E-ISSN: 2962-6110
P-ISSN: 2964-7843
JSRET
jsret.knpub.com
Journal of Scientific Research, Education, and Technology

K-Nearest Neighbor–Based Recommendation System for Informatics Student Concentration

Puja Mei Siswanti, Sutarman
Yogyakarta University of Technology, Yogyakarta, Indonesia

ABSTRACT

The selection of a study concentration is a critical stage in a student's academic journey, as it determines their future expertise. However, many students struggle to choose a path that aligns with their abilities. This study aims to develop a recommendation system for Informatics students using the K-Nearest Neighbor (KNN) algorithm. The dataset consists of 446 student academic records from semesters 1 to 4. The classification used an 80:20 split for training and testing data. Experimental results show that the KNN model with $\boldsymbol{k = 15}$ achieved an optimal accuracy of 74.44%. These findings demonstrate that KNN is effective in classifying students into Web and Mobile (WEM) or Intelligent Systems (SCR) concentrations based on academic patterns. Beyond its technical performance, the system serves as a web-based decision-support tool designed to enhance academic advising efficiency and provide institutional benefits by supporting student success. While the system offers objective, data-driven recommendations, it is intended to complement personal interests and professional goals. This research provides a structured approach for institutions to assist students in making more informed academic decisions.

Keywords: *Student Concentration, K-Nearest Neighbor, Recommendation System.*

Corresponding author
Name: Puja Mei Siswanti
Email: pujasiswanti@gmail.com

INTRODUCTION

The selection of a study concentration is a critical aspect of a student's academic journey, as it shapes the direction of specialization and determines the relevance of graduates' competencies in the workforce (Hidayat et al., 2025). In the Informatics Study Program, students are offered several concentration options, namely Web and Mobile (WEM) and Intelligent Systems (SCR). The concentration selection process is conducted in the fifth semester, after students have completed the prerequisite courses. However, determining a concentration is not easy for most students, as many of them do not yet fully understand their interests and abilities. As a result, some students struggle to choose the appropriate course concentration, which creates obstacles for them to graduate on time, as they must retake courses in a different concentration (Raharjo et al., 2022). The alignment of a chosen study concentration can influence students' learning motivation and

their ability to achieve strong academic performance (Verawati, 2015). Sometimes, students who have difficulty recognizing their own abilities and strengths tend to follow the choices made by their peers (Verawati, 2015). Given these challenges, it is essential to develop a system that can guide students in selecting the concentration or field of study that aligns best with their academic strengths. In many educational institutions, the process of selecting a study concentration is often carried out subjectively and lacks structured support, which may lead to a mismatch between students' interests, abilities, and labor market demands (Hidayat et al., 2025). Therefore, this study aims to develop a recommendation system using the K-Nearest Neighbor (KNN) algorithm to provide personalized concentration suggestions based on students' academic performance. A recommendation system is a system designed to provide useful information or suggest actions that a user is likely to take to achieve their goals, such as selecting a particular product (Fasya et al., 2025). In this context, the system assists students in making informed decisions by analyzing their academic records and identifying patterns similar to those of other students. The K-Nearest Neighbor (K-NN) algorithm, which is widely used in machine learning, enables this process by classifying data based on the proximity or distance between a data point and other points in the dataset (Zaidah et al., 2025).

Research on decision support systems for determining students' concentrations has been conducted by several previous studies. A study by Hidayanti et al. (2020) utilized the C4.5 and Naïve Bayes algorithms to compare the accuracy levels in classifying student concentrations. The results showed relatively low accuracy, 48.06% and 42.79% respectively, and did not develop a support system that could be directly implemented. Meanwhile, Wibowo et al. (2024) employed the K-Nearest Neighbor (KNN) algorithm to assist in determining students' concentrations; however, the dataset used was still limited, resulting in suboptimal accuracy. Similarly, the study by Prasetyo et al. (2019) although successfully applying KNN for concentration recommendations, used a relatively small dataset and the system has not yet been implemented practically. Based on these three studies, a research gap can be identified in the need for a KNN-based recommendation system with a larger dataset that can be practically implemented by students and the study program. Therefore, this study focuses on applying the KNN algorithm with an increased dataset and developing a web-based system to provide more accurate recommendations for students' concentrations.

