

IMPLEMENTATION OF K-NEAREST NEIGHBOR ALGORITHM FOR CLASSIFICATION OF CLASS PLACEMENT AT JUNIOR HIGH SCHOOL, PADANG MONTH

¹Dewi Lasmiana Panjaitan, ¹Paska Marto Hasugian

¹Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia

²Rekayasa Perangkat Lunak, STMIK Pelita Nusantara, Medan Indonesia

dewilasmiana@gmail.com

Abstract

Article Info

Received : 02 September 2021

Revised : 28 October 2021

Accepted : 02 November 2021

Superior class is a number of students who have outstanding abilities or achievements in these students, who are grouped in one particular class. One way to do this is through the class placement process. However, during class placement there are problems that arise, namely during the class determination process, whether students enter the superior class or the ordinary class. Students who have certain abilities will later occupy the superior class and students who do not have certain abilities will not enter the superior class. With this research, it will help the school in determining the superior class and the ordinary class, so that no one is harmed, there should be students who deserve to be the superior class. The purpose of this study is to implement the principles of data mining on Class Placement Classification using the K-Nearest Neighbor Algorithm. Where the K-Nearest Neighbor Algorithm will classify objects based on the learning data that is closest to the object. Based on the results of trials carried out using the K-NN algorithm with 64 data tested and training data as many as 82 data, then the results obtained from class placement with students occupying class A as many as 26 students, students occupying class B as many as 20 students and students who occupy class C as many as 18 students. Where the K-Nearest Neighbor Algorithm will classify objects based on the learning data that is closest to the object. Based on the results of trials carried out using the K-NN algorithm with 64 data tested and training data as many as 82 data, then the results obtained from class placement with students occupying class A as many as 26 students, students occupying class B as many as 20 students and students who occupy class C as many as 18 students. Where the K-Nearest Neighbor Algorithm will classify objects based on the learning data that is closest to the object. Based on the results of trials carried out using the K-NN algorithm with 64 data tested and training data as many as 82 data, then the results obtained from class placement with students occupying class A as many as 26 students, students occupying class B as many as 20 students and students who occupy class C as many as 18 students.

Keywords: Data Mining, Class Placement, K-Nearest Neighbor

1. Introduction

GKPI Padang Bulan Junior High School (SMP) is one of the private educational institutions where teaching and learning activities take place. Teaching and learning activities are the core of overall educational activities to awaken the potential that exists in students. This GKPI Junior High School accepts new students every year, to shape students into intelligent, skilled, achievement, broad-minded, and pious human beings to God Almighty. In accordance with the regulations of the GKPI Padang Bulan Junior High School that new students will be divided into several classes, namely class A, B, C and so on, based on the number of students registering, with an average class of 30 people per room. But at the time of the new teaching, Class placement will be carried out based on student report card scores or rankings, the problem is when placing the class whether the student enters the superior class or the regular class. With this research, it can help schools that implement a class division system based on grades, so that it is easier to place students according to the grades obtained. Where students who

enter class A will be used as an independent class or referred to as a superior class, the superior class in question is a class that has outstanding abilities or achievements in students. to make it easier to place students according to the value obtained. Where students who enter class A will be used as an independent class or referred to as a superior class, the superior class in question is a class that has outstanding abilities or achievements in students. to make it easier to place students according to the value obtained. Where students who enter class A will be used as an independent class or referred to as a superior class, the superior class in question is a class that has outstanding abilities or achievements in students.

Research conducted (Sebastian, 2019) with the topic "Implementation of the K-Nearest Neighbor Algorithm to Classify Products from several E-market places". Based on the results of the tests carried out, the K-Nearest Neighbor method can classify products from e-marketplaces, especially Tokopedia and Bukalapak. The accuracy resulting from test 1, the selection of values of $k=1$, 5, or 10 is 78%, 97.33% and 92%. Based on test 1, it is concluded that the optimal k value for this case is 5. In test 2, multi-brand, the resulting accuracy is 96.67%. This accuracy value can be said to be good because it exceeds 90%. The accuracy of the K-Nearest Neighbor algorithm is strongly influenced by the training data. The more complete the training data, the better the accuracy. Have not tested the product description content data,

