Grouping Of Outstanding Students At Abdi Negara Vacational School Using The K-Nearest Neighbor Method

Erika Br Milala[1], Relita Buaton[2], Zira Fatmaira[3]

[1,2,3]Information System STMIK Kaputama, Binjai Indonesian
Jl. Veterans No. 4A-9A, Binjai, North Sumatra, Indonesia
erika366milala@gmail.com[1], bbcbuaton@gmail.com [2], zirafatmaira0301@gmail.com[3]

Abstract

Abdi Negara Vocational School is a Vocational High School in the city of Binjai. To improve the competence of their students, and so that students are more focused on developing their own potential and interests. As for the basis for classifying students, namely values, academic potential, academic ability, academic achievement, ethics and attitudes, as well as the assessment of teachers and homeroom teachers. Currently the system running at Abdi Negara Vocational School for grouping students is through school management and teacher meetings. Where in the meeting it will be discussed which students are nominated as outstanding students. So the homeroom teacher will provide names for consideration. The grouping process at Abdi Negara Vocational School has several problems, including first, a lot of data will take up time, energy and requires extra precision. Second, the accuracy of the recapitulated processed data is often wrong because the components used as parameters are quite a lot, as well as the data to be grouped. From these problems, we need a technique that can help in grouping high achieving students in State Vocational Schools precisely and accurately so as to reduce the risk of errors. Algorithms such as K-Nearest Neighbor and a website-based system can make it easier to input grades and group students. From the results of trials on existing cases, it was obtained that a student named Rahma was classified as an Outstanding Student. These results are obtained from the results of calculations with the distance between the selected K values and then with the most classification, namely the outstanding students with grades 0.929 and 0.976.

Keywords: *K-Nearest Neighbor Algorithm, grouping of outstanding students, Euclidian Distance*

1. Introduction

The rapid development of information technology makes information an important element. There is also a lot of data produced by information technology, including in the field of education [1]. Outstanding students are students who have succeeded in achieving an achievement in both academic and non-academic fields pursued at school so they should be proud of [2]. Abdi Negara Vocational School is a Vocational High School in the city of Binjai. To improve the competence of its students, and so that students are more focused on developing their own potential and interests, State Servant Vocational School conducts groupings of students. Grouping is carried out when children move up from class X to class XI. The basis for grouping students is grades, academic potential, academic ability,

Currently the system running at Abdi Negara Vocational School for grouping students is through school management and teacher meetings. Where in the meeting it will be discussed which students are nominated as outstanding students. So the homeroom teacher will provide names for consideration. The grouping process at SMK Abdi Negara has several problems including the first, a lot of data will take time, effort and require very extra accuracy. Second, the accuracy of the process data that has been recapitulated is often wrong because there are quite a lot of components used as parameters, as does the data that will be grouped.

From this problem, a technique is needed that can help in grouping outstanding students at State Vocational Schools precisely and accurately so as to reduce the risk of errors. Simply put, data mining is mining or discovering new information by looking for certain patterns or rules from a very large amount of data [3]. Data mining consists of various techniques that can be used to make predictions and classifications, where this technique estimates the possibility that will occur in the future by looking at some of the existing information and data patterns [4].

2. Literature

Previous research can be used as reference material to increase references for research that will be carried out. Some of the references found by researchers as reference material for the author are as follows: This research is strengthened by research, with the title Classification of Determining Credit Card Submissions Using K-Nearest Neighbor. The research results show that the application

effectively helps credit card analysts classify customers who get credit cards according to the criteria [5]. This research is also reinforced by research, with the research title Classification of Local Orange Quality Based on Texture and Shape Using the K-Nearest Neighbor (K-NN) Method. The results of the research, the k-Nearest Neighbor Method succeeded in identifying oranges based on shape and texture with an accuracy of 93.33% [6]. The following is also reinforced by research (Rozzi Kesuma Dinata, 2020: 50), entitled Classification of Junior High Schools/Equivalent in the Bireuen Region Using the Web-Based K-Nearest Neighbors Algorithm. The results of the research, the k-Nearest Neighbor Method, succeeded in identifying oranges based on shape and texture. 93.33% accuracy [7]. Reinforced by research, with the title Application of the KNN Method to Determine Betta Fish by Extraction of Shape and Canny Features. The results of the K-Nearest Neighbor research can classify objects with an average accuracy of 70% [8].

2.1 Data Mining

DataMining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and assembled knowledge from various large databases [9].

