Classification of Students' Interest Patterns in Sebiru-Biru Senior High School to Continue Their Education to Higher Education with CRISP-DM

Enda Suran Barus, Gerry Firmansyah, Budi Tjahjono and Agung Mulyo

March 5, 2025

CLASSIFICATION OF STUDENTS' INTEREST PATTERNS IN SEBIRU-BIRU SENIOR HIGH SCHOOL TO CONTINUE THEIR EDUCATION TO HIGHER EDUCATION WITH CRISP-DM

Enda Suran New [1] , Gerry Firmansyah [2] , Budi Tjahjono [3] , Agung Mulyo Widodo [4]
[1,2,3,4] Esa Unggul University, Jakarta, Indonesia
[1]endasurany@student.esaunggul.ac.id, [2]gerry@esaunggul.ac.id, [3]budi.tjahjono@esaunggul.ac.id, agung.mulyo@esaunggul.ac.id

Abstract

The quality of education greatly affects the quality of human resources in the future. In Indonesia, students' interest in continuing their education to college is often influenced by various factors, such as economic conditions, personal aspirations, family influences, and information obtained regarding career choices, and so on. The quality of education in our country compared to other countries is very worrying, as we all know and see. SMA Masehi Sebiru-biru is a high school that has successfully graduated an average of 100 students each year. With a percentage of only ±20% of students who graduate who continue their education to college, which means that there is still a lack of awareness of SMA Masehi Sebiru-biru students about the importance of continuing their education to college. Based on this, the author wants to find out what factors influence the interest of SMA Masehi Sebiru-biru students to continue their education to college using the Decision Tree and Naïve Bayes methods.

Keywords : Data Mining , Decision Trees , Naïve Bayes.

1. Introduction

In the era of globalization and rapid development of information technology, the quality of education greatly affects the quality of human resources in the future. The younger generation with good education has greater ability to face global challenges and increase the competitiveness of the future nation (Rahman, 2020). In Indonesia, students' interest in continuing their education to college is often influenced by various factors, such as economic conditions, personal aspirations, family influences, and information obtained regarding career choices, and so on. The quality of education in our country compared to other countries is very worrying, as we all know and see. The current problem of education is the low quality at various levels of education, both formal and informal, due to a lack of human resources (Khoirunnisa, 2021) . Thus, although the importance of higher education is increasingly recognized, there are still challenges in understanding the pattern of student interest in continuing their education to college level and what factors influence students in making this decision.

Sebiru-biru Christian High School is one of the high schools located Jalan deli tua - penen km 36.5 periaria village, Sibiru-biru district, Deli Serdang Regency, North Sumatra Province. This high school has 100 students in each class, the students come from various family backgrounds and different economic levels. This high school has also succeeded in graduating an average of 100 students each year. With a percentage of only 100 students who graduate who continue their education to college, which means that there is still a lack of awareness of $\pm 20\%$Sebiru -biru High School students about the importance of continuing their education to college. From the results of the author's brief discussion with several students at this school, it was found that some of them still do not understand the importance of continuing their education to college. Several students expressed that there were limited funds as an obstacle, even though there were many government scholarships that could be a solution. In addition, there were also those who stated that their parents did not allow them, who considered that high school education was sufficient to find a job. Based on this discussion, the author was interested in analyzing the interest patterns of high school students in continuing their education to college, as well as understanding what factors most influenced students' decisions to continue their education to college. To find out these factors, it is necessary to carry out data analysis and processing using data mining (Nas, 2021).

Data mining, as one of the data analysis techniques, offers an effective approach to extracting information from a dataset. Through this technique, researchers can identify patterns and factors that may not be visible with conventional analysis methods (Mardi, 2020). By using data mining, this study aims to analyze the pattern of student interest in continuing their education to college, as well as identify the factors that influence these students' decisions.

