



*Corresponding author: Zulkifli Rais,
Statistics Study Program, Faculty of
Mathematics and Natural Sciences,
Makassar State University, Indonesia.

E-mail: zulkifli.rais89@unm.ac.id

RESEARCH ARTICLE

Classification of Family Welfare Card Recipients in Makassar City Using Decision Tree Algorithms

Zulkifli Rais*, Muhammad Fahmuiddin S, Musfira

Statistics Study Program, Faculty of Mathematics and Natural Sciences,
Makassar State University, Indonesia.

Abstract: This study aims to analyze the factors influencing the determination of recipients of the Family Welfare Card (KKS) program in Makassar City and evaluate the level of accuracy of the decision tree model in the classification process. The KKS program is a government effort to accelerate poverty alleviation, so it is important to ensure that the selection process for program recipients is carried out on target. The decision tree method is used in this study because of its ability to simplify the decision-making process through an easy-to-understand tree structure. This study utilizes KKS recipient data with various variables, such as income, number of dependents, employment status, asset ownership, and education level, to build a classification model. The results of the study indicate that the variable of the Head of Household's (KRT) Highest Education Level (X4) has the highest level of importance in determining KKS recipients, followed by the variable Number of Family Members (X1), and the variable Ownership of Residential Buildings (X5). The decision tree model that was built has an accuracy level of 84.21%, which states the model's ability to classify KKS recipients effectively. This study also provides insight into the description of factors influencing KKS receipts, which can be used as a basis for formulating more efficient and targeted policies.

Keywords: Decision Tree, Family Welfare Card, Classification, Accuracy, Social Policy.

1. INTRODUCTION

Data mining is a process that utilizes statistical methods, mathematics, artificial intelligence, and machine learning to find patterns or information in a particular data set (Mardi, Y 2017). The goal is to extract valuable information from a large database. These patterns are identified through the use of analytical tools that can provide in-depth data insights, which can then be further used with the help of a decision support system (Almufqi & Voutama, 2023). Said, & Ruliana, 2022).

The classification method is the process of analyzing data to obtain a model to describe the classification class. perform modeling derived from training data which is then used to classify data in testing (Rais, Said, & Ruliana, 2022). Decision trees and decision rules have the main benefit of simplifying the decision-making process, making it easier for decision makers to understand and interpret solutions to a problem (Rizmayanti, Hidayat, Nugraha, & Gata, 2021). In this study, KKS

Used as an indicator to identify underprivileged families who are entitled to receive various forms of social assistance, including the Prosperous Family Savings Program (PSKS). The KKS program is a development of the previous program known as the Social Protection



Card (KPS), which was launched during the administration of President Susilo Bambang Yudhoyono. Currently, the program has undergone a gradual transformation into KKS. In addition to being a marker of the family's economic status, KKS also acts as an identity card to access assistance from PSKS (Nurjoko & Yuliawati, 2015).

Several previous studies that applied decision trees, including; (Ridho & Hendra, 2021) on the classification of Covid-19 disease diagnosis using the decision tree method which produced a decision tree model that had an accuracy of 90%. (Depari, Widiastiw, & Santonii, 2022) on the comparison of decision tree models, Naive Bayes and Random Forest for predicting heart disease classification where the random forest model is the best method used in the results of the highest classification accuracy of 75% followed by the decision tree model 72% and lastly with another Bayes model of 71%.

2. Literature Review

2.1. Definition of Decision Tree

Statistics is a branch of science that discusses various methods for collecting, managing, presenting, analyzing, and interpreting data in numerical form. Thus, statistics plays a role as a science that supports a more precise decision-making process through the collection, organization, presentation, analysis, and interpretation of data into useful information (Hidayati, Handayani, & Ikasari, 2019). One approach in classification analysis is the decision tree method, which is an interesting classification technique because it forms a tree structure with decision nodes that are interconnected through branches, starting from the root to reaching the leaf node (terminal) (Larose & Daniel, 2005).