Data Miningis a process of finding meaningful relationships, patterns and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques.
The extraordinary progress that continues in the field of data mining is driven by several factors, including
1. Rapid growth in data set.
2. Data storage in the data warehouse, so that all companies have access to a reliable database.
3. There is an increase in data access through web and internet navigation.
4. Business competency pressure to increase market share in economic globalization.
5. Development of software technology for data mining (technology availability).
6. Substantial developments in computing capabilities and expansion of storage capacity.
From the definitions that have been presented, the important things related to data mining are
1. *Data mining*is an automatic process of existing data.
2. The data to be processed is very large data.
3. The goal of data mining is to find relationships or patterns that might provide useful indications.
   Based on the opinions of the experts above, it can be concluded that data mining is a process of collecting important information and data in large quantities or big data. This process often utilizes several methods, such as mathematics, statistics and the use of artificial intelligence (AI) technology .

2.2 K-Nearest Neighbor Method

The K-Nearest Neighbor method is a non-parametric method that can be used for classification based on k-nearest neighbors and regression. The K-Nearest Neighbor algorithm is a method for classifying objects closest to an object based on learning data. Determining the value of k in the K-Nearest Neighbor classification algorithm can be searched based on the nearest k sample values k1, k2…,ks. The moreLotsthe smaller the number of k selected, whereas if the dimensions of the existing data are larger, the number of k selected must be higher. In determining the value of k, it is better to use odd numbers such as k = 1, 3, 5, …, etc. The value of k must meet the requirements, namely k < N where N is the number of training datasets, because the value of k is used to find the majority number of classes/labels in the training data, the value of k cannot be more than the number of training datasets. There are 5 ways to find the nearest neighbor or distance in the K-NN algorithm, namely Euclidean Distance, Manhattan Distance, Cosine Distance, Correlation Distance, Hamming Distance. The distance between the two nearest k neighbors based on similarity values can be calculated using the Euclidean distance which is defined as follows:

$$Dist\ (X,Y) = \sqrt{\sum_{i=1}^{D}(X_i - Y_i)^2}$$

....(1)

with:
Dist(X,Y): distance between objects (Euclidean Distancing)
Xi: sample data
Yi: test data
D: data dimension
i: data variables

2.3 Achievement Student

An outstanding student is a student who has succeeded in achieving an achievement in both academic and non-academic fields pursued at school and is worthy of being proud of. Participating students are students who always follow the rules set by the school or teacher who educates them, and always have the obligation as a student to uphold their honor and dignity as an exemplary student.

2.3.1 Academic Potential

Academic Potential Test (TPA) is a test that aims to determine a person's talents and abilities in the scientific (academic) field. This test is also often associated with intelligence. This Academic Potential Test is also synonymous with the GRE (Graduate Record Examination) test which has become an international standard.

2.4.2 Academic Ability

The term academic ability consists of two words, namely ability and academic. In the Big Indonesian Dictionary, ability has the meaning of ability, skill or strength, while academic has the meaning related to academics (education). According to Krishnawati and Suryani academic ability is part of intellectual ability which is generally reflected in academic achievement (learning outcome scores). The concept of academic ability is an individual's beliefs and self-evaluation regarding academic traits related to the individual's skills and abilities.

2.4.3 Academic Achievement

Defines academic achievement as the results obtained in the form of impressions that result in changes in the individual as the final result of learning activities.

2.4.4 Ethics and attitudes

Ethics is a critical and rational reflection regarding moral values and norms that determine and are manifested in the attitudes and behavioral patterns of human life, both individually and as a group.

2.4.5 Assessment of teachers and homeroom teachers

This is how the teacher assesses the overall behavior of students in a learning atmosphere. How is the behavior, manners, norms and the ability of a student to receive lessons that have been given by an educator, namely the teacher.

3. Results And Discussion

3.1. Application of the Method

Problem can be defined as a question that can be solved. Therefore, at the system analysis stage, the first step in the analysis is to identify problems by examining existing problem objects.

The steps of the KNN algorithm are:
1) Determine the K parameter (number of closest neighbors. The K parameter in testing is determined based on the optimum K value during training. The optimum K value is obtained.
2) Normalize the value of X in the available data.
    With the following formula:
$$Xnew = \frac{Xold - Xmin}{Xmax - Xmin}$$ ....(2)
3) Calculate the squared Euclidean distance of each object to the given sample data.

$$Dist\,(X,Y) = \sqrt{\sum_{i=1}^{D}(X_i - Y_i)^2}$$ ....(3)

with: *Dist*(X,Y): distance between objects (Euclidean Distancing)
   Xi: sample data
   Yi: test data
   D: data dimension
   i: data variables

4) Sort these objects into groups that have the smallest Euclidean distance.
5) Collecting category Y (nearest neighbor classification).
6) By using the majority category, the value of the query instance that has been calculated can be predicted.

