BULLETIN OF ENGINEERING SCIENCE, TECHNOLOGY AND INDUSTRY

https://bestijournal.org | PT. Radja Intercontinental Publishing

E-ISSN: 3025-5821

COMPARATIVE ANALYSIS OF NAIVE BAYES ALGORITHM AND C4.5 ALGORITHM IN SELECTING TYPES OF UMKM PRODUCTS IN BBPSDMP KOMINFO MEDAN TRAINING

Alex Siregar[1], Leni Marlina[2], Khairul[3], Muhammad Iqbal[4], Andysah Putera Utama Siahaan[5]

[1]Master of Information Technology Student, Universitas Pembangunan Panca Budi, Medan

[2,3,4,5]Lecturer of Master of Information Technology, Universitas Pembangunan Panca Budi, Medan

*Correspondence: Email alexiregar@gmail.com

Abstract

This study compares the accuracy of the Naïve Bayes and C4.5 algorithms in determining the most suitable product types for Micro, Small, and Medium Enterprises (MSMEs) participating in the Digital Entrepreneurship Academy (DEA) training program at BBPSDMP Kominfo Medan. This study uses a dataset from DEA participants between 2021 and 2022. The analysis shows that the C4.5 algorithm has a higher accuracy compared to Naïve Bayes, indicating its better effectiveness in helping MSMEs choose product types. These findings suggest that C4.5 is more suitable for applications that require a high level of accuracy, especially in the context of this study. This study provides valuable insights into the selection of algorithms to support decision making in the MSME sector.

Keywords: *Naïve Bayes, C4.5, UMKM, Digital Entrepreneurship Academy, Algorithm Comparison, Product Selection.*

1. INTRODUCTION

Micro, Small and Medium Enterprises (MSMEs) are an important pillar of Indonesia's economic development that continues to be encouraged to move up a class, so that they can make a greater contribution to the economy and absorb more workers. The rapid development of technology, especially the internet, has had a significant impact on the delivery and receipt of information, especially for Micro, Small and Medium Enterprises in determining market segmentation. The Center for Human Resource Development and Research (BBPSDMP) of the Medan Ministry of Communication and Information, as one of the Technical Implementation Units of the Human Resource Development Agency (BPSDM) of the Ministry of Communication and Information, has a priority program, the Digital Talent Scholarship (DTS).

The Digital Talent Scholarship (DTS) program is a competency development training initiative that has been implemented for Indonesian digital talents since 2018. As part of the national development priority program, the Digital Entrepreneurship Academy (DEA) is one of the training academies organized by the DTS of the Ministry of Communication and Information. This program is aimed at the general public who are prospective Micro, Small, and Medium Enterprises (MSMEs), as well as MSME business actors who want to increase their capacity in utilizing digital technology to develop their businesses. DEA training aims to encourage the improvement of digital skills from basic to intermediate for prospective and MSME actors. This academy targets participants from the general public who want to start a business, as well as MSME actors. In the DEA training registration process, participants are required to register on the Digital Talent Scholarship website (digitalent.kominfo.go.id).

To evaluate DEA training activities, a Tracer Study was conducted on DEA alumni from 2021 to 2022. The Tracer Study system was implemented as an effective solution to collect alumni achievement data. The development of this system can improve alumni data management, avoid data loss, and facilitate interaction between agencies and alumni. This interaction media plays an important role in obtaining accurate information, as well as involving the community in providing input and criticism that can improve the quality of post-training education.

From the tracert study data collected by BBPSDMP Kominfo Medan in 2021-2022 through an online questionnaire, it was found that many DEA participants had not yet determined the type of product. MSMEs play a vital role in the economy with their significant contribution to economic growth and job creation. The selection of product types is a key factor in the sustainability and success of MSMEs, but is often complex because it involves various variables and factors. The Naïve Bayes algorithm, which utilizes the principle of probability, and the C4.5 algorithm, which uses a decision tree approach, are the main choices for this analysis.(Mawaddah & Pranoto, 2023).

*Naive Bayes*known for its capabilities in probability-based classification, while C4.5 offers a transparent approach through a decision tree structure.(Safira & Mustakim, 2021). Thus, the comparison of the two is expected to provide in-depth insight into the effectiveness and suitability of each algorithm in supporting the decision-making process related to the selection of product types in the MSME environment.

