International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM)

Septian Girendra Wardhani[1], Ana Kurniawati[2]
[1]Business Information System, Faculty of Technology and Engineering, Gunadarma University, Central Jakarta, Indonesia-10430
[2]Faculty of Technology and Engineering, Gunadarma University, Central Jakarta, Indonesia-10430

Abstract**—** *The rapid growth of the banking industry today is evident in the increasing number of digital banks. The core business of banking lies in credit disbursement, which involves the risk of customers failing to repay their loans. This factor poses potential losses for the banking industry. However, the risk of default may be mitigated using computer technology. This study aims to analyze creditworthiness using the K-Nearest Neighbor (K-NN) algorithm. The research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, comprising the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The study utilized a dataset of 1,000 records, categorized into 30% bad and 70% good classes. The tested dataset achieved the highest accuracy of 0.78 with an AUC value of 0.80 at k = 5. Based on this performance, the K-Nearest Neighbor algorithm shows promising results in classifying creditworthiness data.*

Keywords**—***Banking, Creditworthiness, Cross-Industry Standard Process for Data Mining, K-Nearest Neighbor.*

I. INTRODUCTION

Computer technology as a supporting tool in banking services is essential for economic activities. Banking has a role in boosting the growth and development of the economy in a country, serving as the backbone of the nation's financial system [1]. The banking industry is experiencing rapid growth and is moving toward the digital era, as evidenced by the rising number of digital banks. With such rapid growth, competition in the banking industry will involve innovation, flexibility, and the implementation of technology in banking operations [2]. Banking serves as a place to store and manage money. Additionally, it provides financial services such as credit, investment, and transactions [3]. Credit can be disbursed, but it carries the risk that customers may be unable to repay after the credit has been granted. Non-performing loans, categorized as substandard, doubtful, or bad, can result in reduced profitability, leading to losses for banks due to poor credit quality[4]. However, the risk of default can be mitigated by leveraging computer technology.

Computer technology is currently shifting toward the application of artificial intelligence (AI). AI can be used to support economic activities in daily social life. One of the algorithms that can be applied in AI implementation is K-Nearest Neighbor (K-NN). K-NN functions as a classification technique for data that share the same or closest classification to an object[5].

K-NN classifies a set of data based on predefined classes. K-NN is a type of machine learning known as supervised learning, where the training data has predetermined outcomes. The new data is then classified based on the majority proximity to the categories within the K-NN algorithm[6].

A study that utilized the K-NN algorithm was conducted by Sularno, focusing on heart disease classification. The research used a sample dataset of 1,025 instances, divided into 820 training data and 205 testing data. The results showed an accuracy of 92%, a precision of 90%, and a recall of 92%[7]. Research by Rusda Wajhillah on credit eligibility at the AKU Cooperative achieved an accuracy of 79.45% with k=1k=1k=1. The Receiver Operating Characteristic (ROC) curve formed showed the highest Area Under Curve (AUC) values of 0.804 at k=5k=5k=5 and 0.811 at k=9k=9k=9 [8]. Research by Firna Yenila on the prediction and classification process for loan status using the K-NN algorithm achieved an accuracy of 90%[9]. Another study related to the classification of underprivileged communities using the K-NN algorithm was conducted by Ahmad Khairi. The study used 681 data points and achieved the highest accuracy of 98.68% at k=7[10].

Based on the outlined background, this study conducts an accuracy analysis of the implementation of the K-Nearest Neighbor algorithm for creditworthiness analysis using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology.

CRISP-DM is used as a structured and well-documented framework with clear steps for conducting data mining analysis [11]. Data mining issues can be resolved using the CRISP-DM method, which is capable of analyzing business problems and current conditions, providing appropriate data transformations, and delivering models that can assess effectiveness and document the results obtained [12].

II. METHODOLOGY

The methodological approach used in this research to predict creditworthiness utilizes the Cross-Industry Standard Process for Data Mining (CRISP-DM), as shown in Figure 1.

Septian Girendra Wardhani and Ana Kurniawati, "Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM)," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 152-157, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

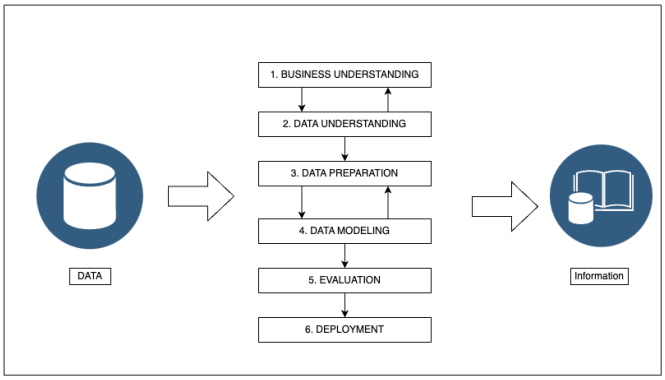

Fig. 1. The Methodology Research.

