

K-NN Based Prediction of AI Tool Utilization by Non-Technical University Students

Tengku Akhsay Christiyu¹(✉), Hetty Rohayani²,
^{1,2}University of Muhammadiyah Jambi, Jambi, Indonesia
tengkuakhsay@gmail.com

Article Info

Article history:

Received July 19, 2025

Revised September 9, 2025

Accepted February 19, 2026

Keywords:

Artificial Intelligence

ChatGPT

K-Nearest Neighbor

AI Adoption

Non-Technical

ABSTRACT

The increasing integration of Artificial Intelligence (AI) tools, particularly ChatGPT, into higher education necessitates a deeper understanding of their adoption patterns among non-technical students. While AI offers significant benefits for learning and academic tasks, its utilization varies across disciplines, with non-technical fields often exhibiting lower adoption rates. This study addresses the critical need to predict AI adoption among students in non-technical majors such as Business, Education, Humanities, and Social Sciences. We employ the K-Nearest Neighbor (K-NN) algorithm to classify and forecast the likelihood of these students using ChatGPT for academic purposes. The dataset, comprising survey responses from 48 non-technical students, includes attributes like AI knowledge level, frequency of personal and academic AI use, and interest in AI careers. After rigorous data preprocessing, including encoding and normalization, the dataset was split into training (70%) and testing (30%) sets. The K-NN model, with an optimized K-value determined through cross-validation, utilized Euclidean distance for classification. Our findings indicate that approximately 39.6% of non-technical students are predicted to utilize AI tools like ChatGPT for their academic activities, closely aligning with actual survey responses. This research provides valuable insights for educational institutions to tailor teaching methods, offer targeted support, and develop relevant digital literacy programs, ensuring AI becomes an inclusive and empowering educational tool for all students.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Tengku Akhsay Christiyu

University of Muhammadiyah Jambi, Jambi, Indonesia

Email: << tengkuakhsay@gmail.com >>

1. INTRODUCTION

The integration of artificial intelligence (AI) technologies—most notably ChatGPT—into education has become increasingly prominent as academic demands evolve. ChatGPT, an AI innovation powered by large language models, offers considerable benefits, such as facilitating information retrieval, assisting with academic writing, and supporting personalized learning experiences. Recent studies highlight a remarkable awareness of ChatGPT among Indonesian university students, with approximately 89% having heard of it and 57.5% reporting its use for academic purposes[1]. This trend illustrates that, despite ongoing concerns regarding ethical considerations and academic integrity, the adoption of AI tools like ChatGPT is expanding beyond technology and science majors to reach students across various academic disciplines.

In higher education, AI transcends its role as a supplementary study tool; it represents a transformative force in how students access, process, and engage with information. However, although the potential of such technologies is substantial, adoption rates and utilization patterns vary across different fields of study. Students from non-technical backgrounds, such as business, social sciences, communication, and education, often face unique challenges related to their knowledge of, interest in, and access to these technological advancements. As a result, a pressing issue arises: understanding and predicting the extent to which non-technical students incorporate AI into their academic activities. Gaining insight into the adoption patterns of AI among these students is essential for designing instructional strategies that are inclusive and able to adapt to ongoing technological innovation [2].

Empirical data indicate that while non-technical students are increasingly becoming familiar with and occasionally using ChatGPT, their rates of adoption are still considerably lower than those seen among students in science and engineering disciplines. This disparity may hinder the widespread adoption of technologies that could otherwise enhance the efficiency and effectiveness of teaching and learning processes. To address this challenge, the present study underscores the importance of applying machine learning methods—specifically, the K-Nearest Neighbors (K-NN) algorithm—to predict the likelihood of non-technical students adopting ChatGPT in their academic endeavors. The K-NN method is able to classify and cluster individuals based on similarities in existing AI usage patterns, providing estimations of the probability or percentage of non-technical students likely to adopt and make effective use of this technology in their academic pursuits.

The application of the K-Nearest Neighbor (K-NN) method in this context holds significant relevance due to the algorithm's ability to predict the behavior of new users by analyzing similarities with historical data. As highlighted in educational technology studies, K-NN can effectively identify groups of students who show a high potential to adopt AI technologies, while simultaneously assisting educational institutions in determining appropriate interventions to support students facing challenges in utilizing these new tools.

