DETERMINING LECTURAL EVALUATION IN FACULTY OF ENGINEERING UNIVERSITAS MALIKUSSALEH USING K-NN

Asrianda[1], Risawandi[2], Gunarwan[3]
Informatics, Malikussaleh University
gunkhair5@gmail.com[3]

Abstract

Abstract — K-Nearest Neighbor is a method that can classify data based on the closest distance. In addition, K-NN is one of the supervised learning algorithms with learning processes based on the value of the target variable associated with the value of the predictor variable. In the K-NN algorithm, all data must have a label, so that when a new data is given, the data will be compared with the existing data, then the most similar data is taken by looking at the label of that data. Filling and processing many questionnaires to determining the results of lectural evaluation from the performance of lecturers certainly requires a lot of time and process. Therefore, it is necessary to apply the K-NN Manhattan Distance method. In this study, the testing data is taken from one of the training data and has a classification result that is "Very Good". After going through the K-NN Manhattan Distance method with k being the closest / smallest neighbor, then the following results are obtained: Distance 5.4, the classification result is "Very Good" and 74.03% of similarity value. Based on the results obtained, the result of the classification from K-NN Manhattan Distance method show similarities with the results of the pre-existing classification.

Keywords : Classification, K-Nearest Neighbor, Manhattan Distance, Similarity

1. Introduction

Every human being needs education, up to when and wherever he is. In addition, to achieve the quality of human resources that are reliable, superior and highly competitive, the quality of education is also needed.

The issue of quality in higher education is also inseparable from the role of the lecturer. So, it is appropriate for each college to give awards to lecturers whose performance is very good. Therefore, for lecturers who have been assigned to give lectures, an evaluation of their learning performance is needed (Miarso, 2004).

The performance evaluation of each lecturer in the Faculty of Engineering, University of Malikussaleh (UNIMAL) can be taken from the results of the lecture evaluation evaluation questionnaire at the Faculty of Engineering.

Based on the data contained in the lecture evaluation questionnaire, it appears that there are several questions and answer choices in it. The questions and answer choices will then be filled out by the students. The completed questionnaire will be an evaluation material to determine whether the lecturer concerned has the right to get a good performance evaluation result or need to be repaired again.

From the way the questionnaire works, it can be seen that before getting the performance evaluation results of each lecturer, each lecturer must distribute the questionnaire to his students to be filled in first. Filling and processing so many questionnaires to determine the results of evaluations of each lecturer certainly requires a short time and process.

Thus, in this study, the authors felt interested in applying the Manhattan Distance K-Nearest Neighbor (K-NN) method. K-NN is one of the techniques in data mining that can be used to determine evaluation.

Lectures by classifying the performance results of a lecturer based on existing questionnaire data. The reason the authors chose the K-NN method is because the method is suitable to be applied in this case, besides because the data consists of a series of

questions and answer choices, the K-NN method also makes it easier for researchers to classify the performance evaluation results of each lecturer based on the similarity of the new questionnaire data with pre-existing questionnaire data. In addition, it also makes it easier for the Quality Assurance Group at the Faculty of Engineering, University of Malikussaleh to collect, evaluate and develop the learning process in undergraduate programs, especially in determining the results of lecture evaluations. The previous research using similar methods is (Rahmawati & Adnan, 2016).

Discussing the Determination of BPJS Insurance Products Based on Customer Profiles with the K-Nearest Neighbor Manhattan Distance Approach. The results of the study indicate that the prediction of the product to be promoted can be done by the K-NN algorithm and the calculation of the distance of Manhattan Distance to existing customer data, the accuracy of the MATLAB K-NN program is 54% of the prediction of consumer interest in the product. Other studies with similar cases, but using different methods have been done by (Ari Jayanti, 2016).

Discussing the Implementation of the SAW and AHP Methods in the Lecturer Performance Assessment Information System. This study aims to produce lecturer performance reports through a multicriteria decision support system such as Analytical Hierarchy Process (AHP) and Simple Additive Weighting (SAW). The results of the research were lecturer performance appraisal systems, which were developed based on a structured software development approach. As for the shortcomings of this study, the author only provides ranking information based on the best performance without calculating accuracy or similarity in order to achieve the applicable standard values.