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a framework used in data mining. CRISP-DM was developed by five companies: Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation, and OHRA [13]. CRISP-DM aims to perform a strategic analysis process that solves research problems or addresses business issues within a company [14]. The CRISP-DM method comprises six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [15].

A. *Business Understanding*

This phase aims to comprehend the objectives from a business perspective and subsequently translate them into a data mining problem statement. The problem in this study is the classification of creditworthiness for a bank in Germany. Credit eligibility is determined based on several factors, including savings amount, loan duration, credit history, loan purpose, credit amount, marital status and gender, length of residence, property ownership, age, residence status, occupation, and number of dependents.

B. *Data Understanding*

The data understanding phase is carried out after the data collection stage. In this phase, the process involves understanding and identifying the quality of the data. Data understanding refers to the German Credit dataset obtained from the machine learning repository provided by Professor Dr. Hans Hofmann from the Institut für Statistik und Ökonometrie, Universität Hamburg. The dataset used consists of 1,000 records.

C. *Data Preparation*

Data preparation is the stage of preparing data, which is typically repeated until the data is fully ready for modeling. This phase includes data selection, data preprocessing, data transformation, and normalization using the min-max method [16]. The equation 1 is presented for the min-max method in equation (1).

$$N = \frac{x0-xmin}{xmax-xmin} \quad (1)$$

Information :
N : Min-Max Normalization.
xmax : The maximum value of the attribute being compared.
x0 : The value of the actual data.
xmin : The minimum value of the attribute being compared.

D. *Data Modeling*

Data modeling is selecting and applying modeling techniques based on the available parameters. This study employs the K-Nearest Neighbor method to classify creditworthiness at a bank in Germany. K-NN operates on the principle of determining and finding the closest distance with the k-nearest neighbors in the training data. The result of the k value depends on the data values, k with a higher value can reduce the effect of errors or noise in the classification process. Still, the boundaries between classifications may become suboptimal[17], calculating the closest distance using Euclidean distance, as shown in equation (2).

$$euc = \sqrt{\sum_{i=1}^{p}(x2i - x1i)^2} \quad (2)$$

Information :
euc : Distance
x2 : Sample data/training data
x1 : Test data
p : Data dimensions
i : Data variables

E. *Evaluation*

The evaluation stage is conducted on the results of the model implementation to assess its accuracy, after which a decision is made regarding the use of the results from data mining [18]. In this study, the evaluation uses the confusion matrix method. The confusion matrix is a method commonly used to calculate accuracy in the context of data mining concepts. The confusion matrix performs calculations with four outputs: recall, precision, accuracy, and error rate [19]. The confusion matrix is a commonly used matrix for evaluating the accuracy of data mining, as shown in Table I.

TABLE I. Confusion matrix

Predicted class		**Good**	**Bad**
	Good	True Positive (TP)	False Negative (FN)
	Bad	False Positive (FP)	True Negative (TN)

Information:
True Positive : Data count from the correct "good" class.
True Negative : Data count from the correct "bad" class.
False Positive : Data count from the incorrect "bad" class.
False Negative : Data count from the incorrect "good" class.

The accuracy calculation from the confusion matrix is shown in equation (3).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

The results of the confusion matrix will be evaluated using the ROC Curve method. The ROC Curve is a method used to illustrate and classify categories in a statistical model. The ROC Curve can also determine the threshold of a model. The ROC Curve is plotted on the confusion matrix, with the false positive rate on the horizontal axis and the true positive rate on the vertical axis [16]. The value of the ROC Curve is then used to calculate the Area Under Curve (AUC). AUC represents the area under the ROC Curve or the integral of the ROC function, which is used to compare the performance between two algorithms [20]. The performance categories of the model based on the AUC value range are shown in Table II.

