Implementation of Data Mining for Predicting Student Graduation Using the K-Nearest Neighbor Algorithm at Jambi Muhammadiyah University

[1]Shandy Amandha, [2]Hetty Rohayani, [3]Kevin Kurniawansyah
[1,2,3]Fakultas Sains dan Teknologi, Universitas Muhammadiyah Jambi, Jambi, Indonesia
Email: [1]shandyamandha@gmail.com, [2]hettyrohayani@gmail.com, [3]kevin.kurniawansy4h@gmail.com

Article Info	ABSTRACT
	Graduation is one of the assessment items in the accreditation process of a tertiary institution. So that if students graduate on time, it will help assess the accreditation of a tertiary institution. The method used is K-Nearest Neighbor (K-NN). This method is used to classify objects based on learning data closest to the object. This research aims to predict the graduation of Jambi Muhammadiyah University students, whether it is worth graduating on time or not graduating on time. In research using K-NN to predict student graduation, the results were that the K-NN approach in this study produced an accuracy value of 93.33%. The result is testing with a value of K = 5 using 50 training data for Jambi Muhammadiyah University students who have graduated in 2022; then data testing is tested with training data that has been tested before, 4 students who graduated not on time and who graduated on time were 46 students. *Copyright © 2024 Puzzle Research Data Technology*

Corresponding Author:
Shandy Amandha,
Faculty of Sains and Technology,
Universitas Muhammadiyah Jambi,
1Perum Puri Masurai Blok. E No. 10, Jambi, Indonesia.
Email: shandyamandha@gmail.com

1. INTRODUCTION

Education is one of the most important and essential parts of people's lives. It is used to enhance individual growth academically and financially. An educated individual must contribute to his family, society, and community [1]. The highest educational institution is a tertiary institution that organises academic education for students. Students are often referred to as people with broader intellectual characteristics compared to their age group who are not students or other age groups below them. With their intellect, students will be able to face and look for problems systematically which will later be applied in everyday life so they can compete in the world of work [2].

Universities must produce competent graduates. This can be assessed from the graduation rate of students. In addition, graduating on time is also the dream of students. Students no longer need to pay tuition fees and can work faster [3]. But what happened in the field differed greatly from expectations because students may need more time to complete their studies. Several factors can affect late student graduation, such as student marriages, student status, and student understanding of the course material, which can be seen from student GPA scores. More importantly, student graduation rates also greatly affect the accreditation of a tertiary institution, so universities try or have to help students as much as possible so they can graduate on time.

Graduation is one of the assessment items in the accreditation process of a tertiary institution. One of the assessment elements for college graduates is that universities have an ideal educational efficiency rate (BANPT, 2007). So if students graduate on time, it will help assess the accreditation of a tertiary institution[4]. The percentage of ups and downs in a student's ability to complete studies on time is one of the elements of evaluating higher education accreditation [2].

A high on-time graduation rate is a boon for both parties. First, the students, students will get jobs more easily because companies tend to look for new graduates. Second, for the university, because of the exact timing of student graduation will help advance the university's quality, such as accreditation [5]. Muhammadiyah Jambi University is one of the private tertiary institutions in Jambi, which BAN-PT has very well accredited. It has 5 Study Programs: Development Economics Study Program is accredited B, and Management is accredited Very Good. Good accredits Informatics. Information Systems is accredited Good. Forestry is accredited Good.

With a total number of active students of 2775 in the odd year 2022, every year, Jambi Muhammadiyah University graduates more than 100 students consisting of 5 Study Programs. Jambi Muhammadiyah University should have and implement policies to manage and evaluate these standards, namely student graduation, that reflect tertiary institutions' performance in quality improvement. One of the efforts to manage student graduation is to achieve a student study period according to the scheduled length of study. However, Jambi Muhammdiyah University is experiencing difficulties regarding how students can reach their study period on time according to the schedule set by this tertiary institution, as evidenced by the decline in the number of graduates in 2022, as many as 134 students [6].

K-NN is a supervised method that aims to get a new data pattern by connecting the previous data pattern with the new one to classify data into several classes based on existing attributes [7]. The K value used here represents the number of nearest neighbours that can be involved in determining the predictions on the class label of the test data. As many as the K's closest neighbours are elected later, a vote will be carried out from the K's nearest neighbours. The class with the most neighbouring votes will be considered the predicted class label in the test data [8]. K-Nearest Neighbor (K-NN) algorithm is a classification algorithm that solves various classification problems [9]. The K-NN algorithm is a method for classifying objects based on learning data closest to the object [10].