2.2. Model Decision Tree

The Decision Tree method is a structure like a flowchart that resembles a tree shape, where each internal node shows a test of an attribute, each branch shows the results of the test, and the leaf node describes the class or class distribution. The decision-making process is carried out by following the path from the root node to the leaf node that presents the prediction results (Aprilla Dennis, 2013). The selection of attributes as the root of the tree is based on the attribute that has the highest gain value. Before determining the gain value

From an attribute, the entropy value is first calculated using a specific formula:

$$H(S) = - \sum_{i=1}^n P_i \log_2 P_i$$

Information:

H : Entropy

S : Group (set data)

n : number of record

Pi : the proportion Si si to S

$$IG(A) = H(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} H(S_i)$$

Information:

IG : Gain

H : Entropy

S : Group (data set)

A : attribute of attributes of partition i

n : number of attributes of partition i

|S_i| : number of cases on prtition i

|S| : number of cases in s



2.3. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) is a process to find hidden, unknown, and potential information from a data set. The stages in KDD include the data mining process, which is the process of finding patterns or tendencies in data, which are then converted into information that can be easily understood and is more informative (Andayani, 2007).

2.4. Decision Tree Based Classification

Decision trees are one of the classification methods that are widely used because they are easy for humans to understand and interpret. This model works by utilizing a tree-like structure to make predictions (Romindo, 2021). In general, a decision tree is a predictive model built in the form of a branching structure, where the structure consists of nodes and branches that describe the results of the classification process.

2.5. Determining the Level of Model Accuracy

The assessment of the level of accuracy of a model is done by looking at the highest percentage of predictions or close to 100%. To ensure the stability of the results, a robustness check is carried out to evaluate whether the final results remain consistent despite changes in assumptions, and to ensure that the main findings remain valid (Gelman, A., 2017). The accuracy of the prediction model is analyzed based on the extent to which the model's estimated results match the interest coverage ratio value. This condition is important, especially in identifying companies in financial distress (Claessens et al., 2003).

2.6. Family Welfare Card Assistance Program (KKS)

The Family Welfare Card (KKS) is one of the government's initiatives to accelerate poverty alleviation efforts, as stipulated in Presidential Regulation Number 166 of 2014. This program is a continuation and development of the previous program, namely the Social Protection Card (KPS), which was first implemented during the administration of President Susilo Bambang Yudhoyono. Over time, the program has undergone gradual changes to become the KKS. The KKS functions as an identity for underprivileged families who are entitled to receive various forms of social assistance, including the Family Welfare Savings Program. This assistance is provided in non-cash form in the form of savings worth IDR 200,000 per family every month, which is intended for around 15.5 million underprivileged families throughout Indonesia (Andela, 2018).

3. Research Method and Materials

3.1. Types of research

The type of research used is quantitative research. The quantitative approach is research that using the decision tree method with classification accuracy for recipients of the Prosperous Family Card (KKS) program in Makassar City.

3.2. Data Source

The data used in this study are secondary data such as micro data from the National Social Survey (SUSENAS) in March 2023 accessed from <https://silastik.bps.go.id/v3/index.php/site/index#tta>, namely data on recipients of the Prosperous Family Card (KKS) program in Makassar City.

3.3. Operational Definition Of Variable

The operational definition of the variable is as follows:

- (1). Recipients of the Prosperous Family Card Program (KKS) (Y)
- (2). Number of Household Members (ART) (X1)
- (3). Gender of Head of Household (KRT) (X2)
- (4). Age of Head of Household (KRT) (X3)
- (5). Highest Education Level of Head of Household (KRT) (X4)

- (6). Ownership of Residential Building (X5)

3.4. Research Procedures

The data analysis techniques used in this study are:

- (1). Collect references, data and information needed for research.
- (2). Process data using decision trees using the R program.
- (3). Make conclusions based on the research results that have been discussed
- (4). Compile research results reports.

