CLASSIFICATION OF PROSPECTIVE SCHOLARSHIP RECIPIENTS KARTU INDONESIA PINTAR (KIP) WITH DECISION TREE ALGORITHM AND NAÏVE BAYES

Asnimar[1*], Andani Achmad[2], Yuyun[3], Akbar Iskandar[4], Mansyur[5]
Bisnis Digital, Universitas Teknologi Akba Makassar, Makassar, Indonesia[1,4]
Sistem Komputer, Universitas Handayani Makassar, Makassar, Indonesia[1,3]
Penelitian dan Evaluasi Pendidikan, Universitas Negeri Makassar, Makassar, Indonesia[5]
Teknik Elektro, Universitas Hasanuddin, Makassar, Indonesia[2]
E-mail address: asnimar@email.com[1]

Received: 26, August, 2022 ***Revised: 15, October, 2022*** ***Accepted: 31, December, 2022***

ABSTRACT

The purpose of this study is to produce a model that can assist in determining prospective new students of STMIK AKBA who receive scholarships. The algorithm used is decision tree and nave Bayes to classify the graduation of prospective recipients of the Indonesian Smart Card (KIP) scholarship. Based on the results of the classification of the decision tree algorithm with the confusion matrix, the accuracy value is 44.12% and the F1-Score is 34.48%. If you use the Naive Bayes algorithm, it produces an accuracy value of 76.47% using data on diploma scores and average report cards. Furthermore, for accuracy without using diploma value data and the average report card is 79.41%. The results of this study show that nave Bayes has a better performance even though it does not use diploma scores and average report cards. Measurement of the results of the nave Bayes classification with the confusion matrix showed low sensitivity with values of 66.67% and 58.33% for the first and second scenarios. Based on the evaluation results, the Naïve Bayes algorithm has a better performance than the Decision Tree algorithm in classifying KIP scholarship recipients.

Keywords: Classification, Decision Tree, Smart Indonesia Card, Naïve Bayes.

1. INTRODUCTION

In February 2020 the Ministry of Education and Culture issued Permendikbud 10 of 2020 concerning the Smart Indonesia Program (PIP). The Smart Indonesia Program (PIP) through the Smart Indonesia Card (KIP) is assistance in the form of cash, expansion of access, and learning opportunities from the government given to students and students who come from poor or vulnerable families to finance education. Minister of Education and Culture Regulation 10 of 2020 concerning the Smart Indonesia Program (PIP) reaches students from the elementary level to the university level. It is hoped that this program can develop self-potential through the learning process available at the basic education level to the higher education level.

The Smart Indonesia Card (KIP) program provides very broad expectations in solving educational problems in Indonesia. This is one solution to a complex problem. If implemented properly, there will be a reduction in the dropout rate which is quite high in Indonesia, as well as being able to reach the Indonesian people from Sabang to Merauke. Smart Indonesia Card

(KIP) is a cash assistance provider for education up to high school graduation with school children aged 6-21 years who come from poor families (underprivileged) or who are registered as participants in the Family Hope Program (PKH) or the Prosperous Family Card (Kartu Keluarga Prosperous). PSC).

The process of determining KIP scholarship recipients at STMIK AKBA is based on criteria including ownership of the Smart Indonesia Card (KIP), Prosperous Family Card (KKS), Poor Card (KTM), parental salary, number of dependents of parents, parent's occupation, status graduated from high school or equivalent, has passed a maximum of 2 years, the value of a diploma and the average value of student report cards. There are many scholarship applicants, for example, there are 110 scholarship applicants in 2020, which causes decision makers to find it difficult to determine scholarship recipients quickly and accurately. Therefore, In the process of receiving scholarships, a system is needed that can classify scholarship recipients which aims to reduce subjectivity in decision making and increase the effectiveness of the process of determining scholarship recipients. In addition, the existence of a scholarship determination system will be more targeted to students who really need it and with the system it can also reduce the rate of policy determination errors.

