

STUDENT ACHIEVEMENT CATEGORY WITH K-MEANS AND K-NEAREST NEIGHBORS

Desy Lusiyana¹, Dadan Dasari²

¹Universitas Muhammadiyah Cirebon, Jl.Tuaprev No.70, Indonesia

²Universitas pendidikan Indonesia, Jl.Setia Budi, Indonesia

e-mail: dadan.dasari@upi.edu

Abstract

One of the objectives of organizing learning implementation is to determine the achievement of students. Learning at the university level, one of the student learning achievements can be seen from the cumulative achievement index (GPA). GPA is obtained from IP every semester for each subject. This research will determine the categorization of achievement based on mathematics scores. Categories of student achievement are needed to make decisions including finding ways to improve student achievement and facilities needed by students. The purpose of this research is to categorize student achievement so that policy holders get a reference for decision making. Data processing techniques use K-Means and K-NN to determine student achievement categories. There are three categories of achievement namely high, medium and low. The object of this research is students of Universitas Muhammadiyah Cirebon semester 1 in 2022/2023 with certain categories. The data retrieval stage is by downloading score documents on the campus assessment system. Then data preprocessing is carried out to eliminate outliers and data normalization followed by data processing. Data processing stages are generally carried out twice, namely data processing using K-Means then the results are processed again using K-Nearest Neighbors with the aim of predicting better. The results of data processing obtained AUC value 1,000 CA value 0.992, F1 value 0.992, Precision value 0.992 and recall 0.962.

Keywords: student achievement, K-Means, KNN

Abstrak

Salah satu tujuan dari penyelenggaraan pelaksanaan pembelajaran yaitu untuk mengetahui prestasi peserta didik. Pembelajaran pada tingkat universitas prestasi belajar mahasiswa salah satunya dapat dilihat dari nilai Indeks prestasi kumulatif (IPK). IPK diperoleh dari IP setiap semester untuk setiap matakuliah. Pada penelitian ini akan menentukan kategorisasi prestasi berdasarkan nilai matematika. Kategori prestasi mahasiswa dibutuhkan untuk membuat keputusan diantaranya adalah untuk menemukan cara meningkatkan prestasi siswa dan fasilitas yang dibutuhkan oleh mahasiswa. Tujuan penelitian ini yaitu melakukan kategori prestasi mahasiswa sehingga pemegang kebijakan mendapatkan rujukan untuk pengambilan keputusan. Teknik pengolahan data menggunakan K- Means dan K-NN untuk menentukan kategori prestasi siswa. Terdapat tiga kategori prestasi yaitu tinggi, sedang dan rendah. Objek Penelitian ini adalah mahasiswa Universitas Muhammadiyah Cirebon semester 1 pada tahun 2022/2023 dengan kategori tertentu. Tahapan pengambilan data yaitu dengan melakukan pengunduhan dokumen nilai pada sistem penilaian yang dimiliki kampus. Kemudian dilakukan preprosessing data untuk menghilangkan outlier dan normalisasi data dilanjutkan dengan pengolahan data. Tahapan pengolahan data secara garis besar dilakukan 2 kali yaitu dilakukan pengolah data menggunakan K-Means kemudian hasilnya olah Kembali menggunakan K-Nearest Neighbors dengan tujuan memprediksi menjadi lebih baik. Hasil pengolahan data diperoleh nilai AUC 1.000 Nilai CA 0.992, Nilai F1 0.992, Nilai Precision 0.992 dan recall 0.962.

Kata kunci: prestasi belajar, K-Means, K-NN

INTRODUCTION

Based on the decision of the Director General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia Number 84/E/KPT/2020 concerning Guidelines for Implementing Compulsory Courses in the Higher Education Curriculum, it was decided that every college curriculum must contain courses on religion, Pancasila, citizenship and Indonesian. This course is mandatory because it has the function of forming the character and civility of students with dignity. The minimum study load for each course is 2 semester credit units. These four courses were developed by every university in Indonesia. The importance of these four courses is that the government under the Director of Learning and Student Affairs provides assistance in developing mandatory curriculum courses (MKWK) to develop MKWK tools and conduct MKWK implementation surveys at each university. Apart from MKWK at Cirebon Muhammadiyah University there are university courses.

One of the subjects that is becoming common at universities is mathematics. Mathematics is a mandatory university subject because mathematics is seen to play a large role in everyday life. Typical mathematical thinking patterns such as systematic, scientific, logical and critical make mathematics courses a provision for students in decision making. Low achievement in learning mathematics not only has an impact on academic achievement but also affects cognitive development and logical thinking. Therefore, various efforts are made by lecturers to be able to improve student achievement. One of them is by knowing the student Achievement Index which is used to measure the ability to think mathematical reasoning because mathematical reasoning can contribute to student learning achievement as done by (Gde Somatanaya, 2017).