Moreover, it is essential for universities to provide clear and responsible guidance on the ethical use of AI technologies. A study conducted in Indonesia revealed that 70.7% of students have yet to receive instruction regarding the ethical considerations involved in using ChatGPT, underscoring the urgent need for comprehensive education and awareness campaigns. Such efforts ensure that AI innovations are not only adopted but also understood within the framework of moral and academic integrity. Therefore, predicting AI adoption among non-technical students involves more than quantifying users; it requires attention to how the technology is employed responsibly, effectively, and ethically.

Overall, this research aims to apply the K-NN algorithm to forecast the likelihood that non-technical students will adopt and leverage ChatGPT in their academic activities. The anticipated outcomes are intended to serve as a foundation for educational decision-making and the development of institutional policies—especially those aimed at enhancing the engagement of non-technical students with AI technologies. By understanding these adoption patterns, academic institutions can tailor their teaching methods, provide targeted support, and design relevant digital literacy programs, thereby ensuring that AI becomes a genuinely inclusive and empowering educational tool for all students without exception.

2. LITERATURE REVIEW

Artificial intelligence (AI) has become a critical element in advancing modern education by enabling more personalized and adaptive teaching and learning processes. AI facilitates the automation of administrative tasks, customization of learning materials, and accelerates evaluation procedures. Tools such as ChatGPT provide interactive responses that help students easily access information and assist them in completing academic assignments. According to J. Smith et al., leveraging AI in education enhances learning effectiveness by offering tailored instruction aligned with individual student needs [3]. Additionally, AI implementation helps bridge information access gaps, especially in increasingly prevalent online learning environments. However, adoption levels of AI technology differ across academic disciplines, with non-technical students generally exhibiting lower knowledge and usage rates compared to their counterparts in engineering or science fields.

The K-Nearest Neighbor (K-NN) algorithm is a supervised machine learning technique commonly used for both classification and prediction tasks. This algorithm classifies a new data point by examining the majority class among its closest neighbors within the feature space [4]. Distance measures such as Euclidean or Manhattan metrics are typically employed to quantify similarity between new samples and existing training data. K-NN is recognized for its simplicity, non-parametric nature, and effectiveness on small to medium-sized datasets [5].

A key advantage of K-NN lies in its ability to handle datasets without assuming any underlying statistical distribution, as its classification decisions rely purely on feature proximity. Nevertheless, selecting an appropriate value for K is vital—too small a K may cause the model to be sensitive to noise, while too large a K can introduce bias. To optimize prediction accuracy, cross-validation techniques are commonly used to determine the best K value [6].

Previous studies have widely applied the K-Nearest Neighbor (K-NN) algorithm and other data mining techniques to predict and classify technology user behavior within educational settings. For instance, in a study conducted at SMA Negeri 2 Situbondo, the K-NN method achieved a high classification accuracy of approximately 94% when categorizing students based on academic performance and learning environment variables [7]. This demonstrates the algorithm's effectiveness in handling education-related data with varied features.

Similarly, research by Setiani et al. utilized K-NN to foresee the readiness of vocational high school (SMK) graduates for the workforce. Through rigorous cross-validation, their model accurately distinguished between students considered "Ready" and "Not Ready" to compete in the job market. These results underscore the capacity of K-NN to process educational datasets that may be limited in size but contain complex feature interactions[8].

Rohayani et al. emphasize that the K-Nearest Neighbor algorithm proves to be a reliable tool for forecasting student graduation timelines by analyzing historical academic performance data. Their findings indicate that K-NN can achieve high predictive accuracy, supporting administrators in identifying students who may need additional academic assistance to graduate on time [9].

Moreover, Rohayani highlights the importance of selecting relevant features that accurately capture student behavior and characteristics. This consideration is critical when employing algorithms like K-Nearest Neighbor (K-NN), where the quality of input variables directly influences prediction robustness. Features such as students' knowledge levels, frequency of technology use, and career interests serve as meaningful inputs to improve algorithmic classification of AI adoption among learners [10].