Decision Tree and Naïve Bayes algorithms have the advantage of being able to model non-linear relationships in a dataset. In addition, these two algorithms can also be used for data with categorical and numeric data types, such as the KIP scholarship candidate dataset. This dataset consists of attributes that have categorical and numeric data types. In order to produce a decision tree to select the right scholarship recipients.

2. THEORY

1. Smart Indonesia Program

The Smart Indonesia Program, which is abbreviated and known as PIP, is a government assistance program that is implemented to ease the financing of prospective school students or university students who are underprivileged to be able to gain knowledge.(Maimon & L, 2005)without the need to think about costs by using the Indonesia Smart Card as a condition for getting cash assistance and being able to withdraw KIP funds at banks that have collaborated with government assistance programs (Retnaningsih, 2019).

Education is very important to improve the quality of a person. Getting an education is certainly inseparable from the educational problems that are often faced by someone, namely, the cost of education. The cost of education is one of the important factors to determine the success of educational goals. STMIK AKBA as one of the Information Technology-based Colleges that is developing in Eastern Indonesia. One of the programs launched by STMIK AKBA is the provision of scholarships to students who need financial assistance for the continuation of their studies. Scholarships are not only offered for outstanding students, but also for underprivileged students in order to assist these students in pursuing their education.

Smart Indonesia Card (KIP) is a cash assistance provider for education up to high school graduation with school children aged 6-21 years who come from poor families (underprivileged) or who are registered as participants in the Family Hope Program (PKH) or the Prosperous Family Card (KKS). The Smart Indonesia Card is given as a marker/identity to guarantee and ensure that children receive assistance from the Smart Indonesia Program if the child is registered or enrolled in a formal education institution (school/madrasah) or non-formal institution (Islamic Boarding School, Community Learning Activity Center/PKBM, Package A). /B/C, training institutions/courses and other non-formal education institutions under the Ministry of Education and Culture and the Ministry of Religion).

2. Data Mining

The definition of data mining is a very important work step in extracting information using data. Data Mining includes the extraction of different patterns, in other words, attracting large amounts of data, known as big data (Abduallah, 2020). The work steps of data mining depend on the case to be solved, one of which is by using data mining classification as a solution in solving a case. Such as going through the process of finding some common characteristics of a class or group (Agustiani, 2016). Classification in data mining is also known as one of the most important methods with the aim of being able to estimate the class of an object where the label is not known (Gyssens & Paredaens, 1984). Data mining is an important step in the knowledge discovery process in databases which has the main objectives consisting of data cleaning, integration, selection, transformation, evaluation of patterns, and presentation of knowledge. Since data is often interspersed with missing values and noise which makes it incoherent, data pre-processing becomes an important step before data mining to improve data (Dulli et al., 2009).

3. Decision Tree Algorithm

Decision tree is the most intuitive and rule-based data classification and prediction method. The most critical concept in understanding a decision tree is entropy (Zhou, 2020), where entropy as a measure of energy in a unit temperature system is called a thermodynamic unit that has symmetrical properties (Godlewski & Raviart, 2013). One way of classification in the decision tree is the C4.5 algorithm. This algorithm is a basic algorithm that recursively goes to the decision node and performs an optimal split. The method of calculating the C4.5 Algorithm is to use Information Gain, which is a calculation that functions to select attributes using information entropy theory. Suppose there is a variable X, which has k possible values with probabilities p1, p2, …, pk. Then the entropy X is the least number of bits (average per symbol) needed to transmit a stream of symbols representing the value of X and can be expressed by the following equation:

$$H(X) = -\sum_{j} p_j \log_2 (p_j) \quad (1)$$

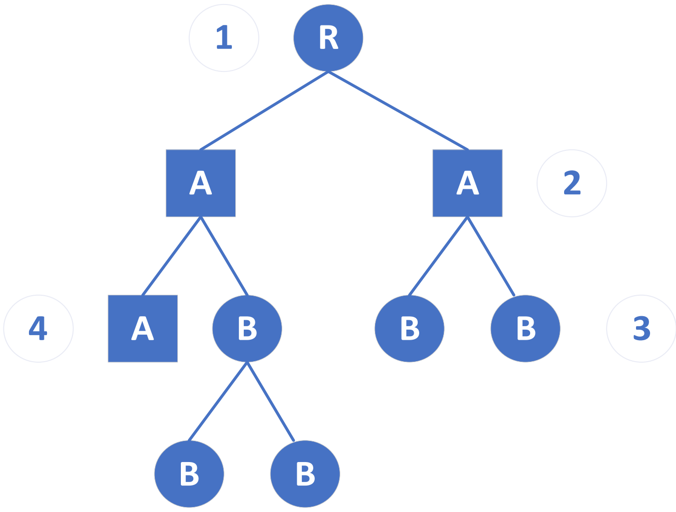
C4.5 uses the concept of entropy as follows. Suppose we have a candidate split S, which partitions the training data T into several subsets, T1, T2, ...Tk. The information requirements of importance can be calculated:

$$H_s(T) = \sum_{i=1}^{k} P_i H_s(T_i) \quad (2)$$

Pi represents the proportion of records in subset i. So the information gain can be defined:

$$GAIN(S) = H(T) - H_s(T \quad (3)$$

A decision tree or known as a decision tree is one method to make it easier for users to read the steps of the algorithm and can be used to display algorithms that only contain conditional control statements. A decision tree is a flowchart-like structure in which each internal node represents a possible attribute, each branch represents the outcome of that possibility, and each leaf node represents a class label. The descriptions in Figure 2.1 are 1) Root Node: The topmost node in the tree, 2) Each branch represents the test results on the training data set. 3) Leaf node: store numerical predictions. 4) Internal nodes: show tests on attributes. You can see the following picture:



Figures. 1. Decision Tree For Simple Disjunction
(Buntine, 2020)

4. Naïve Bayes Algorithm

The definition of Naïve Bayes is a classification method based on simple probabilities which has the advantage of being able to work well even with complex conditions in the real world and being able to work using nave Bayes models without Bayesian (Iswanto et al., 2015). The nave Bayes method uses Bayes' rules to calculate several probabilities of each possible value contained in the target attribute (Maimon & L, 2005). Suppose X = {x1, x2, x3, x4..., xn) is a data that has a value consisting of n attributes. In Bayes' terms, X is referred to as evidence. Suppose H is a hypothesis which states that data X belongs to a certain class C. In the classification problem, it is determined P(H|X) which is the probability that data X belongs to class C or the probability of obtaining a result of H if the data X is known. P(H |X) is called the posterior probability (Han & Micheline, 2006).

P(H) is the prior probability of H, that is, the initial probability. For example, student X will continue his studies abroad without regard to TOEFL and GPA scores. P(X|H) is the posterior probability of X to state H. In this example, P(X|H) is the probability that a student X with a TOEFL score of 560 and a GPA of 3.75 is known to continue his studies abroad. P(X) is the prior probability of X. In this example, P(X) is the probability that someone from the student data set has a TOEFL score of 560 and a GPA of 3.75. In the Naïve Bayes classification, the probability of X data belonging to class C, given the attribute value of X, can be calculated using Bayes' theorem as follows (Kabir et al., 2011):

$$P(H|X) = \frac{P(H)\, P(X|H)}{P(X)} \qquad (4)$$

$$P(X \in C_i|x_1, x_2, \cdots, x_n) = \frac{P(C_i)\, P(x_1, x_2, \cdots, x_n|C_i)}{P(x_1, x_2, \cdots, x_n)} \qquad (5)$$

In simple terms, the equation can be written as follows:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \qquad (6)$$

The above equation can only be used for discrete data. While the data that is continuous is calculated using the Gaussian formula.