Research ever done (Kartika & Santoso, 2017) with the topic "Determination of Achievement Students Using the K-Nearest Neighbor and Weighted Product Methods". The problem at SMP Negeri 3 Mejayan is that it has not been balanced in conducting assessments because academic scores are still the main consideration compared to non-academic scores. Non-academic scores are only used as supporting data that is not clear in the weighting of values. So it feels unfair in determining the students who excel. The first ranking test has an accuracy rate of 66.67%. The second ranking test has an accuracy rate of 88.89%. The third ranking test has an accuracy rate of 100%. With report cards, extracurricular, personality, and discipline are the main priorities in determining the outstanding students. The data used are 30 test data and 30 training data.

Research conducted (Fitrianti, 2018) conducted "Analysis of Sentiments on Restaurant Reviews with Indonesian Texts Using the K-Nearest Neighbor Algorithm". Tested using a confusion matrix and got the accuracy results from sentiment analysis on restaurant reviews with Indonesian text using K - Nearest Neighbor with a percentage of 96.61%, k value = 1 and an error rate value of 3.39%.

Research conducted (Badu, 2016) do "Application of the K-Nearest Neighbor Algorithm for Classification of Village Funds". Based on the results of the evaluation of the k -nearest neighbor model using the Confusion Matrix, the use of the k -nearest neighbor model of the dataset used in the research object obtained the highest accuracy of 78.95% or including Fair by using the $k=2$ parameter. Then the Precision value is 100% and Recall is 100%. Based on these results, it can be stated that the classification system built can be used to make decisions in classifying village funds.

Research conducted (Kurniawan & Saputra, 2019) conducting "The Application of K-Nearest Neighbors in Accepting Students With the Zoning System". Based on the tests carried out with the experimental calculation of determining the acceptance of new students using the Confusion matrix and Rapidminer model calculations, the proposed accuracy value lies in K5 with an accuracy rate of 83.36%.

Based on the research that has been done previously, it will be used as a reference or guide in this research. The difference between this research and previous research is in terms of the attributes used and in terms of the data used. The attributes used are name, class, knowledge, skill and rank. This study aims to facilitate the school in classifying class placements.

2. Method

The research carried out can be described in a flow of research work activities as shown below:

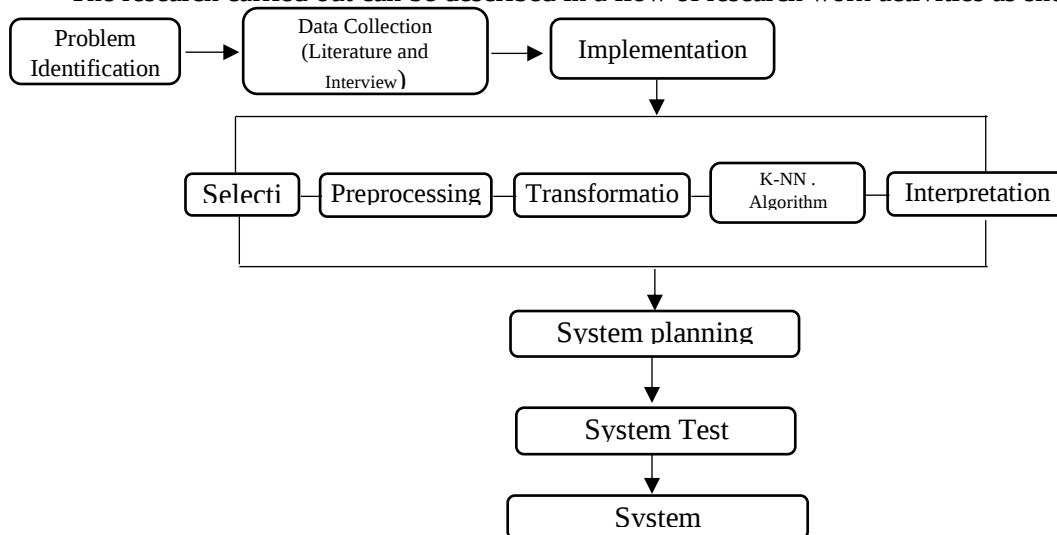


Figure 1. Thinking Framework

At this stage, we learn about the introduction of problems that exist in SMP GKPI Padang Bulan. Part of the problem is at the time of class placement from class VII to class VII whether later students will occupy class A, B or C, and so on. Therefore, SMP GKPI Padang Bulan requires a system capable of classifying class placements. One of the data collection carried out was interviews, which were conducted at SMP GKPI Padang Bulan. Researchers conducted interviews with direct communication at the Student Affairs Section of SMP GKPI Padang Bulan with the aim of digging up certain information. With the data obtained as many as 154 data and 84 data used as training data and 70 data used as testing data, but previously carried out several stages of data mining. The stages of data mining (Mardi, 2017) the first Selection is the selection (selection) of data from a set of operational data needs to be done before the stage of extracting information in Knowledge Discovery in Database (KDD) begins. Pre-processing, before the data mining process can be carried out, it is necessary to carry out a cleaning process on the data. The cleaning process includes, among others, removing duplicate data, checking for inconsistent data, and correcting errors in data, such as printing errors. Transformation Coding is a transformation process on the data that has been selected, so that the data is suitable for the data mining process. Data mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods. Then the method that will be used is the K-Nearest Neighbor algorithm to classify class placement.

3. Results and Discussion

3.1 K-Nearest Neighbor Algorithm Analysis

K-Nearest Neighbor (K-NN) is a method that uses a supervised algorithm where the results of the new test sample are classified based on the majority of the categories in K-NN. The purpose of this algorithm is to classify new objects based on attributes and training samples (Yahya & Puspita Hidayanti, 2020). Before performing the steps of the K-Nearest Neighbor algorithm, first determine the training data and testing data to be tested, below are the training and testing data that will be tested with the K-Nearest Neighbor Algorithm.

Training data is data that will be applied when using the algorithm. The training data used is class VIII data in 2019, with the attributes used are name, class, knowledge (P), Skills (K), and ranking attributes. For training data can be seen in Table 1 below.

Table 1. Training Data

No	Student's name	Class	Knowledge	Skills	Rank
1	Agrita Manurung	VIII-A	820	822	30

2	Clarisa Napitupulu	VIII-A	886	871	7
3	Daniel Siregar	VIII-A	827	824	29
4	Pomegranate Lafao	VIII-A	867	858	15
5	Ancient Elvia	VIII-A	875	862	12
...	...	...	...	...	...
82	Virel Esekiel	VIII-C	808	801	23

Test data is the data used or the data to be tested with training data. With the data used as many as 64 data with the attributes used, namely the attributes of name, class, knowledge (P), Skills (K), and ranking as in table 2 below.

Table 2. Testing Data

No	Name	Class	Knowledge	Skills	Rank
1	Adinda M manalu	VII-A	884	887	2
2	Adrian C Harahap	VII-A	836	827	22
3	Albert P Sihotang	VII-A	821	827	25
4	Amoren Naibaho	VII-A	871	861	8
5	Arianto Hulu	VII-A	801	818	30
...	...	...	...	...	...
64	Manulu Varsi Zone	VII-B	819	839	15

The first step to perform the stages of the K-Nearest Neighbor algorithm is to determine the value of K (nearest neighbor). K here is intended as the point of the object to be measured. For example, 5 nearest neighbors have been determined (K=5), meaning that the object points to be taken are 5 object points. After determining the value of K, the next step is to calculate the Euclidean distance of the object to the given training data using the formula:

Table 3. Euclidean Distance

No	Name	Euclidean Distance	Class
1	Adinda M Manalu	9,433	A
		9,433	A
		9,949	A
		14,899	A
		15,842	A
2	Adrian C Harahap	7	A
		10,862	A
		11,789	A
		13,674	A
		14,071	C
...	...	...	...
64	Manulu Varsi Zone	10,049	B
		12,688	C
		14,177	C
		14,730	A
		14,866	C

The table above is the result of the calculation of the Euclidean distance of the testing data object to the training data, where the k closest neighbors are equal to 5. Based on the results of the classification carried out by applying the steps of the K-Nearest Neighbor algorithm, the results of the class placement with those occupying Class A as many as 26 students, class B as many as 20, and class C as many as 18 students as follows.