This research uses mathematics course grades because apart from the objectives of the MKWK, mathematics influences other achievements. This can be seen from the bibliometric results obtained based on a search via Scopus with the keyword achievement during 2018 to 2023, 2782 articles were obtained with the theme of learning achievement. Visualization can be shown in the following image.

the data processed is the final student grades from 7 study programs for the same student. The 7 study programs are Nutrition, Government Science, Mathematics, Early Childhood Education, PGSD, Industrial Engineering and Information Engineering.

The data obtained used secondary data originating from the academic system at Muhammadiyah University of Cirebon. This system is called a great system. The data taken is data from MKWK and Mathematics course grades. The MKWK data that can be downloaded are those from which the data is taken, namely Mathematics, Indonesian, Religion and Pancasila. Citizenship courses cannot be downloaded because there is a system migration from SIMAKU to the great system.

The number of students studied was 369 students. The scores for four courses were taken from 369 students. Then data processing was carried out using K-Means and KNN. The achievement categories that will be achieved are high, medium and low categories.

The data processing technique in this research uses K-Mean and K-NN calculations using orange software. The stages used are uploading files, uploading data, carrying out preprocessing such as avoiding outlier data and carrying out normality tests after that processing the data using K-mean. After completing the K-mean steps, the K-NN process is carried out. The stages carried out are the same as in K-Means processing, the data uploaded is based on the K-Means results obtained.

The data processing technique in this study uses K-Mean and K-NN calculations using orange software. (Purnamasari & Widiastuti, 2017) states that K-Means is one of the unsupervised learning methods. The K-Means method is included in the non-hierarchical clustering category which is easy and simple, finds convergence quickly, and can adapt to data distribution. The research data is unsupervised data that needs to be clustered so it requires the K-Means method. K-Means is used to determine the level of learning achievement based on the scores obtained by students in mathematics courses. (Cholil et al., 2021) stated that the KNN algorithm has several advantages, including robustness against training data that has a lot of noise and a large amount of data. In this study, the data taken is secondary data consisting of hundreds of data so that it requires data processing with KNN. The combination of both can produce good data interpretation as done by (Aditra Pradnyana et al., n.d) on the division of college classes based on learning quality. The data obtained can clearly map the results of the learning process carried out. In this study, the stages used are uploading files,

uploading data, doing preprocessing such as avoiding outlier data and doing normality tests after that processing data using K-mean. After the stages in K-mean are completed, the K-NN process is carried out. The stages carried out are the same as in K-Means processing, the uploaded data is based on the results obtained by K-Means.

RESULTS AND DISCUSSION

Results of data processing using orange software on secondary data with a total of 369 values at the Muhammadiyah University of Cirebon. The course scores used are mathematics, Pancasila, religion and Indonesian in 7 study programs, namely the 7 study programs, namely Nutrition, Government Science, Mathematics, PAUD, PGSD, Industrial Engineering and Information Engineering. The data tabulation is as follows.

Tabel 1. Data Tabulation

| Mahasiswa | Prodi | Agama | Pancasila | Bahasa Indonesia | MTK |
|-----------|--------|-------|-----------|---------------------|-------|
| 1 | Gizi A | 83 | 89,45 | 88,75 | 85,1 |
| 2 | Gizi A | 83 | 90,05 | 90,3 | 89,6 |
| 3 | Gizi A | 82,3 | 90,45 | 91,5 | 85,9 |
| 4 | Gizi A | 83 | 91,55 | 84,3 | 84,2 |
| 5 | Gizi A | 91,3 | 0,00 | 6,3 | 6 |
| 6 | Gizi A | 83 | 89,35 | 95,5 | 79,7 |
| 7 | Gizi A | 83 | 91,05 | 90,35 | 88,6 |
| 8 | Gizi A | 86,5 | 93,45 | 94,05 | 80,95 |
| 9 | Gizi A | 83 | 90,95 | 85,15 | 83,75 |
| 10 | Gizi A | 89,5 | 89,25 | 88,1 | 77,5 |
| 11 | Gizi A | 83 | 89,70 | 87,9 | 89,45 |
| 12 | Gizi A | 77,9 | 89,25 | 91,05 | 85,15 |
| 13 | Gizi A | 83 | 88,95 | 91,85 | 89,65 |
| 14 | Gizi A | 74,65 | 89,85 | 87,4 | 82,05 |
| 15 | Gizi A | 85,4 | 0,00 | 0 | 1,95 |
| ... | ... | ... | ... | ... | ... |
| 369 | TIF C | 90,5 | 81 | 88,1 | 76,5 |

the data tabulation results above were then processed using orange software. the steps for using preprocessing are as follows.

