PASCAL Journal of Computer Science and Informatics
https://jurnal.devitara.or.id/index.php/komputer
E-ISSN: 3047-5074 Volume 2 Nomor 1 Tahun 2024

Application of the Decision Tree Method as a Factor Causing Students to Drop Out (Case Study: SMK Tunas Pelita Kota Binjai)

[1]Intan Darma Yanti, [2]Marto Sihombing, [3]I Gusti Prahmana
[1,2,3] STMIK Kaputama, Binjai, Indonesia

ARTICLE INFO

Keywords:
Factor Causing Students,
Drop Out,
Decision Tree,
C4.5

ABSTRACT

The problem of the phenomenon of dropping out of school for students at the Senior High School or Vocational Education level is increasingly widespread. In fact, education is a very important field to develop the quality of Human Resources (HR) for the country. Ayu et al said that one of the important sectors that directly contributes the most in developing the quality of Human Resources (HR) is the education sector. Meanwhile, SMK Tunas Pelita Binjai is one of the vocational high schools in Binjai City that also faces the problem of dropping out of school. Based on internal school data, in 2023 there are 32 students or around 5% of the total students at SMK Tunas Pelita Binjai who cannot complete their education until the end. This figure is quite high and is a serious concern for the school. Some of the factors such as students' interest in learning have decreased so much that they make students inactive or often alpha, parents with low levels of education tend to lack understanding and appreciation for the importance of formal education for their children's future, they may not have high aspirations for the achievement of children's education, economic impacts that are very binding on the family so that children prefer to work to help the family economy. Seeing the very complex problem of dropping out of school, a comprehensive and systematic approach is needed to analyze what factors are the causes. The application of the Decision Tree method in identifying the factors that cause school dropouts at SMK Tunas Pelita Binjai, is expected to produce new information that can be used as a basis for schools to design more effective strategies and interventions in preventing and reducing school dropout rates. The author uses the help of RapidMiner software to see the factors that cause students to drop out of school by using the C4.5 Decision Tree method to create a decision tree from existing student data. So that from the RapidMiner calculation process, it can be concluded that what are the factors that cause the student to drop out of school.

Email : intandarmayanti07@gmail.com

INTRODUCTION

The problem of the phenomenon of dropping out of school for students at the Senior High School or Vocational Education level is increasingly widespread. In fact, education is a very important field to develop the quality of Human Resources (HR) for the country. Ayu et al. said that one of the important sectors that directly contributes the most in developing the quality of Human Resources (HR) is the education sector (Ayu et al. 2014). Thus, education cannot be considered a trivial problem, because education is the main factor in the progress of a nation. Dropout is one of the problems in the world of education that must be looked for to the root cause. Especially during the period of vocational education and equivalent, it is the final phase of the government program of compulsory learning for 12 years. Only then is someone considered worthy to enter the world of work.

Judging from data from the Data and Information Technology Center of the Ministry of Education and Culture, the number of students dropping *out* of school has increased again in the 2022/2023 school year. The dropout rate at various levels of education reached 76,834 people, with details of the number at the high school level 10,091 people, and vocational school 12,404 people. Of this figure, the highest dropout rate is at the vocational school level, which is 0.24%. Statistics and Indicators of Gender-Oriented Education show that the number of male dropouts is greater than that of women. The comparison of the number of male students who drop out of school compared to female students reaches 15.29%. Meanwhile, the largest gender ratio of school dropouts is found in vocational schools, which is 25.99. This means that there are 25.99% more male dropouts compared to female school dropouts.

Meanwhile, SMK Tunas Pelita Binjai City is one of the vocational high schools in Binjai City that also faces the problem of dropping out of school. Based on internal school data, in 2023 there were 32 students or around 5% of the total students at SMK Tunas Pelita Kota Binjai who could not complete their education until the end. This figure is quite high and is a serious concern for the school. As for some of the factors such as students' interest in learning has decreased so much that it makes students inactive or often alpha, parents with low levels of education tend to lack understanding and appreciation of the importance of formal education for the future of their children, they may not have high aspirations for the achievement of children's education, The economic impact is very binding on the family so that children prefer to work to help the family economy. Seeing the very complex problem of dropping out of school, a comprehensive and systematic approach is needed to analyze what factors are the causes.