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} e^{\frac{-(x-\mu)^2}{2\,\sigma^2}} \qquad (7)$$

The symbol is the average value and$\sigma$is the standard deviation calculated using the following equation(Kabir et al., 2011):

$$\sigma = \sqrt{\left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2\right)} \qquad (8)$$

5. Confusion Matrix

The Confusion Matrix serves as a performance measurement tool or a clearer performance in explaining the case of machine learning classification. Figure 2 shows the Confusion Matrix table model.

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive	False Negative
	Negative	False Positive	True Negative

Figures. 1. Confusion Matrix

The confusion matrix table is divided into 2 rows and 2 columns showing the number of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)(Chicco et al., 2021). This allows for a more detailed analysis than an analysis that only looks at the number of correct classifications or accuracy. True Positive (TP) is the number of Positive data that the model has successfully recognized or classified as Positive. False Positive (FP) is the amount of Negative data classified by the model as Positive. True Negative (TN) is the number of Negative data that the model has successfully recognized or classified as Negative. And False Negative (FN) is the number of positive data recognized or classified by the model as Negative. With the Confusion Matrix, the model can also be analyzed further by calculating the

Accuracy, Precision, Sensitivity, Specivity, and f1-score values which can be calculated using the following formula.

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+FP+TN+FN} \qquad (9)$$

$$Precision\ PPV\ (Positive\ Predictive\ Value) = \frac{TP}{TP+FP} \qquad (10)$$

$$Precision\ NPV\ (Negative\ Predictive\ Value) = \frac{TN}{TN+FN} \qquad (11)$$

$$Sensitivity = \frac{TP}{P} = \frac{TP}{TP+FN} \qquad (12)$$

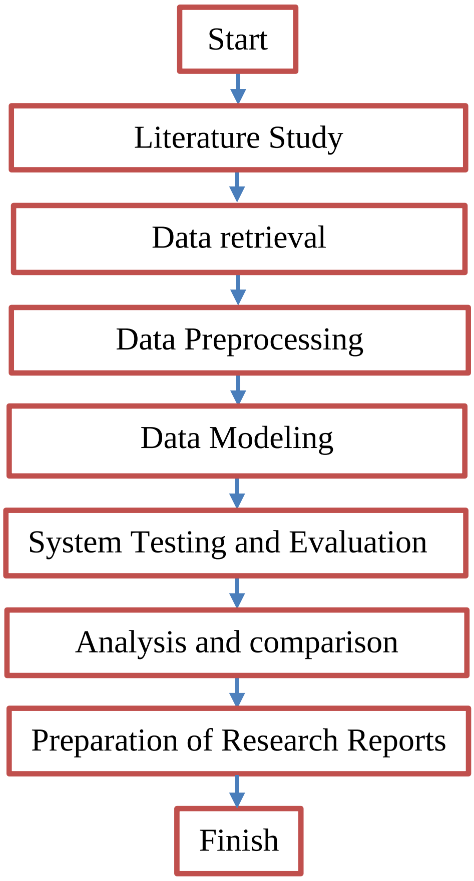
$$Specificity = \frac{TN}{N} = \frac{TN}{TN+FP} \qquad (13)$$

$$F1 - score = \frac{2TP}{2TP+FP+FN} \qquad (14)$$

3. METHOD

Research Stages

The stages carried out in this study can be seen in Figure 3.



Figures. 3. Research Stages

The explanation for each stage of the research to be carried out is as follows.

1. Literature Study

The initial stage carried out in this research is a literature study on the Smart Indonesia program, the requirements that must be met When a student registers for the Smart Indonesia program, the algorithms used are Decision Tree and Naïve Bayes, as well as other information that can later support the process of implementing this research.

2. Data retrieval

The data collection used in this study came from the information provided by the participants of the Smart Indonesia Card to STMIK AKBA for the purposes of registration for KIP and lectures. The data taken for this study include academic information and student achievements. Data taken from registration in 2020 and 2021.

3. Data processing

After the data is collected, the next stage is data processing which consists of 3 processes, namely data selection, data cleaning, data transformation and data sharing.

