017, doi: 10.5194/isprs-archives-XLII-2-W5-177-2017.

[4] "Statue Face Recognition Using the Sift and Suft Method Based on Figure Extraction." .

[5] H. Wang, Z. He, Y. He, D. Chen, and Y. Huang, "Average-face-based virtual inpainting for severely damaged statues of Dazu Rock Carvings," *Journal of Cultural Heritage*, vol. 36. pp. 40–50, 2019, doi: 10.1016/j.culher.2018.08.007.

[6] S. Temple, *Relief and Statue*. .

[7] N. A. Rasheed and M. J. Nordin, "Classification and reconstruction algorithms for the archaeological fragments," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 8, pp. 883–894, 2020, doi: 10.1016/j.jksuci.2018.09.019.

[8] A. Y. Rahman, S. Sumpeno, and M. H.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

Purnomo, “Arca detection and matching using Scale Invariant Feature Transform (SIFT) method of stereo camera,” *Proc. - 2017 Int. Conf. Soft Comput. Intell. Syst. Inf. Technol. Build. Intell. Through IOT Big Data, ICSIIT 2017*, vol. 2018-Janua, pp. 66–71, 2017, doi: 10.1109/ICSIIT.2017.45.

[9] H. Wang, “SPCDet : Enhancing Object Detection with Combined Feature Fusing,” no. 2017, pp. 236–251, 2019.

[10] D. Papp, F. Mogyorósi, and G. Szucs, “Image matching for individual recognition with SIFT, RANSAC and MCL,” *CEUR Workshop Proc.*, vol. 1866, no. 1, 2017.

[11] L. A. Martín, “A Similarity Measure of Gaussian Process Predictive Distributions,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12882. pp. 150–159, 2021, doi: 10.1007/978-3-030-85713-4_15.

[12] B. Renoust *et al.*, “Buda.art: A multimodal content-based analysis and retrieval system for Buddha statues,” *MM 2019 - Proc. 27th ACM Int. Conf. Multimed.*, pp. 1062–1064, 2019, doi: 10.1145/3343031.3350591.

[13] F. Stanco, L. Tenze, G. Ramponi, and A. De Polo, “Virtual restoration of fragmented glass plate photographs,” *Proc. Mediterr. Electrotech. Conf. - MELECON*, vol. 1, no. May 2014, pp. 243–246, 2004, doi: 10.1109/melcon.2004.1346819.

[14] B. Majumder, “Face Recognition Based on Human Sketches Using Fuzzy Minimal Structure Oscillation in the SIFT Domain,” *Communications in Computer and Information Science*, vol. 1241. pp. 325–335, 2020, doi: 10.1007/978-981-15-6318-8_27.

[15] A. D. Narhare and G. V. Molke, “Trademark detection using SIFT features matching,” *Proc. - 1st Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2015*, pp. 684–688, 2015, doi: 10.1109/ICCUBEA.2015.140.

[16] W. Hua, M. Hou, S. Yang, and Y. Dong, “DISCRIMINATION of CULTURAL RELICS SIMILARITY BASED on PHASH ALGORITHM and SIFT OPERATOR,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, no. 2/W15. pp. 571–575, 2019, doi: 10.5194/isprs-archives-XLII-2-W15-571-2019.

[17] W. Hua, M. Hou, Y. Qiao, X. Zhao, S. Xu, and S. Li, “Similarity Index Based Approach for Identifying Similar Grotto Statues to Support Virtual Restoration,” *Remote Sens.*, vol. 13, no. 6, p. 1201, 2021, doi: 10.3390/rs13061201.

[18] G. Shi, X. Xu, and Y. Dai, “SIFT Feature Point Matching Based on Improved RANSAC Algorithm,” no. 1, pp. 1–4.

[19] B. Renoust *et al.*, “Historical and Modern Features for Buddha Statue Classification,” pp. 23–30, 2019.

[20] L. Marlinda, U. D. Nuswantoro, U. D. Nuswantoro, U. D. Nuswantoro, and U. D. Nuswantoro, “Matching Images On The Face Of A Buddha Statue Using The Scale Invariant Feature Transform (SIFT) Method,” pp. 164–167, 2020.

[21] K. Indra Gandhi, S. Janarthanan, R. Sathish, and A. Surendar, “Dominant Feature Prediction by Improved Structural Similarity Computation,” 2020, doi: 10.1109/ICITIIT49094.2020.9071534.

[22] E. Rublee, W. Garage, and M. Park, “ORB : an efficient alternative to SIFT or SURF,” pp. 2564–2571, 2011.

[23] F. Marton, M. B. Rodriguez, F. Bettio, M. Agus, A. J. Villanueva, and E. Gobbetti, “IsoCam: Interactive visual exploration of massive cultural heritage models on large projection setups,” *J. Comput. Cult. Herit.*, vol. 7, no. 2, Jun. 2014, doi: 10.1145/2611519.

[24] S. Rao, S. Shekhar, and A. Kum, “Fe eature Ex xtractionu using ORB B-RANSA AC for Fa ace Recog gnition,” vol. 70, pp. 174–184, 2015, doi: 10.1016/j.procs.2015.10.068.

[25] E. Karami, S. Prasad, and M. Shehata, “Image Matching Using SIFT , SURF , BRIEF and ORB : Performance Comparison for Distorted Images.”

[26] H. Chien, C. Chuang, C. Chen, and R. Klette, “When to Use What Feature ? SIFT , SURF , ORB , or A-KAZE Features for Monocular Visual Odometry,” no. 1, pp. 0–5.

[27] E. Karami, S. Prasad, and M. Shehata, “Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images,” *arXiv*, 2017, [Online]. Available: http://arxiv.org/abs/1710.02726.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 8. NO. 2 FEBRUARY 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v8i2.4086