This research aims to contribute to the development of data analysis methods in the context of education. The application of Data Mining techniques, one of the methods that can be used is *the Decision Tree* built with the C.45 algorithm. In this study, the researcher chose *the Decision Tree* method because it is based on the advantages of the *Deision Tree C4.5* algorithm which can manage various data and has such a high level of accuracy. According to Fatma and Rochmawati, the C4.5 algorithm and *the Decision Tree* are two algorithm models that cannot be separated, therefore, in order to create a decision tree, the *C4.5 Decision Tree algorithm is needed* (Fatma & Rochmawati, 2024).

METHOD

The research method is an overview of the steps so that research can be carried out in a structured manner, so a framework is prepared from the beginning to the achievement of the final result as follows:

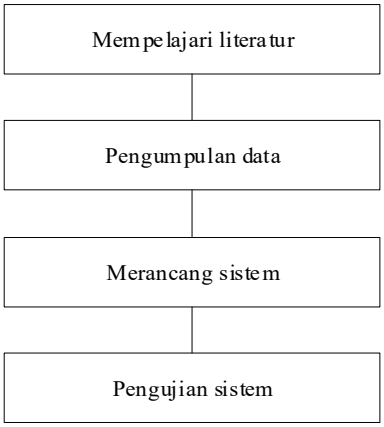

Figure 1. Research Methods

Based on the research framework that has been described above, the discussion of each stage in the research can be described as follows:

a. Studying Literature

At this stage, a search for theoretical foundations obtained from various books and the internet is carried out to complete the treasury of concepts and theories, so that they have a good and appropriate foundation and science.

b. Data Collection

Data collection by using or collecting written sources, by reading, studying and recording important things sourced from books, journals and the internet related to the problem being discussed in order to obtain a theoretical picture.

c. Designing the System

At this stage, the system is designed using *RapidMiner 9* software.

d. System Testing

At this stage, the application that has been previously designed has been completed and the system review stage is carried out whether there are *errors* or damage to the system that has been designed.

RESULTS AND DISCUSSION

The system to determine the factors that cause students to drop out of school uses the C4.5 method in the process using different student data datasets using datasets based on *Microsoft Excel* files. The dataset of student data is as follows:

Table 1. Attributes to be used

Code	Attribute
A1	Attendance
A2	Father's Education
A3	Father's Work
A4	Maternal Education
A5	Mother's Work
A6	Parents' Income
A7	Number of Children