3.2.1 Case Study

In discussing the calculation of the K-Nearest Neighbor Method, we can take 10 samples from those with 5 variables that will become the group of outstanding students. Calculation of the K-Nearest Neighbor Method in the system, if searched manually, we can see the solution as follows:

Table 1: Student/Student Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5
1	20180101	Lambok MT	93	81	85	85	92
2	20180102	Legawathi	67	82	80	70	73
3	20180103	M. chairul	76	89	70	90	85
4	20180104	Martaulina S	62	77	80	85	78
5	20180105	Mhd. Rifqi	84	98	95	90	100
6	20180106	Norita sinurat	88	80	72	74	85

Information :
X1 : Academic Potential
X2: Academic Ability
X3 : Academic Achievement
X4 : Ethics and Attitude
X5 : Teacher and homeroom teacher assessment

Based on alternative data from 5 various types of Student/Learner data from the table above, the suitability of each alternative is obtained, as in the following table

Table 2: Training Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
1	20180101	Lambok MT	93	81	85	85	92	Achievement
2	20180102	Legawathi	67	82	80	70	73	No Achievement
3	20180103	M. chairul	76	89	70	90	85	No Achievement
4	20180104	Martaulina S	62	77	80	85	78	No Achievement
5	20180105	Mhd. Rifqi	84	98	95	90	100	Achievement
6	20180106	Norita sinurat	88	80	72	74	85	Achievement

Given new data that we will classify, namely X1=80, X2=95, X3=75, X4=78 and X5.98

Table 3: Test Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
1	20180110	Rahma	80	95	75	78	98	?

Solution:
The following are the steps for calculating the K-Nearest Neighbor method in data classification.
1. Determine the value of K=3
2. Normalize the X values in student data.

Table 4: Table of Normalization of X Values in Training Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
1	20180101	Lambok MT	1.0	0.2	0.6	0.8	0.7	Achievement
2	20180102	Legawathi	0.2	0.2	0.4	0.0	0.0	No Achievement
3	20180103	M. chairul	0.5	0.6	0.0	1.0	0.4	No Achievement
4	20180104	Martaulina S	0.0	0.0	0.4	0.8	0.2	No Achievement
5	20180105	Mhd. Rifqi	0.7	1.0	1.0	1.0	1.0	Achievement
6	20180106	Norita sinurat	0.8	0.1	0.1	0.2	0.4	Achievement

Table 5: Table of Normalization of X Values in Test Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
1	20180110	Rahma	0.6	0.9	0.2	0.4	0.9	?

3.Calculate the Euclidean distance of the test data.
This process is to calculate the overall value in the Student data to produce a value*Euclidean distance.*

Table 6: Euclidean Distance

No	NAME	X1	X2	X3	X4	X5	Euclidean Distance (0.6 , 0.9 , 0.2 , 0.4 , 0.9)
1	Lambok MT	1.0	0.2	0.6	0.8	0.7	$\sqrt{\begin{array}{l}(1,0-0,6)^2+(0,2-0,9)^2+(0,6-0,2)^2\\+(0,8-0,4)^2+(0,7-0,9)^2\end{array}}$ =0.976
2	Legawathi	0.2	0.2	0.4	0.0	0.0	$\sqrt{\begin{array}{l}(0,2-0,6)^2+(0,2-0,9)^2+(0,4-0,2)^2\\+(0,0-0,4)^2+(0,0-0,9)^2\end{array}}$ =1,271
3	M. chairul	0.5	0.6	0.0	1.0	0.4	$\sqrt{\begin{array}{l}(0,5-0,6)^2+(0,6-0,9)^2+(0,0-0,2)^2\\+(0,1-0,4)^2+(0,4-0,9)^2\end{array}}$ =0.854
4	Martaulina S	0.0	0.0	0.4	0.8	0.2	$\sqrt{\begin{array}{l}(0,0-0,6)^2+(0,0-0,9)^2+(0,4-0,2)^2\\+(0,8-0,4)^2+(0,2-0,9)^2\end{array}}$ =1.335
5	Mhd. Rifqi	0.7	1.0	1.0	1.0	1.0	$\sqrt{\begin{array}{l}(0,7-0,6)^2+(1,0-0,9)^2+(1,0-0,2)^2\\+(1,0-0,4)^2+(1,0-0,9)^2\end{array}}$ =1.021
6	Norita sinurat	0.8	0.1	0.1	0.2	0.4	$\sqrt{\begin{array}{l}(0,8-0,6)^2+(0,1-0,9)^2+(0,1-0,2)^2\\+(0,2-0,4)^2+(0,4-0,9)^2\end{array}}$ =0.929

a) Sorts grades by K Grade

Table 7: Sequence of Distance with Ascending

No	NAME	Euclidean Distance (0.6 , 0.9 , 0.2 , 0.4 , 0.9)	Distance Order
1	Lambok MT	0.976	3

2	Legawathi	1,271	-
3	M. chairul	0.854	1
4	Martaulina S	1.335	-
5	Mhd. Rifqi	1.021	-
6	Norita sinurat	0.929	2

b) Determine the classification of nearest neighbors. Rows 2 and 5 in table 3.12 of the Yes classification are taken if the K value <=3. the rest No.