Table 4. Class A . Students

No	Student's name	P	K	Rank	Euclidian Distance	Class
1	Adinda M manalu	884	887	2	9,433	A
2	Adrian C Harahap	836	827	22	7	A
3	Albert P Sihotang	821	827	25	7.141	A
4	Amoren Naibaho	871	861	8	5.385	A
5	Gracia JE Siburian	896	899	1	7.874008	A
6	Gresa Listani	857	865	10	5.656854	A
7	Hadasa F Silalahi	838	854	17	10.48809	A
8	Princess Hanna	855	878	7	15.93738	A
9	Jonathan Ferdinand	882	885	3	8.3666	A
10	Joshua A Migo	841	855	16	8.306624	A
11	Joe KY Sitinjak	865	876	5	8.602325	A
12	Kezia A Gurusinga	847	857	15	6.78233	A
13	Leonardo P Hutagalung	832	843	20	10.48809	A
14	Marchello Febrian	827	853	19	18.78829	A
15	Pearl of Asa	819	831	24	9.899495	A
16	Nadia Ulina	854	862	13	7	A
17	Romaito Hutajulu	880	873	3	7.483315	A
18	Silviana Putri	840	867	14	18.3303	A
19	Sinta Daniel	860	872	9	8.660254	A
20	Vebi Harmika Marbun	835	848	18	13.2665	A
21	Verawati Br. Nasution	855	866	12	6.403124	A
22	Yohannes Ndruru	828	823	23	6.164414	A
23	Judeo Benedict	833	838	21	7.28011	A
24	Yusus S Sihombing	864	875	6	8.774964	A
25	Alforis Talaumbanua	887	882	2	5	A
26	Kennedy Coal	899	886	1	6.324555	A

After the classification by applying the steps of the K-Nearest Neighbor algorithm, the results of the classification of class placement are obtained, as many as 20 people occupy class B.

Table 5. Table of Class B Siswa Students

No	Name	P	K	Rank	Euclidian	Class
1	Arianto Hulu	801	818	30	13.41641	B
2	Benny H Sitanggang	790	795	33	9.69536	B
3	Dicki Erta Pasaribu	802	815	29	11.44552	B
4	Immanuel Manurung	783	801	34	13.96424	B
5	Marchell Nainggolan	718	818	31	15.29706	B
6	Steven Hutabarat	791	804	32	12.08305	B
7	Astri Hutapea	844	843	9	9.848858	B
8	Fariacy	797	794	25	8.185353	B
9	Freno Divo Lubis	833	838	13	8.062258	B
10	Gidion Surbakti	766	781	29	9.539392	B
11	Glory A Sitohang	867	854	3	11.22497	B
12	Jessica Amelia	850	845	6	4.472136	B
13	Joyce Elisa Manuela	837	839	11	6.708204	B
14	Lita Tia	771	777	30	5.09902	B

15	Lovandro Tampubolon	777	782	28	3.741657	B
16	Margareta Tampubolon	822	844	8	9.899495	B
17	Nova Lumbangaol	781	779	27	6.164414	B
18	Randy Glen Ginting	781	790	26	7.348469	B
19	Rosintan Sidabutar	831	834	14	6.324555	B
20	Sarah Chaniago	841	833	12	8.062258	B

From the classification carried out by applying the steps of the K-Nearest Neighbor Algorithm, the results obtained from the classification of class placements that occupy class c as many as 18 people, can be seen in the table below.

Table 6. Class C . Student Table

No	Name	P	K	Rank	Euclidian Distance	Class
1	Jesika A Bancin	851	869	10	11.6619	C
2	Joel Limbong	805	827	26	13.92839	C
3	Julianto Francisco	813	822	26	9.539392	C
4	Steven Andrianto	805	829	27	15.77973	C
5	Teresa's Flowers	853	855	4	7.211103	C
6	Christian JA Sinaga	813	797	24	6.480741	C
7	Cristian R Simanjuntak	815	811	21	2.828427	C
8	Cici Chaniago	843	841	10	7.615773	C
9	Daniel F Suhombing	810	812	22	4.358899	C
10	Febrian Pardede	816	817	18	3.316625	C
11	Frengdi Solin	855	851	5	8.544004	C
12	Kayla Vera Fitri	846	854	7	6.708204	C
13	Kelvin Erlangga	824	820	16	3.741657	C
14	Manaek Samuel	822	806	19	7.874008	C
15	Willy Simanullang	811	803	23	3.605551	C
16	Yesimonika Ndruu	821	824	17	5.09902	C
17	Yohannes Hutagalung	812	813	20	2.44949	C
18	Manulu Varsi Zone	819	839	15	12.68858	C