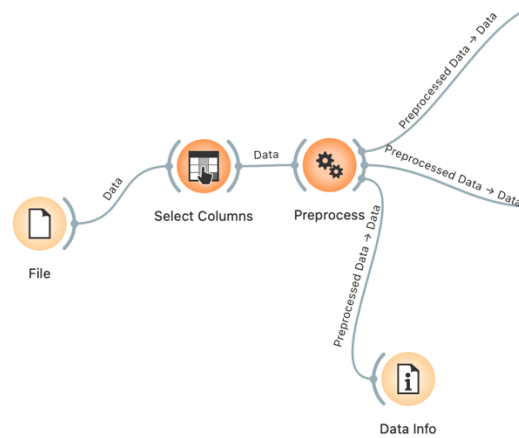


Figure 2. Preprocessing data

In Figure 1 is the stage to process data using orange software by inputting data in the file. Then do the colom selection and preprocessing process. The results of data preprocessing obtained the following data

| | Agama | Pancasila | Bahasa Indonesia | MTK |
|----|----------|-----------|------------------|----------|
| 1 | 0.12691 | 0.51134 | 0.40030 | 0.34826 |
| 2 | 0.12691 | 0.53938 | 0.47575 | 0.53998 |
| 3 | 0.09419 | 0.55807 | 0.53416 | 0.38235 |
| 4 | 0.12691 | 0.60947 | 0.18369 | 0.30992 |
| 5 | 0.51489 | -3.66844 | -3.61298 | -3.02165 |
| 6 | 0.12691 | 0.50667 | 0.72886 | 0.11821 |
| 7 | 0.12691 | 0.58611 | 0.47818 | 0.49738 |
| 8 | 0.29051 | 0.69825 | 0.65828 | 0.17146 |
| 9 | 0.12691 | 0.58144 | 0.22507 | 0.29075 |
| 10 | 0.43075 | 0.50200 | 0.36866 | 0.02448 |
| 11 | 0.12691 | 0.52303 | 0.35892 | 0.53359 |
| 12 | -0.11148 | 0.50200 | 0.51225 | 0.35039 |
| 13 | 0.12691 | 0.48798 | 0.55119 | 0.54211 |
| 14 | -0.26340 | 0.53004 | 0.33459 | 0.21832 |
| 15 | 0.23910 | -3.66844 | -3.91964 | -3.19420 |
| 16 | 0.12691 | 0.53004 | 0.54389 | 0.32057 |
| 17 | 0.23910 | 0.48798 | 0.36866 | 0.44838 |
| 18 | 0.08952 | 0.55574 | 0.35892 | 0.12886 |
| 19 | 0.32324 | 0.56742 | 0.71912 | 0.35465 |
| 20 | 0.12691 | 0.59078 | 0.40760 | 0.71465 |
| 21 | 0.12691 | 0.55807 | 0.61690 | 0.54211 |
| 22 | 0.05680 | 0.47863 | 0.61934 | 0.40578 |
| 23 | 0.12691 | 0.55574 | 0.51469 | 0.53572 |
| 24 | 0.31155 | 0.40621 | 0.33215 | 0.47394 |
| 25 | 0.26714 | 0.35013 | 0.27374 | 0.48885 |
| 26 | 0.29519 | 0.38284 | 0.22993 | 0.50164 |
| 27 | 0.28350 | 0.41088 | 0.25914 | 0.47181 |
| 28 | 0.34193 | 0.41088 | 0.22263 | 0.49738 |
| 29 | 0.27182 | 0.39453 | 0.24697 | 0.49738 |
| 30 | -0.15121 | 0.06276 | 0.27131 | 0.42069 |
| 31 | 0.23208 | 0.35013 | 0.24697 | 0.46542 |
| 32 | 0.22741 | 0.34546 | 0.30538 | 0.51442 |
| 33 | 0.26714 | 0.35948 | 0.22750 | 0.47394 |
| 34 | 0.24377 | 0.33378 | 0.22263 | 0.47394 |
| 35 | 0.32791 | 0.33378 | 0.35406 | 0.49311 |
| 36 | 0.28584 | 0.41088 | 0.27861 | 0.48246 |
| 37 | 0.22741 | 0.37817 | 0.22750 | 0.47607 |
| 38 | 0.25546 | 0.38284 | 0.23724 | 0.48885 |
| 39 | 0.17366 | 0.30808 | 0.25184 | 0.41217 |
| 40 | 0.26714 | 0.38752 | 0.24454 | 0.45051 |

Figure 3. sample result Prepoessing data

the data obtained was tabulated in excel and then imported without changing the format because orange software can read data in xls format. then select outlier data and carry out the normalization testing stage. after that, we enter the next stage, namely processing the preprocessing data to be processed using k-means. this stage can be explained in the following picture.