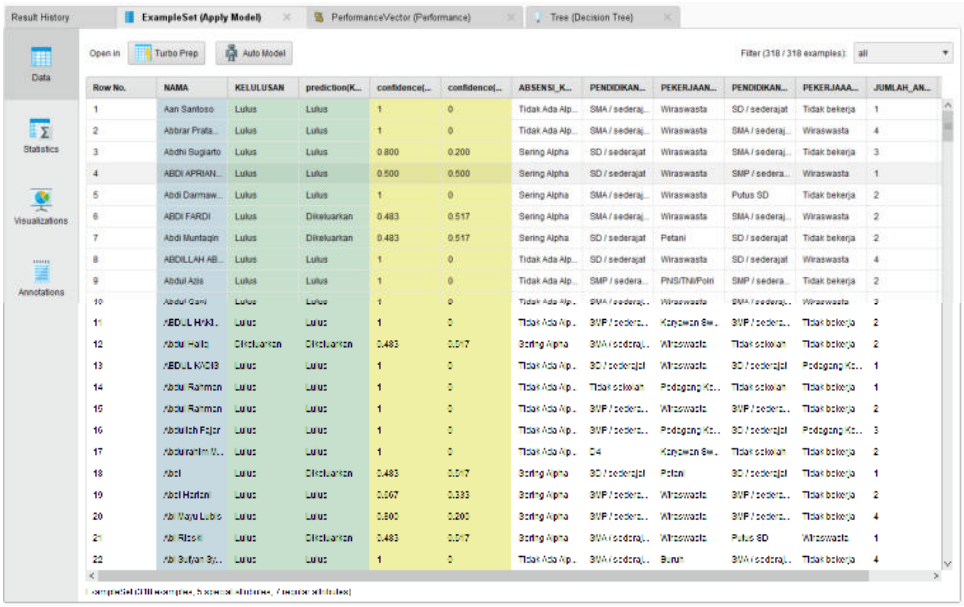

Figure 2. ExampleSet

The RapidMiner shown refers to the dataset or sample used to train and evaluate the C4.5 Decision Tree model. Each row in an ExampleSet represents a single piece of data or example, while each column represents an attribute or feature of that data. In the image, it can be seen that the ExampleSet contains information regarding the individual's name, graduation status, attendance frequency, education level, occupation, and number of children. This data set will be processed by the C4.5 Decision Tree algorithm to build a model that can predict a person's graduation status based on other attributes. The results obtained from *the RapidMiner software* using the C4.5 method can produce rule associations based on the input student data dataset, along with the relationship display of each result item obtained from the application.

PASCAL Journal of Computer Science and Informatics
https://jurnal.devitara.or.id/index.php/komputer
E-ISSN: 3047-5074 Volume 2 Nomor 1 Tahun 2024

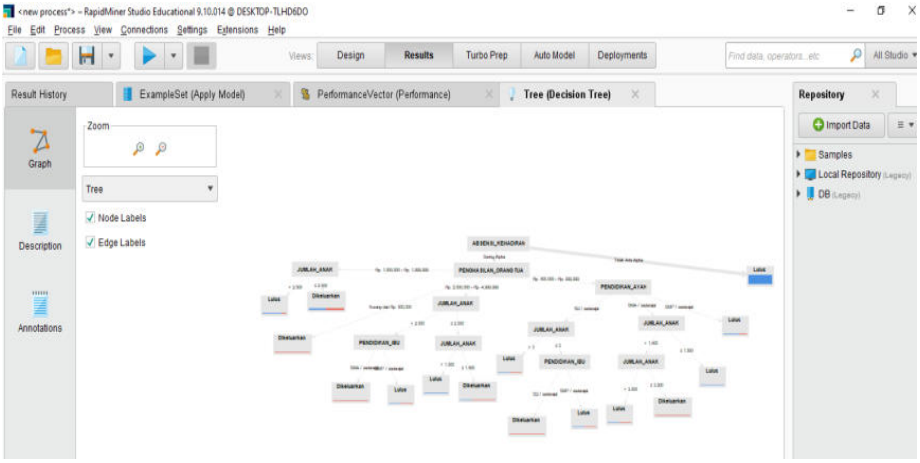

Figure 3. Decision Tree Results

The image above represents the prediction model built by *RapidMiner* to predict whether a student will graduate or not. This model uses several factors as considerations, such as the frequency of student absenteeism, father's income, number of children, and parents' education level. Each box on the chart represents a decision taken based on the value of a particular attribute. The line connecting these boxes indicates the flow of decisions. The blue color on the end box (leaf) indicates a "Pass" prediction, while the red color indicates a "Not Pass" prediction. By following the decision path from top to bottom, it is possible to see how the model makes predictions based on a combination of values from various attributes.