From the results above, it can be concluded that the K-Nearest Neighbor algorithm is effective in classifying. Of the 146 data tested with 82 training data and 64 data testing data using the K-Nearest Neighbor algorithm, the results of class placement are 26 students occupying class A, 20 students occupying class B and 18 students occupying class B. class C. Where students who occupy class A are the superior class.

4. Conclusion

Based on the research conducted, it can be concluded that in the research application of Data Mining to Classify Class Placements with the K-Nearest Neighbor Method, several conclusions were obtained, namely: The data used for the application of the K-Nearest Neighbor algorithm to classify class placement is more focused in classifying class placements and past data can also be applied in doing class placement. The K-Nearest Neighbor algorithm can be applied to classify class placements so as to produce class placement classifications with class A, class B and class C and so on depending on a lot of data which will be processed. The application or system created can make it easier for the school to classify Class Placements with the K-Nearest Neighbor Algorithm.

Reference

- [1] Badu, ZS (2016). APPLICATION OF THE K-NEAREST NEIGHBOR ALGORITHM. November.
- [2] Budiman, A. (2017). Development of Coating Inefficiency Calculation System for Web-Based Ceramic Products at PT Satyaraya Keramindoindah. 7(2).
- [3] Fitrianti, RP (2018). Sentiment Analysis of Restaurant Reviews with Indonesian Text Using the K-Nearest Neighbor Algorithm. 27–32.
- [4] Habibi. AM & Santika. (2020). Implementation of the K-Nearest Neighbor Algorithm in Determining Majors Using the Web-Based Euclidean Distance Method at Setia Gama Middle School, 3(4), 7-14.
- [5] Hermawati Astuti Fajar 2018. Data Mining. Yogyakarta. Publisher CV ANDI.
- [6] Informatics, T., & Akakom, S. (2018). K-NN And C4 Algorithm Performance Analysis. 5 On. 2(2).
- [7] Kartika, JI, & Santoso, E. (2017). Determination of Achievement Students Using the K-Nearest Neighbor and Weighted Product Methods (Case Study: SMP Negeri 3 Mejayan). 1(5).
- [8] Mardi, Y. (2017). Data Mining: Classification Using the C4.5 Algorithm. Journal of Informatics Education, 2(2), 213–219.
- [9] Rahayuningtyas, C., Winarso, DS, Informatics, T., Students, P., & Introduction, I. (2017). Implementation of the k-Nearest Neighbor Algorithm for Majoring High School Students. 6(02), 50–57.
- [10] Rosa, AS, and M. Saladin. 2019. Structured and Object Oriented Software Engineering. Bandung : Informatics Bandung.
- [11] Sebastian, D. (2019). Implementation of the K - Nearest Neighbor Algorithm to Classify Products from several E - marketplaces. 5(April), 51–61.
- [12] Setiawan, MPH, & Masya, F. (2020). Analysis of the Design of Public Facilities Reporting Information Systems and Information for Making E-KTPs for the Community. Rabbit : Journal of Technology and Information Systems Univrab, 5(1), 1–8. <https://doi.org/10.36341/rabit.v5i1.810>
- [13] Vulandari Tri Retno 2017. Data Mining Theory and Rapidminer Application. Surakarta. Publisher Gava Media
- [14] Wulandari, DA (2017). *Implementation of Data Mining to Determine High School Majors Using the K-Nearest Neighbor Classification Method at Sman 1 Pace*. 1(10).
- [15] Yahya, Y., & Puspita Hidayanti, W. (2020). Application of the K-Nearest Neighbor Algorithm for Classification of the Effectiveness of Vape Sales (Electric Cigarettes) on “Lombok Vape On.” Infotek : Journal of Informatics and Technology, 3(2), 104–114. <https://doi.org/10.29408/jit.v3i2.2279>