Septian Girendra Wardhani and Ana Kurniawati, "Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM)," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 152-157, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

TABLE II. Range of AUC values

No	The AUC value range	Categories
1	0.90 – 1.00	Excellent Classification
2	0.80 – 0.90	Good Classification
3	0.70 – 0.80	Fair Classification
4	0.60 – 0.70	Poor Classification
5	0.50 – 0.60	Failure

F. Deployment

The preparation for deployment involves the creation of reports or presentations derived from the modeling and evaluation phases in data mining. The results will be presented as information or knowledge that can support the creditworthiness analysis process, helping to mitigate potential credit risk issues.

III. Result and Discussion

A. Business Understanding

The business understanding phase focuses on the background of the problem, namely the risk of credit default, which constitutes the primary issue in the credit granting process. This issue can be mitigated through information technology, aligning with the objective of this study, which is to analyze the K-Nearest Neighbour algorithm in determining creditworthiness, thereby reducing the occurrence of such problems.

B. Data Understanding

This study utilizes the German Credit dataset from the UCI Machine Learning Repository, accessible at https://archive.ics.uci.edu/dataset/144/statelog+german+credit+data. The dataset comprises 21 variables and 1 class, with 1,000 records. A detailed description of the dataset is presented in Table III.

TABLE III. Dataset

No	Attribute Names	Categories
1	Status of existing checking account (V1)	1) A11 : ... < 0 DM 2) A12 : 0 <= ... < 200 DM 3) A13 : ... >= 200 DM /annual salary 4) A14 : no checking account
2	Duration in month (V2)	4 - 72 month
3	Credit history (V3)	1) A30 : no credits taken/ all credits paid back duly 2) A31 : all credits at this bank paid back duly 3) A32 : existing credits paid back duly till now 4) A33 : delay in paying off in the past 5)A34: critical account/ other credits existing (not at this bank)
4	Purpose (V4)	1) A40 : car (new) 2) A41 : car (used) 3) A42 : furniture/equipment 4) A43 : radio/television 5) A44 : domestic appliances 6) A45 : repairs 7) A46 : education 8) A47 : (vacation - does not exist?) 9) A48 : retraining 10) A49 : business 11) A410 : others
5	Credit amount (V5)	250 DM - 18424 DM
6	Savings account/bonds (V6)	1) A61 : ... < 100 DM 2) A62 : 100 <= ... < 500 DM 3) A63 : 500 <= ... < 1000DM 4) A64 : .. >= 1000 DM 5) A65 : no savings account
7	Present employment since (V7)	1) A71 : unemployed 2) A72 : ... < 1 years 3) A73 : 1 <= ... < 4 years 4) A74 : 4 <= ... < 7 years 5) A75 : .. >= 7 years
8	Installment rate in percentage of disposable income (V8)	1- 4 Percentage Unit (%)
9	Personal status and sex (V9)	1) A91 : male : divorced/separated 2) A92 : female : divorced/separated/married 3) A93 : male : single 4) A94 : male : married/widowed 5) A95 : female : single
10	Other debtors / guarantors (V10)	1) A101 : none 2) A102 : co-applicant 3) A103 : guarantor
11	Present residence since (V11)	1-4 in years
12	Property (V12)	1)A121 : real estate 2)A122 : if not A121 : building society savings agreement/life insurance 3)A123 : if not A121/A122 : car or other, not in attribute 6 4)A124 : unknown / no property
13	Age in years (V13)	19 - 75 in years
14	Other installment plans (V14)	1) A141 : bank 2) A142 : stores 3) A143 : none
15	Housing (V15)	1) A151 : rent 2) A152 : own 3) A153 : for free
16	Existing credits at bank (V16)	1-4
17	Job (V17)	1) A171 : unemployed/ unskilled 2) A172 : unskilled - resident 3) A173 : skilled employee / official 4) A174 : management/ self-employed/ highly qualified employee/ officer
18	Number of dependents (V18)	1-2
19	Telephone (V19)	1) A191 : none 2) A192 : yes, registered
20	Foreign worker (V20)	1) A201 : yes 2) A202 : no
21	Target (V21)	1) 1 : Good 2) 2 : Bad

Table III presents the dataset, with 21 variables and different data types. The input variables comprise 12 categorical types and numeric types 8. The output variable has a single type, which is numerical. The dataset contains one class, which is divided into two categories: (1) Good and (2) Bad, as illustrated in Figure 2.

Figure 2 presents the percentage distribution of the target variable in the German Credit dataset, with 70% of the data classified as good (creditworthy) and 30% classified as bad (less creditworthy).

