Classification of High School History Questions Based on Cognitive Level Revised Bloom's Taxonomy Using K-Nearest Neighbor Method

Farah Sherina Salma
Fakultas Rekayasa Industri
Telkom University
Bandung, Indonesia
farahsherinasalma@student.telkomuniversity.ac.id

Oktariani Nurul Pratiwi
Fakultas Rekayasa Industri
Telkom University
Bandung, Indonesia
onurulp@telkomuniversity.ac.id

Riska Yanu Fa'rifah
Fakultas Rekayasa Industri
Telkom University
Bandung, Indonesia
riskayanu@telkomuniversity.ac.id

Abstract— Education plays an important role in transmitting knowledge to its students and to measure how well the student understands, testing is needed based on the cognitive level of knowledge. In measuring the cognitive level, it can be applied with reference to the Revised Bloom's Taxonomy which explains the regulation of learning processes and targets. Then by testing knowledge through the questions that have been made, it is necessary to classify the questions into several cognitive levels according to Revised Bloom's Taxonomy to determine the learning process and understanding of each individual. Many types of questions that are formed make classification difficult because the method is still done manually, therefore machine learning is needed. This study will focus on the classification of questions from the History subject in high school. The dataset used is downloaded from internet searches of USBN, PAS, PTS, and other exams. This study focuses on RBT C4 to C6 only. This study uses the K-Nearest Neighbor algorithm to obtain accuracy and with the imbalance of data in the dataset, an oversampling method using SMOTE will also be used. The accuracy results obtained are precision is 76%, recall is 76%, f1-score is 74%, accuracy is 76%.

Keywords—question classification, KNN, high school history, SMOTE, RBT, oversampling

I. INTRODUCTION

Education is the most important pillar in the development of a nation for its future. With the development of the times, knowledge also develops over time. School education is made with efforts to educate the nation by developing abilities and building a good personality [1]. Then the understanding of each individual will be tested with the aim of understanding the level of understanding of the student who has come to understand the teaching material they have received so far. Giving exams will usually be divided into practical tests that test how students can apply their knowledge directly by working and written exams that test the knowledge they receive.

The making of the exam must also be based on the provisions that have been given from the government. In Indonesia, based on the provisions of Government Regulation No. 19 of 2005 which discusses National Education Standards, it is arranged so that there are minimum criteria that can be used as benchmarks that must be applied by every educational institution throughout Indonesia [2]. In accordance with Law No. 20 of 2003; Article 35, Paragraph (1) which explains the content of the national standard, reads "The national standard of education consists of the standard of content. processes, graduate competencies, educational staff, facilities and infrastructure, management, financing, and educational assessment that must be improved on a planned and regular basis." [3].

In supporting the development of understanding of each individual, the number of questions that are made is not only from the exams that are held but also the number of questions that are made for mere practice during the teaching process. Therefore, classification is needed. The teacher must first think about the cognitive level from which students will be tested when they design a list of questions on the test questions [4]. In this problem, there are many ways to classify the question bank and the author will raise this topic based on Bloom's Taxonomy classification at cognitive level. In short, the understanding of Bloom's Taxonomy is a way of regulating the learning process and targets of students by classifying the cognitive level of learning targets [5]. The reason why Bloom's Taxonomy is the classification choice for this research is because it is a measurement tool according to the cognitive level, especially the structure of RBT which describes the representation of alignment between educational standards and objectives, objectives, products, and activities, and is still in accordance with de facto standards compared to classification theory and other developed hierarchical systems [6]. The number of question banks that keep popping up results in classification that takes a long time because the way it works is still done manually [7].

This research is a form of further research regarding the classification of questions based on cognitive level on RBT. The following are details of some previous studies results. A study [8] using the KNN and Chi-Square methods to test the combination of the KNN and Chi-Square methods in the preparation of Biology texts obtained a percentage accuracy rate of 79.36%. In research [7] which determines the effectiveness of performance in classifying questions based on the cognitive level of Bloom's Taxonomy with the Naïve Bayes and TF-IDF methods, the results are 85% precision and 80% recall. Then in research [9] which analyzed the suitability of student-made questions based on aspects of the PISA and RBT levels, the results showed that the research proved that student-made questions did not meet the PISA and RBT level conformity standards. The final result of this research is to get the level of accuracy and prediction to get the classification of questions from Indonesian History subjects at the high school level using the Euclidean distance approach KNN method. The KNN method with the Euclidean distance approach is considered to show the effectiveness of machine learning techniques, especially for domain classification of cognitive level skill levels and a collection of questions can be arranged covering all types of taxonomic levels and types [10]. This research was conducted to find the accuracy results of Euclidean distance approach of KNN method using the dataset of History subject.