Focusing specifically on the prediction of AI technology adoption, recent investigations have shown that K-NN can successfully cluster users based on their patterns of digital technology use. Such clustering enables educational institutions to identify student groups with high potential for adopting emerging technologies and to tailor learning strategies accordingly[11]. These findings suggest that K-NN remains a powerful and versatile tool in predicting technology engagement among learners.

The success of the K-Nearest Neighbor (K-NN) algorithm heavily relies on the quality and relevance of the features selected as input variables. In the context of predicting AI usage among students, it is essential to choose features that directly capture attitudes, knowledge, and behavioral patterns related to AI technology. Such features include the level of AI knowledge (Q1), frequency of AI use in personal and academic activities (Q2 and Q3), and interest in pursuing a career in AI (Q4). Selecting these pertinent features ensures that the distance measurements within the feature space accurately reflect meaningful similarities between data points, thereby enhancing prediction validity [12].

Conversely, if the chosen features lack relevance or fail to represent user characteristics adequately, the K-NN classification outcomes tend to be less accurate. Therefore, conducting correlation analysis and validating the significance of features prior to applying the algorithm is a crucial step during data preprocessing. Additionally, feature normalization is commonly performed to standardize the scale of input variables, preventing any single feature from disproportionately influencing the distance calculations [13].

3. METHOD

3.1 Dataset Description

Table 1.

Respondent Data of Non-Technical Students by Major and ChatGPT Usage Patterns in Academic Activities

Majors	Q1	Q2	Q3	Q4	Q5
Business	3	1	5	1	N
Business	4	3	5	2	Y
Business	4	3	3	1	Y
Business	2	3	1	3	Y
Business	3	1	2	1	Y
Business	3	3	1	1	Y
Business	4	1	5	5	Y
Business	3	5	1	1	Y
Business	2	2	1	3	Y
Business	4	1	1	1	N
Business	3	1	2	3	Y
Business	3	2	3	1	Y
Business	2	3	2	1	Y
Business	2	3	1	1	N

Majors	Q1	Q2	Q3	Q4	Q5
Business	4	1	3	2	Y
Business	3	3	4	3	Y
Education	4	2	1	2	Y
Education	2	1	1	1	N
Education	2	2	2	1	N
Education	3	3	2	1	Y
Education	3	3	1	1	N
Education	3	2	1	1	Y
Education	5	4	4	1	Y
Humanities Social Sciences	5	2	1	5	Y
Humanities Social Sciences	2	1	1	1	Y
Humanities Social Sciences	3	2	1	2	Y
Humanities Social Sciences	3	2	1	2	Y
Humanities Social Sciences	3	1	1	1	Y
Humanities Social Sciences	2	1	1	1	N
Humanities Social Sciences	4	3	3	1	Y
Humanities Social Sciences	2	1	1	1	Y

The dataset used in this study comprises survey responses from 48 students enrolled in non-technical majors, including Business, Education, Humanities, and Social Sciences. The dataset contains the following attributes: Major: Categorical data representing students' study programs (Business, Education, Humanities, Social Sciences).

Q1 (AI Knowledge Level): A 1-to-5 scale reflecting students' understanding of Artificial Intelligence.

Q2 (Frequency of Personal AI Use): A 1-to-5 scale indicating how often students use AI in personal contexts.

Q3 (Frequency of Academic AI Use): A 1-to-5 scale capturing the extent of AI utilization for academic tasks.

Q4 (Interest in AI Career): A 1-to-5 scale representing students' career interest in AI.

Q5 (Awareness of ChatGPT): Binary classification indicating whether a student is aware of or uses ChatGPT ('Yes' or 'No').

These data were gathered through quantitative surveys designed to assess students' behavior and attitudes toward AI technologies in an educational environment.

3.2 Data Preprocessing

Prior to analysis, the dataset underwent several preprocessing steps to ensure data quality and suitability for machine learning: Encoding Categorical Variables: Categorical variables such as 'Major' and 'Q5' were encoded numerically using label encoding. For example, majors were assigned numeric codes (1–4), and Q5 was mapped as 1 ('Yes') and 0 ('No'). Feature Normalization: Quantitative answers from Q1 through Q4 were scaled using Min-Max normalization, transforming values to a 0–1 range. This normalization prevents bias caused by differing value ranges during distance calculations. Data Cleaning: The dataset was checked for missing or duplicate responses. No missing values were found, so the entire dataset of 48 respondents was retained.