Septian Girendra Wardhani and Ana Kurniawati, “Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM),” *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 152-157, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

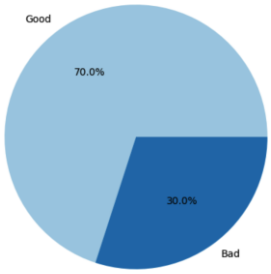

Fig. 2. Target Variable Percentage

C. Data Preparation

The data intended for training undergoes a preparation phase to achieve higher accuracy. This study's data preparation process consists of three stages: data selection, data preprocessing, and transformation and normalization using the Min-Max method.

As part of the data selection stage, the variables V8, V10, V14, V16, V19, and V10 are excluded from the dataset. The data used for training undergoes data preprocessing to enhance accuracy. The data preprocessing conducted in this study involves identifying missing values within the dataset. The analysis confirms that no data is missing, all values are assigned appropriately to their respective attributes, and the values are consistent with their corresponding variables. Data transformation involves converting categorical data into a format that can be processed by a computer system. This transformation is performed after the data preprocessing stage. The process represents categorical data as numerical values. The results of this transformation are in Table IV.

TABLE IV. Data Transforming

No	V1	V2	V3	V4	V5	V6	V7	V9	V11	V12	V13	V15	V17	V18	Target
1	0	24	3	0	4870	0	2	2	4	3	53	2	2	2	2
2	1	30	4	0	5234	0	3	3	2	2	28	1	3	1	2
.		.	.	.	.	.	.	.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.	.	.	.	.	.	.
.		.	.	.	.	.	.	.	.	.	.	.	.	.	.
999	3	24	2	9	4844	0	3	2	2	2	33	0	3	1	2
1000	2	15	0	9	6289	0	2	0	1	1	33	1	2	1	1

The data in Table IV is then subjected to the normalization stage using the Min-Max method. The results of the overall normalization are presented in Table V.

TABLE V. Normalization using the Min-Max method

No	V1	V2	V3	…	V18	Target
1	0.00	0.29	0.75	…	1.00	1.0
2	0.33	0.38	1.00	…	0.00	1.0
.	.	.		…	.	.
.	.	.			.	.
.	.	.			.	.
999	1.00	0.65	0.50	…	0.00	1.0
1000	0.67	0.56	0.00	…	0.00	0.0

D. Data Modelling

The modeling process employs the K-Nearest Neighbor (K-NN) method. The good column represents creditworthy data, while the bad column indicates less creditworthy data. The K-NN model uses four values: k = 5, k = 7, k = 9, and k = 11. The dataset is split in an 80:20 ratio, as shown in Table VI.

TABLE VI. Dataset split

	Good	Bad	Data count
Training data	563	237	800
Testing data	137	63	200
Data count	700	300	1000

E. Evaluation

The results of applying K-NN to the German Credit dataset were evaluated using a confusion matrix. For example, with k = 5, the obtained values are presented in Table VII.

Table VII presents the confusion matrix values for k = 5 and is calculated for the accuracy k = 5 in the K-NN model, as shown in the following computation.

TABLE VII. Confusion Matrix Results

	Good	Bad
Good	129	8
Bad	36	27

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{129+27}{129+127+36+38} = \frac{156}{200}$$

$$Accuracy = 0.78$$

The value 0.78 represents the performance of the K-NN algorithm at k = 5. Table VIII presents the overall calculation results.

TABLE VIII. Accuracy result for K values

No	K Value	Accuracy
1	K = 5	0,78
2	K = 7	0,76
3	K = 9	0,76
4	K = 10	0,76

Based on Table 8, the highest accuracy is obtained at k = 5. The k = 5 value, representing the highest accuracy of the K-NN algorithm in creditworthiness analysis using the German Credit dataset, is illustrated as a ROC curve, as shown in Figure 3.

The ROC Curve (Receiver Operating Characteristic) visualization in Figure 3 shows an Area Under the Curve (AUC) value of 0.80, indicating the accuracy of the K-Nearest Neighbor (K-NN) model in classifying creditworthiness. According to Table II, an AUC range between 0.70 – 0.80 classifies the model as a fair model. A fair model maintains a balanced performance and does not exhibit bias toward specific groups within the data.

