

INTELLIGENT LEARNING ACHIEVEMENT PREDICTION SYSTEM USING K-MEANS ALGORITHM AT UPT EDUCATION UNIT SMPN 5 BLITAR

Ahmad Farid Arwani¹

¹University Students, Faculty of Informatics Engineering, Universitas Islam Balitar, Indonesia

Abstract

This research aims to develop an intelligent system that can predict student performance using the K-Means clustering algorithm in the UPT teaching unit SMPN 5 Blitar. In data collection methods, student scores are taken as input into intelligent systems by educators teaching these subjects. This kind of research is done using quantitative research. The study collected data from a sample of a total of 227 students, with student performance outcomes as input variables. The K-means clustering method is used to classify students into subject grades: subject grades (Religious Education, Indonesian Language, English, Mathematics, Science, Social Studies, Cultural Studies, PJOK, and Computer Science) and behavioral scores (attitude and discipline). After testing using the Silhouette Coefficients data test, the accuracy of the data reached 86%. The conclusion of this study is that the implementation of the K-Means clustering method improves the efficiency and accuracy of identifying bright students.

Keywords: *Intelligent System, Learning Achievement Prediction, K-Means clustering algorithm, Clustering, silhouette coefficient.*

1. INTRODUCTION

One of the public secondary education institutions in Blitar City, namely the UPT Education Unit of SMPN 5 Blitar, is a secondary school that has a vision to become an advanced school and become a competitive force and interest by students in Blitar City. Educators at UPT Education Unit SMPN 5 Blitar mostly still use academic grades as a benchmark to see the level of achievement of their students. Every year the teacher collects a list of grades using the ms excel application. After that, at the end of the semester, the total score of all students will be added up, from here the teachers at the UPT Education Unit of SMPN 5 Blitar can determine the outstanding students. If the grades have all been processed, they will be deposited into the curriculum to find out whether the student is outstanding or not. So that making the assessment takes a long time to determine outstanding students. The assessment is deemed ineffective and inefficient.

¹*Corresponding author, email: andiabrianto.05@gmail.com

Citation in APA style: Arwani, A, F. (2024). Intelligent Learning Achievement Prediction System Using K-Means Algorithm At Upt Education Unit Smpn 5 Blitar. *JOSAR (Journal of Students Academic Research)*, Vol. 9. No. (1):215-222.

Received:
March, 1st 2024

Revised:
March, 15th 2024

Published:
March, 31st 2024

DOI: <https://doi.org/10.35457/josar.v9i1.3100>

Based on existing problems, therefore using the application of Data Mining to make it easier for the school curriculum to select students to enter superior classes, this study uses the K-Means Algorithm Clustering Method to determine the results of cluster grouping of outstanding students because this algorithm is an algorithm that has a fairly high accuracy on the size of the object, so this algorithm is relatively more scalable and efficient for processing large numbers of objects.

There are several previous studies that have been conducted by other researchers regarding superior classes but using other algorithms such as in research (Hutasuhut, 2019) with the Iterative Dichotomiser 3 (ID3) algorithm, then in research using the K-Means Clustering algorithm and the same dataset, namely superior classes, but using a computerized Academic Information System in research (Kusuma, 2019). While research using the same algorithm but with different datasets used such as the K-Means Clustering algorithm in research (Sadewo, 2017). K-Means Clustering algorithm in research (Yulianti, 2019). Fuzzy C-Means and K-Means algorithms in research (Butarbutar, 2016). And the K-Means Clustering Algorithm in research (Praja, 2019). So based on the background above, the author is interested in making a research entitled "Implementation of the K-Means Algorithm to Determine Superior Classes Based on Student Grades at SMPN 1 BOJONG".