In this study, the author will use classification with the aim of predicting whether the student has the potential to continue college or not based on the data and factors that have been analyzed. In addition, this study is expected to provide new insights into how student interest in college can be influenced by certain factors for the field of education, and through this study it is also expected to be able to help schools to find out what factors influence students' interest in making decisions to continue their education to college, so that schools can design more targeted coaching or guidance programs, in order to increase students' interest in continuing their education to college in the following years.

Based on the background described above, the author will conduct research on "Analysis of Student Interest Patterns to Continue Education to Higher Education with Data Mining", the author hopes that this research can provide a significant contribution to the development of education in Indonesia.

2. Literature Review

From the results of the Systematic Literature Review (SLR) that the author has conducted, it can be seen that there are several methods used by previous researchers in predicting interest patterns, namely research conducted by (Nas, 2021), (Ayunda et al., 2024), (Dina et al., 2024), (Situmorang, 2024), (Doahir & Annisa, 2022), (Sari & Imam, 2023), (Sadat et al., 2023) these researchers used the Decision Tree method in the classification process, and from the five studies an average accuracy value of 85.6% was obtained, which means that the Decision Tree method is good at predicting existing data sets, so that if used, the results obtained will be in accordance with reality. The next research is research conducted by (Kriestanto & Femmy, 2021), (Handoko & Muhammad, 2021), (Lizar et al., 2023), Ninosari & Jhoanne, 2022) these studies use the Naïve Bayes method in the classification process of the data set to be analyzed, and from this study an average accuracy value of 84.5% was obtained, which means that the Naïve Bayes method is also a good method for predicting data.

According to the results obtained from several previous studies that have been conducted, it was found that the Decision Tree method and the Naïve Bayes method are both good methods in predicting the data set to be analyzed, where the average accuracy value for the Decision Tree method was 85.6% and the average accuracy value obtained for the Naïve Bayes method was 84.5%.

There are several previous studies that compare the Decision Tree method, Naïve Bayes, and other methods, namely research conducted by (Khoirunnisa et al., 2021), (Anam et al., 2022) in these studies the best accuracy value was obtained by the Decision Tree method, but in the study (Kunjumon et al., 2023) which compared the Decision Tree method, Naïve Bayes, and other methods the best accuracy value was obtained by the Naïve Bayes method. In the study (Budiman & Zatin, 2021) obtained the same accuracy value between the Decision Tree and Naïve Bayes methods.

Based on this, the researcher is interested in comparing the Decision Tree and Naïve Bayes methods to predict the analysis of the interest patterns of Sebiru-biru High School students to continue their education to college.

3. Methods

3.1 Data Mining Classification

Classification in data mining is a technique for predicting categories or class from data based on patterns that have been found in historical data that has been labeled (labeled data). This technique falls into the category of supervised learning , where the model is trained using data that already has clear labels or classes. Once the model is trained, it can be used to predict the class of unlabeled data (new data) (Novitasary et al., 2024).

3.2 CRISP-DM (Cross Industry Standard Process For Data Mining)

CRISP-DM (Cross Industry Standard Process for Data Mining) is a process model used in data mining to provide guidance in managing and running data mining projects (Wirth & Jipp, 2000). CRISP-DM is one of the most popular methodologies in both industry and academia, due to its flexibility that can be applied across industries and types of data (Mariscal et al., 2010). This model helps in planning and managing the data mining project life cycle systematically. The benefits of using CRISP-DM are reducing costs and time, as well as minimizing the need for knowledge for data mining projects. In addition, accelerating training, knowledge transfer, documentation, and capturing best practices are also benefits of using CRISP-DM (Chapman et al. 1999). The stages that will be carried out in CRISP-DM are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Basiri et al. , 2024). For more details on the stages in CRISP-DM, see Figure 1 below.