II. RESEARCH METHODS

This study uses a systematic completion through several stages, namely initialization, data processing, then extracting the results, and conclusions. This process is broken down into 6 process parts. The initiation process consists of problem identification and literature study which is the beginning of determining the topic of writing, then along with determining the boundaries of the problem will determine the purpose of this research. The second process, namely data preparation, consists of collecting data, selecting data attributes, merging datasets, and labeling data. In this study, data labeling was done manually. The third process is data preprocessing which consists of case folding, tokenizing, filtering, and stemming on the dataset. The fourth process is feature extraction using the TF-IDF method. The fifth process is in the form of data classification and prediction which consists of implementing SMOTE, K-Fold Cross Validation, and calculating algorithm accuracy. K-Fold Cross Validation was applied twice to the dataset without and with SMOTE implementation. The last process is the calculation of the confusion matrix as well as the analysis and conclusions of the research results which are illustrated in Figure 1.

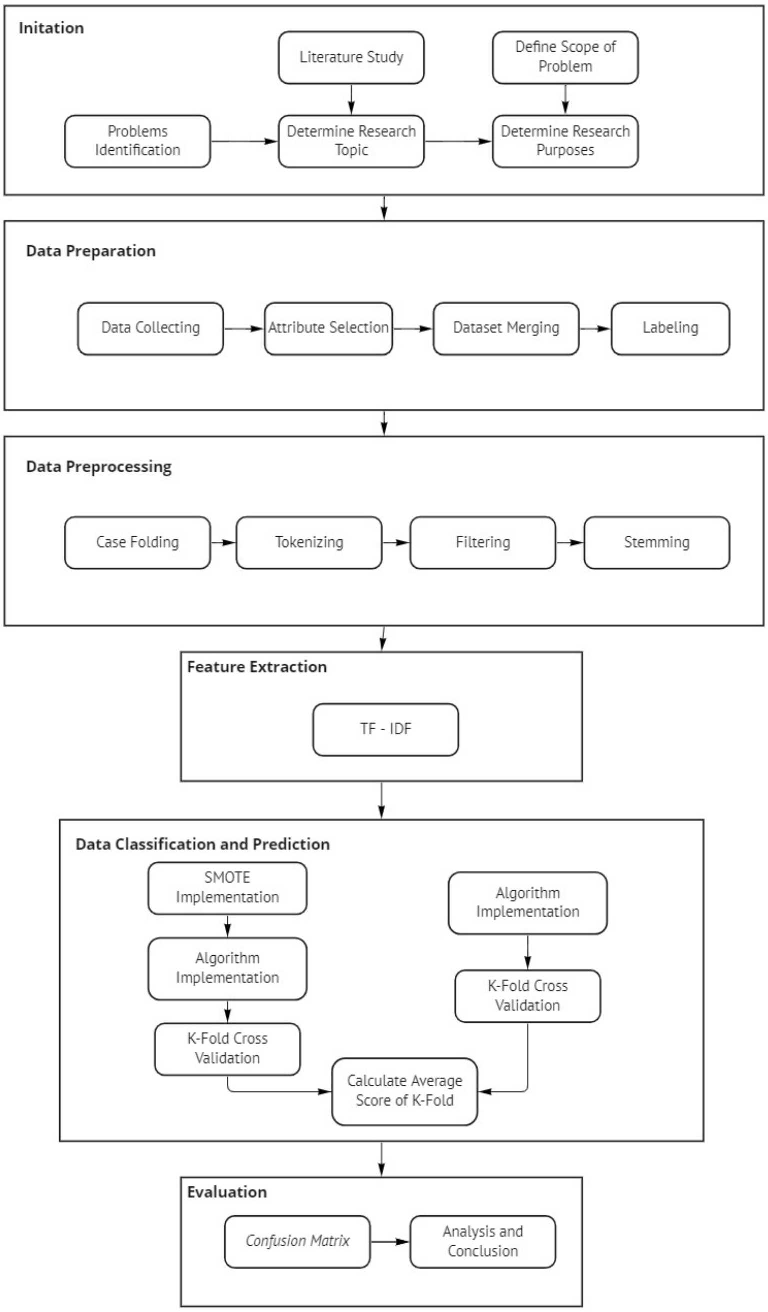

Figure 1. Problem Solving Systematic Diagram

A. Initiation

This research begins by identifying the problem. The researcher then determines the research topic along with conducting a study of the solution and thus determine the purpose of the study along with the boundaries of the problem.