Septian Girendra Wardhani and Ana Kurniawati, “Implementation of K-Nearest Neighbor Algorithm for Creditworthiness Analysis Using Methods Cross-Industry Standard Process for Data Mining (CRISP-DM),” *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 1, pp. 152-157, 2025.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

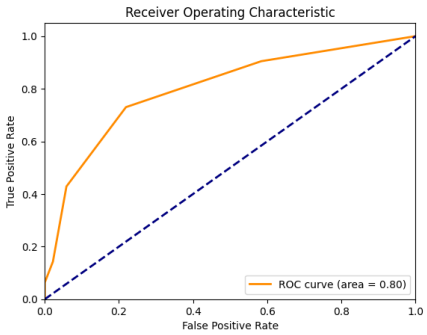

Fig. 3. ROC Curve Result

F. Deployment

The issue identified in the business understanding phase is the Credit repayment risk. Therefore, this study was conducted to support creditworthiness analysis using the K-Nearest Neighbor (K-NN) algorithm. After testing, the highest accuracy was with k = 5, a value of 0.78 (78%), and an AUC of 0.80 (80%). Based on the business objectives outlined in the business understanding phase, the research objectives have been successfully achieved. Figure 4 displays the plot of the K-NN algorithm.

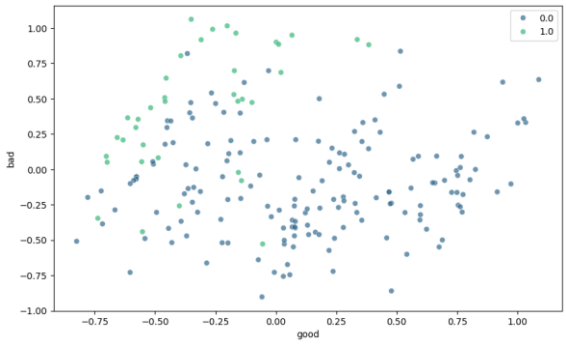

Fig. 4. K-NN algorithm plot

Figure 4 illustrates the creditworthiness classification results using the K-NN algorithm. The classification is represented by two colors: green and blue. The green class represents the bad label (1.0), while the blue class represents the good label (0.0).

Figure 4 displays two overlapping regions representing the creditworthiness classification results. The 'bad' class (1.0) has fewer data points compared to the 'good' class (0.0), as indicated by the plot.

Figure 4 shows green points positioned among the blue points, where the green points represent misclassifications of the bad class (1.0). Similarly, the blue points among the green points represent instances of the good class (0.0) that were incorrectly predicted as bad (1.0). The points that do not align with their correct class reflect the 78% accuracy rate.

IV. Conclusion

Based on the research, the author successfully developed the K-Nearest Neighbor (K-NN) algorithm for creditworthiness analysis and demonstrated its ability to classify the target variable. The dataset used in this study was sourced from the UCI Machine Learning Repository and contained 20 input variables and one output variable. Of these, 14 input variables were utilized.

The research utilized the Cross-Industry Standard Process for Data Mining (CRISP-DM), which provides a structured yet flexible framework. CRISP-DM consists of six stages that support the development of the K-NN model: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

The training process utilized 800 training data points and 200 testing data points. Based on the evaluation results using the confusion matrix method, the accuracy achieved for k = 5 was 0.78, while k = 7, k = 9, and k = 11 yielded an accuracy of 0.76. The highest accuracy, 0.78 at k = 5, corresponds to an AUC value of 0.80, which falls into the fair model category.

References

[1] S. Ilfa Dianita., Irawan, Heri., Mulya, Andi Deah Salsabila., "Peran Bank Syariah Indonesia Dalam Pembangunan Ekonomi Nasional," *Jurnal Lembaga Keuangan, Ekonomi dan Bisnis Islam*, vol. 3, issue 2, pp. 148-158, 2021.

[2] R. F. Irsyad., F. A. Siregar., J. Marbun and H. Hasyim., "Menghadapi Era Baru : Strategi Perbankan Dalam Menghadapi Perubahan Pasar Dan Teknologi Di Indonesia," *Journal of Economics and Business Management*, vol. 3, pp. 29-46, 2024.

[3] Judijanto, Loso., Putri, PA Andiena Nindya., Syamsuri., Dewantara, Billy., Alfiana., "Impact of Financial Technology (Fintech) Innovation on Traditional Banking and Finance Business Models," *Management Studies and Entrepreneurship Journal*, vol. 5, pp. 1020-1025, 2024.

[4] Permatasari, L., & Dasman, S. "Pengaruh NPL (Non Performing Loan), EPS (Earning Per Share), dan PER (Price Earning Ratio) Terhadap Harga Saham BBRI Per-Quartal Tahun 2012-2022" in *Proceedings SEMANIS: Seminar Manajemen Bisnis* ,Vol. 2, No. 1, pp. 98-109, 2024.