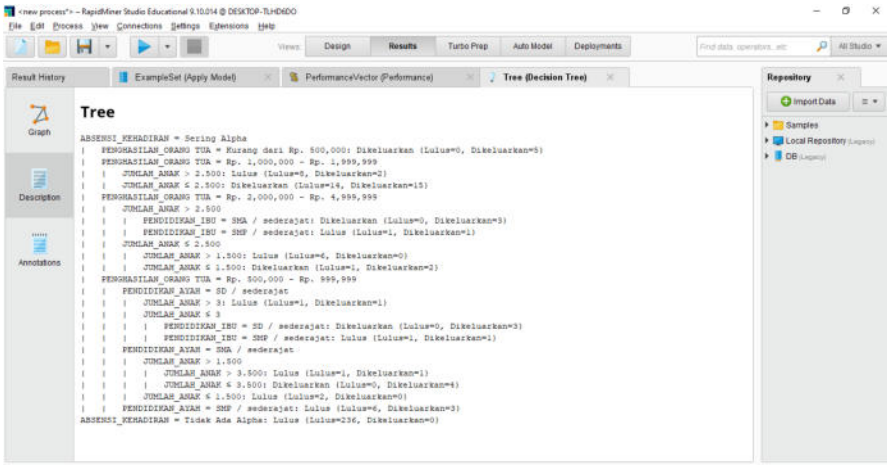

Figure 4. Result Tree Description

The Tree (*Decision Tree*) in the image above is a description of the prediction model that has been built by the *C4.5* Decision Tree algorithm on *RapidMiner*. This model is shaped like an inverted tree, where each branch represents a decision based on the value of a certain attribute, such as the frequency of absences, the father's income, the number of children, and the level of education. Each leaf on the tree shows the result of the prediction, that is, whether a student will pass or not.

Tabel 2. Performance Vector

	True Pass	True Issued	Class Precision
Pred. Pass	262	9	96.68%
Pred. Issued	15	32	68.09%
Class Recall	94.58%	78.05%	

The Performance Vector in the table provides an idea of how well the C4.5 Decision Tree model has been built in predicting student graduation. *This PerformanceVector* presents a variety of performance metrics, such as accuracy, precision, and recall. Accuracy shows the overall percentage of correct predictions, whether they are passable or non-passable predictions. Precision measures how accurate the model is in predicting students who actually graduate. Recall measures how many students who actually graduated are successfully identified by the model.

Based on the results obtained by using *the RapidMiner software*, several factors that cause students to drop out of school are obtained, including:

Table 3. Causative Factors

No	Factor
1	If a student is often alpha and his parents' income is less than Rp. 500,000, then the student is expelled.
2	If a student is often alpha, his parents' income is between Rp. 1,000,000 - IDR 1,999,999, and the number of child members is more than 2,500, then the student graduates.
3	If a student is often alpha, his parents' income is between Rp. 1,000,000 - IDR 1,999,999, and the number of children is less than or equal to 2,500, then the student is expelled.
4	If a student is often alpha, his parents' income is between Rp. 2,000,000 - IDR 4,999,999, the number of children is more than 2,500, and the education of high school mothers or equivalent, then the students are expelled.
5	If a student is often alpha, his parents' income is between Rp. 2,000,000 - IDR 4,999,999, the number of children is more than 2,500, and the education of junior high school mothers or equivalent, then the student graduates.
6	If a student is often alpha, his or her income is between Rp. 2,000,000 - IDR 4,999,999, the number of children is less than or equal to 2,500, and the number of children is more than 1,500, then the student passes.
7	If a student is often alpha, his parents' income is between Rp. 2,000,000 - IDR 4,999,999, the number of children is less than or equal to 2,500, and the number of children is less than or equal to 1,500, then the student is expelled.
8	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is elementary or equivalent, the number of children is more than 3, then the student graduates.
9	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is elementary school or equivalent, the number of children is less than or equal to 3, and the mother's education is elementary school or equivalent, then the student is expelled.
10	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is elementary school or equivalent, the number of children is less than or equal to 3, and the mother's education is junior high school or equivalent, then the student graduates.
11	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is high school or equivalent, the number of children is more than 1,500, and the number of children is more than 3,500, then the student graduates.
12	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is high school or equivalent, the number of children is more than 1,500, and the number of children is less than or equal to 3,500, then the student is expelled.
13	If a student is often alpha, his parents' income is between Rp. 500,000 - IDR 999,999, the father's education is high school or equivalent, and the number of children is less than or equal to 1,500, then the student graduates.
14	If a student is often alpha, and his parents' income is between Rp. 500,000 - IDR 999,999, and his father's education is junior high school or equivalent, then the student graduates.
15	If a student is not often alpha, then the student graduates.