3.5. Data Analysis Techniques

The data techniques used in this research are:

- (1). Entering data on recipients of the Prosperous Family Card (KKS) as the dependent variable and the number of household members, gender of the household head, age of the household head, highest level of education of the household head, ownership of residential building.
- (2). Conducting data exploration from recipients of the Prosperous Family Card (KKS) program in Makassar City.
- (3). Conduct analysis using a decision tree with the following steps;
 - (a). Perform dataset processing.
 - (b). Making with decision tree
 - (c). Creating a classification model
- (4). The classification accuracy results have the highest value as the appropriate method.
- (5). Conclusion and suggestions.

4. Results and Discussion

4.1. Descriptive Analysis

4.1.1. Number of KKS Recipients

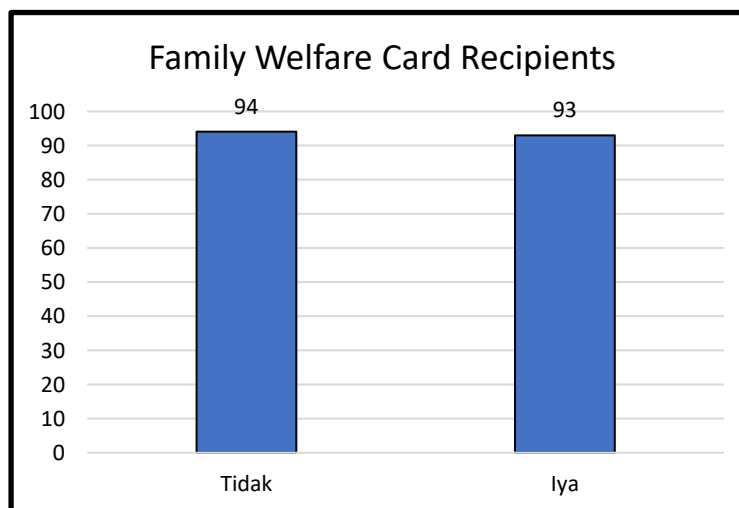


Figure 1. Number of KKS Recipients

Figure 1 shows an equal distribution between recipients and non-recipients of the Family Welfare Card (KKS). Of the total 187 individuals, each category consists of 93 individuals (49%) for KKS recipients and 94 people (51%) for those who did not receive KKS. This shows a balanced proportion between those who received KKS and those who did not receive it.

4.1.2. Distribution of KKS Recipients Based on the Number of Household Members (ART)

Table 1. Distribution of KKS Recipients based on the number of ART

Y	No	Of	Total	Percentage
ART \leq 4	69	47	116	62%
ART $>$ 4	25	46	71	38%
Total	94	93	187	100%

Based on table 4.1 above, recipients of the Prosperous Family Card (KKS) are more common in households with less than four members, which is 47 people, compared to households with more than four members, which is 46 people. Meanwhile, the group that did not receive the KKS was dominated by households with less than four members, which was 69 people, while the other 25 people came from households with more than four members. This shows that households with more members tend to have a higher proportion of KKS recipients than smaller households.

4.1.3. Distribution of KKS Recipients Based on Gender of Head of Household (KRT)

Tabel 1. Distribution of KKS Recipients based on KRT Gender

Y	No	Of	Total	Percentage
Woman	17	20	37	20%
Man	77	73	150	80%
Entire	94	93	187	100%

In the table above, KKS recipients are more male than female, with a total of 73 males and 20 females respectively. In contrast, the group that did not receive KKS was dominated by males with a total of 77 people, then females only 17 people. This shows that proportionally, KKS recipients are more male than female.

4.1.4. Distribution of KKS Recipients with the Head of Household (KRT) age variable

Table 2. Distribution of KKS Recipients with KRT age variable

Y	No	Of	Total	Percentage
\leq 50 Years	56	58	114	61%
$>$ 50 Years	38	35	73	39%
Total	94	93	187	100%

Based on the table above, KKS recipients are more in the group of Heads of Households (KRT) aged under 50 years, which is 58 people, compared to KRT aged over 50 years, which is 35 people. On the other hand, the group that did not receive KKS was also dominated by KRT aged under 50 years, as many as 56 people, while KRT aged over 50 years only numbered 38 people. This shows that both recipients and non- recipients of KKS are more from KRT aged under 50 years.