The primary objective of this research is to develop and evaluate a web-based recommendation system that utilizes the K-Nearest Neighbor (KNN) algorithm to assist Informatics students in selecting a study concentration based on their academic performance. To achieve this objective, the study addresses the following research questions:

1. How does the K-Nearest Neighbor algorithm perform in terms of accuracy when classifying student concentrations using a larger dataset?
2. What is the optimal value of k that provides the most reliable recommendations for students?

3. To what extent can the developed web-based system provide practical decision support for both students and academic advisors?

By answering these questions, this study is expected to provide significant benefits. For students, it reduces the risk of academic mismatch, while for the study program, it serves as a data-driven tool to enhance academic advising and institutional efficiency.

METHOD

The research employs a quantitative method with a supervised learning approach to develop a recommendation system for student concentration selection, as K-Nearest Neighbors (K-NN) is categorized under supervised learning algorithms (Zaidah et al., 2025). The method used in this study is the K-Nearest Neighbor (KNN) algorithm, which belongs to distance-based classification methods. The K-Nearest Neighbor (KNN) algorithm is suitable for this recommendation system because it operates based on similarity between data instances. Its simplicity, interpretability, and effectiveness in handling numerical academic data make KNN appropriate for identifying patterns in students' academic performance and generating concentration recommendations. To provide a comprehensive overview of the research process, the study was carried out through several interconnected stages, ranging from a literature review to system evaluation. Each stage was designed to support the subsequent steps and to achieve the research objectives effectively. The overall flow of the research process can be seen in figure 1.

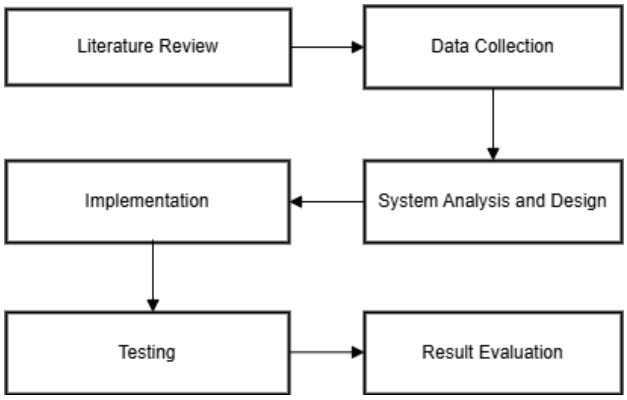

Figure 1. Research stage flow

Based on the figure above, the research process begins with a literature review to examine relevant theories, previous studies, and references, providing a solid foundation for selecting the appropriate method and developing the system (Zaidah et al., 2025).

The next stage, data collection, involves gathering information from each student. Ethical considerations were strictly applied in this study. Student academic data were anonymized by removing personal identifiers to protect privacy. The data were used solely for research purposes and analyzed without revealing individual identities. The data used in this study consist of academic grades of Informatics Study Program students from

semesters 1 to 4. A total of 446 students from the 2022 cohort were included, covering two concentrations: Web and Mobile (WEM) and Intelligent Systems (SCR). The attributes used in this study are detailed in Table 1.

Table 1. Research data structure

Attribute	Description
Name	Respondent Identity (anonymized)
Concentration	Web and Mobile (WEM) / Intelligent Systems (SCR)
Course 1- Course 20	Grades of 20 compulsory courses from semesters 1–4

Each student has 20 attributes representing the grades of compulsory courses, which are used as features in the classification process. The list of courses used can be seen in table 2.

Table 2. List of course

NO	Courses	Semester
1	Programming Algorithms	1
2	Programming Algorithms Practice	1
3	Linear Algebra	2
4	Computer Architecture & Organization	2
5	Database Practice	4
6	Databases	3
7	Calculus 2	2
8	Artificial Intelligence	3
9	Mathematics & Statistics	1
10	Discrete Mathematics	4
11	Object-Oriented Programming Practice	3
12	Web Programming	3
13	Web Programming Practice	3
14	Web and Mobile Development	1
15	Software Engineering	4
16	Number Systems & Information Logic	1
17	Operating Systems	3
18	Statistics & Probability	2
19	Data Structures & Algorithms	2
20	Data Structures Practice	2

Before performing the classification, the students' academic grade data are preprocessed to ensure their quality and suitability for analysis. In this stage, the data are initially processed to prepare them for classification using the K-Nearest Neighbor (KNN) algorithm. Data preprocessing is an initial stage in data processing that involves a series of

steps to clean, organize, and prepare the data for further analysis (Khasanah et al., 2025). The first step involves removing attributes that do not affect the classification, specifically the students' name column. Next, letter grades are converted into numerical values according to the following scheme: A = 4, B = 3, C = 2, D = 1, and E = 0. This conversion allows the KNN algorithm, which requires numeric input, to process the data effectively.