Based on the above problems, a research title was made about "CERDASY SYSTEM PREDICTION OF STUDENT PRACTICE USING K-MEANS ALGORITHMMA ON CLASS 7 STUDENTS IN UPT SATUAN EDUCATION SMPN 5 BLITAR". So it can be identified into several problems, namely that the teacher only relies on one academic score to find out which students are outstanding and the lack of efficiency of student value data processed using the Ms. Excel application.

2. METHOD

The research location was at SMPN 5 Blitar from February to August 2023. For the type of research used in this study is quantitative research in which the main data used is obtained from the recap of grades from all teachers who teach in grades 7A to 7H in odd and even semesters at UPT Education Unit SMPN 5 Blitar City.

For data collection techniques, researchers use interview techniques, namely by conducting observations, interviews and literature studies. Meanwhile, the type of data used is primary data, namely the original source or first party.

The data collection instrument uses data collection and problem solving, namely from the teaching teacher of each class subject 7A to 7H. For the first stage of the first research is by collecting data, data cleaning, data transformation (attitude and discipline value data), testing K-Means Clustering data. After calculating K-Means Clustering in determining the superior class, it is obtained that the group of outstanding students (Cluster 2) and the group of non-achieving students (Cluster 1) with reference to entering grades from Religion, Indonesian Language, English, Mathematics, Science, Social Studies, Cultural Arts, PJOK, Informatics, attitude and discipline scores. The value is obtained from the recap of the student teacher who teaches in class 7A to class 7H.

Furthermore, the first Silhouette Coefficient test carried out is the calculation of $a(i)$. $a(i)$ is to find the average distance between points in one cluster. Next, the Silhouette Coefficient test is carried out by calculating $b(i)$. $b(i)$ is to find the average distance between points in the overall data.

From the results of Silhouette Coefficient K-Means clustering and each distance calculation method, the results of each distance are obtained, after obtaining the results of $s(i)$, the average value will be taken and used as the Silhouette Coefficient value. The average result of the calculation of the average distance in one cluster ($a(i)$) and the average distance between all points ($b(i)$) obtained the result $s(i) = 1$ which means that object i is in the right cluster, which means the clustering results are correct.

3. RESULTS AND DISCUSSION

The data used in this study is a collection of student values from class 7A to class 7H obtained from the subject teacher in charge of the class totaling 227 students. The following is the data that has been obtained.

Table 1 Data on student grades from grade 7A to grade 7H

No	Nama Siswa	Kelas	Agama	PKN	Indo	Ing	Mat	IPA	IPS	Seni	pjok	Info sika r p	disip lin
1.	Adelia Wahyu Setyoningsih	7A	81	81	81	80	83	80	80	81	82	84	B B
2.	Aji Eka Saputra	7A	84	79	79	77	78	79	77	79	85	83	D A
3.	Andi Adrian	7A	79	80	80	78	78	77	78	80	85	80	C A
4.	Andrean Titan Ridho Pangestu	7A	92	95	92	97	93	90	93	94	95	92	A A
5.	Anggun Gunawan	7A	85	84	84	85	83	82	80	84	84	86	B A
6.	Armando Junian Nugroho	7A	93	94	95	92	97	82	78	84	83	94	A A
7.	Asyraf Pratama	7A	81	82	92	95	92	97	78	82	85	92	A A
8.	Ayumi Talitha Bernadine	7A	82	78	77	76	77	76	76	79	82	79	D D
9.	Bagus Permadi	7A	80	81	80	77	78	78	78	81	82	83	C B
10.	Cinta Bella Oktavia	7A	85	84	84	85	83	82	80	84	84	86	B A

--													
223	Nilam Rhamadani	7H	80	80	81	77	78	66	77	82	82	81	B B
224	Revalina Hafiz Azalia	7H	80	83	84	78	78	79	78	82	82	83	C A
225	Rivalno Septiansah	7H	82	80	80	78	79	80	78	82	81	82	C A
226	Satria Briant Ardinata	7H	79	78	77	66	76	76	76	78	83	79	B B
227	Savira Putri Felisha	7H	83	82	84	79	81	81	79	84	82	83	C C