Discussion

System Preparation

Before proceeding to configure the system of factors that cause students to drop out of school using the C4.5 method, the first step that needs to be taken is to install *the RapidMiner software*. *The RapidMiner software* used is version 9. After the installation is complete, the main display of *the RapidMiner software will appear.*

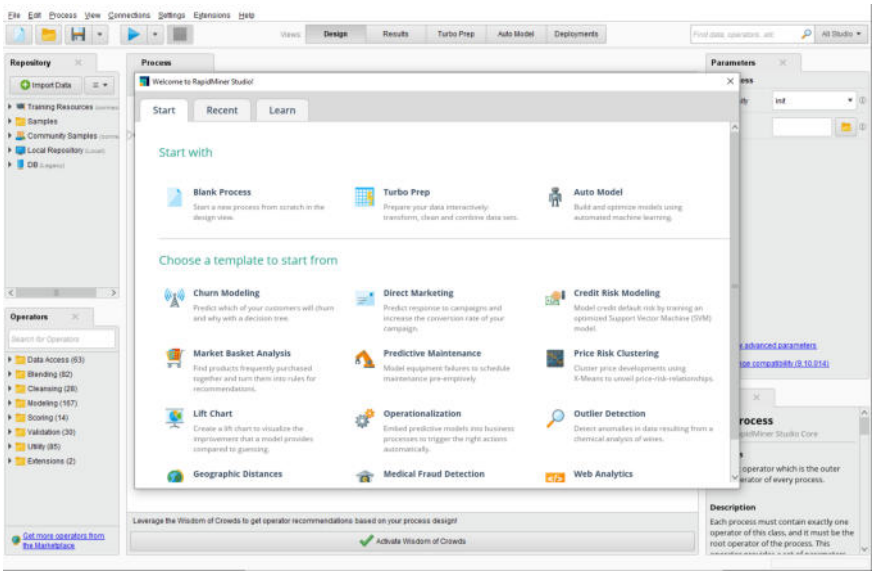

Figure 5. RapidMiner Display

System Design

After the software preparation has been prepared, it is followed by system design to find out the factors that cause students to drop out of school. The requirements of the design to make the C4.5 method are as follows:

a. Read Excel

This operator reads the dataset from the specified *microsoft excel* file . This operator can read data from Excel 95, 97, 2000, XP, and 2003.

b. Set Role

A feature used to define the role or function of each attribute (column) in the dataset. By defining roles, machine learning models can better understand how each attribute contributes to the learning process. Common roles that can be assigned to attributes include: target (the variable you want to predict), input (the variable used to predict), and ID (a unique attribute for each piece of data).

c. Decision Tree

One of the machine learning algorithms used to create prediction models in the form of diagrams such as trees. Each branch on a tree represents a rule or condition, and each leaf represents the result of a prediction. This algorithm works by dividing data repeatedly based on the value of certain attributes, thus generating a decision tree that can be used to classify new data or predict the value of a target variable.

d. Apply Model

operator used to apply the pre-trained model to the new dataset. Once a model is built through a machine learning process, this operator allows us to use the model to make predictions or classifications on data that the model has never seen before.

e. Performance

Performance in *RapidMiner* refers to how well a machine learning model can perform a given task. Model performance is typically evaluated based on certain metrics, such as accuracy, precision, recall, and F1-score. These metrics are used to measure how often the model gives correct predictions, how

often the model gives correct positive predictions, how many positive cases the model successfully identifies, and the balance between precision and recall.

