Application of C4.5 Classification in Improving Recitation Fluency in Students

Sisca Yuliantina[1], Betha Nurina Sari[2]

[1,2] Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang, Indonesia

ARTICLE INFORMATION	ABSTRACT
	Fluency in reciting the Koran is learning the recitation of the Qur'an in a tartil way. Based on the observations of researchers, the learning of tajwid in recitation and at school has not been effective so far. Because of this, both teachers and students at recitation or at school need improvement by finding out what can increase fluency in reciting the Koran and what has the most influence on improving fluency. So this study aims to improve the fluency of the Koran in students. The research method used is the Decision Tree data mining classification method with the C4.5 algorithm. The results of data processing with the C4.5 algorithm using the Rapidminer tools are attribute C1(fluency) being the most influential attribute for increasing students' reading fluency and performance data obtained with an accuracy of 83,33%.

Corresponding Author:

Sisca Yuliantina
Universitas Singaperbangsa Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Indonesia
Email: 1910631170045@student.unsika.ac.id

INTRODUCTION

The fluency of reciting the Koran in reading the Qur'an is studying the recitation of the Qur'an in a tartil manner. As for what is meant by tartil, namely reading by paying attention to the characteristics of the letters and their tajwid. In learning to recite the Koran, it is not just reading the Qur'an or studying with the Iqra' method, but it is necessary to pay attention to how well and correctly the students' ability to read the Qur'an or iqra' is in accordance with their tajwid. Therefore, learning the science of recitation is very important because it can make it easier to know the short length of reading lapadz along with the rules for reciting the Qur'an. Based on the observations of researchers, the learning of tajwid in recitation and at school has not been effective so far. Because of this, both teachers and students at recitation or at school need improvement by finding out what can increase fluency in reciting the Koran and what is most influential in increasing fluency in reciting.

In formal schools there are indeed Islamic Religious Education subjects and in them there is sub-discussion about reading the Qur'an but what is taught in schools is only the basics such as regarding reading hijaiyah letters and being taught about Tajwid. Whereas the subject of Islamic Religious Education is only given 2 hours per week so that it does not maximize the learning process about the Al-Qur'an or reciting the Koran using the iqra' method. As for the recitation, they also learn about the science of recitation, limited to its basics and its application when reciting the Koran is constrained by time constraints. So that many students can read the Al-Qur'an or Iqra' but in their reading it is still not precise, such as the reading is not in accordance with tajwid, and the pronunciation of hijaiyah letters is not quite right. Even though it has been studied for quite a long time, many students experience problems reading the Qur'an or iqra'. This of course causes fluency in reciting, namely

reading the Qur'an or iqra' to decrease and must be improved.

According to (Ardiansyah et al., 2021)conducted research on a comparative analysis of the accuracy of the Naïve Bayes and C4.5 algorithms for the classification of Diabetes. The results of the tests performed show that the C4.5 algorithm (scenario 4) has good results in the classification of diabetes compared to the Naïve Bayes algorithm (scenario 2) where the performance of the C4.5 algorithm has 99.03% accuracy, 100% precision, and 98.18% recall for 88.35% accuracy, 92.16% precision, and 85.45% recall. Research conducted by (Yani Hakim Tanjung, 2021)regarding the comparison of the ID3 and C4.5 algorithms to ATM money filling data, resulted in the highest score achieved by the C4.5 algorithm with a Precision value of 100% and the Recall value tends to be the same, namely 93, 33%. From the average value, it can be concluded that the C4.5 algorithm is proven to be a potentially effective and efficient classifier algorithm.
A comparative study of the K-Nearest Neighbor,

Decision Tree (C4.5) and Naïve Bayes algorithms to determine creditworthiness by (Muryono et al., 2021)resulted in the fact that the decision tree algorithm (C4.5) obtained the highest accuracy results with a value of 98.00% compared to K-Nearest Neighbor which obtained a value of 93.33% and naïve bayes of 86.67%. Based on this, it can be concluded that the performance of Decision Tree (C4.5) is better than K-Nearest Neighbor and Naive Bayes. Furthermore, research that tries to compare the use of 3 data mining classification methods, namely Decision Tree, Naïve Bayes and Logistic Regression to get the most accurate algorithm and find out the results of the classification of the dangers or not of using electric cigarettes. From the results of data processing and testing by measuring the performance of the three algorithms using the confusion matrix procedure, operator cross validation and ROC curves, the decision tree algorithm produces the highest level of accuracy value of 81.00% (Argarini Pratama & Hellyana, 2022). And research conducted by (Fanny Irnanda & Perdana Windarto, 2020)regarding the application of the C4.5 classification in improving English proficiency in society obtained calculation results that stated attribute C2 (Practice) was the variable that most influenced the increase in English proficiency in society. Tests were also carried out to prove whether the C4.5 method could be applied to cases of increasing English proficiency in the community using the Rapidminer software and an accuracy of 95% was obtained.