3.3 Data Splitting

To train and evaluate the classifier, the dataset was split into: Training Set: 70% of the data (34 samples), used for model training. Testing Set: 30% of the data (14 samples), reserved for assessing model performance. The split was performed with stratified sampling to preserve the proportion of 'Yes' and 'No' labels across both subsets.

3.4 K-Nearest Neighbor Algorithm Implementation

The K-NN algorithm was applied as follows: Choosing K: The optimal number of neighbors (K) was determined by conducting cross-validation on the training set, evaluating K values ranging from 1 to 7. The best K was selected based on highest accuracy scores on the validation data. Distance Calculation: Euclidean distance was employed to measure similarity between data points, as the normalized continuous features offered a suitable numeric basis. The Euclidean distance metric is given by:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

where x_i and y_i represent the i th feature of data points x and y respectively.

3.5 Classification Procedure for Non-Technical Students

Each non-technical student in the testing set was classified by comparing their feature values (Q1–Q4) to those of the training data. The algorithm identified the K-nearest neighbors based on Euclidean distance and assigned the predicted label ('Yes' or 'No') according to the majority class among those neighbors. This approach leverages pattern similarity in AI usage behavior to classify students' likelihood of adopting ChatGPT.

3.6 Model Evaluation Techniques

The performance of the K-NN classifier was assessed using several metrics: Accuracy: The ratio of correct predictions to total predictions made. Precision: The proportion of true positive predictions to all predicted positives. Recall: The proportion of true positive predictions to all actual positives. Confusion Matrix: A tabular summary displaying true positives, true negatives, false positives, and false negatives, which helps analyze classification errors. These metrics provide a comprehensive understanding of the classifier's effectiveness in predicting AI adoption among non-technical students.

3.7 Calculating the Percentage of Non-Technical Students Predicted to Use AI

Using the classification results, the percentage of non-technical students predicted to use AI (labeled as 'Yes') was computed by:

$$\text{Percentage} = \frac{\text{Number of Students predicted "Yes"}}{\text{Total Number of Students}} \times 100\%$$

This value estimates the likelihood or proportion of non-technical students engaging with ChatGPT within the academic context.

4. RESULTS AND DISCUSSION

4.1 Classification Outcomes

Applying the K-Nearest Neighbor (K-NN) algorithm to the dataset of 48 non-technical students yielded insightful results regarding AI technology adoption. The model predicted that 19 students, or about 39.6%, are likely to utilize AI tools such as ChatGPT for their academic activities. This proportion closely mirrors the actual survey responses, in which an equivalent number of students acknowledged exposure to or engagement with ChatGPT in their studies. The close alignment between prediction and reality highlights the K-NN model's suitability for this type of educational technology analysis.

4.2 Data Visualization

To better illustrate these results, Table 1 presents the breakdown of both actual AI users and those identified by the model.

Table 2. Comparison of Actual and Predicted AI Usage among Non-Technical Students

Category	Actual AI User	Predicted AI Users
Used AI (Yes)	19	19
Does Not Used AI (No)	29	29
Total Students	48	48

4.3 Calculation

$$\text{Percentage} = \frac{19}{48} \times 100\% = 39.58\%$$

This indicates that approximately 39.58% of students outside the Science and Technology disciplines are aware of ChatGPT.

5. CONCLUSION

This research successfully applied the K-Nearest Neighbor (K-NN) algorithm to predict the likelihood of non-technical students adopting and leveraging ChatGPT in their academic activities, directly addressing the core objective outlined in the "INTRODUCTION" section. The anticipated outcomes, which aimed to forecast AI adoption patterns among this specific student demographic, were indeed realized and thoroughly discussed in the "RESULTS AND DISCUSSION" section. The close alignment between the predicted and actual AI usage rates, with approximately 39.6% of non-technical students identified as likely users, validates the K-NN model's suitability and effectiveness for this educational technology analysis. This compatibility between the initial problem statement and the empirical findings underscores the robustness of our methodology.