[5] N. Nopiyanto and B. Ulum., "Penerapan Metode Clusteringdengan Algoritma K-Means pada Pengelompokan Jurusan Data Calon Siswa Baru Di SMK Al-Ishlah Cikarang Utara," *Journal Transformation of Mandalika*., vol. 3, pp. 5-17, 2022.

[6] Cahyanti, D., Rahmayani, A., and Husniar , S. A., "Analisis performa metode KNN pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, issue 2 pp. 39-43, 2020.

[7] Sularno, M. F., Wiyanto, W., Ardiatma, D., & Zy, A. T., "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, issue 4, pp. 850-860, 2023.

[8] Wajhillah , R., Ubaidallah, I. H., & Bahri, S., "Analisis Kelayakan Kredit Berbasis Algoritma K-Nearst Neighboar (Studi Kasus: Koperasi AKU)," *Jurnal Nasional Informatika Dan Teknologi Jaringan*, vol. 4, issue 1,pp. 121-125, 2019.

[9] Yenila, F., Marfalino, H., & Defit, S., "Model Analisis Machine Learning dengan Pendekatan Deep Learning dalam Penentuan Kolektabilitas," *Jurnal Sains dan Teknologi*, vol. 12, issue 2, pp. 403-414, 2023.

[10] K. Ahmad., A. F. Ghozali and A. D. N. Hidayah., "Implementasi K-Nearest Neighbor (Knn) Untuk Klasifikasi Masyarakat Pra Sejahtera Desa Sapikerep Kecamatan Sukapura," *Jurnal Ilmu Teknologi, Kesehatan, dan Humaniora*, vol. 2, issue 3, pp. 319-323, 2021.

[11] F. Salsabila., I. Fitrianti., Y. Umaidah and N. Heryana., "Penerapan Metode Crisp-Dm Untuk Analisa Pendapatan Bersih Bulanan Pekerja Informal Di Provinsi Jawa Barat Dengan Algoritma K-Means," *Jurnal Dinamik*, vol. 28, issue 2, pp. 97-104, 2023.

[12] Sari, Ni Ketut Ayu Purnama., Candiasa, I Made., Aryanto, Kadek Yota Ernanda., "Sistem Pendukung Keputusan Pengembangan Ekowisata Pedesaan Menggunakan Metode Fucom-Moora Dan Fucom-Vikor," *Jurnal Sains dan Teknologi*, vol. 10, issue 2, pp. 112-126, 2021.

[13] M. A. Wiratama and W. M. Pradnya., "Optimasi Algoritma Data Mining Menggunakan Backward Elimination Untuk Klasifikasi Penyakit Diabetes," *Jurnal Nasional Pendidikan Teknik Informatika*, vol. 11, issue 1, pp. 1-12, 2022.

[14] R. D. Fitriani., H. Yasin and T. Tarno., "Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes

(Studi Kasus: Status Peserta Kb Iud Di Kabupaten Kendal)," *Jurnal Gaussian*, vol. 10, issue 1, pp. 11-20, 2021.
[15] N. C. Sastya and I. Nugraha., "Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution," *Jurnal Pendidikan Dan Aplikasi Industri*, vol. 10, issue 2, pp. 103-115, 2023.
[16] M. R. Givari., M. R. Sulaeman and Y. Umaidah., "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *Jurnal Nuansa Informatika*, vol. 16, issue 1, pp. 141-149, 2022.
[17] A. Tangkelayuk and E. Mailoa., "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, issue 2, pp. 1109-1119, 2022.
[18] S. Amri., "Perbandingan Kerangka Model Klasifikasi untuk Pemilihan Metode Kontrasepsi dengan Pendekatan CRIPS-DM," *Jurnal Information Science and Library*, vol. 1, pp. 14-23, 2020.
[19] R. D. Mendrofaa., M. H. Siallagan., D. P. Pakpahanc and J. Amalia., "Credit Risk Analysis dengan Algoritma Extreme Gradient Boosting dan Adaptive Boosting," *Journal of Information System, Graphics, Hospitality and Technology*, vol. 5, issue 1, pp. 1-7, 2023.
[20] Yunitasari, S.. Hopipah and R. Mayasari., "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-Nearest Neighbor (k-NN) dan Naïve Bayes," *Technomedia Journal (TMJ)*, vol. 6, issue 1, pp. 99-110, 2021.