Missing data are then handled by replacing absent values with the mode of each column, as most attributes are categorical. After the data are cleaned and made consistent, the dataset is split into 80% training data and 20% testing data. The split is performed using a stratified method to maintain proportional balance between the Web and Mobile (WEM) and Intelligent Systems concentrations. These steps ensure that the data are ready for classification and subsequent model evaluation.

FINDING AND DISCUSSION

RESEARCH RESULT

The evaluation of the KNN-based recommendation system demonstrates its effectiveness in classifying student concentrations. Based on the experimental results, the model achieved its optimal performance with a peak accuracy of 74.44% at $k = 15$. Furthermore, a 10-fold cross-validation confirmed the stability of the model with an average accuracy of 0.6484 and a standard deviation of 0.0550. The following subsections detail the experimental stages, performance metrics, and the final system implementation.

The k parameter, which indicates the number of closest neighbors considered (Yustanti, 2012). Choosing an appropriate k value significantly affects the model's performance. A k value that is too small can make the model overly sensitive to noise, whereas a k value that is too large may reduce the model's ability to capture local patterns in the data (Khasanah et al. 2025).

The dataset was divided into training and testing sets with an 80:20 proportion. After training and testing, the model's performance was evaluated using a confusion matrix, as shown in table 3.

Table 3. Confusion Matrix Result

Actual Class	WEM Prediction	SCR Prediction
WEM	52	9
SCR	19	10

Based on table 3, the testing results of the K-Nearest Neighbor (KNN) model indicate that the model was able to correctly classify the majority of the data. Out of the total testing data, 52 WEM class records were correctly classified, while 10 SCR class records were accurately identified. However, some misclassifications still occurred, with 9 WEM records incorrectly predicted as SCR and 19 SCR records misclassified as WEM. The model's accuracy was then calculated by comparing the number of correct predictions to the total testing data, resulting in an accuracy of 68.89%.

Following the initial experiment, further testing was conducted using various values of k ranging from 1 to 25 to determine the optimal value (best k) that yields the highest accuracy.

Figure 2. Comparison chart of accuracy against k value

Based on the testing of various k values (1–25), it was found that $k = 15$ yielded the highest accuracy of 74.44%. The accuracy comparison chart is presented in Figure 5. To further ensure the stability of the model's performance, a 10-fold cross-validation was conducted. The results indicated an average cross-validation accuracy of 0.6484 with a standard deviation of 0.0550, demonstrating that the model exhibits good performance consistency.

After training the KNN model with $k = 15$ (the optimal value based on the evaluation results), the system was used to predict the test data. Some of the test data, namely the 1st, 2nd, and 6th students, are presented in table 4.

Table 4. Test data

Courses																			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
B	B	B	B	A	A	B	B	A	B	A	B	A	B	B	B	A	A	A	A
C	B	D	B	B	B	B	E	C	D	A	B	A	B	D	C	B	B	C	B
B	B	B	A	A	A	B	B	C	B	A	B	B	B	A	A	B	B	A	A

The prediction results of the system, compared with the actual labels from the dataset, are presented in table 5.

Table 5. Test results

No	Name	Actual Class	System Prediction	Match
1	1st student	WEM	WEM	Yes
2	2nd student	WEM	WEM	Yes
3	6th student	SCR	WEM	No

Out of a total of 90 test data points, the system correctly predicted 67 instances and incorrectly predicted 23 instances. These results indicate that the system achieved an accuracy rate of 74.44%, which is consistent with the results of the previous model evaluation.

To make these predictive capabilities accessible to users, the model was integrated into a functional application. The system features a web-based interface designed to facilitate users in entering data and viewing recommendation results. An input form is provided for entering the required data, namely students' academic scores, while a results popup displays the output generated by the system. The interface for these two components is illustrated in figures 3 and 4.