Data mining is a series of processes to obtain knowledge or patterns from data sets [11]. Data Mining is a method used for processing large-scale data. Therefore, data mining has become a very important role in data processing. Data mining can be interpreted as a series of activities to obtain large amounts of data and can be stored in databases, data warehouses or other information stores. Data mining techniques include data analysis, signal processing, neural networks and pattern recognition [12].

The following are some of the previous studies that the author made as support in this study, namely:
1. Research Nurul Khasanah, Agus Salim, Nurul Afni, et al. In 2022. About predicting student graduation using the Naïve Bayes method. The goal of predicting student graduation rates uses 379 data, with details of 303 training data and 76 data testing data [3].
2. Early research by Destiani Siti Fatimah, Elita Rahmawati. In 2022. Regarding using the decision tree method in the design of prediction systems for student graduation. The research objective was to apply a timely graduation prediction system built to provide information to high schools about the quality of each student and predictions about timely graduation [13].
3. Research by Manarul Hidayat, Ahmad Faqih, and Tati Suprapti. In 2022. Regarding the implementation of the k-nearest neighbour algorithm for predicting graduation accuracy. The research objective is to apply the K-NN method using cross-validation to predict student graduation rates at STMIK IKMI[14].
4. Research by Robi Sepriansyah, Susan Dian Purnamasari, Kiky Rizky Wardani, et al. In 2023. Regarding predicting students' graduation at the Faculty of Engineering, Bina Darma University, using the Naïve Bayes algorithm [15].
5. Research by Hendrik Saputra, Fika Trisnawati. In 2021. Regarding using the Fuzzy K-Nearest Neighbor method in the student graduation rate prediction system. The aim is to change the score from the first-semester grade point index to the fourth-semester grade point index into three sets of fuzzy values, namely satisfactory, very satisfactory, and cum laude [16].
6. Research by Hamdani, Torkis Nasution. In 2020. Regarding implementing the K-Nearest Neighbor algorithm for determining student graduation on time. The aim of overcoming the problem is to apply the K-Nearest Neighbor Algorithm for clustering data on the results of prospective new students' exams. The procedure determines the number of data clusters and the cluster's centre point and calculates the distance between objects and centroids, group objects [17].
7. Research by Aldo Sabathos Mananta, Green Arter Sandag. In 2021. About predicting student graduation in choosing a master's program using the K-NN algorithm. The aim is to assist decision-making in determining entry to the next higher education level [18].

From the problem description above, the writer is interested in predicting student graduation time using the k-nearest neighbour method. Prediction is estimating something that will happen in the future[19].

From this research, it is hoped that it can predict student graduation rates which can assist academics in managing and evaluating quality and quality related to student graduation within Jambi Muhammadiyah University. If the study period can be predicted and, from the prediction, it does not match or exceed the time specified, the academic can take precautions, assist, and provide solutions in overcoming obstacles to achieving graduation on time.

2. RESEARCH METHOD

In this research, the object of research is Jambi Muhammadiyah University, which is one of the private universities in the city of Jambi. In this study, several stages of research will be carried out, as figure 1 follow.
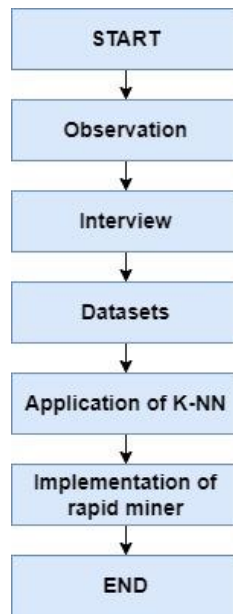
Figure 1. Stages of Research Methods

Observations were made by directly observing an object to be studied, namely at the Jambi Muhammadiyah University. It is conducting direct interviews with one of the staff in the relevant academic field to obtain information verbally to obtain data that can explain or answer a problem in this research. This is Literature by searching for supporting literature and studying previous research, which can provide the information needed to complete this research and can help reinforce existing theories.

2.1. Dataset

Data is a set of operational data that is needed before it is carried out before the information mining stage begins[20]. The data used in processing is data on students at Jambi Muhammadiyah University who have graduated as much as 50 data per study program as training data. And semester VI student data as data testing. With attributes, name, gender, semester 1-6 GPA and graduation status.