4.1.5. Distribution of KKS Recipients with the Head of Household's Last Education Level Variable

Table 3. Distribution of KKS recipients with the variable of KRT's last education

Y	No	Of	Total	Percentage
$<$ SMA	29	57	86	54%
\geq SMA	65	36	101	46%
Total	94	93	187	100%

Based on the table above, KKS recipients are more often found in Heads of Households (KRT) with a last level of education below high school, which is 57 people, compared to KRT who have an education equivalent to high school or higher, which is 36 people. On the other hand, the group that did not receive KKS is dominated by KRT with an education equivalent to high school or higher, as many as 65 people, while KRT with an education below high school only numbered 29 people. This shows that KKS recipients proportionally come more from KRT with a lower level of education.

4.1.6. Distribution of KKS Recipients with Residential Ownership Ownership Status Variabel

Tabel 4. Distribution of KKS Recipients with Residential Ownership status Variable

Y	No	Of	Total	Percentage
One's own	31	33	64	34%
Not my own	63	60	123	66%
Total	94	93	187	100

Based on the table above, KKS recipients are more often found in the group of Heads of Households (KRT) who do not have their own residence, which is 60 people, compared to KRT who have their own residence, which is 33 people. On the other hand, the group that does not receive KKS is dominated by KRT who do not have their own residence, as many as 63 people, while KRT who have their own residence only number 31 people. This shows that KKS recipients are more often found in the group of KRT who do not have their own residence compared to those who have their own residence.

4.2. Data Processing

Data processing is an important stage in data analysis and modeling, where the goal is to prepare data so that it is ready to be used in building a model. In this case, several steps are taken in the data processing process for the application of the decision tree method. The following is an explanation of the stages of data processing that are carried out:

4.2.1. Cleaning Data

Through the data obtained in Figure 4.1, it shows that the research data is balanced because the respondents who received KKS numbered 93 and did not receive KKS numbered 94, in addition, the contents of the research data to be processed are also complete, so there is no need for a data cleaning process . By improving data balance, the decision tree model can obtain more accurate predictions in all classes, which in turn can provide better results in real applications, especially in cases involving classification with very unbalanced classes.

4.2.2. Split Data Training and Data Testing

Next, the data is divided into training data (90%) and testing data (10%) randomly for training purposes. and model evaluation. Here is a table of training and testing data distribution:

Tabel 5. Distribution of Data Training dan Testing

Data	Number of KKS Recipients	
	No	Of
Data integrity	94	93
Data Training	84	84
Data Testing	10	9

Splitting the data into a training set and a testing set is a crucial step in building a model. The training set is used to help the model learn existing patterns, while the testing set is used to see how well the model can make predictions based on new data that it has never seen before. This division ensures that the model not only performs well on data that it already knows, but can also produce accurate results in real-world situations.

4.3. Formation of Decision Tree Moel

The classification model is built using the decision tree method with a model formula that uses variables X1, X2, X3, X4, dan X5 and X5 to predict variable Y. The following is a table of node divisions in the decision tree model:

Tabel 6. Results of the decision tree model

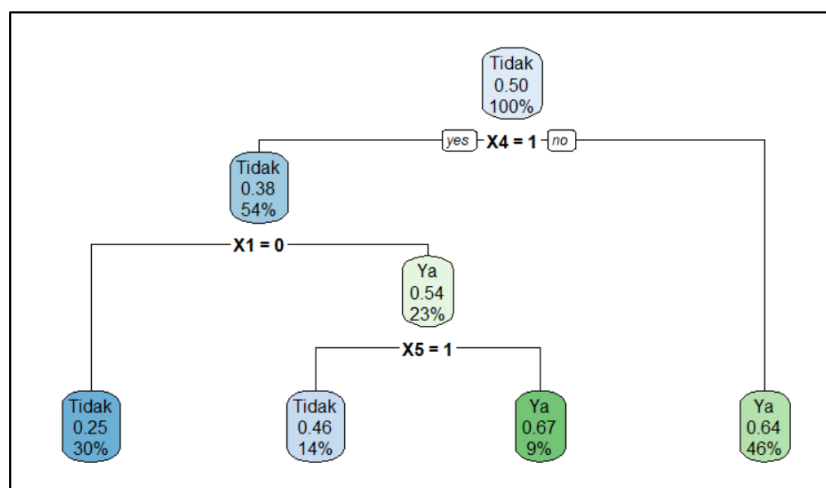
Node	Condition	Amount Data	Class Prediction	Level Error
1	Root	168	Tidak	0,500
2	$X_4 = 1$	90	Tidak	0,378
3	$X_4 = 0$	78	Ya	0,359
4	$X_4 = 1, X_1 = 0$	51	Tidak	0,255
5	$X_4 = 1, X_1 = 1$	39	Ya	0,462
10	$X_4 = 1, X_1 = 1, X_5 = 1$	24	Tidak	0,458
11	$X_4 = 1, X_1 = 1, X_5 = 0$	15	Ya	0,333

Based on the results of the decision tree obtained, the X_4 variable has a major role in predicting KKS recipients. At the first node (root), the data is evenly divided between the "No" and "Yes" classes with an error rate of 0.500, indicating that the initial model has not been able to separate the data effectively. At the second node, data with a value of $X_4 = 1$ is more likely to be predicted as "No", with the error rate decreasing to 0.378.

On the other hand, at the third node, data with a value of $X_4 = 0$ is more dominantly predicted as "Yes", with a lower error rate, which is 0.359. This indicates that the X_4 variable is a fairly good initial separator. Further division occurs at nodes four and five based on the X_1 variable. At node four, with the conditions $X_4 = 1$ and $X_1 = 0$, the prediction is more dominant to the "No" class with a lower error rate, which is 0.255, indicating that this combination is quite effective in predicting this class. Meanwhile, at node five, with the conditions $X_4 = 1$ and $X_1 = 1$, the prediction changes to "Yes", but the error rate increases to 0.462, indicating that the data in this branch is more difficult to separate. In further division at nodes ten and eleven, the X_5 variable also affects the prediction results. At node ten, with the conditions $X_4 = 1, X_1 = 1,$ and $X_5 = 1$, the prediction remains "No" with an error rate of 0.458. On the other hand, at node eleven, with conditions $X_4 = 1, X_1 = 1,$ and $X_5 = 0$, the prediction changes to "Yes", with a lower error rate of 0.333.

Although the amount of data at this node is relatively small, variable X_5 makes an additional contribution to improving the prediction. Overall, this decision tree shows that variable X_4 is a significant initial separator, followed by X_1 which smoothes the data division, and X_5 has an additional influence on certain branches. The combination of these three variables produces varying levels of error, reflecting the effectiveness of the division in predicting KKS recipients. Based on the table above, this decision tree can be visualized as follows:

Figure 2 Decision Tree Results



Based on the visualized decision tree results, variable X_4 is the main factor in predicting KKS recipients. At the root node, the data is evenly divided between the "No" and "Yes" classes, with a probability of 0.50 each. This shows that without considering other variables, the

chances of predicting both classes are the same. At the second node, if $X_4 = 1$, the data is more likely to be predicted as "No" with a probability of 0.38, covering 54% of the total data. Conversely, if $X_4 = 0$, the data is more dominantly predicted as "Yes" with a probability of 0.64, covering 46% of the total data. Furthermore, on the branch with $X_4 = 1$, variable X_1 becomes the next divisor. If $X_1 = 0$, the probability of predicting "No" decreases to 0.25, covering 30% of the total data. However, if $X_1 = 1$, the data is more likely to be predicted as "Yes" with a probability of 0.54, covering 23% of the total data. On the branch with $X_1 = 1$, variable X_5 becomes the last divisor. If $X_5 = 1$, the data is more likely to be predicted as "Yes" with a probability of 0.67, even though it only covers 9% of the total data. Conversely, if $X_5 = 0$, the probability of predicting "No" increases to 0.46, covering 14% of the total data.