B. Data Preparation

At this stage, practice questions collected from various website platforms are downloaded in PDF format. The questions are arranged in Indonesian and collected from question banks scattered on the internet. The subject of the question is history subject in high school. The form of the question consists of multiple choices and does not contain an equation symbol. The researcher then selects the attributes of the dataset until leaving one attribute in the form of the 'Question' column. After that the questions were combined into a single dataset which was then manually labeled based on Revised Bloom's Taxonomy.

C. Data Preprocessing

The stage after data preparation is data preprocessing which contains several dataset processing processes. The process is divided into several parts, including case folding, tokenizing, filtering, and stemming. This stage uses libraries in Python programming languages such as NLTK and Sastrawi.

D. Feature Extraction

At this stage, the word weighting process is carried out on the dataset using TF-IDF. The process that occurs is that each question occurrence will be counted (Term Frequency), and then the words will be weighted (Inverse Document Frequency). All words in the document will be weighted for classification calculations using a classifier.

E. Data Classification and Prediction

The initial stage of classification and prediction is to divide the dataset into two different types of experiments. The first experiment was the application of the SMOTE oversampling method, while the second experiment did not use any oversampling method in order to find out how imbalance data may effect the accuracy. After that both datasets were tested for classification using K-Fold Cross Validation with the KNN algorithm.

Basically the KNN algorithm works by finding the shortest distance in the training data which is used as an evaluation with the value of k closest neighbors [11]. A more detailed explanation is, first, the data used will be projected into a multidimensional space and then this space will be divided into sections based on the data classification groupings, where point c (space point) marks if it is a classification that is found in the k closest neighbors [12].

F. Evaluation

The last stage of this research is an evaluation by applying the Confusion Matrix to compare the results of the algorithm with or without the oversampling method. The classification accuracy test using the confusion matrix can be seen in Table 1 [13].

Table 1. Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

This stage will also calculate the level of accuracy using the following formula [14].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

III. RESULT AND DISCUSSION

A. Data Preparation

1) Labeling

The labeling stage is carried out after the data is collected into one column in an xlsx format file. The dataset is a collection of questions on Indonesian History subjects in high school such as practice questions, UH, PTS, PAS, and USBN which are downloaded and obtained for free from the question bank website. Then researcher manually sorts the types of questions only for multiple choice without pictures and bullet points before getting into labeling stage. The total number of questions gathered are 58 items. The labeling stage is done manually based on the table per cognitive level of RBT. The end result will be like the following example.

Table 2**.** Labeling Result

Questions	Label
Salah satu penyebab kacaunya kondisi perekonomian Indonesia pada masa awal kemerdekaan karena kas negara kosong. Upaya pemerintah Republik Indonesia mengisi kas negara yang kosong pada awal Kemerdekaan adalah	C4 (Analyzing)
Indonesia harus dapat mengatasi permasalahan ekonomi yang dihadapi pada masa awal kemerdekaan. Salah satu upaya bangsa Indonesia dalam melakukan perbaikan ekonomi pada awal kemerdekaan dilakukan dengan cara	C5 (Evaluating)
Proklamasi Kemerdekaan 17 Agustus 1945 dijelaskan dengan menguraikan berbagai aspek seperti sosial, ekonomi, dan politik dikenal dengan cara berfikir	C6 (Creating)

2) Data Preprocessing

Data preprocessing consists of several stages consisting of case folding, tokenizing, filtering, and stemming with the aim of changing the dataset that has been inputted into a word per word which is given a weighted value at the next stage using the TF-IDF method. The following table is an example of the results of preprocessing data.