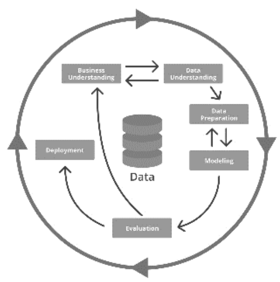

Figure 1 : Tahapan CRISP-DM (Cross Industry Standard Process for Data Mining)

3.3Research Stages

In this study, the author will use the stages of the Cross-Industry Standard Process for Data Mining (CRISP DM). The stages that will be carried out in the Cross-Industry Standard Process for Data Mining are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Basiri et al. , 2024). The stages that will be carried out in the Cross-Industry Standard Process for Data Mining (CRISP DM) can be seen in Figure 2 below.

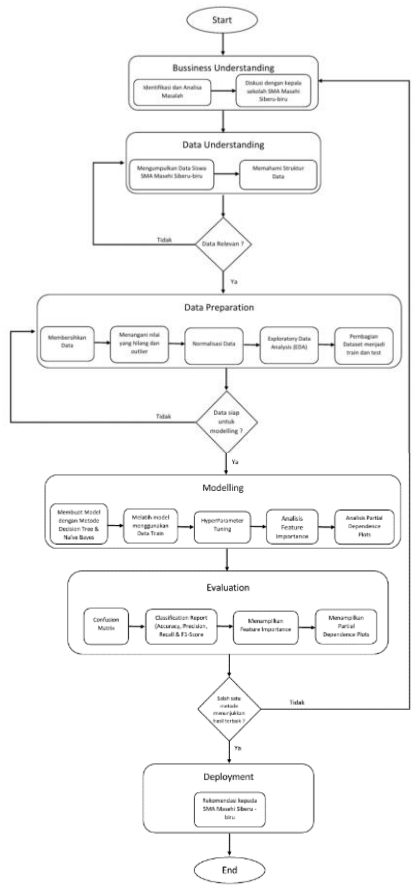

Figure 2 : Research Flow Diagram

3.3.1 Business Understanding

This stage is to understand customer needs in depth. Activities carried out at this stage are determining business objectives, assessing the situation of resource availability, determining data collection objectives, and producing project plans (Basiri et al. , 2024). In this study, the author found that SMA Masehi Sebiru-biru needed an analysis related to student interests in continuing their education to college.

3.3.2 Data Understanding

The next stage is the data understanding stage, namely identifying, collecting, and analyzing data sets that can help achieve project objectives (Basiri et al. , 2024). Activities at this stage are collecting initial data, explaining data, exploring data, and verifying data quality. At this stage, the researcher saw that only 20% of SMA Masehi Sebiru-biru students were interested in continuing their education to college, and held discussions with the principal of SMA Masehi Sebiru-biru.

3.3.3 Data Preparation

This phase is often called " data mining ", which is preparing the final data set for modeling. The activity in this phase is to improve the quality of the data to suit the modeling process that will be carried out next (Basiri et al. , 2024). At this stage, the author will pre-process the data using phyton so that the data to be analyzed becomes a data set that is ready to be used in analysis or modeling with Decision Tree and Naïve Bayes. The pre-processing steps are as follows (Safdara et al., 2024):

a. Data Cleaning, is the process of identifying and fixing data problems, such as input errors, duplication, and cleaning up missing or empty values. This stage has the main goal of producing consistent and accurate data sets.

b. Data Integration, identifying and resolving inconsistencies between data from multiple sources, such as differences in units of measurement or date formats, then combining disparate data sets while maintaining consistent data structures and formats.

c. Data Transformation (Data Transformation), this section includes

- Normalization is adjusting the scale of data values to a certain range, such as 0-1, to facilitate comparison.
- Standardization, which is changing the data so that it has a mean of 0 and a standard deviation of 1. This helps some machine learning algorithms work more effectively.
- Encoding, which is converting categorical variables into numeric ones through methods such as one-hot encoding or label encoding. This allows machine learning algorithms to process categorical data.

d. Data Reduction, is a data reduction process to reduce the volume of data without losing important information. This can speed up processing and analysis.