4.4. Influence of Variables on KKS Recipient Classification

The influence of variables on the classification of recipients of the Prosperous Family Card (KKS) is measured through the variable value importance in the decision tree model. This value shows how much each variable contributes to determining the classification results. Here is a table of importance variables :

Tabel 7. Influence of Variables on KKS Recipient Classification

Variabel	Value of Interest
X_4	5,791453
X_1	3,628196
X_5	0,801282
X_2	0,222748

Based on the previous table, variable X_4 has the highest importance value of 5.791453, indicating that this variable has the greatest influence on the classification of KKS recipients. Variable X_1 is in second place with a value of 3.628196, which also makes a significant contribution. In contrast, variables X_5 (0.801282) and X_2

(0.222748) has a much smaller influence on the model. This indicates that variables X_4 and X_1 are more important to consider in the classification process, while X_5 and X_2 have a more minor role.

4.5. Model Based Prediction

After modeling with decision tree The next step is to predict the testing data to see how accurate the previously obtained model is. Here is a table of the results of the testing data prediction:

Tabel 8. Prediction Results

Prediction/ Current	No	Of
Tidak	9	2
Ya	1	7

The table presented shows the results of the model prediction compared to the actual data. Out of a total of 19 observations, the model successfully predicted 9 cases correctly as "No" and 7 cases correctly as "Yes". However, there were some prediction errors, where the model predicted 2 cases as "Yes" which were actually "No", and 1 case that was predicted as "No" but should have been "Yes". This gives an idea of the level of accuracy of the model in predicting different classes, as well as the potential errors that can occur. Further evaluation can be done to see how improvements can be made to this prediction model.

4.6. Model Accuracy Test

Model Accuracy Test is a step to evaluate how well the model created can predict the target value (variable Y) based on the given data. In the context of decision trees, accuracy testing is done to assess whether the model classification results are in accordance with the actual data. Accuracy is calculated as the percentage of the number of correct predictions to the total test data. The following are the results of the model accuracy obtained:



$$\text{Curation} = \frac{\text{large number of predictions}}{\text{Total amount of data}} \times 100\%$$

$$\text{Accuracy} = \frac{9 + 7}{19} \times 100\% = 84,21\%$$

Tabel 9. Accuracy Results

	Accuracy Results
Model <i>Decision Tree</i>	84,21%

Based on the table above, it shows that the decision tree model has an accuracy lever of 84.21%. This means that the model is able to correctly predict 84.21% of the total data used in the test.

In other words, about 15.79% of the data predicted incorrectly by the model. This level of accuracy shows that the decision tree model is quite reliable in classifying.

5. Conclusion

5.1. Conclusion

Based on the results of research on the application of the decision tree model in predicting recipients of the Card Program Prosperous Family (KKS) in Makassar City, the following conclusions were obtained:

- (1). Dominant Influential Variables The decision tree results show that variable X4 has the most dominant influence in predicting KKS recipients. This variable is the main factor in dividing data at the root node, followed by variables X1 and X5 which contribute to the following branches.
- (2). Based on the variable importance analysis, variable X4 has the greatest influence in the classification with a value of 5.791453, followed by X1 (3.628196) and X5 (0.801282). On the other hand, variable X2 has a very small influence (0.222748), so it is less significant in determining the data division.
- (3). The performance of the decision tree model that was built has an accuracy level of 84.21%, which states that
- (4). This model is quite reliable in predicting KKS recipients. However, there are several prediction errors that indicate that the model still needs improvement, perhaps by adding variables or adding research data.