Table 3. Data Preprocessing Result

Data	ASEAN merupakan organisasi regional Asia Tenggara yang dibentuk berdasarkan Deklarasi Bangkok tanggal 8 Agustus 1967. Dilihat dari segi politik organisasi tersebut bertujuan
Case Folding	asean merupakan organisasi regional

	asia tenggara yang dibentuk berdasarkan deklarasi bangkok tanggal 8 agustus 1967. dilihat dari segi politik organisasi tersebut bertujuan
Tokenizing	“asean”, “merupakan”, “organisasi”, “regional”, “asia”, “tenggara”, “yang”, “dibentuk”, “berdasarkan”, “deklarasi”, “bangkok”, “tanggal”, “8”, “agustus”, “1967”, “dilihat”, “dari”, “segi”, “politik”, “organisasi”, “tersebut”, “bertujuan”
Filtering	“asean”, “merupakan”, “organisasi”, “regional”, “asia”, “tenggara”, “yang”, “dibentuk”, “berdasarkan”, “deklarasi”, “bangkok”, “tanggal”, “8”, “agustus”, “1967”, “dilihat”, “dari”, “segi”, “politik”, “organisasi”, “tersebut”, “bertujuan”
Stemming	“asean”, “organisasi”, “regional”, “asia”, “tenggara”, “bentuk”, “dasar”, “deklarasi”, “bangkok”, “tanggal”, “8”, “agustus”, “1967”, “segi”, “politik”, “organisasi”, “tuju”

3) Feature Extraction

TF-IDF is the next process after completing the data preprocessing stage with the final result of the data in the form of word fragments that have no relationship between other words. TF-IDF works by converting text data in the form of a string to numeric so that the algorithm can process it at a later stage. The resulting output can be seen for example in the table below.

Table 4. Feature Extraction Result

Stemming	TF-IDF
“asean”, “organisasi”, “regional”, “asia”, “tenggara”, “bentuk”, “dasar”, “deklarasi”, “bangkok”, “tanggal”, “8”, “agustus”, “1967”, “segi”, “politik”, “organisasi”, “tuju”	“asean:0.27687398379081984”, “organisasi:0.4138933156618686”, “regional:0.27687398379081984”, “asia:0.23275475325503828”, “tenggara:0.27687398379081984”, “bentuk:0.20694665783093433”, “dasar:0.21855153342075778”, “deklarasi:0.27687398379081984, “bangkok:0.27687398379081984”, “tanggal:0.18113856240683032”, “agustus:0.18113856240683032”, “1967:0.27687398379081984”, “segi:0.27687398379081984”, “politik: 0.20694665783093433”, “tuju: 0.1971348753245809

B. Data Classification, Prediction, and Evaluation

1) SMOTE Implementation

The next research stage is to implement the SMOTE method. This is due to the discovery of a large data imbalance in the dataset of high school history questions between C4, C5, and C6. One of these oversampling methods works by establishing a replication known as data synthesis from minority data [15]. In the question dataset, it can be seen that there are far differences in the number of each label in the table below.

Table 5.
Total Number of Questions Before SMOTE

Label	Total Number of Questions
C4	37
C5	15
C6	6
Total	58

From table 5, it can be seen that the highest composition is found on label C4 which has a total of 37 questions and has a gap so far from label C6 which only has 6 questions. In this research, it can be seen that SMOTE increases the total number of questions of C5 and C6 following the one that has the highest total number of questions which is label C4. The next result after applying the SMOTE method will be like the following table 6.

Table 6.
Total Number of Questions After SMOTE

Label	Total Number of Questions
C4	37
C5	37
C6	37
Total	111

As seen in table 6, SMOTE increases the number of data on minority data. In the dataset used, the minority class is on the labels C5 and C6 so that when SMOTE is implemented, the amount of data will be added to a total of 111 items, while before the implementation of the SMOTE method there were only 58 questions. The total data on each label is 37 items.

2) Implementation of K-Fold Cross Validation and KNN Algorithm

The next stage is implementing K-Fold Cross Validation with KNN algorithm on two datasets, namely the output of SMOTE implementation and without SMOTE implementation. The function of K-Fold is to avoid overlapping in the testing data [16], [17]. The K-Fold used in this study is the value of k = 10. The 10-fold model in general attempts to produce an estimate of the error rate from tests that do not experience high bias or variance [18]–[20].

Table 7. K-Fold Without SMOTE Result

k-fold	1	2	3	4	5	6	7	8	9	10
Accuracy	0,67	0,17	0,83	0,5	0,67	0,67	0,33	0,67	0,6	0,8

Table 8. K-Fold With SMOTE Result

k-fold	1	2	3	4	5	6	7	8	9	10
Accuracy	1,0	0,82	0,45	0,73	0,91	0,55	0,73	0,91	0,73	0,73

It can be seen from table 7 and table 8 that there are quite far differences from the folding results that have been implemented with SMOTE compared to the results without SMOTE. The average accuracy result without SMOTE is 59% while the accuracy with SMOTE is 75%. Thus, information can be obtained that the use of SMOTE can produce higher accuracy.