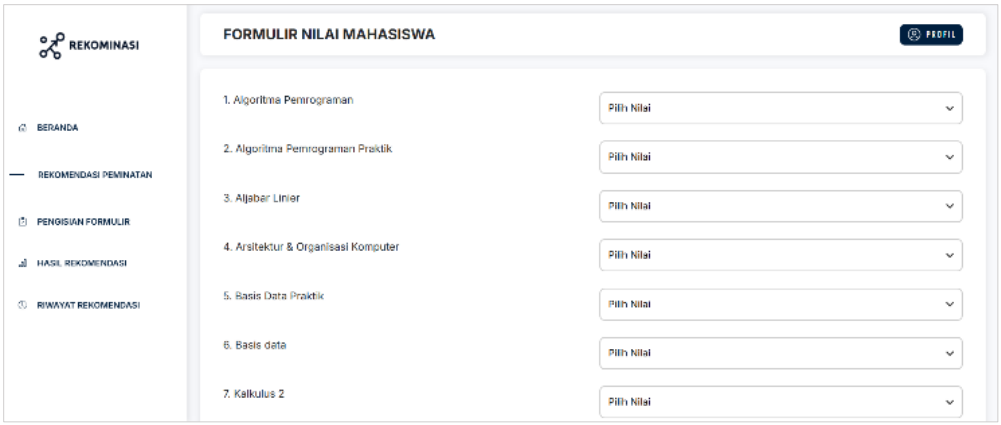

Figure 3. Form filling page view

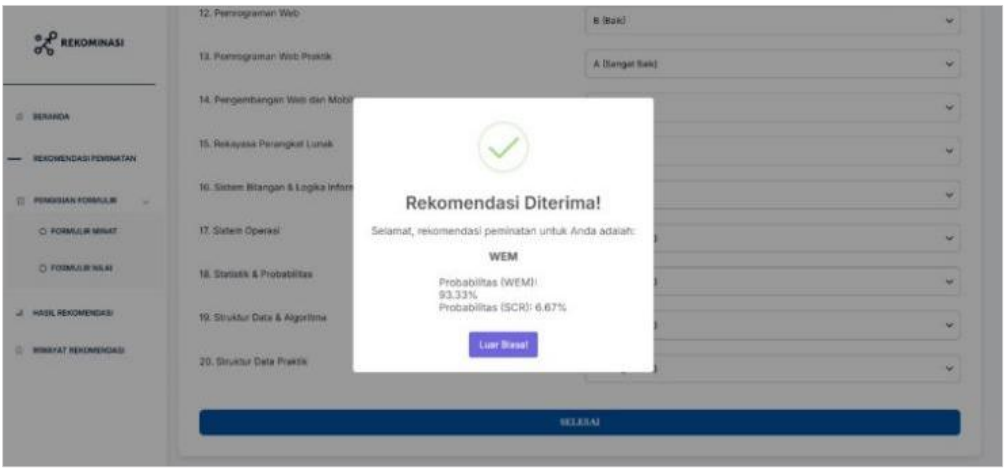

Figure 4. Pop-up display of recommendation results

DISCUSSION

The results of the KNN-based recommendation system indicate that the model is effective in classifying students according to their academic performance patterns. With an optimal $k = 15$, the system achieved an accuracy of 74.44%, which demonstrates that students' past academic scores can be a reliable indicator for predicting the most suitable concentration. The correct classification of 67 out of 90 test instances shows that the system can provide meaningful guidance for students in making concentration choices.

When compared to previous studies, the proposed system shows improvements in prediction accuracy. For instance, Hidayanti et al. (2020) and Wibowo et al. (2024) reported lower accuracy due to smaller datasets and limited optimization of algorithm parameters. This study addresses those issues by using a larger dataset and systematically determining the best k value, resulting in a more reliable recommendation system.

Despite the promising results, several limitations should be noted. The system relies solely on students' academic grades from semesters 1 to 4 and does not incorporate other factors such as personal interest, career goals, or soft skills, which may also influence concentration suitability. Additionally, the model has been tested only on one cohort (2022), so the generalizability of the system to other cohorts remains to be validated.

It is important to emphasize that this system is intended solely as a decision-support tool to provide students with an objective overview of their potential concentration paths. The recommendations are generated based on historical data from previous students who have already selected their concentrations. However, it must be acknowledged that some students in the historical dataset may have chosen their paths without full certainty of their own suitability. Consequently, the system's output should serve as a supplementary consideration rather than a definitive mandate. The final decision remains with the students, who should weigh the system's data-driven insights against their personal interests and professional goals.