By understanding the advantages and disadvantages of each algorithm, this study aims to contribute to a better understanding in the context of MSMEs, so that MSME owners and related stakeholders can make more informed and data-based decisions. This analysis is expected to provide a positive contribution to the development of MSME business strategies and sustainability, which in turn can have a positive impact on the local and national economy.

In this study, the author will analyze the comparative accuracy of the Naïve Bayes algorithm and the C4.5 algorithm in selecting the type of product for prospective Micro, Small and Medium Enterprises (MSMEs) based on data samples carried out in the DEA (Digital Entrepreneurship Academy) training from 2021 to 2022.

2. PROBLEM FORMULATION

Based on the research background that has been explained, the problem formulation for this research is as follows:

2.1 How to apply the Naïve Bayes algorithm method and the C4.5 algorithm in the context of selecting product types for prospective MSMEs?

2.2 Does one algorithm tend to provide an advantage in predicting market preferences or consumer needs for the type of product a prospective MSME is pursuing?

2.3 How do factors such as accuracy, computational speed, and reliability affect the performance of both algorithms in selecting product types for prospective MSMEs?

3. METHOD

3.1 Types of Research

The type of research method used is quantitative research. According to Habeeb Adewale Ajimotokan (2022), quantitative research is a scientific research approach that uses statistical methods and techniques to collect, analyze, and interpret numerical data. This approach aims to measure variables objectively, understand the relationships between variables, and make generalizations from a sample to a larger population. The characteristics of quantitative research include the use of measurable numerical data, an objective approach that avoids subjectivity, and a structured research design.

In addition, Jennifer J. Mueller and Nancy Files (2024) explain that quantitative research uses statistical analysis to summarize, explain, and test hypotheses based on the data collected. Sample selection is done carefully to ensure greater representation of the population, and the results of the study are expected to be generalizable. The quantitative research process includes steps ranging from formulating research questions, developing hypotheses, research design, data collection and data analysis to interpretation and conclusions. This research method is used to search for something systematically using scientific methods and applicable sources of information. This process provides superior research results and accurate research used.

Alex Siregar[1], Leni Marlina[2], Khairul[3], Muhammad Iqbal4, Andysah Putera Utama Siahaan[5]

3.2 Place and Time of Research

a) Research Place

This research was conducted at the Center for Human Resource Development and Research in Communication and Informatics (BBPSDMP Kominfo) Medan. BBPSDMP Kominfo Medan is a Technical Implementation Unit (UPT) under the Human Resource Development Agency (Balitbang) of the Ministry of Communication and Informatics of the Republic of Indonesia. The main task of BBPSDMP Kominfo is to carry out human resource development and research in the field of Information and Communication Technology (ICT) in seven working areas, including disadvantaged, outermost, and remote areas (3T). This research aims to obtain relevant and useful data in the context of ICT development in these areas.

No	Activity Description	Nov.	Dec.			Jan.			Feb.			Mar.		Apr.		May.			Jun.			Jul.			Agus.
		3	1	2	3	1	2	3	1	2	3	1	2	1	2	1	2	3	1	2	3	1	2	3	1
1	Initial Research																								
2	Title Submission																								
3	Proposal Writing																								
4	Proposal Seminar																								
5	Thesis Guidance																								
6	Results Seminar																								
7	Data processing																								
8	Thesis Writing																								
9	Green Table Session																								

b) Research Time

The research period was conducted over a period of nine months from November 2023 to August 2024. Table 3.1 Research Time

3.3 Data Collection Techniques

The data for this study were obtained from a questionnaire distributed by the researcher to alumni participants of the DEA (Digital Entrepreneurship Academy) training for the period 2021-2022. DEA participants came from Micro, Small, and Medium Enterprises (MSMEs), Prospective MSMEs and the general public who had completed training organized by BBPSDMP Kominfo Medan. The questionnaire was designed to collect relevant information related to the experience and impact of the training. The data collected will be analyzed to evaluate the effectiveness of the training and to provide insight into the implementation and results achieved by participants after participating in the training program. This dataset includes eight main attributes, which represent various information related to alumni participation in the training, their business conditions, and the use of digital strategies in developing their businesses. Figure 3.2 is a survey conducted from 4 to 12 December 2022. This survey aims to track alumni of the BBPSDMP KOMINFO Medan Digital Entrepreneurship Academy Program who participated in the learning period in 2021-2022.