3) *Performance Model Measurement*

After going through the process of testing the algorithm using K-Fold Cross Validation, it is necessary to evaluate the performance with the aim of measuring the performance of the KNN classifier model. In the information test, both the results of using SMOTE and without using SMOTE will be tested again. This is done to see the difference in the percentage results produced at the end of the test. The results before the application of balancing can be seen in table 9 below.

Table 9. Question Prediction Before SMOTE

		Predicted			
		C4	C5	C6	All
Actual	C4	32	4	1	37
	C5	13	2	0	15
	C6	6	0	0	6
	All	51	6	1	58

From the observations in the table 9 above, with a total of 58 questions, information was obtained that the correctness of the C4 label according to the prediction was 32 questions. For the prediction of the label C5, it is obtained that there are 2. Meanwhile, for C6, it is not predicted to have questions. It is known from the table, the comparison with the predictions made by the author has many differences between the labeling results. Furthermore, the results after balancing can be seen in table 10.

Table 10. Question Prediction After SMOTE

		Predicted			
		C4	C5	C6	All
Actual	C4	16	14	7	37
	C5	5	31	1	37
	C6	0	0	37	37
	All	21	45	45	111

In table 10, with the total number of questions increased to 111 items by using SMOTE, it can be seen that C4 is counted as 16 items. Then it has the prediction of C5 labels are 31 and C6 has 37 items. Thus after calculating the percentage of the results of the confusion matrix, it is further described in table 11 which shows at the same time the results of the accuracy of the performance of the classifier model used in this study.

Table 11. Confusion Matrix Result

	Before SMOTE	After SMOTE
Precision	49%	76%
Recall	59%	76%
F1-Score	51%	74%
Accuracy	59%	76%

Based on table 11, it can be seen that the comparison results before and after the use of SMOTE can increase the level of accuracy of the performance of the classifier model. The precision result before SMOTE is 49% and after SMOTE is 76%. Recall before SMOTE was 59% and after SMOTE was 76%. F1-Score before SMOTE was 51% and after SMOTE was 74%. Accuracy before SMOTE is 59% and after SMOTE is 76%.

Based from past research [21], the conclusion that has been drawn from this research is that the use of the KNN algorithm with the Euclidean distance approach can work quite well for classifying and predicting questions based on the cognitive level of RBT as it stated the accuracy results after SMOTE are all above 70%. In addition, the results of this study also prove that the existence of data imbalances can affect the performance level of a classifier model.

IV. CONCLUSION

From the results of this study, several conclusions can be drawn. The results of automatic classification on Indonesian History Education subjects at the high school level with cognitive levels of C4-C6 using the KNN algorithm from a total of 58 questions (before balancing) answered that there were 32 C4, 2 C5. Then from a total of 111 questions (after balancing) answered that there were 16 C4, 31 C5, 37 C6. The accuracy level taken was the result after SMOTE implementation. The results obtained are precision is 76%, recall is 76%, f1-score is 74%, accuracy is 76%. Handling unbalanced data in this study using the SMOTE oversampling method resulted in higher accuracy than without the use of SMOTE. So it can be interpreted that the use of SMOTE can improve performance and be a good solution to deal with data imbalances. The KNN algorithm with the Euclidean distance approach has a poor performance used for classifying questions based on RBT cognitive levels C4 to C6. There are several inputs that can be developed for further research, namely increasing the number of datasets to retest the KNN algorithm with the Euclidean distance approach aimed at finding more information if the number of datasets can affect the accuracy generated from the algorithm. Furthermore, testing can be done using different algorithm methods or from KNN with different distance approaches or other algorithms such as Support Vector Machine, Naïve Bayes, Neural Network, and so on. Testing can also be done with different feature extraction and oversampling methods.

REFERENCES

[1] I. W. C. Sujana, “FUNGSI DAN TUJUAN PENDIDIKAN INDONESIA.,” *Jurnal Pendidikan Dasar*, pp. 30–31, 2019.

[2] E. Poerwanti, *Standar Penilaian Badan Standar Nasional Pendidikan (BNSP)*. 2008.