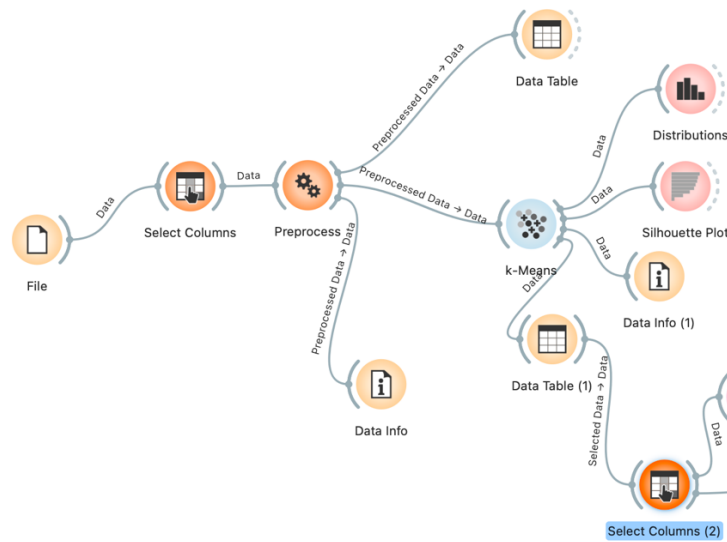


Figure 4. process K-Means

processing results using k-means. the data obtained as a result of data mining transformation is obtained as follows.

| Data Table | | Mon Dec 18 23, 12:30:34 | | | | |
|---------------------|---------|-------------------------|----------|-----------|------------------|----------|
| Data instances: 369 | | | | | | |
| Features: 4 | | | | | | |
| Meta attributes: 2 | | | | | | |
| | Cluster | Silhouette | Agama | Pancasila | Bahasa Indonesia | MTK |
| 1 | C1 | 0.722091 | 0.12691 | 0.51134 | 0.40030 | 0.34826 |
| 2 | C1 | 0.719764 | 0.12691 | 0.53938 | 0.47575 | 0.53998 |
| 3 | C1 | 0.71835 | 0.09419 | 0.55807 | 0.53416 | 0.38235 |
| 4 | C1 | 0.718775 | 0.12691 | 0.60947 | 0.18369 | 0.30992 |
| 5 | C2 | 0.528955 | 0.51489 | -3.66844 | -3.61298 | -3.02165 |
| 6 | C1 | 0.708568 | 0.12691 | 0.50667 | 0.72886 | 0.11821 |
| 7 | C1 | 0.719058 | 0.12691 | 0.58611 | 0.47818 | 0.49738 |
| 8 | C1 | 0.709447 | 0.29051 | 0.69825 | 0.65828 | 0.17146 |
| 9 | C1 | 0.719853 | 0.12691 | 0.58144 | 0.22507 | 0.29075 |
| 10 | C1 | 0.714139 | 0.43075 | 0.50200 | 0.36866 | 0.02448 |
| 11 | C1 | 0.721541 | 0.12691 | 0.52303 | 0.35892 | 0.53359 |
| 12 | C1 | 0.714697 | -0.11148 | 0.50200 | 0.51225 | 0.35039 |
| 13 | C1 | 0.719121 | 0.12691 | 0.48798 | 0.55119 | 0.54211 |
| 14 | C1 | 0.70883 | -0.26340 | 0.53004 | 0.33459 | 0.21832 |
| 15 | C2 | 0.543877 | 0.23910 | -3.66844 | -3.91964 | -3.19420 |
| 16 | C1 | 0.718788 | 0.12691 | 0.53004 | 0.54389 | 0.32057 |
| 17 | C1 | 0.72423 | 0.23910 | 0.48798 | 0.36866 | 0.44838 |
| 18 | C1 | 0.716637 | 0.08952 | 0.55574 | 0.35892 | 0.12886 |
| 19 | C1 | 0.713754 | 0.32324 | 0.56742 | 0.71912 | 0.35465 |
| 20 | C1 | 0.716363 | 0.12691 | 0.59078 | 0.40760 | 0.71465 |
| 21 | C1 | 0.71628 | 0.12691 | 0.55807 | 0.61690 | 0.54211 |
| 22 | C1 | 0.717006 | 0.05680 | 0.47863 | 0.61934 | 0.40578 |
| 23 | C1 | 0.718758 | 0.12691 | 0.55574 | 0.51469 | 0.53572 |
| 24 | C1 | 0.725748 | 0.31155 | 0.40621 | 0.33215 | 0.47394 |
| 25 | C1 | 0.726233 | 0.26714 | 0.35013 | 0.27374 | 0.48885 |
| 26 | C1 | 0.725557 | 0.29519 | 0.38284 | 0.22993 | 0.50164 |
| 27 | C1 | 0.72589 | 0.28350 | 0.41088 | 0.25914 | 0.47181 |
| 28 | C1 | 0.724971 | 0.34193 | 0.41088 | 0.22263 | 0.49738 |
| 29 | C1 | 0.725669 | 0.27182 | 0.39453 | 0.24697 | 0.49738 |
| 30 | C1 | 0.714536 | -0.15121 | 0.06276 | 0.27131 | 0.42069 |
| 31 | C1 | 0.726219 | 0.23208 | 0.35013 | 0.24697 | 0.46542 |
| 32 | C1 | 0.725631 | 0.22741 | 0.34546 | 0.30538 | 0.51442 |
| 33 | C1 | 0.726084 | 0.26714 | 0.35948 | 0.22750 | 0.47394 |
| 34 | C1 | 0.726003 | 0.24377 | 0.33378 | 0.22263 | 0.47394 |
| 35 | C1 | 0.725644 | 0.32791 | 0.33378 | 0.35406 | 0.49311 |
| 36 | C1 | 0.725825 | 0.28584 | 0.41088 | 0.27861 | 0.48246 |
| 37 | C1 | 0.72571 | 0.22741 | 0.37817 | 0.22750 | 0.47607 |
| 38 | C1 | 0.725786 | 0.25546 | 0.38284 | 0.23724 | 0.48885 |
| 39 | C1 | 0.725896 | 0.17366 | 0.30808 | 0.25184 | 0.41217 |
| 40 | C1 | 0.726259 | 0.26714 | 0.38752 | 0.24454 | 0.45051 |
| 41 | C1 | 0.72641 | 0.29519 | 0.33378 | 0.25427 | 0.47181 |
| 42 | C1 | 0.723239 | 0.14795 | 0.13285 | 0.27374 | 0.43134 |
| 43 | C1 | 0.724415 | 0.15729 | 0.18425 | 0.26644 | 0.39087 |