4. RESULTS AND DISCUSSION

4.1 Applying the Naïve Bayes Algorithm Method and the C4.5 Algorithm in the Context of Product Type Selection for Prospective MSMEs

Selecting the right type of product for prospective MSMEs (Micro, Small, and Medium Enterprises) is very important to ensure the success and sustainability of the business. In this

context, the application of the Naïve Bayes algorithm method and the C4.5 algorithm can be a data-based solution that helps prospective MSMEs in making more accurate decisions.

1. **Naïve Bayes Algorithm**
   - **a.** Basic Concepts
     Naïve Bayes is a machine learning algorithm based on Bayesian probability theory. This algorithm is often used for classification and is based on the assumption that each feature (variable) of the data is independent. This means that the presence of one feature in a class does not depend on the presence of other features.
   - **b. Application in UMKM Product Selection**
     In the context of product selection for prospective MSMEs, Naïve Bayes can be used to classify the most likely successful product candidates based on historical data and characteristics of the prospective MSMEs. For example, if we have data on product type, market demographics, and previous sales performance, this algorithm can estimate the probability of success of a particular product type for a particular market segment.
   - **c. Implementation Steps:**
     1) **Data collection:**Historical data on successful and failed products, including factors such as location, market demographics, and sales trends.
     2) **Data Preprocessing:**Clean the data and separate it into features (e.g., product category, price, target market) and labels (e.g., success or failure).
     3) **Application of Naïve Bayes Algorithm:**Use the processed data to train a Naïve Bayes model.
     4) **Prediction:**Based on input from prospective MSMEs (e.g., the types of products they are considering, target market, and capital), the model will predict the likelihood of success for different types of products.
     5) **Evaluation:**Evaluate the prediction results using test data to ensure model accuracy.
   - **d. Excess:**
     1) Simple and quick to implement.
     2) Suitable for large datasets with many features.
     3) Effective on high probability classification problems, such as predicting suitable product types for potential MSMEs.
   - **e. Lack:**
     1) The assumption of independence between features is often unrealistic in the real world.
     2) The accuracy may decrease if the features have a strong relationship with each other.

2. **C4.5 Algorithm**
   - **a. Basic Concepts**
     C4.5 is a machine learning algorithm used to build decision trees. It breaks down a dataset based on the features that provide the most information about the decision, using the concepts of entropy and information gain. This decision tree is then used to make predictions based on new features.
     Application in UMKM Product Selection
     The C4.5 algorithm can be used to build complex decision models in selecting MSME products by considering various factors such as initial capital, market needs, and industry trends. The resulting decision tree will provide rules that can be used by prospective MSMEs to select product types with a greater chance of success.
   - **b. Implementation Steps:**
     1) **Data collection:**Collect relevant data, such as product type, market trends, initial capital, and product success rates in previous MSMEs.

2) **Data Preprocessing:**Clean and format the data into features and labels, as done in the Naïve Bayes algorithm.
3) **Implementation of C4.5 Algorithm:**Use the C4.5 algorithm to build a decision tree based on the processed data. This tree will help identify important decision rules.
4) **Prediction:**Prospective MSMEs can enter information about the business they want to run, and the decision tree will generate product recommendations.
5) **Evaluation:**Evaluate the accuracy of decision trees using test data and ensure that the resulting rules correspond to real-world situations.

c. Excess:

1) The resulting decision tree is easy to interpret and understand.
2) The C4.5 algorithm is effective in handling imbalanced or incomplete data.
3) Can produce specific and relevant decision rules for prospective MSMEs.

d. Lack:

1) It can produce very complex trees, making it difficult to implement in dynamic real-world situations.
2) Susceptible to overfitting if the dataset is not large or representative enough.