4. Data Selection

In this process, the data needed is selected. Privacy information such as student name, residential address, parent name, parent address and so on will be excluded from the dataset. Meanwhile, data on ownership of the Smart Indonesia Card (KIP), Prosperous Family Card, high school graduation status or equivalent, diploma value, parental occupation, parental salary, number of dependents and average student report cards will be selected as attributes used in the data modeling process.

5. Data Cleaning

After data selection, the next process is to clean data that has incomplete or incomplete information on one or more attributes that will be used.

6. Data Transformation

After the data that has incomplete attributes are removed, the next stage is data transformation where the data will be normalized. This concept aims to change the range of data values that are too far on the attributes of the number of dependents, the average value of student report cards and the value of diplomas.

7. Data Sharing

The last stage for data processing is to divide the dataset into two parts, namely training data and testing data. The comparison ratio used is 80% for training data and 20% for testing data.

8. Data Modeling

After the data is divided into training data and testing data, the next stage is data modeling using the Decision Tree and Naïve Bayes Algorithm to classify prospective Indonesian Smart Card (KIP) recipients. The data used at this stage is training data. The modeling process for each algorithm is carried out with different parameter values.

9. System Testing and Evaluation

The resulting Decision Tree and Naïve Bayes models will be evaluated with data testing to see the performance or performance of each algorithm. Algorithm performance will be measured by looking at Accuracy, Precision, Sensitivity, Specificity, and F1-score which is calculated using the confusion matrix of the classification results.

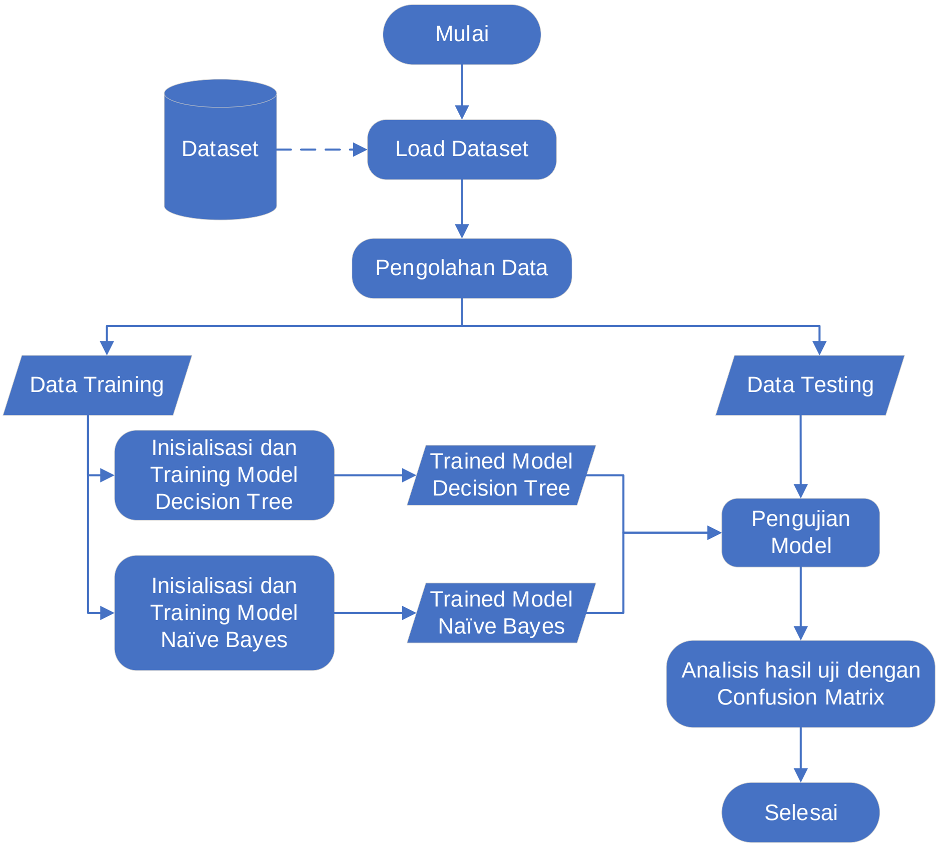
10. Analysis and Comparison of Algorithm Results

At this stage, a performance analysis is carried out to see the parameters with the best test results from each algorithm. After getting the test results from each algorithm, the next stage is a comparative analysis of the results between the Decision Tree and Naïve Bayes algorithms.