Figure 5. Sample result K-Means

the second data processing was carried out using k-nn. based on the results of k-means, three categories were obtained: c1 in the high category, c2 in the medium category and c3 in the low category. this data can be presented in the following graphical form.

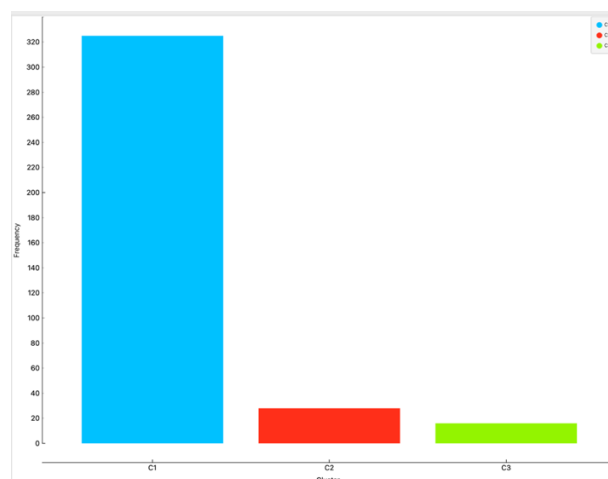


Figure 6. Distribution categories

Based on the figure above, 3 categorizations are obtained, namely C1, C2 and C3. C1 shows the highest number and the lowest number is obtained by C3.

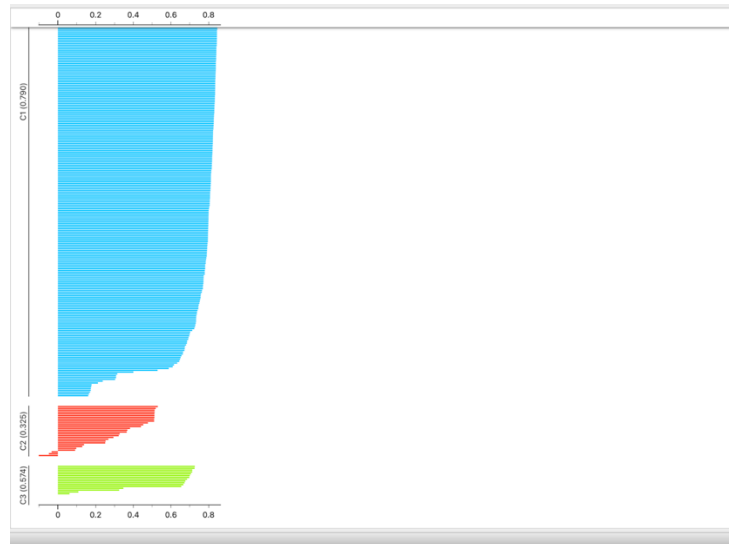


Figure 7. Grafic silhouette K-NN

Based on this graph, there is a very large gap. The achievements of students in the high category are more dominant when compared to other categories. Next, the data is carried out. The stages of the K-NN process are as follows.