Table 1. Data Training

No	Name	Gender	S1	S2	S3	S4	S5	S6	Information
1	Agus Kurniawan	Male	3,58	3,52	3,39	3,31	3,28	3,12	On time
2	Diana Yunita	Female	3,48	3,61	3,46	3,39	3,79	3,57	On time
3	Ade Anggraini	Female	4	3,91	3,75	3,7	3,79	3,85	On time
4	Dimas Andrean	Male	3,57	3,07	2,9	3	2,76	3	On time
5	Randi Alan Fikar	Male	3,13	2,74	2,7	2,8	2,62	3	On time
6	Irwandi	Male	2,91	2,93	2,87	2,84	2,6	2,79	On time
7	Ike Pramida	Female	2,87	2,93	2,97	3,03	2,8	2,9	Not on time
8	Imam Frasdiansyah	Male	3	3,04	2,96	2,97	2,71	2,7	Not on time
…	…	…	…	…	…	…	…	…	…
50	Salima Barutu	Female	3,68	3,04	3,04	3,61	3,32	3,33	On time

The following are 9 data from 50 data from Jambi Muhammadiyah University students who have graduated, which will be used as training data.

Table 2. Data Testing

No	Name	Gender	S1	S2	S3	S4	S5	S6	Information
1	Meilani Putri	Female	3,44	3,43	3,39	3,4	3,36	3,7	On time
2	Jessica Azra	Female	3,45	3,7	3,75	3,8	3,89	3,9	On time
3	Tri Erma Vitarani	Female	3,82	3,83	3,74	3,88	3,9	4	On time
4	Wahyu Saputri	Female	3,88	3,89	3,85	3,91	3,9	3,92	On time
5	Aldi Nurmansyah	Male	3,70	3,72	4	3,50	3,94	3,9	On time
6	Wahyudi Pengestu	Male	3,42	3,65	3,5	3,78	3,80	3,85	On time
7	Sartika Apriyanti	Female	3,51	3,49	3,32	3,34	3,36	3,88	On time
8	Hayatul Munawaroh	Female	3,94	3,89	3,8	3,76	3,79	3,9	On time
…	…	…	…	…	…	…	…	…	…
50	Silvy Gusmiarti	Female	3	3,16	3,68	3,67	3,29	3,64	On time

The following are 9 data from 50 semester VI students of Muhammadiyah Jambi University who have not graduated used as testing data.

2.2. K-Nearest Neighbor Algorithm

The K-Nearest Neighbor algorithm is a classification method that groups new data based on the distance of the new data to some of the closest data/neighbors. The K-Nearest Neighbor technique with the following steps [2]:

1. For all testing data, calculate the distance to each training data
2. Determine the k training data closest to the data
3. testing
4. Examine the labels of this k data
5. Determine the label with the most frequency
6. Enter testing data into the class with the most frequency
7. Stop

$$d_i = \sqrt{\sum_{i=1}^{p}(x_{2i-x_{1i}})^2} \qquad (1)$$

2.3. Rapid Miner Implementation

Rapid Miner is a software/data processing device. In this study, the authors plan to use Rapid Miner to test data from the calculation results of the K-NN method. Data mining is a process that employs one or more computer learning techniques to analyze and extract knowledge automatically [21].

3. RESULTS AND ANALYSIS

In this study, the K-Nearest Neighbor (K-NN) method was applied to predict the graduation time for Jambi Muhammadiyah University students. In processing data and testing the K-NN method, the author uses RapidMiner software for its implementation. In this process, there are 3 operators: External data reader from Microsoft Excel Read, K-NN for the implementation of the classification, and Validation. After creating the K-NN Classification model in the Validation operator, it is divided into 2: the training data and the data testing section. The training data section only has the K-NN operator, while the data testing section has 2 Apply Model and Performance operators.