3.3.4 Modeling

Creating and evaluating various models based on several different modeling techniques. At this stage, there are four tasks, namely choosing a modeling technique, generating a test design, building a model, and finally evaluating the model (Basiri et al. , 2024). In this study, the data set that has been pre-processed will be processed using machine learning techniques. The data set will be analyzed using the Decision Tree and Naïve Bayes methods.

3.3.5 Evaluation

This evaluation phase looks more broadly at the model that best suits the business and what to do next. There are three activities that represent the evaluation phase, namely evaluation of results, review process, and determination of next steps (Basiri et al. , 2024). From the results of modeling and processing of the data set that has been carried out, an evaluation of the results will be carried out, namely the accuracy value obtained using the Decision Tree and Naïve Bayes methods. Furthermore, the author will find out what factors have the most influence on the interest of SMA Masehi Sebiru-biru students to continue their education to college. After these factors are found, the author will consult with experts to provide solutions to SMA Masehi Sebiru-biru to help teachers take the next steps so that the interest of SMA Masehi Sebiru-biru students to continue their education to college increases.

3.3.6 Deployment

This is the last and most important stage of the CRISP-DM process. Planning for deployment begins with the business understanding phase and must incorporate not only the generation of model values, but also how to convert decision scores and the integration of those decisions. Once the accuracy value is obtained, the factors influencing the interest of SMA Masehi Sebiru-biru students to continue their education to college are known, and the solutions provided by experts are obtained, the school will take further action to increase students' interest in continuing their education to college.

4. Results & Discussion

4.1 Data Attributes

In this study, the data used are data from students of SMA Masehi Sebiru-biru majoring in science and social studies, where the attributes used include student biodata, such as name, address, parents' occupation, parents' income per month, and so on, as well as students' subject scores from semester 1 to semester 6.

From the data obtained, statistics show that parental income varies quite significantly between students, with most parents of students having an income of around 4 million. The year of graduation tends to be distributed around 2022-2024. The label 'Continue to College' shows that most students do not continue to college because class 0 is much larger than class 1, namely students who continue to college. With a percentage of 86.3% for students who do not continue to college, and 13.6% who continue to college for science classes, and for social studies classes with a percentage of 92.3% for students who do not continue to college and 7.6% for students who continue to college.

4.2 Modeling for Data Class IPA and IPS

In this study, the classification models used will be Decision Tree and Naïve Bayes.

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt
from sklearn import tree


# Membuat dan melatih model Decision Tree
model = DecisionTreeClassifier(max_depth=5, random_state=42)
model.fit(X_train, y_train)

# Memprediksi hasil pada data uji
y_pred = model.predict(X_test)

# Menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print(f"Akurasi: {accuracy:.2f}")

# Menampilkan classification report
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Figure 3 : Modelling for Data Class IPA with Decision Tree Method

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import LabelEncoder
import numpy as np

model = GaussianNB()
model.fit(X_train, y_train)

# Prediksi dengan data uji
y_pred = model.predict(X_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy * 100:.2f}%")

# Menampilkan classification report (precision, recall, f1-score)
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Figure 4 : Modelling for Data Class IPA with Naïve Bayes Method

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt
from sklearn import tree

# Membuat dan melatih model Decision Tree
model = DecisionTreeClassifier(max_depth=3, random_state=42, min_samples_split=5, min_samples_leaf=3)
model.fit(X_train, y_train)

# Memprediksi hasil pada data uji
y_pred = model.predict(X_test)

# Menghitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print(f"Akurasi: {accuracy:.2f}")

# Menampilkan classification report
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Figure 5 : Modelling for Data Class IPS with Decision Tree Method

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, classification_report
from sklearn.preprocessing import LabelEncoder
import numpy as np

model = GaussianNB()
model.fit(X_train, y_train)