Table 2 Comparison Results of 1st to 7th Iteration

Iterasi	Cluster	Titik Centroid												Jumlah Anggota Cluster
		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	
Iterasi ke-1	C1	79,84	79,87	79,39	76,59	77,08	77,56	77,00	79,61	82,86	80,78	3,56	3,82	170
	C2	87,44	88,11	88,98	89,65	88,52	87,65	86,97	86,58	85,91	90,32	4,47	4,95	57
Iterasi ke-2	C1	79,99	79,92	79,46	76,75	77,30	77,76	77,12	79,76	82,90	80,94	3,57	3,86	176
	C2	87,80	88,90	89,86	90,61	89,13	88,17	87,74	86,88	86,12	90,86	4,53	4,94	51
Iterasi ke-3	C1	80,01	79,94	79,45	76,85	77,33	77,78	77,12	79,79	82,90	80,96	3,58	3,86	177
	C2	87,90	89,02	90,10	90,57	89,26	88,30	87,94	86,93	86,19	90,99	4,54	4,94	50
Iterasi ke-4	C1	80,32	80,21	79,75	77,40	77,70	78,07	77,34	80,09	83,00	81,31	3,61	3,94	190
	C2	89,10	90,78	92,31	92,52	91,51	90,51	90,62	87,90	86,84	92,74	4,73	4,92	37
Iterasi ke-5	C1	80,35	80,25	79,79	77,51	77,76	78,11	77,39	80,13	83,01	81,38	3,61	3,95	192
	C2	89,41	91,20	92,82	92,78	92,01	91,00	91,13	88,11	87,03	93,00	4,77	4,91	35
Iterasi ke-6	C1	80,35	80,25	79,79	77,51	77,76	78,11	77,39	80,13	83,01	81,38	3,61	3,95	192
	C2	89,41	91,20	92,82	92,78	92,01	91,00	91,13	88,11	87,03	93,00	4,77	4,91	35
Iterasi ke-7	C1	80,35	80,25	79,79	77,51	77,76	78,11	77,39	80,13	83,01	81,38	3,61	3,95	192
	C2	89,41	91,20	92,82	92,78	92,01	91,00	91,13	88,11	87,03	93,00	4,77	4,91	35

When viewed from the table above, it can be concluded that the results between the 5th, 6th and 7th iterations are the same and the iteration process has stopped. From the above data it can be concluded that the results of clustering using K-Means conclude as follows:

C1 (Non-achieving students): 192 students

C2 (Achievement students): 35 students

The following data presents the overall data from these results.

Table 3 Overall data of clustering results

No	Nama Siswa	Kelas	Agama	PKN	Indo	Ing	Mat	IPA	IPS	Seni	pjok	Info	Sika	Disi	plin	Hasil
1.	Adelia Wahyu Setyoningsih	7A	81	81	81	80	83	80	80	81	82	84	4	4		Tidak Prestasi
2.	Aji Eka Saputra	7A	84	79	79	77	78	79	77	79	85	83	2	5		Tidak Prestasi
3.	Andi Adrian	7A	79	80	80	78	78	77	78	80	85	80	3	5		Tidak Prestasi
4.	Andrean Titan Ridho Pangestu	7A	92	95	92	97	93	90	93	94	95	92	5	5		Prestasi
5.	Anggun Gunawan	7A	85	84	84	85	83	82	80	84	84	86	4	5		Tidak Prestasi
6.	Armando Junian Nugroho	7A	93	94	95	92	97	82	78	84	83	94	5	5		Prestasi