2. Literature Review
2.1 Definition of Questionnaire

Questionnaire or questionnaire as a set of questions used to obtain information from respondents related to their personal or other matters related to research material (E. Nugroho, 2018).

2.2 Definition of Data Mining

Data Mining is a term used to describe knowledge discovery in a database. Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases (Nofriansyah, 2014).

3. Research Methods
3.1 Research Framework

The first step is to select the data set that will be used in the research process, in this study using data set questionnaires from the Faculty of Engineering, Malikussaleh University. The data set is then divided into two training data and testing data.

After dividing the data into 2 parts, namely training and testing, then the Class selection step (classification) is carried out using the K-NN Algorithm with the Manhattan Distance approach to get the best Class for testing data.

After the K-NN classification results are carried out, the next step is to calculate the similarity value between the testing data and the training data that have been selected through the K-NN classification results.

The results that have been obtained from the performance of K-NN are the results of classification of testing data and similarity values.

The following are training data obtained from the results of the questionnaire collected by the Quality Assurance Group at the Faculty of Engineering, University of Malikussaleh:

Table 1. Questionnaire Training Data

No.	X (Questionnaire Descriptions)																Classifi-cation
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	4.2	4.4	4.4	4.2	4.3	3.4	4.5	4.5	4.3	3.3	4.2	4.2	4.2	4.4	4.3	4.4	Baik
2	4.2	4.4	4.4	4.2	4.3	3.4	4.5	4.5	4.3	3.3	4.2	4.2	4.2	4.4	4.3	4.4	Baik
3	3.6	4.1	4.1	3	3.5	4.8	4.1	4.2	3.4	3.2	4.6	4.2	4.3	3.8	3.8	4.6	Cukup
4	4.1	4.2	4.5	4.4	4.6	4.3	4	4.3	4.5	4.4	4.7	4.6	4.3	4.4	4.5	4.5	Sangat Baik
5	4.6	4.5	4.6	4.3	4.3	4.2	4.5	4.5	4.4	4.5	4.4	4.5	4.6	4.6	4.4	4.5	Sangat Baik
6	5	5	5	5	4.9	5	5	5		4.9	4.8	4.8	4.9	5	5		Sangat Baik
7	3.9	3.3	4.8	3	4.9	4.6	4.3	3.6	4.6	4.6	4.9	4.7	3.7	4.8	4.1	3.4	Baik
8	4.4	4	4.8	4.9	4.9	4.7	4.6	4.9	4.6	4.7	4.7	4.6	4.6	4.9	4.4	4.9	Sangat Baik
....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
218	4.8	2.7	4.7	4.8	3	4.7	3.3	2.7	3	3.3	4.6	2.8	2.2	4.9	2.7	3.2	Cukup

Source: Quality Assurance Group of the Faculty of Engineering, University of Malikussaleh

Then, one of the questionnaire data is taken as testing data to be tested later using the Manhattan Distance K-NN method. The following is a table of testing data:

Table 2. Questionnaire Testing Data

No.	X (Questionnaire Descriptions)																Classifi-cation
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	4	4	4	5	5	5	4	5	5	5	4	5	4	5	4	5	Cukup

Source: Quality Assurance Group of the Faculty of Engineering, University of Malikussaleh

3.2 K-Nearest Neighbor Manhattan Distance Method

K-Nearest Neighbor (KNN) is a method that calculates the level of similarity (distance) of a case to another case based on

several attributes defined based on a certain weighting and then the level of similarity (distance) of the overall attribute will be summed.

Steps of the KNN (Rahmawati & Adnan, 2016):

Determine the parameter K (number of closest neighbors).

Calculate the square of Euclid distance / query instance of each object against the sample data provided.

Then sort the objects into groups that have the smallest Euclid distance.

Collect Y categories (nearest neighbor classification).

By using the most majority nearest neighbor category, we can predict the calculated query instance value.