[3] BPSI, “ndang-Undang Republik Indonesi No. 20 Tahun 2003 Tentang Sitem Pendidikan Nasional,” *http://lppro.pancabudi.ac.id/new/pages/read/undangundang-republik-indonesi-no-20-tahun-2003-tentang-sitem-pendidikan-nasional*, 2014.

[4] N. D. Ahmad, W. A. W. Adnan, M. A. Aziz, and M. Y. Yusof, "Automating Preparation of Exam Questions: Exam Question Classification System (EQCS)," in *2011 International Conference on Research and Innovation in Information Systems (ICRIIS)*, 2011, pp. 1–6.

[5] P. Armstrong, "Bloom's Taxonomy," *Vanderbilt University Center for Teaching*, 2016.

[6] M. Forehand, "Bloom's Taxonomy," *http://epltt.coe.uga.edu/index.php?title=Bloom%27s_Taxonomy*, 2017.

[7] A. Aninditya, M. A. Hasibuan, and E. Sutoyo, "Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy," in *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, 2019, pp. 112–117.

[8] I. Listiowarni and N. P. Dewi, "Pemanfaatan Klasifikasi Soal Biologi Cognitive Domain Bloom's Taxonomy Menggunakan KNN Chi-Square Sebagai Penyusunan Naskah Soal," *Digital Zone: Jurnal Teknologi Informasi & Komunikasi*, vol. 11, no. 2, pp. 185–195, 2020, doi: 10.31849/digitalzone.v11i2.4798ICCS.

[9] Y. Pantiwati and F. H. Permana, "ANALISIS BUTIR SOAL OLEH MAHASISWA S1 PENDIDIKAN BIOLOGI UNIVERSITAS MUHAMMADIYAH MALANG (UMM) BERDASARKAN PISA DAN TAKSONOMI BLOOM REVISI," in *Seminar Nasional Kedua Pendidikan Berkemajuan dan Menggembirakan (The Second Progressive and Fun Education Seminar)*, 2017, pp. 707–716.

[10] S. K. Patil and M. M. Shreyas, "A Comparative Study of Question Bank Classification based on Revised Bloom's Taxonomy using SVM and K-NN," in *2017 2nd International Conference On Emerging Computation and Information Technologies (ICECIT)*, 2017, pp. 1–7.

[11] E. Budianita, Jasril, and L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi," *Jurnal Sains, Teknologi dan Industri*, vol. 12, no. 2, pp. 242–247, 2015.

[12] R. N. Whidhiasih, N. A. Wahanani, and Supriyanto, "KLASIFIKASI BUAH BELIMBING BERDASARKAN CITRA RED-GREEN-BLUE MENGGUNAKAN KNN DAN LDA," *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, vol. 1, no. 1, pp. 29–35, 2013.

[13] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann Math Artif Intell*, vol. 81, no. 3–4, pp. 429–450, Dec. 2017, doi: 10.1007/s10472-017-9564-8.

[14] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf Sci (N Y)*, vol. 340–341, pp. 250–261, May 2016, doi: 10.1016/j.ins.2016.01.033.

[15] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbour," *Journal Information System Development (ISD)*, vol. 3, no. 1, pp. 44–49, 2018.

[16] G. Gumelar *et al.*, "Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance," in *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, 2021, pp. 250–255.

[17] N. Noorhalim, A. Ali, and S. M. Shamsuddin, "Handling Imbalanced Ratio for Class Imbalance Problem Using SMOTE," in *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*, 2019, pp. 19–30. doi: 10.1007/978-981-13-7279-7_3.

[18] A. S. Callista, O. N. Pratiwi, and E. Sutoyo, "Questions Classification Based on Revised Bloom's Taxonomy Cognitive Level using Naive Bayes and Support Vector Machine," in *Proceedings - 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021*, 2021, pp. 260–265. doi: 10.1109/IC2IE53219.2021.9649187.

[19] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. 2012.

[20] G. James, D. Witten, T. Hastie, and R. Tibshirani, "Data for an Introduction to Statistical Learning with Applications in R.," *Package 'ISLR.'CRAN.*, 2017.

[21] S. Santosa and R. Yuliantara, "MODEL PREDIKSI POLA LOYALITAS PELANGGAN TELEKOMUNIKASI MENGGUNAKAN NAIVE BAYES DENGAN OPTIMASI PARTICLE SWARM OPTIMIZATION," *Jurnal Teknologi Informasi*, vol. 13, no. 2, pp. 154–169, 2017, [Online]. Available: http://research.