11. Research Report Preparation

At this stage, research reports and theses are prepared for the classification of KIP scholarship recipients using the Decision Tree and Naïve Bayes algorithms.

Research design



Figures. 4. Research Design

The research design to be carried out for the classification of KIP scholarship recipients can be seen in Figure 3.2. In the process of loading the dataset, the processed data will be divided into two, namely, training data and testing data. After determining the training data and testing data, the next stage is the initialization process and the training process for the Decision Tree and Naïve Bayes models using training data.

After the Decision Tree and Naïve Bayes model training process, the next stage is testing using data testing. At this stage, the model will be given data on prospective recipients to classify whether the prospective recipients will pass or not. The results of the classification will be compared with the actual status of scholarship recipients. The test results will be analyzed using the Confusion Matrix to calculate the Accuracy, Precision, Sensitivity, Specivity and F1-Score values in the Decision Tree and Naïve Bayes models.

Research Attributes

This study uses attributes or variables of student academic data (i) prospective KIP recipients who register as prospective students (i) STMIK AKBA from 2020-2021. The attributes used in this study can be seen in Table 1.

Table 1. Research Variables/Attributes

Attribute Name	Data Type
Ownership of Smart Indonesia Card (KIP)	Categorical
Ownership of a Prosperous Family Card (KKS)	Categorical
Possession of a Poor Card (KTM)	Categorical
Parents Salary	Categorical
The number of dependents	Numerical
Parents' job	Categorical
High school graduation status or equivalent	Categorical
Has passed a maximum of 2 years	Categorical
Certificate Value	Numerical
Average Student Report Score	Numerical

4. RESULTS AND DISCUSSION

Research result

The data used in this study is data from the admissions process for the AKBA College of Management Computer Science (STMIK AKBA) in 2020 and 2021. The amount of data obtained is 212 data. The next stage is data preprocessing and data modeling with Naïve Bayes algorithm and Decision Tree.

Data Preprocess

The stages in data preprocessing consist of data selection, data cleaning and data transformation.

1. Data Selection

The data selection process is the process of selecting data or attributes that will be used in this study. In order to maintain the confidentiality of personal data, information such as name, student identification number, and mobile number will not be selected. The data used in this study is the ownership status of the Smart Indonesia Card, the ownership status of the Prosperous Family Card, and the ownership of the Certificate of Disability. In addition, this study also uses information from parents of prospective recipients such as parental income, number of dependents of parents and the profession or occupation of parents.

The data for the selected recipients are the status of high school graduates, vocational high schools or the equivalent, have been declared high school graduates or equivalent, diploma grades, and the average value of the report cards of participants starting from class X to class XII. After all data is collected, The data are grouped into two classes, namely Passed and Failed which will be the target data for the classification of prospective Indonesian Smart Card (KIP) scholarship recipients.

2. Data Cleaning

In the data cleaning stage, data or attributes that have empty values are deleted, data for prospective recipients who resign, changes in KIP status and others. From the data cleaning stage, 174 data were obtained with the number of Passed data as many as 106 data, while the number of Disability data was 68 data.