Then there is Manhattan or City Distance which is used to calculate the distance of closeness or similarity between the current case and the other case on a case basis. The following is the formula for the distance between Manhattan (Latifah, 2015):

$$d_{ij} = \sum W \mid X_i - C_j \mid \qquad (1)$$

Where:

dij   = The distance between case i and j with all its parameters
W   = Amount of weight
X   = New cases that will be reduced by C later
C   = History (case in Case Base)

In determining the proximity or similarity value of a Nearest Neighbor (K-NN), the following equations are used (Mariana, Redjeki, & Razaq, 2015):

$$Similarity = \frac{S_1 \times W_1 + S_2 \times W_2 + \ldots + S_n \times W_n}{W_1 + W_2 + \ldots W_n} \qquad (2)$$

Where:

S   = Similarity (1 similarity) and 0 (different)
W   = Weight (weight given)

The value of proximity or similarity of Nearest Neighbor (K-NN) can also be known through pseucode, the way it works is by looking for the difference in the error rate value first. The error rate that has been obtained through pseucode is the value of dissimilarity. After the error rate is found, the similarity value can also be found by reducing 100% of the similarity values with the dissimilarity values that have been searched through pseucode. The following table looks for the error rate through pseucode.

Table 3. Pseudocode Of K-NN

|	
Step 1:	Load the training data set train_data
Step 2:	Get y = feature vector to be classified
Step 3:	Select k value
Step 4:	Calculate Manhattan Distance between the y and train_data
Step 5:	Search train_data for the closest feature vector to y
Step 6:	Classify each point in test_data using the k nearest neighbors in test_data
Step 7:	Assign a new label to y
Step 8:	error_rate = 100*(number of images misclassified)/(number of images in test_data)
Step 9:	end

Source: Book of Advancing Technologies and intellegence in healthcare and clinical Environment: Breakthroughs, 2012

4. Results and Discussion

The system to be built will calculate the distance between testing data (test) and training data (training) obtained through questionnaire data from the Quality Assurance Group at the Faculty of Engineering, University of Malikussaleh. The output of this system is the result of lecture evaluations that are calculated and processed using the Manhattan Distance K-NN method. The

system to be built can show the results of classifications of lecture evaluation in the form of an assessment of the performance of a lecturer which will facilitate the Quality Assurance Group at the Faculty of Engineering, University of Malikussaleh in collecting, evaluating and developing the learning process in undergraduate programs, especially in determining the results of lecture evaluations.

4.1 Usecase Diagram
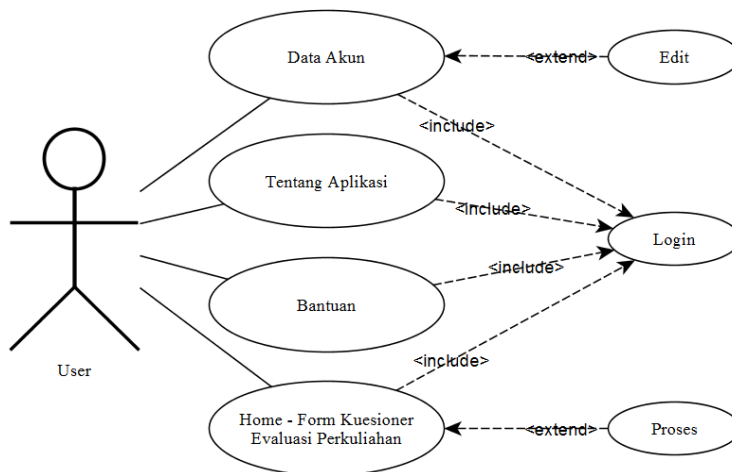
Picture 1. Usecase Diagram

Description:
Users log in to the system. After that, the user can do the lecture evaluation process using the Manhattan Distance K-NN method on the main page (the main page can be accessed by selecting the menu "Home - Lecture Evaluation Questionnaire Form" first). The next step, users can access the system work process by filling in

questionnaire data based on the options that have been set, then the user can see the results of the lecture evaluation based on the data that has been filled. After that, the user can log out of the system.