5.2. Suggestions

Based on the results of this study, several suggestions can be given, namely:

- (1). To improve model accuracy, it is recommended to explore additional relevant variables, such as demographic data, economic conditions, or other information that may affect the status of KKS recipients. Additionally, the use of other machine learning methods such as Random Forest or Gradient Boosting can be considered to compare model performance.
- (2). The results of this study can be used by related parties, such as local governments or social institutions, as a reference in determining policies related to the distribution of KKS. The use of decision tree models can help in making more efficient and data-based decisions.
- (3). Model validation with larger and more diverse data is needed to ensure its reliability under various conditions. Testing with data from other regions can also be done to see if this model can be applied more widely.
- (4). In addition to a data-based approach, decision-making regarding KKS recipients also needs to consider social and policy factors that are not fully reflected in the data. This is important to ensure that the KKS program is truly on target.

References

- Almufqi, FM, & Voutama, A. (2023). Comparison of Data Mining Methods to Predict Academic Achievement Students. *Journal of the Faculty of Engineering, Islamic University of Lamongan*, 61-66.
- Andayani, S. (2007). Cluster Formation in Knowledge Discovery in Database. *Journal of Research Trends Mathematics and Mathematics Education in the Global Era*, 1-10.
- Aprilla Dennis (2013) 'Belajar Data Mining dengan RapidMiner', *Innovation and Knowledge Management in Business Globalization: Theory & Practice*, Vols 1 and 2, 5(4), pp. 1-5. Available at: <https://doi.org/10.1007/s13398-014-0173-7.2>.
- Claessens, S., Djankov, S., & Klapper, L. (2003). Resolution of Corporate Distress in East Asia. *Journal of Empirical Finance*, 10(1-2), 199-216. [https://doi.org/10.1016/S0927-5398\(02\)00023-3](https://doi.org/10.1016/S0927-5398(02)00023-3)
- Depari, DH, Widiastiw, Y., & Santonii, MM (2022). Comparison of Decision Tree, Naive Bayes and Random Models Forest for Heart Disease Classification Prediction. *Journal of Informatics*, 239-248.
- Gelman, A. (2017). What's the Point of a Robustness Check? 04 21, 2021, dari *Statistica Modeling, Casual Inference, and Social Science*. (<http://andrewgelman.com/2017/11/29/whats-pointrobustness-check/>, diakses 21 April 2021) *Basic Statistics*.
- Larose, & T., D. (2005). *Discovering Knowledge In Data: An Introduction To Data Mining*. Jurnal John Wiley and Sons, USA, 11-20.
- Mardi, Y. (2017). Data Mining: Classification Using C4 Algorithm. 5. *Journal of Edik Informatics Informatics Research in Computer Science and Informatics Education*, 213-219.
- Nurjoko, & Yuliawati, D. (2015). Decision Support System for Recipients of the Prosperous Family Card (KKS) Using the Simple Additive Weighting (SAW) Method. *Darmajaya Team Journal*, 203-217.
- Rais, Z., Said, R. N., & Ruliana. (2022). Text Classification on Sentiment Analysis of Marketplace SHOPEE Reviews On Twitter Using K-Nearest Neighbor (KNN) Method. *Journal of Information and Visualization*, 1-8.
- Ridho, R., & Hendra. (2021). on Classification of Covid-19 Disease Diagnosis Using the Decision Tree Method. *Journal of Information Systems, Information Technology and Computers*, 69-75.
- Rizmayanti, AI, Hidayat, NH, Nugraha, FS, & Gata, W. (2021). Application of Data Mining to Predict Student Competence Using the Decision Method (Case Study of SMK Multicomp Depok). *SWABUMI Journal*, 9-18.
- Romindo. (2021). Implementation of Heuristic Techniques and Decision Tree Modeling on Iq Game Design. *Journal of Informatics Management, Ganesha Polytechnic Medan*, 22-27.