Based on the presentation of the results of previous studies regarding the performance of classification algorithms, it can be concluded that the C4.5 algorithm has good data classification performance. This is based on the high accuracy value obtained from the C4.5 algorithm. Therefore, this study will apply the C4.5 algorithm to classify in improving students' reading fluency.

RESEARCH METHOD

The research method is a very important factor in solving a problem and can also determine the success of a study. Basically the research method is a scientific tool in obtaining data with a specific purpose and use. In this study the model used is the Cross-Industry Standard Process for Data Mining Model (CRISP-DM) with data mining techniques used in classification. Classification is a technique by looking at the behavior and attributes of groups that have been defined. The classification method is used to help understand data grouping (Yunus et al., 2021). In this study, researchers also implemented a decision tree with the C4.5 algorithm. Algorithm C4.5 is an algorithm contained in classification techniques to solve cases or problems. The decision tree is the basis of Algorithm C4.5. Algorithm C.45 is a decision tree induction algorithm, namely ID3 (Iterative Dichotomiser 3) (Siahaan et al., 2020)

1. Data Mining

Data mining known as pattern recognition is an algorithm used for data processing to find hidden patterns in the processed data. Data that is processed with data mining techniques then produces new knowledge that is sourced from old data, the results of data processing can be used in determining future decisions (Gading Sadewo et al., 2018). In its application, data mining requires various data analysis software to find data patterns and relationships so that it can be used to make accurate predictions and classifications (Anam & Santoso, 2018).

At the data mining stage, information from a set of data can be determined by using the similarity of the data to one another. techniques that are often used are associations, classification processes, and data clustering (Prasetyowati et al., 2018). According to (Novianti et al., 2016) important things related to data mining are: (1) Data mining is the automatic processing of existing data. (2) The data to be processed has a very large data form. (3) The goal of data mining is to find relationships or patterns that can provide useful indications.

2. Algoritm C4.5

Algorithm C4.5 is one of the algorithms used to perform predictive classification or segmentation. The basis for Algorithm C4.5 is forming a decision tree where the highest attribute is the root and the lowest attribute is called a leaf (P et al., 2019). The data in a decision tree is usually expressed in the form of a table with attributes and records. Attributes declare a parameter that is made as a criterion in forming a tree. Tree pruning is done to simplify the tree so that accuracy can increase. Pruning there are two approaches, namely (Wanto et al., 2020): (1) Pre-pruning, namely stopping the construction of a subtree early (ie by deciding not to partition the training data further). When it suddenly stops, the node turns into a

leaf (end node). This end node becomes the class that occurs most often among the sample subset. (2) Post-pruning, namely simplifying the tree by removing some subtree branches after the tree has been built. The node that is rarely cut will be the leaf (end node) with the most frequently appearing class.

The decision tree is obtained from the results of the Entropy and information gain calculation process, after repeated calculations until all tree attributes have a class and the calculation process can no longer be carried out (Cynthia & Ismanto, 2018). According to (Prasetyowati., 2018) to calculate the gain value using the formula can be seen in the equation below.

$$\text{Gain}(S, A) = Entropy\,(S) - \sum_{i=1}^{n} \frac{|s_i|}{|s|} * Entropy\,(S_i) \;\;\ldots..(1)$$

Information:

S : The set of cases

A : Attributes

N : Number of cases on attribute A

|Si|: The number of cases on the ith partition

|S| : Number of cases in S

After finding the gain value, then a calculation is carried out to find the Entropy value. To determine the level of information from input attributes in producing output attributes. According to (Prasetyowati et al., 2018) below it can be seen through the equation:

$$Entropy\,(S) = \sum_{i=1}^{n} - pi * log2\,pi \ldots\ldots\ldots\ldots..... (2)$$

Information:

S : The set of case

n : Number on partition S

pi : The proportion of Si to S

In general, the C4.5 algorithm for building decision trees is as follows (Hartama et al., 2022): (1) Select the attribute as root. (2) Create a branch for each value. (3) Split cases in a branch. (4) Repeat the process for each branch until all cases on the branch have the same class.

