

Data Mining for Processing of Research and Community Service by Lecturer Using Decision Tree Method

¹Muhammad Fikry, ^{2*}Rizal, ³Fadlisyah, ⁴Nuridin, ⁵Bustami, ⁶Defry Hamdhana, ⁷Mukti Qamal

ABSTRACT--Data on community service and research by the lecturer are important and must be possessed by every lecturer and institution. This is necessary for an effort to ensure quality management of the institution. Therefore needed to get an appropriate algorithm in classifying the data. The Decision Tree algorithm is one of the most powerful and well-known classifications and prediction methods. The decision tree method converts huge facts into decision trees that represent rules. Rules can be easily understood with natural language. Decision Tree algorithm has achieved the main objective of the system to obtain research information and community service at the lecturers. The system work can provide an overview about of the quality of lecturers with data calculated using the Decision Tree method which has an accuracy of 66.67%.

Key words— processing, research, community service, lecturer, decision ,tree method

I. INTRODUCTION

A web-based information system is needed in data processing of lecturers' dedication and research so that the data becomes more structured and efficient so that lecturers can more easily and effectively manage their dedication and research data. With this information system, it can also help lecturers publish their devotion and research reports online for each lecturer so that they can be easily accessed by lecturers, recapitulated by the university and widely accessible on any device.

Data mining is a method of finding new information (knowledge) contained in very large data. in Mining data there are several methods, namely prediction, classification, clustering, and association. Decision tree is one of the most powerful and well-known classification and prediction methods in applying data mining. basically a decision tree converts data into a decision tree and decision rules. the advantage in this method is that it is effective in analyzing a large number of attributes of existing data and easily understood by end users. The problem in this research is how to analyze the feasibility of the Decision Tree algorithm in processing lecturer devotion and research data online.

¹ Informatics Department, Universitas Malikussaleh

^{2*} Informatics Department, Universitas Malikussaleh

³ Informatics Department, Universitas Malikussaleh

⁴ Informatics Department, Universitas Malikussaleh

⁵ Informatics Department, Universitas Malikussaleh

⁶ Informatics Department, Universitas Malikussaleh

⁷ Informatics Department, Universitas Malikussaleh

This research is used for the processing of lecturer devotion and research databases, so that Facilitates lecturers in the process of lecturer data and research archives and makes it easier for the Malikussaleh University to examine and recapitalize lecturer research and research data.

II. LITERATURE REVIEW

Research

Research is a process of inquiry that is carried out actively, diligently, and systematically, where the aim is to find, interpret, and revise facts.

The purpose of research is to find or obtain data for specific purposes and purposes. Therefore, research or research must be done scientifically based on scientific characteristics, namely:

- Rational; the point is that research must be done in ways that are reasonable or rational so that it can be reached by human reason.
- Empirical; the point is that research must be based on the source of knowledge obtained from observations of the human senses. That way, the method can also be observed by others.
- Systematic; the point is that research must be carried out through certain steps that are logical and orderly in accordance with a system that has been arranged so that it can explain the sequence of causation of an object of research.

Community Service

Community service is an implementation of the practice of science, technology and cultural arts directly to the community institutionally through scientific methodologies as the dissemination of the Tri Dharma of Higher Education and noble responsibilities in efforts to develop community capacity, so as to accelerate the growth rate of achieving national development goals.

The objectives of community service are as follows:

1. Increasing the speed of the process of increasing the ability of human resources in accordance with the rate of development growth.
2. Increasing the pace of community development efforts towards the development of a harmonious and dynamic society that is ready to undergo changes towards improvement and progress in accordance with socio-cultural values and norms in the lives of developing communities in the prevailing community life.
3. Increasing the speed of efforts to foster community institutions and professions in accordance with the rate of growth of the modernization process in the life of the community itself.
4. To obtain feedback and input for faculties in order to improve the relevance of education, it is necessary to have experts who have interdisciplinary and multidisciplinary abilities.

Data Mining

Data mining is a process of searching or searching automatically to get useful information in a very large data repository. There are many other names of data mining such as Knowledge Discovery Databases (KDD),

knowledge extraction, data / pattern analysis, data archeology, data dredging, information harvesting, business intelligence.

Decision Tree

Decision Tree Decision tree is one technique that can be used to classify a group of objects or records. This technique consists of a collection of decision nodes, connected by branches, moving down from the root node until it ends at the leaf node. Decision tree development starts from the root node, based on the convention placed at the top of the decision tree diagram, all attributes are evaluated at the decision node, with each outcome that might produce a branch. Each branch can enter either the other decision node or the leaf node. Yusuf, 2007 decision tree is a nonlinear discrimination method that uses a set of independent variables to divide the sample into smaller groups gradually. The procedure is done iteratively in each branch of the tree, i.e. select an independent variable that has the strongest association with the dependent variable according to certain criteria. Requirements that must be met in the application of Joseph's decision tree algorithm, 2007: (1) The decision tree algorithm represents supervised learning so it requires preclassified targets. (2) Training data sets must be rich and varied. (3) The target class attribute must be discrete.

C4.5 algorithm can handle continuous and discrete numerical data. The split for numeric attributes is to sort the examples based on the continuity attribute A, then form the minimum threshold beginning M from the examples from the majority class on each contiguous partition, then join the contiguous partitions with the same majority class. Split for the discrete attribute A has the form value $A \in X$ where $X \subset \text{domain } A$. If a data set has multiple observations with missing values, namely a record with some variable values does not exist, if the number of observations is limited, then attributes with missing values can be replaced with a mean value average of the variable in question. To separate the split object, a test is performed on the attribute by measuring the level of impurities in a node.