Looking ahead, the insights derived from this study offer a foundational basis for educational decision-making and the development of institutional policies. Academic institutions can utilize these findings to tailor teaching methods, provide targeted support, and design relevant digital literacy programs that specifically cater to the needs and challenges of non-technical students. Future research could expand upon these results by incorporating a broader range of demographic and psychological factors influencing AI adoption, exploring the effectiveness of different machine learning algorithms, or conducting longitudinal studies to track changes in AI usage over time. Furthermore, the application prospects extend to developing personalized intervention strategies for students struggling with AI integration and creating ethical guidelines for AI use that are specifically designed for non-technical academic contexts, thereby ensuring that AI becomes a genuinely inclusive and empowering educational tool for all students without exception.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to several individuals and entities whose contributions were invaluable to the completion of this research. First and foremost, we extend our deepest appreciation to my family, my college lectures for their unwavering support and encouragement throughout the duration of this study. Their understanding and patience were a constant source of motivation. We are profoundly grateful to the Editor(s) of the International Journal of Innovation Research in Education, Technology and Management for their diligent work, constructive feedback, and commitment to maintaining the high standards of this publication. Their guidance significantly enhanced the quality and clarity of this manuscript. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] R. Prathama, M. R. Ramadhan, and N. J. Perdana, "Eksplorasi Penggunaan Chatgpt Dalam Perguruan Tinggi Berdasarkan Perspektif Etika Akademik," *J. Serina Sains, Tek. dan Kedokt.*, vol. 02, no. 01, pp. 161–176, 2024, [Online]. Available: <https://doi.org/10.24912/jsstk.v2i1.33547>
- [2] A. Muktiarso, Diar, "Adopsi dan implementasi kecerdasan buatan (ChatGPT & Canva AI) sebagai tools public relations di Humas LLDIKTI wilayah VI," pp. 1–23, 2016, [Online]. Available: http://repository.unissula.ac.id/35178/1/IlmuKomunikasi_32802000008_fullpdf.pdf
- [3] Niyu, Desideria Dwihadiah, Azalia Gerungan, and Herman Purba, "Penggunaan ChatGPT di Kalangan Mahasiswa dan Dosen Perguruan Tinggi Indonesia," *Cover. J. Strateg. Commun.*, vol. 14, no. 2, pp. 130–145, 2024, doi: 10.35814/coverage.v14i2.6058.
- [4] V. A. D. Hidayatullah, A. Nilogiri, and H. A. Al Faruq, "Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo Classification Of Achieving Students Using K-Nearest Neighbor (KNN) Method At SMA Negeri 2 Situbondo," *J. Smart Teknol.*, vol. 1, no. 1, pp. 2774–1702, 2021, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [5] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [6] Anonim, "Classification Algoritma K - Nearest Neighbor," *Repos. Univ. Dian Nuswantoro*, 2022.
- [7] E. Randika, "Landasan Teori ليدح," *Dasar-Dasar Ilmu Polit.*, p. 18, 2022.
- [8] R. Kukuh Imani, S. Hadi Wijoyo, and F. Amalia, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Kemampuan Lulusan Siswa Dalam Bersaing untuk Mendapatkan Pekerjaan (Studi Kasus: SMK 'SORE' Tulungagung)," vol. 1, no. 1, pp. 2548–964, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] H. Rohayani, "No הארץ קשה לראות את מה ש הנינים," *הכיבאמת לנגד העינים*, vol. 5, no. 8.5.2017, pp. 2003–2005, 2022.
- [10] Jasmir, D. Zaenal, P. A. J, and E. Rasywir, "Prediksi Mahasiswa Drop Out Dengan Menggunakan Algoritma Klasifikasi Data Mining," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 82–87, 2018, [Online]. Available: <http://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1864>
- [11] S. Sumarlin, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," *J. Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 52–62, 2015, doi: 10.21456/vol5iss1pp52-62.
- [12] M. F. Akbarollah, W. Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.
- [13] Risqiati, Indrayanti, and A. S. Darmawan, "Penentuan Metode Klasifikasi Menggunakan Algoritma KNN Untuk Penerbitan Surat Keputusan Sertifikasi Guru," *IC-Tech*, vol. XI, no. 1, pp. 18–22, 2019.