3. Decision Tree

The decision tree in (Siahaan et al., 2020) is a classification method that uses a decision tree structure representation where each node represents an attribute, the branches represent the value of the attribute, and the leaves represent the class. The Decision Tree tries to find the right attributes to complete and define classes. In other words, the attribute at the top of the decision tree is the most influential attribute in determining the prediction result. Therefore, the process of making a decision tree model is an interactive process to assess the most influential attributes that will branch off from this tree structure (Susdarwono & Setiawan, 2020).

The advantages of the decision tree are the decision-making areas that were previously complex and very global, can be changed to be more simple and specific, Eliminating unnecessary calculations because when using the decision tree method, samples are tested only based on certain criteria or classes, Flexible to choose features from different internal nodes, the selected features will differentiate a criterion compared to other criteria in the same node. The flexibility of the decision tree method improves the quality of the resulting decisions when compared to using more conventional one-step calculation methods. In multivariate analysis, with a large number of criteria and classes, an examiner usually needs to estimate either a high-dimensional distribution or certain parameters of the distribution. that class (Sang et al., 2021).

Data Collection Techniques

Data collection techniques carried out for this study used observation, library research, and interviews.

1) Observation

Observation in (H. Hasanah, 2016)an activity that involves all the powers of the senses such as hearing, sight, taste, touch, and taste based on the facts of empirical events. The method of data collection carried out by researchers is by observing directly, seeing, and taking data at MDTA Matla'ul Khoir.

2) Literature Study

In this study, the data were obtained directly from human sources, namely by conducting direct tests on the fluency of the Koran, through the observation process. This technique is carried out in order to obtain data related to the research title.

3) Interview

In collecting this data, interviews were conducted to identify problems regarding factors that influence the increase in students' reading fluency. This study also interviewed teachers and principals at MDTA Matla'ul Khoir.

Research Stages

In this research, data analysis used the stages of the CRISP-DM process and analysis. CRISP-DM is a standardization of the data mining process as a general problem-solving strategy from a business or research unit. Cross-Industry Standard Process for Data Mining or CRISP-DM was developed in 1996 by analysts from several industries such as Daimler Chrysler, SPSS, and NCR (Feblian & Daihani, n.d.). CRISP-DM (Cross Industry Standard Process for Data Mining) is a standardization of data mining processing that has been developed where the existing data will go through each structured and clearly defined and efficient phase (M. A. Hasanah et al., 2021)CRISP-DM a data mining project has a life cycle that is divided into six phases. The following is Figure 1 which will provide a general description of the life cycle in CRISP-DM (Feblian & Daihani, n.d.).
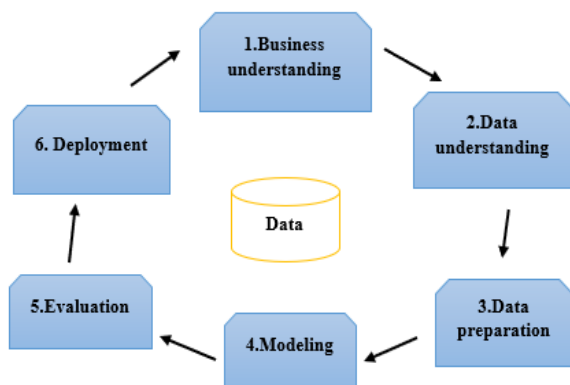
Source: (Yuliantina et al., 2023)

Figure 1. Life cycle in CRISP-DM

From Figure 1 it can be seen that the life cycle of CRISP-DM consists of six phases, namely:

1) Business Understanding

At this stage, it is the determination of project objectives and their needs in detail within the scope of the business or research unit as a whole. In addition to translating the goals and limitations that become the formula for data mining problems.

2) Data Understanding

This stage collects data using data investigation analysis to further identify data and search for initial knowledge.

3) Data Preparation

In this stage, namely building the final dataset in the form of raw data. There are several things that will be carried out including data cleaning, data selection, records and attributes, and data transformation to be used as input in the modeling stage.

4) Modelling

At this stage, statistical methods and machine learning are used to determine the data mining techniques, data mining tools, and data mining algorithms to be applied. After that, apply data mining techniques and algorithms to the data with the help of tools. If data adjustments are needed for certain data mining techniques, then you can return to the data preparation stage.