No	Nama Siswa	Kelas	Agama	PKN	Indo	Ing	Mat	IPA	IPS	Seni	pjok	Info	Sika	Disi	plin	Hasil
7.	Asyraf Pratama	7A	81	82	92	95	92	97	78	82	85	92	5	5		Prestasi
8.	Ayumi Talitha Bernadine	7A	82	78	77	76	77	76	76	79	82	79	2	2		Tidak Prestasi
9.	Bagus Permadi	7A	80	81	80	77	78	78	78	81	82	83	3	4		Tidak Prestasi
10.	Cinta Bella Oktavia	7A	85	84	84	85	83	82	80	84	84	86	4	5		Prestasi

223	Nilam Rhamadani	7H	80	80	81	77	78	66	77	82	82	81	4	4		Tidak Prestasi
224	Revalina Hafiz Azalia	7H	80	83	84	78	78	79	78	82	82	83	3	5		Tidak Prestasi
225	Rivalno Septiansah	7H	82	80	80	78	79	80	78	82	81	82	3	5		Tidak Prestasi
226	Satria Briant Ardinata	7H	79	78	77	66	76	76	76	78	83	79	4	4		Tidak Prestasi
227	Savira Putri Felisha	7H	83	82	84	79	81	81	79	84	82	83	3	3		Tidak Prestasi

From the data that has been clustered above, it is necessary to implement testing using the Silhouette Coefficient method which is used to measure the level of accuracy in the cluster result data.

The following are the steps in calculating the Silhouette Coefficient method:

a) Calculation of getting the value a(i)

a(i) is the result value of the average distance between points in the same cluster. For example, the calculation of a(i) for student 1:

$$a(i) = \frac{((\text{point distance 1} - \text{point distance 2}) + (\text{point distance 2} - \text{point distance 4}) \dots)^2}{(\text{number of data in one cluster} - 1)}$$

$$a(i) = \frac{(0 + 8,86 + \dots)^2}{191}$$

$$a(i) = \frac{2.004,1}{191}$$

$$a(i) = 10,49$$

From the data above, it is concluded that the test on the Silhouette Coefficient method for the 1st student is said to be successful, because the results of s(i) show numbers above 0.5. The following are the results of the Silhouette Coefficient method from 227 students.

Table 4 Silhouette Coefficient calculation b(i)

No	Nama Siswa	a(i)	b(i)	s(i)
1.	Adelia Wahyu Setyoningsih	10,49	34,59	0,70
2.	Aji Eka Saputra	9,50	39,47	0,76
3.	Andi Adrian	7,88	40,56	0,81
4.	Andrean Titan Ridho Pangestu	18,32	44,56	0,59
5.	Anggun Gunawan	15,13	27,98	0,46
6.	Armando Junian Nugroho	23,93	37,22	0,36
7.	Asyraf Pratama	25,42	34,41	0,26
8.	Ayumi Talitha Bernadine	9,92	44,66	0,78
9.	Bagus Permadi	7,77	39,54	0,80
10.	Cinta Bella Oktavia	15,13	27,98	0,46
----- --				
223	Nilam Rhamadani	14,63	45,28	0,68
224	Revalina Hafiz Azalia	9,87	36,43	0,73
225	Rivalno Septiansah	8,59	37,96	0,77
226	Satria Briant Ardinata	14,74	50,21	0,71
227	Savira Putri Felisha	11,23	33,72	0,67

From the data above after the results of the calculation of each distance $s(i)$ which is close to number 1 or the value is above 0.5 is 198 out of a total of 227 and the value closer to number 0 or the value is below 0.5 is 29 out of a total of 227. From testing this data we can conclude that from 227 data we managed to get good accuracy of 87%. So in this case it indicates that the selection of outstanding students is appropriate or good.

4. CONCLUSION

Based on the discussion that has been carried out, it has been successfully implemented using the K-Means Clustering Algorithm Method to determine the grouping of students who excel and do not excel in the UPT of SMPN 5 Blitar Education Unit.

Based on tests carried out using the Silhouette Coefficient, the results obtained from the calculation of the average distance in one cluster ($a(i)$) and the calculation of the average distance between all points ($b(i)$), the accuracy result is 87%, which means that the clustering results of high-achieving and non-achieving students are correct.