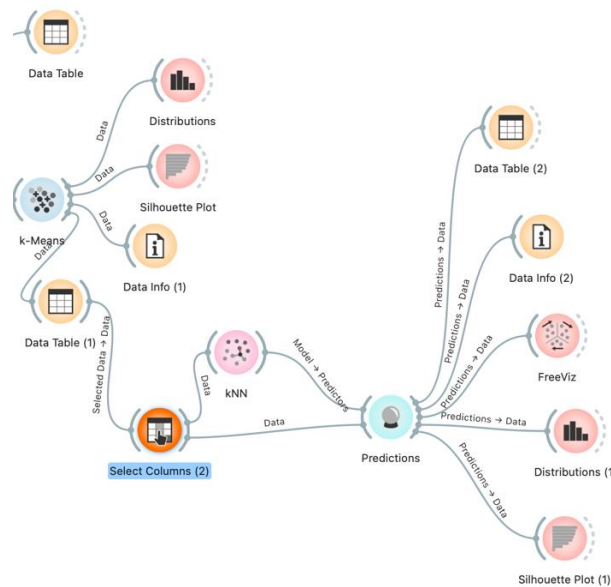


Figure 8. Process K-NN

Based on the image above, it is explained that the data has been transformed in the K-process. Means is then input data to be processed using KNN. The results of K-NN are as follows

Show probabilities for Classes known to the model Show classification errors Restore Original Order

| | kNN | error | Cluster | Silhouette | Agama | Pancasila | Bahasa Indonesia | MTK |
|----|-------------------------|-------|---------|------------|----------|-----------|------------------|----------|
| 1 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.722091 | 0.12691 | 0.51134 | 0.40030 | 0.34826 |
| 2 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.719764 | 0.12691 | 0.53938 | 0.47575 | 0.53998 |
| 3 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.71835 | 0.09419 | 0.55807 | 0.53416 | 0.38235 |
| 4 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.718775 | 0.12691 | 0.60947 | 0.18369 | 0.30992 |
| 5 | 0.00 : 1.00 : 0.00 → C2 | 0.000 | C2 | 0.528955 | 0.51489 | -3.66844 | -3.61298 | -3.02165 |
| 6 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.709568 | 0.12691 | 0.50667 | 0.72886 | 0.11821 |
| 7 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.719058 | 0.12691 | 0.58611 | 0.47818 | 0.49738 |
| 8 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.709447 | 0.29051 | 0.69825 | 0.65828 | 0.17146 |
| 9 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.719853 | 0.12691 | 0.58144 | 0.22507 | 0.29075 |
| 10 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.714139 | 0.43075 | 0.50200 | 0.36866 | 0.02448 |
| 11 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.721541 | 0.12691 | 0.52303 | 0.35892 | 0.53359 |
| 12 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.714697 | -0.11148 | 0.50200 | 0.51225 | 0.35039 |
| 13 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.719121 | 0.12691 | 0.48798 | 0.55119 | 0.54211 |
| 14 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.70883 | -0.26340 | 0.53004 | 0.33459 | 0.21832 |
| 15 | 0.00 : 1.00 : 0.00 → C2 | 0.000 | C2 | 0.543877 | 0.23910 | -3.66844 | -3.91964 | -3.19420 |
| 16 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.718788 | 0.12691 | 0.53004 | 0.54389 | 0.32057 |
| 17 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.72423 | 0.23910 | 0.48798 | 0.36866 | 0.44838 |
| 18 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.716637 | 0.08952 | 0.55574 | 0.35892 | 0.12886 |
| 19 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.713754 | 0.32324 | 0.56742 | 0.71912 | 0.35465 |
| 20 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.716363 | 0.12691 | 0.59078 | 0.40760 | 0.71465 |
| 21 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.71628 | 0.12691 | 0.55807 | 0.61690 | 0.54211 |
| 22 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.717006 | 0.05680 | 0.47863 | 0.61934 | 0.40578 |
| 23 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.718758 | 0.12691 | 0.55574 | 0.51469 | 0.53572 |
| 24 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.725748 | 0.31155 | 0.40621 | 0.33215 | 0.47394 |
| 25 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.726233 | 0.26714 | 0.35013 | 0.27374 | 0.48885 |
| 26 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.725557 | 0.29519 | 0.38284 | 0.22993 | 0.50164 |
| 27 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.72589 | 0.28350 | 0.41088 | 0.25914 | 0.47181 |
| 28 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.724971 | 0.34193 | 0.41088 | 0.22263 | 0.49738 |
| 29 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.725669 | 0.27182 | 0.39453 | 0.24697 | 0.49738 |
| 30 | 1.00 : 0.00 : 0.00 → C1 | 0.000 | C1 | 0.714536 | -0.15121 | 0.06276 | 0.27131 | 0.42069 |