5) Evaluation

At this stage, it is done by looking at the performance level of the pattern produced in the algorithm process. The parameters used in the evaluation of comparison algorithms are the Confusion Matrix with rules that include accuracy, precision, and recall values. The value is obtained by the calculation as follows.

$$Accuracy = \frac{Correct\ classification\ number}{Total\ testing\ samples\ tested} \times 100\%$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100\%$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100\%$$

6) Deployment

At this stage, reports and journal articles are created using the model that has been generated.

RESULTS AND DISCUSSION

The results of this study are expected to be able to determine the accuracy value of the C4.5 Algorithm in increasing the fluency of the Koran for students who meet the criteria, by processing it and selecting the required attributes. After that, testing the data with manual calculations and tools used, namely Rapidminer. The results of this study are presented with a calculation process based on the C4.5 algorithm through the life cycle path of CRISP-DM.

1) Business Understanding

The application of data mining in this study is directly related to data analysis of students' reading fluency in order to gain knowledge about a pattern of applying classification to improve students' reading fluency. As well as to see what parameters affect the fluency of the recitation.

2) Data Understanding

This study uses data originating from MDTA Matla'ul Khoir Class 2 DTA. This data is student test data related to fluency in reciting the Koran in 2022. The dataset obtained was 32 students by classifying their recital ability into increasing and decreasing. The parameters used include fluency (C1), understanding of recitation (C2), memory level (C3), fasohah (C4), and knowledge of Hijaiyah letters (C5). Verification of the assessment of the classification to improve fluency in reciting is categorized as good. enough, and less.

Table 1. Categories of Koran Proficiency Assessment

Range	Information
80-100%	Good
60-79%	Enough
<60%	Less

Source: (Yuliantina et al., 2023)

3) Data Preparation

Data preparation or data preparation is a stage to fix problems in the data before the data enters the modeling stage so as to produce good modeling (Suhanda et al., 2020). In the data cleaning Stage, the dataset which consists of 32 data contains only 32 records that are dichotomous with increasing and decreasing fluency in recitation. So no data is cleaned. The data used has 5 predictors and 1 target attribute with no Missing Value.

Based on the grouping of data that has been done in excel, the attributes that become parameters in

improving fluency in reciting are obtained. These attributes are obtained by looking for factors that cause students' reading fluency to decrease and must be corrected. There are 6 attributes that are of the Polynomial type and the attribute of fluency in reciting the Koran is used as a label as shown below (M. A. Hasanah et al., 2021):

Table 2. Attributes and Values

| Attribute	Type	Value
Fluency	*Polynomial*	Good, Enough, Less
Tajwid understanding	*Polynomial*	Good, Enough, Less
Memory Level	*Polynomial*	Good, Enough, Less
Fashohah	*Polynomial*	Good, Enough, Less
Knowledge of Hijaiyah Letters	*Polynomial*	Good, Enough, Less
Reading fluency	*Polynomials / Labels*	Increase, Decrease

Source: (Yuliantina et al., 2023)

At the data transformation stage, nothing is cleaned.The next stage is the division of the data set using the split data process to produce training data and testing data with a ratio of 80:20.

4) Modelling

In this phase, the process by which the model with the selected classification technique produces a pattern of information that can facilitate interested parties. The classification pattern produced by this data mining technique is used to predict recitation fluency with predictors that influence it. The tools or tools in the processing of data mining are Rapidminer. This stage displays and provides information on the performance of the algorithm in the classification method.

As for the process of testing imported data, researchers used training data and data testing in determining a decision tree model with a ratio of 80% used for training data and 20% for data testing. Figure 2 is the process of finding the accuracy of the C4.5 algorithm using Rapidminer.
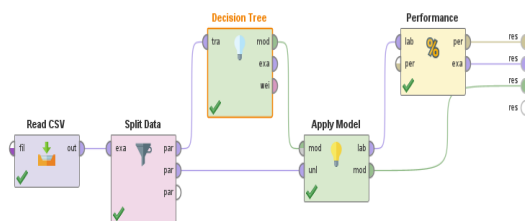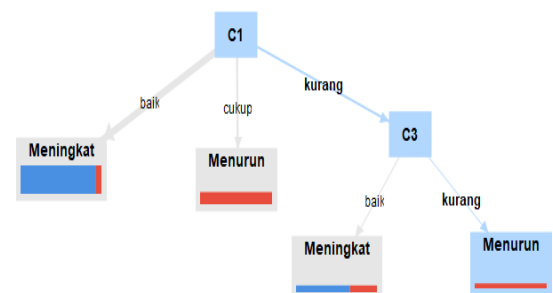