3.1. Performance Vector

Calculation of the closeness of old cases in the training data with new cases in the testing data, it is known that out of 50 data records, 13 are classified as correct, 0 data is classified as correct but turns out not to be, 1 data class is not classified accordingly, and 1 data is classified as incorrect. The accuracy rate of the application of the k-nearest neighbor algorithm is 93.33%, precision 100.00% and recall 92.86%.
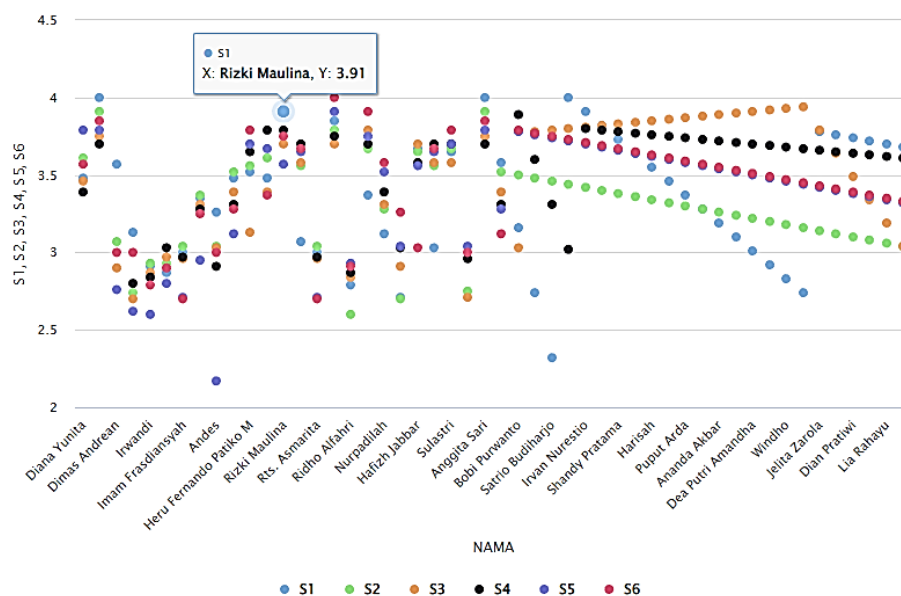
3.2. Visualizations of Timely Graduation Data Distribution

In providing information from the data set processed on the Rapid Miner tool, the following is a graphic display that can be seen based on the value of k 5 to provide an overview of class classification [22]. Graphical visualization of the distribution of graduation on time and not on time (figure 2).

3.3. Design Sub Data Testing Design Based on Training Data

The design of modelling design to carry out tests on data testing based on training data that has been tested before to determine the level of accuracy as a basis. By modeling using rapidminer using 3 operators, namely, external data reader from Microsoft Excel Read, K-NN to implement the classification using K-NN with K=5, Apply Model to apply a previously drilled model using training data on unlabeled data (data

testing). Which produces data that can be seen in table 4 below with predicted results of 46 students graduating on time and four students not graduating on time with an accuracy of 93.33%.

Figure 2. Visualizations of the Distribution of Graduation on Time

3.4. Test Result

The following picture describes that the results of the timely and not timely graduation examinations are known to be in the names of Anisa Dalilah Amsi, Patriana, Siti Murniati, and Ricka Audri Marshella. Did not graduate on time.

4. CONCLUSION

This research was conducted by implementing the K-Nearest Neighbor algorithm model to predict the graduation of Jambi Muhammadiyah University students. Several processes are carried out to get quality data before being implemented into the algorithm. Four variables are used: name, gender, GPA from semesters 1 to 6 and description. In this test, k = 5 was used, and the accuracy reached 93.33%. After going through a series of processes, it was then continued by testing the data of students who had yet to pass as many as 50 data to be tested, and the conclusion was that 4 students graduated not on time and 46 students who graduated on time.

REFERENCES

[1] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax J. Inform.*, vol. 10, no. 01, pp. 46–56, 2021, doi: 10.35706/syji.v10i01.5173.

[2] S. Sumarlin and D. Anggraini, "Implementasi K-Nearest Neighbord Pada Rapidminer Untuk Prediksi Kelulusan Mahasiswa," *High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 10, no. 1, pp. 35–41, 2018, doi: 10.52972/hoaq.vol10no1.p35-41.

[3] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technol. J. Ilm.*, vol. 13, no. 3, p. 207, 2022, doi: 10.31602/tji.v13i3.7312.

[4] T. Nasution, "Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu," *J. Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020.

[5] K. Kartarina, N. K. Sriwinarti, and N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 107–113, 2021, doi: 10.35746/jtim.v3i2.159.

[6] Humas Um Jambi, "Gelar Wisuda ke IV, Universitas Muhammadiyah Jambi Sarjanakan 327 Mahasiswa," *Website*, 2023. https://umjambi.ac.id/gelar-wisuda-ke-iv-universitas-muhammadiyah-sarjanakan-327-mahasiswa/

[7] I. L. Putra, "Implementasi Algoritma Particle Swarm Optimization(Pso) Dan K-Nearest Neighbor(K-

         Nn) Dalam Memprediksi Keberhasilan Anak Smk Mendapatkan Kerja," *Technol. J. Ilm.*, vol. 13, no. 4, p. 339, 2022, doi: 10.31602/tji.v13i4.8167.

[8] S. Nasional, T. Elektro, S. Informasi, and T. Informatika, "Prediksi Kelulusan Mahasiswa dengan Metode K-Nearest Neighbor pada Jurusan Sistem Informasi Institut XYZ," pp. 199–206, 2023.