# Prediksi dengan data uji
y_pred = model.predict(X_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy * 100:.2f}%")

# Menampilkan classification report (precision, recall, f1-score)
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
```

Figure 6 : Modelling Data Class IPS with Naïve Bayes Method

4.3 Evaluation

The following are the evaluation results obtained from the model created above.

4.3.1 Evaluation Model Decision Tree For Science Class

Decision tree model that has been created has an accuracy of 89% , which shows that the model is quite good at predicting overall results. Then, from the results above it can also be seen that class 0 on the label 'Continue College' which means students who do not continue college is predicted very well, this can be seen from the results of precision 96%, recall 92%, f1-score 94%. While for class 1 on the label 'Continue College' which means students are interested in continuing college has a lower performance, this can be seen from the results of precision 60%, recall 75%, f1-score 67%, which shows that the model is more difficult to predict students who will continue college.

4.3.2 Evaluation Of Naïve Bayes Model For Class IPA

Naïve bayes model that has been created has an accuracy of 90% , which shows that the model is very good at predicting overall results. In addition, from the results above it can also be seen that the model can predict class 0 on the label 'Continue College' which means students who do not continue college perfectly, this can be seen from the results of precision 1.00 , recall 0.96 , f1-score 0.98 . However, for class 1 on the label 'Continue College' which means students who continue college has a low precision of 0.50 , but a recall of 1.00 , which means that the model can predict all students who actually continue college, but only 50% of the predictions that say they continue college are correct. For the macro avg and weighted avg values provide an overview of the overall performance of the model.

4.3.3 Features Importance of Class IPA

That there are 5 (five) attributes that contribute to the model prediction results, namely parents' income/month, Indonesian history value, English value and compulsory mathematics value, and student address. For parents' income/month with an importance value of 0.425747 has the highest importance value , meaning that this feature provides the greatest contribution in the model to predict the target. For the values of the courses that provide contributions are Indonesian history value, compulsory mathematics value, and English value.

4.5.6 Evaluation Model Decision Tree For Class IPS

The model that has been created has an accuracy of 98% , which means that most of the predictions are correct. For the prediction of class 0 on the label 'Continue to College' which means students who do not continue to college are predicted very well, both in precision 0.97 , recall 1.00 , and f1-score 0.98 . However, the model cannot predict at all for class 1 on the label 'Continue to College' which means students continue to college, because the precision and recall for this class are 0. This is likely due to the imbalance between class 0 and 1 data which can be seen from the macro avg value showing low performance due to the very extreme class imbalance, with one class that is very dominant.

4.5.7 Evaluation Model Naïve Bayes Class IPS

The naive bayes model that has been created based on the social studies class data has an accuracy of 96% , which shows that the model as a whole is quite good at predicting the results. For the prediction of class 0 on the label 'Continue to College' which means students who do not continue to college are predicted very well, this can be seen from the results of precision 0.93 , recall 1.00 , f1-score 0.96 . However, the model is less good at predicting class 1 on the label 'Continue to College' which means students continue to college, because in recall only 33% are correct even though the model is able to predict 100% in class 1 with a precision of 1.00 .