3. Data Transformation

Data transformation is needed to change the format of the data values used into a form that is easier to use in the modeling process later. The data that will be transformed are data that are included in categorical data, namely:

a) Ownership of an Indonesia Smart Card (KIP), data with this attribute has a value of 1 if the participant has a KIP and 2 if he does not have a KIP card.
b) Possession of a Prosperous Family Card (KKS), data with this attribute has a value of 1 if the participant has KKS and 2 if he does not have KIP.
c) Possession of a Certificate of Disability (KTM), data with this attribute has a value of 1 if the participant has a KTM letter and 2 if he does not have a KTM letter.
d) Parents' income, the data on this attribute are grouped into four groups, namely: (1) For salaries < Rp. 1,000,000 given a value of 1, (2) For salaries between Rp. 1,000,000 - Rp. 1,999,999 is given a value of 2, (3) For salaries between Rp. 2,000,000 - Rp. 2,999,999 given a value of 3, and (4) Salary > Rp. 2,999,999 is assigned a value of 4.
e) Parents' occupations, the data on this attribute are grouped into 6 groups, namely (1) Parents of participants with Entrepreneurial professions or the like are given a category 1.) Parents of participants with the profession of Private Employees or the like are given a category 3 score, (4) For parents of participants with the profession of casual daily laborer or the like are given a category 4 score, (5) For parents of participants with the profession of Housewives or the like are given category 5. (6) For parents of participants with retired or retired professions, category 6.
f) High school graduates or equivalent, the data on this attribute is divided into 3 groups, namely SMA which is given a value of 1, SMK is given a value of 2, and MA is given a value of 3.

The results of the three processes above are used as a dataset that will be modeled with the Naïve Bayes algorithm and Decision Tree.

Discussion

Comparison of the classification of Naïve Bayes algorithms and Decision Tree

In the previous stage, the training process and testing of the Naïve Bayes Algorithm and Decision Tree for the Classification of prospective KIP beneficiaries was carried out. From the testing phase of the two algorithms, the values of accuracy, sensitivity, specificity and F1-score can be seen in Table classification results of the Decision Tree and Naïve Bayes algorithms for each data in the Test Dataset. In the Nave Bayes Classification and Decision Tree columns, the correct classification results will be given a green label and for the incorrect classification results will be given a blue label. Table 2 Comparison of Evaluation Results of the classification of the Naïve Bayes algorithm and Decision Tree in the testing process.

Table 2. Algorithm Naive Bayes and Algorithm Decision Tree

Confusion Matrix Value	Algorithm Naive Bayes(%)	Algorithm Decision Tree(%)S
Accuracy	79.41	44.12
Sensitivity	66.67	41.67
Specificity	86.36	45.45
Positive Predictive Value	72.72	29.41
Negative Predictive Value	82.60	58.82
F1-Score	69.56	34.48

5. CONCLUSIONS AND SUGGESTIONS

Conclusions

Based on the results of the analysis and discussion, the conclusions that can be drawn from the research on the classification of prospective recipients of the Indonesian Smart Card Scholarship (KIP) using the Decision Tree and Naïve Bayes algorithms are as follows:

a) In the Decision Tree training process, numerical attributes such as diploma grades and average report cards do not have a significant influence in determining or classifying pass and fail data. This is indicated by the information gain value of the attribute, which is 0. The information gain value ranges from 0 to 1, and 0 indicates that the attribute has no influence or relationship with the passing status. The information gain value is also lower than the information gain value generated by other attributes such as KIP Ownership, KKS Ownership, KTM Ownership, Parents' Salary, Number of Dependents, Parents' Occupation, High School Graduate Status and equivalent, and Difference in years of graduation and kip list year.

b) The measurement results of the Decision Tree Algorithm classification with the Confusion Matrix showed an accuracy value of 44.12% and F1-Score 34.48%. It can be concluded that the Decision Tree Algorithm used has not found a pattern to distinguish prospective KIP beneficiaries who will pass or fail.

c) The Naive Bayes algorithm produces an accuracy value of 76.47% using data on diploma scores and average report cards. For the accuracy produced without using data on the value of diplomas and the average report card is 79.41%. It can be concluded that Naïve Bayes has a better performance even though it does not use data on diploma scores and average report cards.