Source: (Yuliantina et al., 2023)
Figure 2. The Process of Finding C4.5 algorithm accuracy using Rapidminer

Figure 2 it can be seen in the process of testing data that has been imported into CSV format with rapidminer, which has many operators available. As for this test, the operator needed is split data which is useful for dividing the dataset data partitions training and data testing according to predetermined portions. Then the apply model is used to apply a model that has been previously trained with the aim of obtaining predictions on testing that does not yet have a label. The decision tree is the chosen classification model because it can be determined by making a decision tree for the subtest results. And the performance operator *is* used to evaluate the performance of the model which gives one of the performance criteria automatically based on the given task. The following is figure 3. The results of the decision tree modelling after the testing process are carried out.

Source: (Yuliantina., 2023)
Figure 3. Results of Decision Tree Modeling

The C4.5 decision tree that has been seen in figure 3 is read from top to bottom or from the root (the first top node) to the leaves (the outermost node that no longer has branches). The way to read it is by referring to each node. If the students' fluency in reciting the Koran is good, then it increases without the need to look at other attributes. If the student's reading fluency is sufficient, then it is declared to decrease. If the student's reading fluency is lacking then there are two possibilities that occur, first if the student's memory level is good then it is declared to be increasing. Second, if the level of student memory is less, then it is declared to be decreasing. And if seen based on the results of the classification decision tree in increasing the fluency of reciting the students that the attribute that has the main influence is variable C1 (Smoothness) which occupies the root node, as shown in the following figure 4.

Tree

```
C1 = baik: Meningkat {Meningkat=14, Menurun=1}
C1 = cukup: Menurun {Meningkat=0, Menurun=6}
C1 = kurang
|   C3 = baik: Meningkat {Meningkat=2, Menurun=1}
|   C3 = kurang: Menurun {Meningkat=0, Menurun=2}
```

Source: (Yuliantina et al., 2023)
Figure 4. Decision Tree Result in the form of Text

The decision tree results in the form of text as seen in figure 4 provide information that C1 occupies the root node. Based on that C1 is stated as an attribute

that has the main influence in increasing students' reading fluency.

5) Evaluation

After the classification pattern is obtained in the C4.5 algorithm, the next stage is the evaluation stage of the comparison of the algorithm with the parameters used, namely the *Confusion Matrix* which basically provides information on comparing of the results that have been carried out by the model with the actual results of the classification by looking at the values of accuracy, precision, and recall. The following is the evaluation result of the model that has been created with this algorithm.

accuracy: 83.33%

	true Meningkat	true Menurun	class precision
pred. Meningkat	3	0	100.00%
pred. Menurun	1	2	66.67%
class recall	75.00%	100.00%	

Source: (Yuliantina., 2023)

Figure 5. Confusion Matrix in the C4.5 Algorithm

Based on Figure 5, the results show that the $Accuracy = 0.83$, $recall = 0.75$, and $precision = 0.6667$. So that the accuracy level of the C4.5 algorithm is obtained at 83.33%.

6) Deployment

After the evaluation stage where the result of a model are assessed in detail, the implementation of the entire model that has been built is carried out. In addition, adjustments were also made to the model so that it can produce a result that is in accordance with the initial target of the CRISP-DM stage.

CONCLUSION

This research was carried out using the classification of fluency data in reciting the Koran using the Algorithm C4.5 (*Decision Tree*) method and the CRISP-DM data mining technique. The dataset consists of 6 attributes, namely Fluency (C1), Understanding Recitation (C2), Memory Level (C3), Fashohah (C4), Knowledge of Hijaiyah Letters (C5), and Reciting Fluency (C6) which are used as labels. From this study, it can be concluded that the application of the C4.5 classification to increasing students' reading fluency states that the result of the calculation is that attribute C1 (fluency) is the variable that most influences students' reading fluency. This test was also carried out to prove whether the C4.5 method could be applied to cases of increasing fluency in reciting the Koran to students using the help of *tools* and an accuracy value of 83.33% was obtained.