4.5.8 Features Importance of Class IPS

The most influential attributes are parents' income/month and mandatory math scores. For parents' income/month, it gives an importance value of 0.84 and is the attribute that gives the largest contribution compared to the others. The mandatory math attribute also contributes to the model although it is smaller compared to the parents' income/month attribute.

5. Conclusion and Suggestions

5.1 Conclusion

a. The decision tree and naïve Bayes methods on science and social studies class data are that both decision trees and naïve Bayes have equally good accuracy results, depending on the data set used.
b. The interest of SMA Masehi Sebiru-biru students to continue their studies is very low, only 13.6% for science classes and 7.6% for social studies students.
c. The most influential factor on students' decisions to continue their education to college, both science and social studies students, is their parents' monthly income, although there are other factors such as students' math and English scores, but their influence is very small. With a percentage for science classes of 42.5% of parents' monthly income, 13% of math scores, and 7% of English scores. While for social studies classes, the percentage is 84% for parents' monthly income, and 15% for math scores.
d. As a solution for SMA Masehi Sebiru-biru to increase the interest of its students to continue their education to college, it can be done by conducting socialization to parents about the importance of continuing their education to college, helping students to find information about the availability of government scholarships, establishing cooperation with universities and educational institutions that can provide educational funding assistance for students who want to continue their education, establishing good relations with alumni groups who continue their education, and helping students find information about more affordable education such as online courses. or distance education, for the problem of subject grades, this can be done by the school creating additional classes for mathematics and English subjects for free and helping students find information about online classes that they can take online or at a relatively low cost.

5.2 Suggestion

Here are some things that can be input for further research :

1. Further research can develop this study by collecting more data and more attributes for more complex results.
2. Further research can be conducted using other data mining methods so that the results can be compared with this research.

6. References

Abdelsamea, M. M., Zidan, U., Senousy, Z., Gaber, M. M., Rakha, E., & Ilya, M. (2001). A survey on artificial intelligence in histopathology image analysis. *Wires Data Mining and Knowledge Discovery*.

Anam, K., Bani, N., Christina, J. (2022). *Komparasi Algoritma Klasifikasi Data Mining Menggunakan* Optimize Selection *untuk Peminatan Program Studi*. *Building of Informatics, Technology and Science* (BITS), 4(2), 606-613.

Ayunda, Y. S., Hartama, D., Lubis, M. R., & Gunawan, I. (2024). *Analisis pola minat siswa lulusan SMU/SMK untuk melanjutkan kuliah dengan menggunakan algoritma C4.5*. *Technology and Informatics Insight Journal*, *4*(9), 581-595.

Basiri, M. A., Mohammad, P., & Zahra, S. (2024). Implementing CRISP-DM and Logistic Regression for Predictive Implementing CRISP-DM and Logistic Regression for Predictive Analysis in Financial Transactions: A Case Study Analysis in Financial Transactions: A Case Study. IRAIS

B.M. Monjurul, A. & Matthew, C. (2018). Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia. I.J. *Information Technology and Computer Science,* 2, 1-9.