d) Measurement of the results of the Naïve Bayes classification with Confusion Matrix showed low sensitivity, namely 66.67% and 58.33% for the first and second scenarios. This means that the Naïve Bayes algorithm is difficult to recognize data that goes into the Disqualified class. While the Specificity value generated by this algorithm reaches 86.3% for both scenarios, it means that the accuracy of the Naïve Bayes algorithm to identify prospective recipients who pass is quite high.

e) Based on the evaluation results, the Naïve Bayes algorithm has a better performance than the Decision Tree algorithm in classifying KIP scholarship recipients. The Naïve Bayes algorithm is successful in classifying prospective KIP recipients better by only looking at the effect of attributes with KIP passing status, which is different from Decision Tree, which depends on the relationship or dependency between attributes on KIP pass status.

Suggestions

Suggestions from the author for further research are:

a) Based on the research that has been done, the researcher recommends adding datasets from other campuses that have students who receive KIP assistance.
b) Added more data and information about KIP beneficiaries for use in the training and testing process.
c) It is necessary to develop the Decision Tree and Naïve Bayes algorithms to improve the classification results of prospective KIP beneficiaries.
d) Based on the research that has been done, the researcher recommends for universities to use the Naïve Bayes algorithm as a tool to make it easier to make decisions in the process of determining KIP scholarship recipients.
e) Researchers suggest that socialization be carried out to related parties, namely parents, students and universities regarding the patterns or rules obtained to assist the process of receiving KIP scholarships.

REFERENCES

Abduallah, Y. (2020). *CS 634-001: Data Mining*.

Agustiani, A. (2016). *Klasifikasi Penyakit Karies Gigi Pada Citra Dental Panoramic Menggunakan Metode Binary Decision Tree Support Vector Machine (Bdtsvm)*. Universitas Brawijaya.

Buntine, W. (2020). Learning classification trees. In *Artificial Intelligence frontiers in statistics* (pp. 182–201). Chapman and Hall/CRC.

Chicco, D., Tötsch, N., & Jurman, G. (2021). The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*, 1–22. https://doi.org/10.1186/s13040-021-00244-z

Dulli, S., Furini, S., & Peron, E. (2009). *Data mining: metodi e strategie*. Springer Science \& Business Media.

Godlewski, E., & Raviart, P.-A. (2013). *Numerical approximation of hyperbolic systems of conservation laws* (Vol. 118). Springer Science \& Business Media.

Gyssens, M., & Paredaens, J. (1984). A decomposition methodology for cyclic databases. In *Advances in data base theory* (pp. 85–122). Springer.

Han, J., & Micheline, K. (2006). *Data Mining: Concepts and Techniques* (Second Edi, Issue Second Edition). Morgan Kaufmann Publishers. https://doi.org/10.1007/978-3-642-19721-5

Iswanto, M. H., Permanasari, A. E., Nugroho, H. A., Grafika, J., & Bulaksumur, N. (2015). Pemanfaatan Teknik Data Mining Untuk Diagnosis Penyakit Tuberculosis (TBC). *Seminar Nasional Teknologi Informasi Dan Multimedia (STMIK AMIKOM), Yogyakarta*, 121–126.

Kabir, M. F. M., Rahman, C. C. M., Hossain, A., & Dahal, K. (2011). Enhanced Classification Accuracy on Naive Bayes Data Mining Models. *International Journal of Computer Applications (0975–8887) Volume*, *28*(3), 9–16.

Maimon, O., & L, R. (2005). *Decomposition Methodology for Knowledge Discovery And Data Mining*. https://doi.org/10.1007/978-1-4615-9385-0_4

Retnaningsih, H. (2019). Program Indonesia Pintar: Implementasi Kebijakan Jaminan Sosial Bidang Pendidikan. *Aspirasi: Jurnal Masalah-Masalah Sosial*, *8*(2), 161–177.

Zhou, H. (2020). Excel and Data Mining. In *Learn Data Mining Through Excel* (pp. 1–19). Springer.