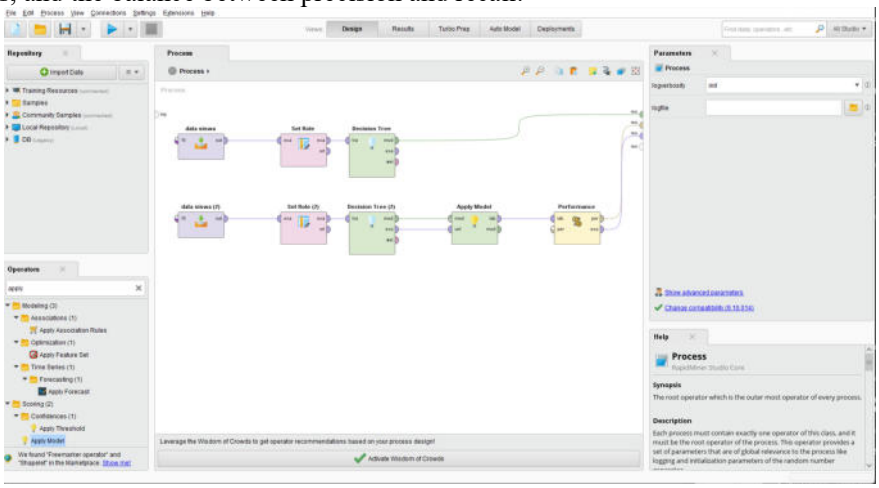

Figure 6. System Design Display

CONCLUSION

Data processing to determine the factors of students dropping out of school uses a Microsoft excel file containing student data and Rapid Miner software using the C4.5 method in the analysis process. By using Rapid Miner software, analysis can be carried out by utilizing the feature of the Decision Tree to generate a decision tree from attributes on each student data, namely attendance, father's work, mother's employment, father's education, mother's education, number of children and parents' income. For now, the variables used in the process of determining the factors of students dropping out of school are attendance, father's work, mother's employment, father's education, mother's education, the number of children and parents' income. In the future, it is necessary to add other variables or attributes so that the analysis becomes more accurate. In the process of knowing the factors of students dropping out of school, they still use the help of Rapid Miner software which only aims to analyze student data. In the future, it is necessary to create a school academic system so that the system not only aims to analyze student data but can support the school's academic process on data in one database.

REFERENCES

Anwar, S. N., Fatkhul Amin, & Nugroho, S. (2014). Desain Uml Aplikasi Navigasi Layanan Kesehatan Berbasis Android. *Seminar Nasional Sistem Informasi Indonesia*, (September), 250–254.

Apriyadi, A., Lubis, M. R., & Damanik, B. E. (2022). Penerapan Algoritma C5.0 Dalam Menentukan Tingkat Pemahaman Mahasiswa Terhadap Pembelajaran Daring. *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, *11*(1), 11–20.

Assa, R., Kawung, E. J. R., & Lumintang, J. (2022). Faktor Penyebab Anak Putus Sekolah Di Desa Sonuo Kecamatan Bolangitang Barat Kabupaten Bolaang Mongondow Utara. *Jurnal Ilmiah Society*, *2*(1), 1–12.

Ayu, Ni, Krisna Dewi1, Anjuman Zukhri1, Ketut Dunia2, dan Jurusan Pendidikan Ekonomi. (2014). "()." Tahun 4(1). www.cetak.kompas.com,.

Bachtiar, L., & Mahradianur, M. (2023). Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai. *Jurnal Informatika*, *10*(1), 28–36.

Bahri, S., & Lubis, A. (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. *Jurnal Sintaksis*, *2*(1), 63–70.

Fatma, Yofi Lailatul, dan Naim Rochmawati. (2024). "Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4.5." Journal of Informatics and Computer Science 05.