Budiman., & Zatin, N. (2021). *Perbandingan Klasifikasi Data Mining Untuk Penelusuran Minat Calon Mahasiswa Baru*. *Jurnal Nuansa Informatika*, 15(2), 37-52.

Budiman., & Zatin, N., (2022). Comparative Analysis of Data Mining Classification Algorithm Performance for Searching Prospective Student Interests. SISTEMASI: *Jurnal Sistem Informasi*, 11(2), 271-290.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz T., & Shearer, C. (1999). DM SIG Workshop in Brussels.

Dasarathy, B. V. (1991). Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques. *Los Alamitos: IEEE Computer Society Press*.

Doahir, A., & Annisa, N. Q. (2022). *Analisis Potensi Siswa Untuk Kuliah Dengan* Classification *Menggunakan Metode* Decision Tree. *Jurnal POROS TEKNIK*, 14(1), 28-32.

Dodo, L., Nenden S., Gerry F., Habibullah A. (2023). Analysis of Tuberculosis Disease Case Growth From Medical Record Data, Viewed Through Clustering Algorithms (Case Study: Islamic Hospital Bogor). Journal of Indonesia Social Science, 4(09), 915-927.

Handoko, W., & Muhammad, I. (2021). *Prediksi Peminatan Program Studi Pada Penerimaan Mahasiswa Stmik Royal Menggunakan* Naïve Bayes. *Journal of Science and Social Research*, 4(2), 231-235.

Hidayatsyah M.R. (2013). *Penerapan Metode* Decision Tree *dalam Pemberian Pinjaman Kepada Debitur dengan Algoritma C4.5*. (Skripsi Sarjana, UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU).

Hidayat, R. (2022). *Pemanfaatan data mining untuk melihat minat siswa setelah menyelesaikan pendidikan sekolah menengah atas (SMA) dengan algoritma* K-Means Clustering. *Technology and Informatics Insight Journal*, *1*(2).

Jadhav, S. D., & Channe, H. P. (2016). Comparative Study of K-NN, Naïve Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR)*, https://doi.org/10.21275/v5i1.nov153131.

K. Ahmed, and T. Jesmin, *"Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach*," *IJSE*, 7(2), 155–160.

Kang, Z. (2019). Using Machine Learning Algorithms to Predict First-generation College Students' Six-year Graduation: A Case Study. I.J. *Information Technology and Computer Science*, 9, 1-8.

Khoirunnisa, Lia, S., Ira, T. R., & Lilis, S. (2021). *Prediksi Siswa SMK Al-Hidayah yang Masuk Perguruan Tinggi dengan Metode Klasifikasi*. *Jurnal Informatika*, 8(1), 26-33.

Kriestanto, D., & Femmy D. A. (2021). *Penerapan Naïve Bayes untuk Analisis Faktor-Faktor Pemilihan STMIK AKAKOM sebagai Tempat Studi. JTKSI*, 4(2), 75-81.

Kunjumon, L.T., Sharon, S., Saffi, T. S., Tahsneem, & N., Neena, J. (2019). An Intelligent System to predict Students academic performance using Data Mining. *International Journal of Information Systems and Computer Sciences*, 8(2), 128-131.

Kusrini & Emha T.L. (2009). Algoritma Data Mining. Yogyakarta : CV. Andi Offset.

Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). A Study on C.5 Decision tree Classification Algorithm for Risk Predictions During Pregnancy. Procedia Technology, 24, 1542–1549. https://doi.org/10.1016/j.protcy.2016.05.128.

Lavindi, E. E., Wijanarto, W., & Rohmani, A. (2019). *Aplikasi* Hybrid Filtering *Dan* Naïve Bayes *Untuk Sistem Rekomendasi Pembelian Laptop*. *JOINS (Journal of Information https://doi.org/10.33633/joins.v4i1.2518.*

Lizar, Y., Alya, S. F., Asriwan, G., & Joko, S., (2023). Data Mining Analysis to Predict Student Skills Using Naïve Bayes Method. *Knowbase : International Journal of Knowledge in Database*. 3(2), 150-159.

Malik, L. A., Mia, K., & Firman, N.H. (2023). Faktor-Faktor Yang Mempengaruhi Minat Calon Mahasiswa Baru Mendaftar Pada Ftii Uhamka Menggunakan Algoritma *K-Nearest Neighbor (K-Nn). Infotech: Journal Of Technology Information*, 9(1), 85-94.

Mardi, Y. (2020). "Data Mining : *Klasifikasi Menggunakan Algoritma* C4.5," *Jurnal Edik Informatika*, 2(2), 213- 219.

Mariscal, G., Oscar, M., & Covadonga, F. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2), 137-166.

McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naïve Bayes Text Classification. *AAAI-98 Workshop on Learning for Text Categorization*, 41-48.

Mienye I. D., Sun Y., & Wang Z. (2021). Prediction performance of improved decision tree-based algorithms: A review, 35(1), 698–703, doi: 10.1016/j.promfg.2019.06.011.

Nas, C. (2021). *Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131-145.

Ninosari, D., & Jhoanne, F. (2022). *Sistem Pendukung Keputusan Hasil Rekomendasi Jurusan Perguruan Tinggi Menggunakan Metode* Naïve Bayes *dan AHP*. *SATIN – Sains dan Teknologi Informasi*, 8(1), 106-117.

Novitasary, D., Yunianita R., & Suprianto. (2024). *Klasifikasi Minat Karir Siswa Sekolah Menengah Atas Menggunakan Algoritma C4.5. Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 13(1), 711 – 724.

Permana A. P., Ainiyah K., Fahmi K., & Holle H. (2021). *Analisis Perbandingan Algoritma* Decision Tree , KNN , dan Naïve Bayes *untuk Prediksi Kesuksesan* 49 Start-up, 6(3), 178–188.

Purboyo, R. A. (2022). Clustering *Data Rekomendasi Penjurusan Ke Perguruan Tinggi Menggunakan Metode* K-Means *Studi Kasus SMA Negeri 2 Palembang*. Skripsi Sarjana. Universitas Bina Darma.

Putri, D. A., Bayu, H., Sarika, A., & A B, P. (2019). *Prediksi Program Studi Berdasarkan Nilai Siswa Dengan Algoritma Backpropagation* (*Studi Kasus Sman 6 Depok Jurusan Ips*). *Jurnal Informatik*, 15(2), 69-78.

Quinnlan, J. R. (1993). *C4.5: Programs for Machine Learning*. *Morgan Kaufmann Publishers.*

Rahman, A. (2020). *Pengaruh metode pembelajaran inovatif terhadap hasil belajar siswa*. *Jurnal Pendidikan dan Kebudayaan*, 15(3), 123-136.

Rohayani, H., & Muhammad, C. U. (2022). *Prediksi Penentuan Program Studi Berdasarkan Nilai Siswa dengan* Algoritma Backpropagation. *Journal of Information System Research* (JOSH), 3(4), 651-657.

Sadat, A. M., Pujiono., Anggun, P., & Sholihul, I. (2023). Comparison Of Algorithm Between Classification & Regression Trees And Support Vector Machine In Determining Student Acceptance In State Universities. *Jurnal Teknik Informatika (JUTIF)*, 4(6), 1589-1604.

Safdara, M. F., Robert, M. N., & Piotr, P. (2024). Pre-Processing techniques and artificial intelligence algorithms for electrocardiogram (ECG) signals analysis: A comprehensivereview. *Computers in Biology and Medicine, 17(2), 1-16.*

Sari, R. N., & Imam, P. (2023). *Data Mining Peminatan Mata Kuliah Pilihan Mahasiswa Tingkat Akhir Jurusan Informatika Menerapkan Algoritma C4.5. Bulletin Of Computer Science Research*, 3(3), 263-269.

Setiawan, I., & Heri, N. (2023). *Data Mining Pengelompokkan Nilai untuk Menentukan Minat Mata Pelajaran pada Siswa SMA Negeri 1 Kota Gajah*. *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, 9(1), 284-292.

Situmorang, S. A. (2024). *Seleksi Calon Siswa Baru Dengan Metode Decision Tree Menggunakan Algoritma C5.0 (Studi Kasus : SMK N. 1 Doloksanggul*). (Skripsi Sarjana, Universitas Medan Area).

SMA Masehi Sibiru-biru. (2023, November 22). Struktur Organisasi. Situs Resmi SMA Masehi Sibiru-biru. https://www.smamasehibiru.sch.id/

Nofitasary, D., Yunianita, R., & Suprianto S. (2024). *Klasifikasi Minat Karir Siswa Sekolah Menengah Atas Menggunakan Algoritma C4.5*. Jurnal Ilmiah Teknik Informatika dan Sistem Informasi, 13(1), 711-724.

Sadat, M. A., Pujiono., Anggun, P., Sholihul, I. (2023). Comparison Of Algorithm Between Classification & Regression Trees And Support Vector Machine In Determining Student Acceptance In State Universities. *Jurnal Teknik Informatika, 4(6).*

Schuster, C., Stebner, F., Leutner, D., & Wirth, J. (2020). Transfer of metacognitive skills in self-regulated learning: an experimental training study. *Metacognition and Learning*, 15(3), 455–477.

Sartika, D., & Dana I. S. (2017). *Perbandingan Algoritma Klasifikasi* Naïve Bayes, Nearest Neighbour, *dan* Decision Tree *pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian*. *Jurnal Teknik Informatika dan Sistem Informasi*, 1(2), 151-161.