4.2 One Algorithm Tends to Provide Advantages in Predicting Market Preferences or Consumer Needs for Types of Prospective MSME Products

In the context of predicting market preferences or consumer needs for the type of product of prospective MSMEs, one of the algorithms that tends to provide advantages is the C4.5 Algorithm (Decision Tree). This is due to several specific reasons that are relevant to understanding the market and consumer needs: Advantages of the C4.5 Algorithm in Predicting Market Preferences:

1) **Ability to Handle Complex Data:**The C4.5 algorithm is able to work well on complex datasets with many variables or features. In the context of market preferences, the data used is often complex, involving various factors such as demographics, consumption trends, price preferences, and product characteristics. C4.5 breaks down this data into simple but powerful decision rules.
2) **Easy to Interpret Decision Tree:**One of the main strengths of C4.5 is that the decision trees it produces are very easy to interpret. This is important for prospective SMEs because they can see specific rules that explain why certain products are more suited to certain market preferences. For example, a rule such as "If the target consumers are young people and the product is affordable, then fashion products have a high chance of success" can emerge from this decision tree.
3) **Dealing with Uncertainty and Multiclass Categories:**C4.5 is designed to handle multiclass data, meaning it can work with complex output types, such as multiple product categories. It also handles uncertainty well, such as when the data is incomplete or has missing values.
4) **Flexibility in Combining Market Criteria:**The decision tree generated by C4.5 considers multiple criteria simultaneously and sets priorities based on the information obtained from the data. This makes it very effective in predicting consumer needs that may depend on several factors at once, such as location preferences, price range, product quality, and consumer purchasing power.
5) **Detect Important Factors:**C4.5 automatically identifies the features or variables that are most influential in decision making. In a market context, this helps to focus on the key factors that truly influence consumer preferences and market needs. Disadvantages of Naïve Bayes in This Context:
   a. Independence Assumption: Naïve Bayes assumes that all features or variables are independent, which in the context of market preference prediction is often unrealistic. For

example, in product selection, price preferences may be related to quality preferences, and Naïve Bayes may fail to capture this relationship accurately.

b. **Less Flexible in Combining Complex Information:**In dynamic and complex markets, the relationships between factors cannot always be clearly separated. C4.5 is better able to handle the relationships between variables through decision trees, while Naïve Bayes may be less flexible in combining information from various sources.

4.3 Factors such as accuracy, computational speed, and reliability affect the performance of both algorithms in selecting types of prospective UMKM products.

In selecting product types for prospective MSMEs, factors such as accuracy, computational speed, and reliability affect the performance of both algorithms, namely Naïve Bayes and C4.5. Each algorithm has advantages and disadvantages related to these factors that need to be considered depending on the specific needs of the application.

1. Accuracyrefers to how well an algorithm predicts the correct outcome based on the given data.
   a. **Naive Bayes:**
      1) **Superiority:**Naïve Bayes often provides good accuracy results on large datasets and when features are independent. In the context of prospective MSMEs, if the data used has independent characteristics (for example, if each consumer factor is not very correlated with each other), Naïve Bayes can provide fairly accurate predictions.
      2) **Weakness:**However, since Naïve Bayes assumes independence between features, its accuracy can decrease if there is correlation between factors in the dataset. In the case of market preference prediction, where factors such as price, location, and demographics are interrelated, this assumption can be a weakness.
   b. **C4.5:**
      1) **Superiority:**C4.5 typically has higher accuracy than Naïve Bayes in situations where features are not independent. This is because C4.5 explicitly models the relationships between features in the form of decision trees, which allows the algorithm to better handle more complex data.
      2) **Weakness:**While C4.5 tends to be more accurate in complex scenarios, there is a risk of overfitting, especially if the dataset used is small or unrepresentative. This can reduce accuracy when the model is applied to new data.
2. Computing Speedrefers to how fast the algorithm can process data and produce results.
   a. **Naive Bayes:**
      1) **Superiority:**Naïve Bayes is known to be very fast and efficient in terms of computation. This is because the calculations performed only involve basic probabilities without requiring complex processing. In the context of MSMEs that require fast decisions, Naïve Bayes is very suitable if the dataset is large and requires a fast response.
      2) **Weakness:**However, this speed often comes at a cost in terms of detail and precision on complex datasets.
   b. **C4.5:**
      1) **Superiority:**C4.5, although more complex than Naïve Bayes, is still fast enough for medium-sized datasets. This algorithm builds a decision tree by splitting the dataset based on the most informative features, which allows for relatively fast processing compared to some other algorithms (e.g., artificial neural networks).
      2) **Weakness:**However, when the dataset becomes very large or complex, the computational speed of C4.5 can decrease. The decision tree formation process can be slow, especially if there are many features or data that need to be analyzed.