Show performance scores Target class: (Average over classes)

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|-------|-------|-------|-------|-------|--------|-------|
| kNN | 1.000 | 0.992 | 0.992 | 0.992 | 0.992 | 0.962 |

Figure 9. Result K-NN

Based on the figure above, it shows that the C1 group, if enumerated, has an accurate classification and value for the categorization that appears. There is a C1 classification which is the most dominant value that appears.

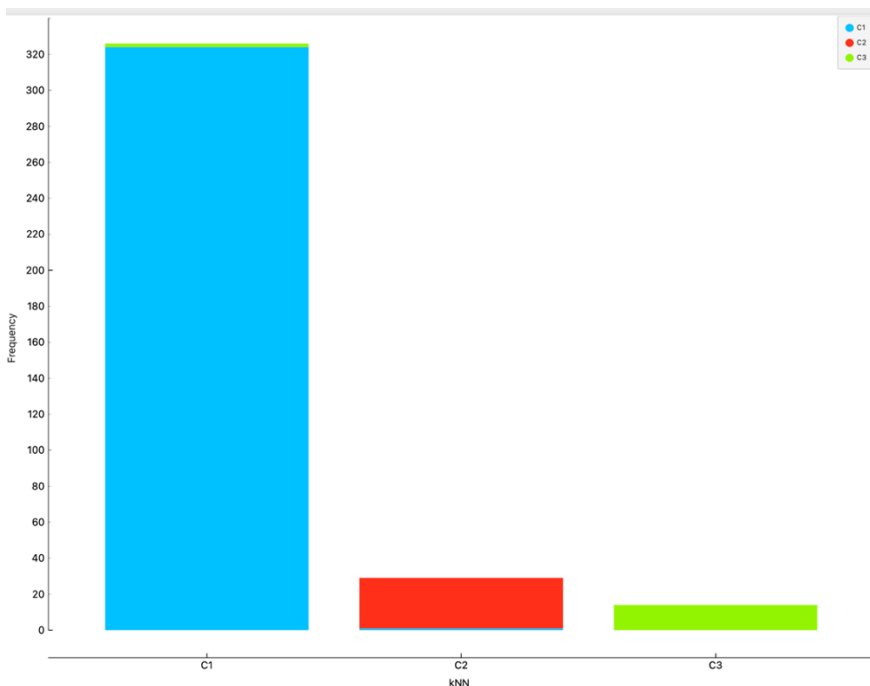


Figure 10. Distribution K-NN

Overall, the stages of data processing using Orange software are as follows.