Table 8: Determination of categories that include K<=3

No	NAME	Euclidean Distance (0.6 , 0.9 , 0.2 , 0.4 , 0.9)	Order Distance	Category K For K-NN
1	Lambok MT	0.976	3	Achievement
2	Legawathi	1,271	-	-
3	M. chairul	0.854	1	No Achievement
4	Martaulina S	1.335	-	-
5	Mhd. Rifqi	1.021	-	-
6	Norita sinurat	0.929	2	Achievement

The yes classification for K-NN is in column 5, covering rows 3, 1 and 6, so that the test data is included in the achievement classification.

c) Results of Rahma values

Judging from the data obtained, the classification that has student scores has two data, then the outstanding student with the name Rahma produces an outstanding student classification score which means that the student is included in the group of outstanding students.

Table 9: Test Data Classification Results

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
7	20180110	Rahma	0.6	0.9	0.2	0.4	0.9	Achievement

Table 10 Results of Classification of Training Data and Test Data

No	NIS	STUDENT'S NAME	X1	X2	X3	X4	X5	Classification
1	20180101	Lambok MT	93	81	85	85	92	Achievement
2	20180102	Legawathi	67	82	80	70	73	No Achievement
3	20180103	M. chairul	76	89	70	90	85	No Achievement
4	20180104	Martaulina S	62	77	80	85	78	No Achievement
5	20180105	Mhd. Rifqi	84	98	95	90	100	Achievement
6	20180106	Norita sinurat	88	80	72	74	85	Achievement
7	20180110	Rahma	80	95	75	78	98	Achievement

4. Conclusion

From the calculation results above using the K-Nearest Neighbor method, the results are obtained with a value0.929 and 0.976 based on determining the value of K = 3 with the classification of outstanding students. So that a student named Rahma is included in the classification of outstanding students.

Reference

[1]   Nugroho, A., Fatmawati, A., & Permata Sari, F. (2015). Clustering Analysis of Provinces in Indonesia Based on Poverty Level Using the K-Means Algorithm. Journal of Information Management and Information Systems, 4(1), 12–20. https://doi.org/10.36595/misi.v4i1.216

[2]   Anni Banjarsari. (2019). Application of K-Optimal in the KNN Algorithm to Predict Timely Graduation for Students of the FMIPA Unlam Computer Science Study Program Based on IP Up to Semester 4. 2(2), 26–40.

[3]   Bhatia, MF, Saputra, R., & Wibowo, A. (2013). Predicting house sale prices using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor. Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017, 17(1), 56–63. https://doi.org/10.1109/ICICOS.2017.8276357

[4]   Santoso. (2014). Comparison of the K-Nearest Neighbor (K-Nn) and Learning Vector Quantization (Lvq) Methods for Classification Problems, 12-24.

[5]   Mustakim, & Oktaviani F, G. (2015). K-nearest neighbor Classification Algorithm as a Student Achievement Prediction System. Journal of Science, [] Technology And Industry, 13(2), 26–33.

[6]   Yogiek. (2020). Classification of Credit Card Application Determination Using K-Nearest Neighbor, 46-54.

[7]   [6] Ratri Enggar Pawening Quality Classification of Local Oranges Based on Texture and Shape Using the K-Nearest Neighbor (K-NN) Method, 10(14), 20–29. Global Journal of Computer Science and Technology. 2020.

[8]   Rozzi Kesuma Dinata. (2020). Classification of Junior/Equivalent Middle Schools in the Bireuen Region Using the Web-Based K-Nearest Neighbors Algorithm. Sainstek : Journal of Science and Technology, 12(2), 50–59. https://doi.org/10.31958/js.v12i2.2483

[9]   Fajar Shidiq. (2021). Application of the KNN Method to Determine Betta Fish by Extraction of Shape and Canny Features. Journal of Management Informatics, 9(1), 89–97. https://doi.org/10.29244/jurnal_mpd.v9i1.27542

[10] Tu'u Lasena, Y. (2014). Application of Data Mining to Determine Daily Rain Potential Using the KNearest Neighbor (KNN) Algorithm. Journal of Information Management and Information Systems, 3(1), 18–21. https://doi.org/10.36595/MISI.V3I1.125 [3] Dwi, G., Yahya, W., & Nurwasito, H. (2022). Implementation of the Dictionary Data Structure for the Internet of Things Device Monitoring System. Journal of Development of Information Technology and Computer Science (J-PTIIK) Universitas Brawijaya, 2(10), 23–33.