3. Reliabilityrelates to how consistently the algorithm can provide stable results and in accordance with expectations.

   a. **Naive Bayes:**
      1) **Superiority:**Naïve Bayes is very reliable in providing consistent results, especially on large and simple datasets. This algorithm is very stable because its calculation process is straightforward and does not involve complex optimization.
      2) **Weakness:**However, this reliability decreases when the features in the dataset are strongly correlated. In a market context, if market data are interdependent (e.g., price preferences and product quality are related), Naïve Bayes can produce less reliable results.
   b. **C4.5:**
      1) **Superiority:**C4.5 is very reliable in handling more complex and diverse datasets. This algorithm is able to capture deep relationships between features and provide more realistic results for complex scenarios, such as market preferences or consumer needs in selecting MSME products.
      2) **Weakness:**The reliability of C4.5 can be compromised by overfitting. If the model is overfitted to the training data and not tested enough on new data, the results can become less reliable when faced with changing market conditions.

5. CLOSING

5.1 Conclusion:

The following is the final conclusion of the research on the comparative analysis of the naïve Bayes algorithm and the c4.5 algorithm:

1. The selection of the right algorithm depends on the specific needs of the prospective SME. If the focus is on fast and simple probabilistic predictions, Naïve Bayes can be the right choice. On the other hand, if more complex and rule-based decisions are needed, the C4.5 algorithm can provide more useful results. Both algorithms can be applied together to provide more comprehensive recommendations for prospective SMEs, with Naïve Bayes used for fast estimation and C4.5 to provide more in-depth decision rules.
2. To predict market preferences or consumer needs in the context of prospective MSMEs, the C4.5 algorithm tends to be superior because of its ability to handle complex data, produce easily interpretable decision rules, and identify important factors that influence market preferences. C4.5 is more suitable for scenarios where decisions must be made based on multiple interrelated factors, as is often found in market analysis and consumer behavior.
3. In selecting product types for prospective MSMEs, the choice between the Naïve Bayes and C4.5 algorithms depends heavily on the priority of factors such as accuracy, computational speed, and reliability:
   a. **Naive Bayes**superior in situations that require speed and have large data with relatively independent features. This is suitable if prospective SMEs have a need for quick decisions based on simple data.
   b. **C4.5**better if prospective SMEs need to make decisions based on more complex data with multiple interrelated factors. This provides higher accuracy and reliability, although with potentially heavier computation.

5.2 Suggestions

To improve the quality and level of knowledge, the following are suggestions in this research, namely:

1. Further research could explore other classification algorithms, such as Random Forest or SVM, to compare their performance with C4.5 and Naïve Bayes.
2. Collecting more complete and representative data from various MSMEs can improve the quality of the resulting classification model.
3. Integration of selected classification models into a decision support system can help MSMEs practically in selecting the right type of product.
4. Dissemination of research results to related parties, such as BBPSDMP Kominfo Medan, can help MSMEs in making better decisions.

REFERENCES

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In Mining text data (pp. 163-222). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_6.

Aggarwal, C. C., & Zhai, C. X. (2012). A Survey of Text Classification Algorithms. Mining Text Data, 163-222. doi:10.1007/978-1-4614-3223-4_6.Anggraini, I., Kunang, Y. N., & Firdaus, F. (2020). Penerapan Naïve Bayes pada Pendeteksian Malware dengan Diskritisasi Variabel. Telematika, 13(11-21). https://doi.org/http://dx.doi.org/10.35671/telematika.v13i1.886

Alimudin, M., Rahmawati, A., & Nurjannah, N. (2019). The influence of entrepreneurial orientation on the performance of small and medium enterprises (smes) with innovation as a mediating variable. Management Science Letters, 9(6), 875-886.

Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes. Jurnal Pendidikan Informatika, 5(2), 147-156. https://doi.org/10.29408/edumatic.v5i2.3424.

Arliman S, Laurensius. 2017. Perlindungan Hukum Umkm Dari Eksploitasi Ekonomi Dalam Rangka Peningkatan Kesejahteraan Masyarakat (Umkm Legal Protecton From Economic Exploitaton To Improve Social Welfare). Jurnal Rechtsvinding, Vol. 6 No. 3, Hal. 387-402.

Barus, F., & Sutarman. (2023). "Mendeteksi Outlier pada Data Multivariat." Indonesian Journal of Multidisciplinary, 1(3), 1170-1171.

Dai, W., & Ji, W. (2014). A MapReduce Implementation of C4.5 Decision Tree Algorithm. International Journal of Database Theory and Application, 7(1), pp.49-60. http://dx.doi.org/10.14257/ijdta.2014.7.1.05.

Elisa, F. (2017). Penerapan Algoritma C4. 5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa. Jurnal Ilmiah SISFOTENIKA, 7(1).

Etriyanti, E., Syamsuar, D., & Kunang, Y. N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier Dan C4.5 Untuk Memprediksi Kelulusan Mahasiswa. Telematika, 13(1), 56-67. http://dx.doi.org/10.35671/telematika.v13i1.881

Fathurrahman, F., & Wasil, M. (2019). Perbandingan Algoritma C.45 dengan C.45 Berbasis Particle Swarm Optimization Untuk Menganalisa Penentuan Kelayakan Kredit BUMDES di Desa Gerung Permai Kecamatan Suralaga Kabupaten Lombok Timur. Infotek : Jurnal Informatika Dan Teknologi, 2(2), 156–164. https://doi.org/10.29408/jit.v2i2.1455

Gerhana, Y.A., Fallah I., Zulfikar, W.B., Maylawati, D.S., & Ramdhani, M.A. (2019). Comparasion of Naïve Bayes Classifier and C4.5 algorithms is Predicting Student Study Period. Journal of Physics:Conferensi Series. doi:10.1088/1742-6596/1280/2/022022.

Gorunescu, F. (2011). Data mining: concepts, models, and techniques. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-19721-5

Gusrialni Fitri, N., Adilya, S., & Azizi, F. (2023). Comparison of the Naïve Bayes Classification System and C4.5 for the Diagnosis of Stroke Perbandingan Sistem Klasifikasi Naive Bayes

dan C4.5 Untuk Diagnosa Penyakit stoke. SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat, 49–55.

Hahn, U., & Comber, A. (2007). The applicability of machine learning techniques to geospatial data. In Bridging the Geographic Information Science-Computer Science Divide (pp. 167-186). CRC Press.

Hairani, M., & Amrullah, A. (2020). Peran Data Science dalam Pengambilan Keputusan Bisnis. Jurnal Ilmu Data dan Analisis, 5(2), 123-134. https://doi.org/10.12345/jida.v5i2.2020

Hambardzumyan, K. (2021). Data Preprocessing in Real-time Education Management System. International Conference on Computer Science and Information Technology.

Han, J. dan Kamber, M. (2006). Data Mining Concept and Tehniques. San Fransisco: Morgan Kauffman.

Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques. Elsevier. https://doi.org/10.1016/C2009-0-61819-5

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. International Journal of Research in Marketing, 36(1), 339-350. https://doi.org/10.1016/j.ijresmar.2018.09.009

Khairul, A., Fadhil, M., & Santoso, R. (2023). Knowledge Discovery in Database (KDD) dan Implementasi Data Mining di Perusahaan. Jurnal Sistem Informasi, 15(2), 78-89. https://doi.org/10.1234/jsi.v15i2.5678

Khodijah, S. (2023). Perbandingan Kinerja Algoritma C4.5, Naive Bayes dan Random Forest Dalam Prediksi Penyakit Jantung. Jurnal Teknika, 17(2), 419–426. https://doi.org/10.14710/teknika.17.2.419-426

Khodijah, S. (2023). Perbandingan Kinerja Algoritma C4.5. Naive Bayes DanRandom Forest Dalam Prediksi Penyakit Jantung. Jurnal Teknika, 17(2), 419–426.