GoodStats. (n.d.). Gender Gap di Indonesia, Angka Putus Sekolah Didominasi oleh Laki-Laki. diakses dari https://goodstats.id/article/gender-gap-di-indonesia-angka-putus-sekolah-didominasi-oleh-laki-laki-uLr22

Hidayat, A. N., Eka M, E. Y., Abdullah, F. M., Akbar, M., & Rosyani, P. (2021). Analisis Perkembangan Kecerdasan Buatan Dalam Industri Game. *Jurnal Kreativitas Mahasiswa Informatika*, *2*(1), 118–120.

Hidayatuloh, M. A., Kartika, K. P., & Permadi, D. F. H. (2022). Implementasi Algoritma C4.5 Untuk Memprediksi Capaian Pembelajaran Daring (Studi Kasus Siswa MAN 3 Blitar). *Jurnal Algoritme*, *3*(1), 33–47.

Nahjan, R., Heryana, N., & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 101–104.

Nufus, N., Ariffin, D. M., Satyawan, A. S., Nugraha, R. A. S., Asysyakuur, M. I., Marlina, N. N. A., … Ema, E. (2021). Sistem Pendeteksi Pejalan Kaki Di Lingkungan Terbatas Berbasis SSD MobileNet V2 Dengan Menggunakan Gambar 360° Ternormalisasi. *Prosiding Seminar Nasional Sains Teknologi Dan Inovasi Indonesia (SENASTINDO)*, *3*(November), 123–134.

Nuraini, P., Tata Hardinata, J., & Pranayama Purba, Y. (2022). Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Kepuasan Pelayanan E-KTP Di Kantor Camat Pematang Bandar. *Resolusi : Rekayasa Teknik Informatika Dan Informasi*, *2*(3), 138–144.

Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(1), 8–17.

Raharjo, J. S. D., Damiyana, D., & Hidayatullah, M. (2016). Sistem Pakar Diagnosa Penyakit Lambung dengan Metode Forward Chaining Berbasis Android. *Sisfotek Global*, *6*(2), 1–8.

Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, *3*(2), 1860–1874.

Rizal, R., Martanto, M., & Arie Wijaya, Y. (2022). Analisa Dataset Software Defined Network Intrusion Menggunakan Algoritma Deep Learning H2O. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 747–757.

Salam, N. F. S., Manap Rifai, A., & Ali, H. (2021). Faktor Penerapan Disiplin Kerja: Kesadaran Diri, Motivasi, Lingkungan (Suatu Kajian Studi Literatur Manajemen Pendidikan Dan Ilmu Sosial). *Jurnal Manajemen Pendidikan Dan Ilmu Sosial*, *2*(1), 487–508.

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Suryasari, Callista, A., & Sari, J. (2012). Rancangan Aplikasi Customer Service Pada PT. Lancar Makmur Bersama. *Jurnal Sistem Informasi (JSI)*, *4*(2), 468–476.

Wijaya, A. E., Swastika, W., & Kelana, O. H. (2021). Implementasi Transfer Learning Pada Convolutional Neural Network Untuk Diagnosis Covid-19 Dan Pneumonia Pada Citra X-Ray. *Sainsbertek Jurnal Ilmiah Sains & Teknologi*, *2*(1), 10–15.

Yahya, & Mahpuz. (2019). Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat. *Jurnal Informatika Dan Teknologi*, *2*(2), 1–19.

Yulia, E. R. (2017). Perancangan Program Penjualan Perhiasan Emas Pada Toko Mas Dan Permata Renny Medan. *Evolusi : Jurnal Sains Dan Manajemen*, *5*(2).

Yulianeu, A., & Oktamala, R. (2022). Sistem Informasi Geografis Trayek Angkutan Umum Di Kota Tasikmalaya Berbasis Web. *JUTEKIN (Jurnal Teknik Informatika)*, *10*(2).

Yunus, M., Ramadhan, H., Aji, D. R., & Yulianto, A. (2021). Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indon[1] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)," Paradig. - J. Komput. dan Inform., vol. *Paradigma - Jurnal Komputer Dan Informatika*, *23*(2).