4.2 Calculation Of K-Nearest Neighbor Manhattan Distance

Examples of training and testing data can be seen in Table 1 and Table 2. The training data is used to test the results of testing data. Then the testing data is calculated using Manhattan Distance for each Training data. After obtaining the closest distance, then select as many as k data to determine the classification of the test data. Examples of calculations between testing data and training data are as follows:

$d1 = |4 – 4.2| + |4 – 4.4| + |4 – 4.4| + |5 – 4.2| + |5 – 4.3| + |5 – 3.4| + |4 – 4.5| + |5 – 4.5| + |5 – 4.3| + |5 – 3.3| + |4 – 4.2| + |5 – 4.2| + |4 – 4.2| + |5 – 4.4| + |4 – 4.3| + |5 – 4.4| = 10.2$

Table 4. Result Of Distance Between Test Data And Training Data

Number	Distance	Classification
1	10.2	Baik
2	10.2	Baik
3	12.1	Cukup
4	7.3	Sangat Baik
5	8.8	Sangat Baik
6	7.1	Sangat Baik
7	10	Baik
8	5.4	Sangat Baik
....	...	...
218.	19.8	Cukup

Table 5. Sorting Result Of Distance Between
Test Data And Training

Number	Distance	Classification
8	5.4	Sangat Baik
19	6.56	Sangat Baik
20	6.63	Sangat Baik
164	6.84	Sangat Baik
24	6.84	Sangat Baik
6	7.1	Sangat Baik
27	7.13	Baik
199	7.2	Sangat Baik

By using k = 3, the 3 closest or smallest distances are 8, 19 and 20. Likewise for the calculation of the values of k to 4 to 218.

Table 6. K Result From Distance Between Test Data And Training

Number	Distance	Classification	K
8	5.4	Sangat Baik	1
19	6.56	Sangat Baik	2
20	6.63	Sangat Baik	3

4.3 Calculation of Similarity from Nearest Distance

To calculate the similarity value from the closest distance that has been found through the Manhattan Distance K-NN method, the similarity formula is used. Then, for the use of the similarity formula to be appropriate and easy, the author will use integers in this study. The following similarity calculation in the closest case is case number 8:

Similarity (Case-8)

$$= \frac{4+4+0+5+5+5+0+5+5+5+0+5+0+5+4+5}{4+4+5+5+5+5+5+5+5+5+5+5+5+5+4+5}$$

$$= \frac{57}{77}$$

$$= 0{,}7402 \times 100\%$$

$$= 74{,}02\%$$

So, the similarity value or the similarity between the 8th case training data and testing data is 74.02%.

5. Conclusions

From the results of the research that has been done it can be seen that the application of the K-NN method with the Manhattan Distance approach method, it can be seen that the results of the classification of the Manhattan Distance K-NN to testing data show similarity with the results of training data training beforehand with a similarity percentage of 74.02%.

Bibliography

Latifah, K. (2015). Kombinasi Algorithma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang. Jurnal Informatika Upgris, 1(1 Juni).

Mariana, N. (2016). Penerapan Algoritma k-NN (Nearest Neighbor) Untuk Deteksi Penyakit (Kanker Serviks).

Marpaung, N., Handayani, M., & Yesputra, R. (2018, September). Sistem Pendukung Keputusan Pemilihan Dosen Terbaik Dengan Metode Weighted Product (WP) Pada STMIK Royal. In Seminar Nasional Royal (SENAR) (Vol. 1, No. 1, pp. 267-270).

Miarso, Y. (2004).Menyemai Benih Teknologi Pendidikan. Edisi ke 2. Jakarta: Kencana Prenada Media Group.

Nofriansyah, Dicky. (2014). Konsep Data Mining Vs Sistem Pendukung Keputusan. Yogyakarta: Deepublish.

Nugroho, Eko. (2018). Prinsip Prinsip Menyusun Kuesioner. Malang: UB Press.

Rahmawati, T. D., & Adnan, F. N. (2016). Penentuan Produk Asuransi BPJS Berdasarkan Profil Pelanggan Dengan Pendekatan K-Nearest Neighbor Manhattan Distance. Journal of Information System, 1(2), 167-180.

Tan, Joseph. (2012). Advancing Technologies and intellegence in healthcare and clinical Environment. United States of America: Medical Information Science.