[9] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *J. Eng. Appl. Technol.*, vol. 2, no. 2, pp. 86–91, 2021, doi: 10.21831/jeatech.v2i2.42777.

[10] G. Mahalisa and Arminarahmah, "Diabetes Classification Analysis Using the Euclidean Distance Method Based on the K-Nearest Neighbors Algorithm," *J. Teknol. Komput. dan Inf.*, vol. 5, no. 3, pp. 178–182, 2022.

[11] A. Rohman and A. Rufiyanto, "Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran," *Proceeding SINTAK 2019*, pp. 134–139, 2019.

[12] R. H. R. Hutagaol, B. A. B. Ardiansyah, I. D. I. Daulay, and ..., "Algoritma K-Means dan Decision Tree untuk Prediksi Penerimaan Calon Mahasiswa Pascasarjana pada Universitas Indonesia: K-Means and Decision Tree …," *SENTIMAS Semin. ...*, pp. 154–161, 2022, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas/article/view/339%0Ahttps://journal.irpi.or.id/index.php/se ntimas/article/download/339/126

[13] D. D. S. Fatimah and E. Rahmawati, "Penggunaan Metode Decision Tree dalam Rancang Bangun Sistem Prediksi untuk Kelulusan Mahasiswa," *J. Algoritm.*, vol. 18, no. 2, pp. 553–561, 2022, doi: 10.33364/algoritma/v.18-2.932.

[14] M. Hidayat, A. Faqih, and T. Suprapti, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOUR UNTUK PREDIKSI KETEPATAN KELULUSAN," *JURSIMA (Jurnal Sist. Inf. dan Manajemen)*, vol. 10, no. 1, 2022.

[15] R. Sepriansyah, S. D. Purnamasari, K. R. N. Wardani, and N. Halim, "Prediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Bina Darma Menggunakan Algoritma Naïve Bayes," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 1, pp. 313–322, 2023, doi: 10.29100/jipi.v8i1.3459.

[16] F. T. Hendrik Saputra, "Penggunaan Metode Fuzzy K–Nearest Neighbor Pada Sistem Prediksi Tingkat Kelulusan Mahasiswa," *J. Dunia Bisnis*, vol. 1, no. 1, pp. 1–11, 2021, [Online]. Available: http://duniabisnis.org/index.php/duniabisnis/article/view/9

[17] T. Nasution, "Implementasi Algoritma K-Nearest Neighbor Untuk Penentuan Kelulusan Mahasiswa Tepat Waktu," *J. Perangkat Lunak*, vol. 2, no. 1, pp. 1–14, 2020, doi: 10.32520/jupel.v2i1.944.

[18] A. Sabathos Mananta and G. Arther Sandag, "Prediksi Kelulusan Mahasiswa Dalam Memilih Program Magister Menggunakan Algoritma K-NN," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 10, no. 2, pp. 90–96, 2021, doi: 10.30591/smartcomp.v10i2.2488.

[19] H. Rohayani, Rico, K. Kurniawansyah, and A. Harahap Saputra, "Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Sains Dan Teknologi Menggunakan Metode Naive Bayes," *J. V-TECH (VISION Technol. e- ISSN 2622-9315*, vol. 5, no. 2, 2022.

[20] H. Rohayani, S. Alam Nur, M. Fauzi, and Rico, "Prediksi Penyebaran Virus COVID-19 Dari Hasil PCR Menggunakan Metode Naïve Bayes," *J. Comput. Syst. Informatics*, vol. 4, no. 1, pp. 109–115, 2022.

[21] C. Anisa and Andri, "Penerapan Algoritma K-Nearest Neighbord Untuk Prediksi Penjualan Obat Pada Apotek Kimia Farma ATMO Palembang," *Bina Darma Conf. Comput. Sci.*, vol. 2, no. 3, pp. 199–208, 2021.

BIBLIOGRAPHY OF AUTHORS

Shandy Amandha is an active student in the Informatics Study Program, Faculty of Science and Technology, Muhammadiyah University, Jambi.

Dr. Hetty Rohayani. ST, M.Kom is an active lecturer at the Informatics Study Program, Faculty of Science and Technology, Muhammadiyah University Jambi. As a writer and researcher who focuses on the fields of artificial intelligence, data mining, decision support systems and expert systems.

Kevin Kurniawansyah, S.Kom., M.Kom currently serves as a lecturer at the Informatics Study Program, Faculty of Science and Technology, Muhammadiyah University, Jambi. The author's research interests lie in intelligent systems, application of artificial intelligence methods or algorithms and technology development