5. ACKNOWLEDGEMENT

The author realizes that the preparation of this thesis will not be completed without the help of various parties. Therefore, on this occasion we would like to thank:

1. Dr. Drs. H. Soebiantoro, M.Si as the Rector of Balitar Blitar Islamic University.
2. Mr. Abdi Pandu Kusuma, S.Kom, M.T as the Dean of the Faculty of Information Technology at Balitar Blitar Islamic University.
3. Mr. Saiful Nur Budiman, S.Kom., M.Kom as Head of the Informatics Engineering Study Program at Balitar Blitar Islamic University.
4. Mrs. Sri Lestanti, S.Kom., M.T. as the main Supervisor for the guidance, advice, and motivation provided.
5. Mr. Rizki Dwi Romadhona, S.S.T., M.Tr.T. as the second Supervisor for direction and correction so that this Thesis Report can be completed.
6. Mr. Saiful Salim, S.Pd. as the principal who has allowed the SMPN 5 Blitar Education Unit as a research site.
7. All Lecturers of the Department of Informatics Engineering at Balitar Islamic University who have provided their knowledge so that this study can be completed.
8. All Employees of Balitar Blitar Islamic University who have provided support for the completion of this Thesis Report.
9. Parents who never stop giving prayers and love and support given so far.
10. The big family of Informatics Engineering class 2019 employees, especially brothers Andi Abrianto, Inun Pujianto and Yusuf Pramono for all their support, enthusiasm, and cooperation.
11. All parties who have helped and provided input for the preparation of this thesis.

6. REFERENCES

- Bernissa. (2020). Implementasi Algoritma K-Means Untuk Menentukan Kelas Unggulan Pada Smpn 1 Bojong. *Eprosiding Sistem Informasi (Potensi)*, 44 - 46.
- Fayyad, U., G. P.-S., & Padhraic, S. (2016, Juni 19). From Data Mining To Knowledge Discovery In Databases. *Doi Foundation*. Diakses dari <https://doi.org/10.1609/aimag.1230>.
- Fitri, S. U. (2018). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteksi (Jurnal Teknologi Dan Sistem Informasi)*, 98 - 103.
- Husein, M. (2022). Penerapan Algoritma C4.5 Dalam Pemilihan Siswa Berprestasi di SMPN 10 Medan. *Digital Transformation Technology (Digitech)*, 4(2), 78 - 80.
- Lutfi. (2021). Optimalisasi Penentuan Centroid Pada Algoritma K-Means Untuk Pemilihan Jurusan Berbasis Genetikal Algoritma. *Jurnal UIN Alaludin*, 3(1), 54 - 55.
- Salsabila . (2021). Penerapan Algoritma Decision Tree Dalam Menentukan Prestasi Akademik Siswa. *Jurnal Informatika dan Sistem Informasi (JIFoSI)* , 44 - 49.

- Nengsih, S. W. (2021). Analisis Pengelompokan Penentuan Jurusan Siswa SMA Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Betrik*, 4(1), 14 - 16.
- Wibowo, A. S. (2022). K-Means Clustering Untuk Klasifikasi Standar Kualifikasi Pendidikan Dan Pengalaman Kerja Guru Smk Di Indonesia. *Jurnal Dinamika Vokasional Teknik Mesin*. 6(2), 97 - 99.
- Xiong, X. (2019). Topic Anallysis of Microblog About “Didi Talxi” Baised On K-Means Algorithm. *Journall of Informaltion Science alnd Technology*, 4(6), 102..
- Zeebalree, M. R., & Siron, R. B. (2017). The Impalct of Entrepreneuriall Orientaltion on Competitive Advalntalge Moderalted by Finalncing Support in SMEs. *Internationall Review of Malnalgement alnd Malrketing*, 6(2), 98 - 106.