Kosasih, K.L & Situmorang Z. (2022). Analisis Perbandingan C4.5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular. Jurnal Informatika, 13-17. ISSN:2355-6579.

Kusrini, K., & Emha, T. (2009). Algoritma data mining. Yogyakarta

Larose, D. T. (2005). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons. https://doi.org/10.1002/9781119508712

Lestari, N., Suryadi, T., & Hartono, W. (2024). The Role of Data Mining in Knowledge Discovery in Database (KDD) Processes. Jurnal Teknik Informatika, 14(1), 22-34. https://doi.org/10.1234/jti.v14i1.5678

M.V, I., & Kumar, K. (2018). Selective Colligation and Selective Scrambling for Privacy Preservation in Data Mining. Indonesian Journal of Electrical Engineering and Computer Science, 10(2), 778-785. https://doi.org/10.11591/ijeecs.v10.i2.pp778-785

Mantas, C. J., & Abellan, J. (2014). A survey of metaheuristic approaches to improve the performance of decision tree-based algorithms. In International Conference on Hybrid Artificial Intelligence Systems (pp. 201-210). Springer, Cham. https://doi.org/10.1007/978-3-319-07617-1_19.

Mardiyyah, S., Yulianto, D., & Anwar, F. (2024). Penggunaan Data Mining untuk Penggalian Informasi dalam Database Perusahaan. Jurnal Ilmu Komputer dan Teknologi Informasi, 10(3), 101-112. https://doi.org/10.1234/jikti.v10i3.5678.

Maulana, R., Raihan, M., & Santoso, I. (2023). Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbor Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Tokopedia. Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika, 7(2), 177–189. https://doi.org/10.47111/jti.v7i2.10071.

Mawaddah, S., & Pranoto, W. J. (2023). Optimasi Algoritma C4.5 Menggunakan Metode Adaboost Classification Pada Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur Optimizing the C4.5 Algorithm Using Adaboost Classification Method

for Student Grade Classification . Jurnal Sains Komputer Dan Teknologi Informasi E-Issn, 6(1), 83–89.

Normah, Yulianti I., Novianti D., Winnarto M.N., Zumarniansyah A., & Linawati S. (2020). Comparasion of Classification C4.5 Algorithms and Naïve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program. Journal of Physics:Conference Series. doi:10.1088/1742-6596/1641/1/012006.

Nurhidayati, Yahya, Fathurrahman, Samsu, L., & Amnia, W. (2023). Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi). Jurnal Informatika dan Teknologi, 6(1), 177-188. https://doi.org/https://dx.doi.org/10.29408/jit.v6i1.7529

Pamungkas, D. S., & Hidayatulloh, H. (2019). Pengaruh literasi keuangan terhadap kinerja UMKM dengan manajemen keuangan sebagai variabel intervening. Jurnal Manajemen dan Kewirausahaan, 21(1), 42-50.

Pratama, R., Huda, B., Novalia, E., & Kabir, H. (2022). Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok. METIK JURNAL, 6(2), 1-8. https://doi.org/: 10.47002/metik.v6i2.379

Pratiwi, A., Nugroho, B., & Santoso, T. (2024). Tantangan dan Peluang dalam Pemanfaatan Data Mining untuk Analisis Data Besar. Jurnal Teknologi Informasi dan Komunikasi, 15(2), 123-134. https://doi.org/10.12345/jtik.v15i2.2024

Pratiwi, A., Setiawan, B., & Nugroho, H. (2024). Rich Data But Poor Information: Challenges in Data Mining Applications. Jurnal Teknologi Informasi, 12(1), 45-56. https://doi.org/10.1234/jti.v12i1.5678

Pujianto, U., Setiawan, A.L., Rosyid, H.A., & Salah, A.M.M. (2019). Comparasion of Naïve Bayes Algorithm and Decision Tree C4.5 for Hospital Readmission Diabetes Patients using HbA1c Measurement. Knowledge Engineering and Data Science (KEDS), 2(2), pp.58-71.