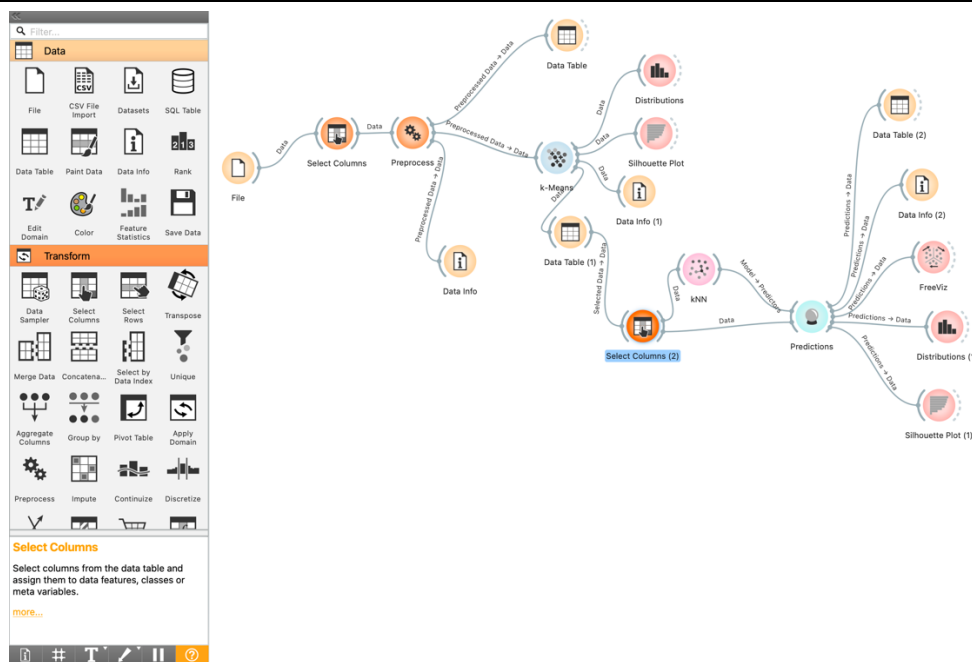


Figure 11. Visualisation K-Mean and K-NN with Oranges

The results of the research have been described in the results above using K-Means and K-NN to obtain an AUC (Area Under The Curve) of 1,000, a CA value of 0.992, an F1 value of 0.992, a precision value of 0.992 and a recall of 0.962. The accuracy results are the best value. The results show that students are currently in category 1, namely high. This shows that performance has been very good.

The results of data processing show that student achievement in mathematics courses at Universitas Muhammadiyah Cirebon in the 7 study programs is good. This is indicated by the existence of a high category in mathematics learning achievement. The university has appropriately made various efforts to improve student achievement. This shows that the factors of learning achievement (Daely et al., 2013) are factors of environmental conditions and parental supervision, financial factors and motivation to learn, learning factors and time distribution as well as factors of lecturer teaching quality and student health. The categories carried out in the research that the quality of lecturers must be further improved. Based on research there are various kinds of efforts to further improve student learning achievement. (Aedi, n.d.) States that by applying problem solving to students can improve learning achievement. In the learning process between students and students is much different. Learning in students can be done with the andragogy learning process so that it is necessary

to give a problem to think broadly. Providing a learning process using problem solving makes students able to recognize problems from various points of view to get solutions. There are other ways to improve learning achievement.

(Fadilah et al., 2021) stated that learning independence and self-regulation can affect mathematics learning achievement, which is related to commitment to learning. So that if there are conditions that do not allow learning such as during the covid outbreak in 2019, students are still aware of learning.

CONCLUSION

The results of the research have been described in the results above using K-Means and K-NN to obtain an AUC (Area Under The Curve) of 1,000, a CA value of 0.992, an F1 value of 0.992, a precision value of 0.992 and a recall of 0.962. The accuracy results are the best value. The results show that students are currently in category 1, namely high. This shows that performance has been very good.

REFERENCES

- Aditra Pradnyana, G., Aan, A., & Permana, J. (n.d.). *PEMBAGIAN KELAS KULIAH MAHASISWA DENGAN METODE K-MEANS DAN K-NEAREST NEIGHBORS UNTUK MENINGKATKAN KUALITAS PEMBELAJARAN*.
- Aedi, W. G. (n.d.). UPAYA MENINGKATKAN PRESTASI BELAJAR MATEMATIKA DENGAN PENDEKATAN PROBLEM SOLVING. *Jurnal Ilmiah Pendidikan Matematika*, 5(1).
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 6, Issue 2).
- Daely, K., Sinulingga, U., & Manurung, A. (2013). *ANALISIS STATISTIK FAKTOR-FAKTOR YANG MEMPENGARUHI INDEKS PRESTASI MAHASISWA* (Vol. 1, Issue 5).
- Fadilah, A. N., Tayeb, T., Nur, F., Suharti, & Sriyanti, A. (2021). PENGARUH KEMANDIRIAN BELAJAR DAN REGULASI DIRI TERHADAP PRESTASI BELAJAR MAHASISWA. *Jurnal Ilmiah Pendidikan Matematika*, 6(1), 100–115.
- Gde Somatanaya, A. (2017). *ANALISIS KEMAMPUAN BERFIKIR NALAR MATEMATIS SERTA KONTRIBUSINYA TERHADAP PRESTASI BELAJAR MAHASISWA (Studi Terhadap Mahasiswa FKIP Pendidikan Matematika Universitas Siliwangi)*.
- Hilda Nur Zerlinda, Isnandar Slamet, & Etik Zukhronah. (2019). *1686-3914-1-SM*. 88–93.
- Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). *Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking* (Vol. 2, Issue 8). <http://j-ptiik.ub.ac.id>
- Purnamasari, K. K., & Widiastuti, N. I. (2017). PERBANDINGAN ALGORITMA K-MEANS DAN K-NEAREST NEIGHBORS PADA SISTEM PERINGKASAN OTOMATIS. *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, 6(2).
- Purwaningsih, E., & Nurelasari, E. (2021). Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa. In *Syntax: Jurnal Informatika* (Vol. 10, Issue 01).