REFERENCES

Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *8*(1), 13 - 19. Retrieved from https://ejournal.upm.ac.id/index.php/energy/article/view/111

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

Argarini Pratama, E., & Hellyana, C. M. (2022). Perbandingan 3 Algoritma Klasifikasi Data Mining Dalam Pro-Kontra Bahaya Rokok Elektrik. In *Jurnal TEKNOINFO, 16*(1), 93-99. DOI: https://doi.org/10.33365/jti.v16i1.1534

Cynthia, E. P., & Ismanto, E. (2018). Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, *3*, 1. https://doi.org/10.30645/jurasik.v3i0.60

Fanny Irnanda, K., & Perdana Windarto, A. (2020). *Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat*. https://prosiding.seminar-id.com/index.php/sainteks

Feblian, D., & Daihani Umar, D. (2016). Implementasi model (Dina F, dkk) Implementasi Model CRISP-DM Untuk Menentukan Sales Pipeline Pada PT X. *Jurnal Teknik Industri* , *6*(1), 1–12. DOI: https://doi.org/10.25105/jti.v6i1.1526

Gading Sadewo, M., Perdana Windarto, A., & Wanto, A. (2018). Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)* , *2*(1), 311–319. http://dx.doi.org/10.30865/komik.v2i1.943

Hartama, D., Daya, K., & Sianipar, R. (2022). Penerapan Algoritma C4.5 Untuk Analisa Tingkat Keberhasilan Mahasiswa Dalam Pembelajaran Praktikum di Masa Pandemi. *Journal of Computer System and Informatics (JoSYC)* , *4*(1), 128–134. https://doi.org/10.47065/josyc.v4i1.2584

Hasanah, H. (2016). *TEKNIK-TEKNIK OBSERVASI (Sebuah Alternatif Metode Pengumpulan Data Kualitatif Ilmu-ilmu Sosial)*. https://doi.org/10.21580/at.v8i1.1163

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2). https://doi.org/10.30871/jaic.v5i2.3200

Muryono, T. T., Taufik, A., & Irwansyah, I. (2021). Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit. *Infotech: Journal of Technology Information*, *7*(1), 35–40. https://doi.org/10.37365/jti.v7i1.104

Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak). In *Jurnal Coding, Sistem Komputer Untan* (Vol. 04, Issue 3). https://doi.org/10.26418/coding.v4i3.17034

P, D. R. S., Windarto, A. P., Hartama, D., & Damanik, I. S. (2019). PENERAPAN KLASIFIKASI C4.5 DALAM MENINGKATKAN SISTEM PEMBELAJARAN MAHASISWA. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1), 593–597. https://doi.org/10.30865/komik.v3i1.1665

Prasetyowati, E., Nilam Ramadhani, dan, Raya Panglegur Km, J., & Timur, J. (2018). Sistem Evaluasi Dan Klasifikasi Kinerja Akademik Mahasiswa Universitas Madura Menggunakan Naive Bayes Dengan Dirichlet Smoothing. In *JUTI: Jurnal Ilmiah Teknologi Informasi* (Vol. 16, Issue 2). https://doi.org/DOI: 10.12962/j24068535.v16i2.a688

Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). *Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine Data Minning Analysis For Classification Of Air Quality Data DKI Jakarta Using Decision Tree Algorthm And Support Vector Machiner Algorithm*. https://doi.org/10.34818/eoe.v8i5.15900

Siahaan, S. W., Sianipar, K. D. R., R.H Zer, P. P. P. A. N. W. F. I., & Hartama, D. (2020). Penerapan Algoritma C4.5 dalam Menentukan Faktor yang Dapat Meningkatkan Kemampuan Bahasa Inggris pada Mahasiswa. *Jurnal Eksplora Informatika*, *10*(1), 59–67. https://doi.org/10.30864/eksplora.v10i1.396

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, *6*(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Susdarwono, E. T., & Setiawan, A. (2020). Penerapan Teori Keputusan Dalam Model Pengambilan Keputusan Terkait Masalah Ekonomi Pertahanankonsep Pohon Keputusan. *Jurnal Ilmu Manajemen Dan Akuntansi Terapan (JIMAT)*, *11*(2), 243–257. https://doi.org/10.36694/jimat.v11i2.239

Wanto, A., Windarto, A. P., Hartama, D., & Prianto Cahyo. (2020). *Data Mining : Algoritma dan Implementasi*. Andi: Yogyakarta. ISBN: 9786230107252

Yani Hakim Tanjung, D. (2021). *Analisis Perbandingan Algoritma ID3 Dan C4.5 Terhadap Data Pengisian Uang ATM*. https://doi.org/10.22303/csrid.13.3a.2021.231-242

Yunus, M., Ramadhan, H., Aji, D. R., & Yulianto, A. (2021). Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP). *Paradigma - Jurnal Komputer Dan Informatika*, *23*(2), 191–196. https://doi.org/10.31294/p.v23i2.11395