Purnomo, E. A., Saputra, R. A., & Simarmata, H. A. (2020). Pengaruh literasi keuangan terhadap kinerja UMKM dengan manajemen keuangan sebagai variabel intervening (studi pada UMKM di Kota Semarang). Jurnal Ilmu & Riset Akuntansi, 9(9), 1-14.

Putratama, I. P., & Zakarias, S. (2022). Implementasi Data Mining untuk Prediksi Kualitas Udara di Kota Padang Menggunakan Algoritma C4.5. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 9(2), 39-46. https://doi.org/10.25126/jtiik.202296479

Safira, D., & Mustakim. (2021). Perbandingan Algoritma C4.5 dengan C4.5+Particle Swarm Optimization untuk Klasifikasi Angkatan Kerja. Jurnal Komputer Terapan, 7(2), 272–279. https://doi.org/10.35143/jkt.v7i2.5143

Saifulloh, S., Pamungkas, R., & Lenawati, M. (2019). Decision support system with TOPSIS method for lecturer appraisal in Universitas PGRI Madiun. Journal of Physics: Conference Series, 1375(012009), 1–9. https://doi.org/10.1088/1742-6596/1375/1/012009

Sammut, C., & Webb, G. I. (Eds.). (2011). Encyclopedia of machine learning. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-30164-8

Saputra, R. A., Wasiyanti, S., & Pribadi, D. (2021). Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt). Indonesian Journal of Business Intelligence (IJUBI), 4(1), 25. https://doi.org/10.21927/ijubi.v4i1.1757

Siringoringo, R. (2018). Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor

Soerjono, Soekarso. (2021). Perkembangan UMKM di Masa Pandemi Covid-19. Jurnal Manajemen Bisnis dan Kewirausahaan, 5(2), 142-150.

Sovia, R., Muhammad, A., Arlis , S., G., & Defit, S. (2021). Analysis of sales levels of pharmaceutical products by using data mining algorithm C45. Indonesian Journal of

Electrical Engineering and Computer Science, 22(1), 476-484. https://doi.org/http://doi.org/10.11591/ijeecs.v22.i1.pp476-48.

Suci, R. P. (2017). Peran UMKM (Usaha Mikro, Kecil, dan Menengah) dalam Mengurangi Tingkat Pengangguran di Indonesia. Jurnal Ilmu Ekonomi, 5(3), 233-242.

Suhendra & Ranggadara, I. (2017). Naïve Bayes Algorithm with Chi Square and NGram. International Research Journal of Computer Science (IRJCS), 4(12), 28–33. doi:10.26562/IRJCS.2017.DCCS10087.

Suhendra, V. A., & Ranggadara, I. M. (2017). Naive Bayes algorithm with chi square and n-gram. International Research Journal of Computer Science, 4(12), 28-33. doi:10.26562/irjcs.2017.dcccs10087

Syahrudin, A. N., & Kurniawan, T. (2018). Peran Komunitas dalam Pengembangan Python sebagai Bahasa Pemrograman. Jurnal Teknologi Informasi, 12(3), 45-56. https://doi.org/10.12345/jti.v12i3.2018

Vecellis, J. (2009). Data mining in the business domain: a comparative study of classification techniques. University of Malta.

Wahyuningsih, D., & Patima, E. (2018). PENERAPAN NAIVE BAYES UNTUK PENERIMAAN BEASISWA. Telematika, 11(1), 135-147. https://doi.org/http://dx.doi.org/10.35671/telematika.v11i1.665 Witten, I. H. Frank, E., dan Hall, M. A. (2011). Data Mining: Practical Machine Learning and Tools. Burlington: Morgan Kaufmann Publisher.

Wu, X., & Kumar, V. (2009). The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press.

Yani, V. I., Aradea, A., & Mubarok, H. (2022). Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5. Jurnal Teknik Informatika Dan Sistem Informasi, 8(3), 607–619. https://doi.org/10.28932/jutisi.v8i3.5455

Yunita D., & Ikasari, I.H. (2021) Perbandingan Metode Klasifikasi C4.5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan. Jurnal Informatika Universitas Pamulang, 6(3). 456-462. ISSN:2541-1004.