In the C.45 algorithm using the gain ratio. Before calculating the acquisition ratio, it is necessary to first calculate 15 information values in units of bits from a collection of objects. How to calculate it is done by using the concept of entropy. $ES = -p_+ \log_2 p_+ - p_- \log_2 p_-$ S is the sample data space used for training, p_+ is the number of positive or supportive solutions in the sample data for certain criteria and p_- is the number of negative or unsupported solutions in the sample data for some criteria. Entropy is equal to 0 if all the examples in S are in the same class. Entropy is equal to 1 if the number of positive and negative examples in S is the same. Entropy is more than 0 but less than 1 if the number of positive and negative examples in S is not the same. Split entropy which divides S with n records into sets S1 with n1 lines and S2 with n2 lines is: $1, 2 = 1 + 2$ Then calculates information acquisition from the output data or dependent variable y which is grouped based on attribute A, denoted by gain yes. Information acquisition, gain y, A, from attribute A relative to output data y is: $\text{Gain } y, A = \text{entropy } y - \text{entropy } y_c$ A value is all possible values of attribute A, and y_c is a subset of y where A has a value c. The first term in the equation above is total entropy y and the second term 16 is entropy after data separation is done based on attribute A. To calculate the acquisition ratio, it is necessary to know a new term called SplitInfo information separation. Separation of information is calculated by that S1 through Sc are c subsets resulting from solving S by using attribute A which has as many as c values.

III. DISCUSSION

Research Methods

Based on research that has been done on the devotion and research data of each lecturer at the University of Malikussaleh environment will fill in the data through the online system so that the data obtained is in accordance with the development of each lecturer, and in this system, each lecturer must upload a document so that the data recapitulation lecturers are also owned and stored on a computerized basis.

The data that has been stored is managed to search for information in the research document and community service of lecturers in order to obtain accurate information for users in terms of reference and linkages between fields of science. The search uses the Decision Tree Method. The following is testing data from the Decision Tree:

Table 1: Data Testing

Jumlah Kata	Nama File	Bidang Terkait	Persamaan Kata
1163	File-1	7	44
2466	File-2	29	69
2466	File-2	29	69
1163	File-1	7	44
1609	File-3	15	69
2466	File-2	29	44
1609	File-3	15	69
1609	File-3	15	44
2466	File-2	29	44
1163	File-1	7	69
1609	File-3	15	69
1609	File-3	15	44
1163	File-1	7	69
1163	File-1	7	44

From the test data above, a gain ratio search is performed. The results of the gain ratio are as follows:

Table 2: Data Gain Ratio

	JML KASUS	44 (S1)	69 (S2)	ENTROPY	INFORMATION GAIN	SPLIT INFO	GAIN RATIO
TOTAL	14	7	7	1			
jumlah kata					0,186601582	1,577406283	0,118296462
1163	5	3	1	0,906564975			
2466	4	3	2	0,811278124			
1609	5	1	4	0,721928095			
Nama File					0,020749575	1,577406283	0,013154237
File-1	5	3	2	0,970950594			
File-2	4	2	2	1			
File-3	5	2	3	0,970950594			
Bidang Terkait					0,367517645	1,046896702	0,351054354
7	5	3	2	0,970950594			
29	4	2	2	1			
15	5	3	2	0,970950594			

To get the level of accuracy of the entropy value is done using the python language, so that obtained by 66.67%.

Accuracy results from Entropy values

Score: [50.0, 50.0, 100.0]

Mean Accuracy: 66.67%

IV. CONCLUSIONS

Based on the results of the study the author can conclude in the form of an application that has reached the main goal of the system to obtain research information and community service lecturers with data calculated using the Decision Tree method which has an accuracy of 66.67%.

REFERENCES

1. Abidin, T. F., Munadi, K., Hasibuan, Z. A., & Fikry, M. (2018). Classification of hadith levels using data and text mining techniques. *Eurasian Journal of Analytical Chemistry*, 13(6), 186-191.
2. Fathansyah, I. (2012). *Basis Data edisi: revisi*. Bandung: CV. Infomatika.
3. Fikry, M. (2016). RANCANGAN BASIS DATA KEPENDUDUKAN BERDASARKAN ASPEK-ASPEK KUALITAS SCHEMA DATABASE. *TECHSI-Jurnal Teknik Informatika*, 8(2).
4. Hasibuan, Z. A. (2007). *Metodologi Penelitian Pada Bidang Ilmu Komputer dan Teknologi Informasi*. Fakultas Ilmu Komputer, Universitas Indonesia.
5. Komputer, W. (2010). *Panduan Belajar MySQL Database Server*. MediaKita.
6. Muhammad, F. (2019). *BASIS DATA*.
7. Permenpan.(2013). *Peraturan Menteri Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi Tentang Jabatan Fungsional Dosen Dan Angka Kreditnya*. Peraturan Menteri Pendayagunaan Aparatur Negara Nomor 17 Tahun 2013.
8. Razali, S. R. (2019). PENDETEKSIAN MADZHAB MENGGUNAKAN TEKNIK TEXT MINING DENGAN ALGORITMA C4. 5 BERBASIS WEB. *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, 6(2), 162-171.