The implications of this study are significant for both students and the study program. For students, the system can reduce uncertainty in selecting a concentration,

helping them make more informed decisions and potentially preventing delays in graduation. For the study program, the system can serve as a decision-support tool in academic advising, curriculum planning, and resource allocation. Future research can explore the integration of additional features, such as student preferences and personal traits, as well as real-time deployment within academic portals to further enhance the system's applicability and effectiveness.

CONCLUSION

This study successfully developed a K-Nearest Neighbor (KNN)-based recommendation system to assist Informatics students in selecting a suitable concentration. The system utilizes students' academic scores from semesters 1 to 4 as input data and provides recommendations for two concentrations: Web and Mobile (WEM) and Intelligent Systems (SCR). Experimental results show that the KNN model with an optimal k value of 15 achieved an accuracy of 74.44%, demonstrating that students' academic performance can be a reliable predictor for determining the most appropriate concentration.

The system provides meaningful guidance for students and serves as a decision-support tool for the study program, enhancing academic advising and curriculum management. Furthermore, integrating this system into academic advising processes could provide significant institutional benefits by offering a data-driven tool for faculty, improving the efficiency of student placement, and supporting the institution's goals of academic success and graduation rates.

However, the study has limitations, including reliance solely on academic grades and testing on a single cohort, which may affect generalizability. Future research can enhance the system by incorporating additional factors such as students' personal interests, career goals, and soft skills, as well as integrating real-time deployment within academic portals to improve accessibility and effectiveness.

REFERENCES

Farkhatun Zaidah, & Supatman. (2025). Implementasi Metode K-Nearest Neighbor Dalam Menentukan Klasifikasi Strata Posyandu Di Kabupaten Brebes. *JEKIN - Jurnal Teknik Informatika*, *5*(1), 181–192. https://doi.org/10.58794/jekin.v5i1.1124

Hidayanti, I., & Basuki Kurniawan, T. (2020). Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan. *Jurnal Ilmiah Informatika Global, 11*(1), 16-21.

Hidayat, A. T., Wahyuni, D. S., Gede, I., & Subawa, B. (2025). Sistem Pemilihan Konsentrasi Mahasiswa Berbasis PSI pada Pendidikan Teknik Informatika: Studi Kasus pada Program Studi Pendidikan Teknik Informatika Universitas Pendidikan Ganesha. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, *14*(2).

Khasanah, N., Uki Eka Saputri, D., Aziz, F., & Hidayat, T. (2025). Studi Perbandingan Algoritma Random Forest dan K-Nearest Neighbors (KNN) dalam Klasifikasi Gangguan Tidur. *Computer Science (CO-SCIENCE)*, 5(1).

Muhammad Ridho Fasya, Muhamad Alda, & Adnan Buyung Nasution. (2025). Rancang Bangun Sistem Informasi Rekomendasi Produk pada Supermarket Menggunakan Content Based Filtering Berbasis Web. *JUMINTAL: Jurnal Manajemen Informatika Dan Bisnis Digital*, *4*(1), 28–37. https://doi.org/10.55123/jumintal.v4i1.5114

Prasetyo, A., & Rudyanto Arief, M. (2019). Penerapan Algoritma K-Nearest Neighbor untuk Rekomendasi Minat Konsentrasi di Program Studi Teknik Informatika Universitas PGRI Yogyakarta. *Jurnal Informasi Interaktif*, *4*(1).

Raharjo, A. I., Ramsari, N., & Munawar, Z. (2022). Sistem Pendukung Keputusan untuk Pemilihan Konsentrasi Peminatan Menggunakan Metode Naive Bayes (Studi Kasus: Program Studi Teknik Informatika Universitas Nurtanio Bandung). *FIKI: Jurnal Teknologi Informasi dan Komunikasi.*

Tang, S., Yuan, S., & Zhu, Y. (2020). Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis Towards Rotating Machinery. *IEEE Access, 8*, 149487–149496. https://doi.org/10.1109/ACCESS.2020.3012182

Verawati, I. (2015). Sistem Pakar Penentuan Konsentrasi Penjurusan Mahasiswa Menggunakan Algoritma Bayes. *Jurnal Ilmiah DASI*, *16*(4), 31–36.

Wibowo, A. M., Kasih, P., & Farida, I. N. (2024). Sistem Bantu Penentuan Konsentrasi Mahasiswa Menggunakan Metode K-Nearest Neighbor Classification. *Prosiding Seminar Nasional Teknologi dan Sains, 1*.

Yustanti, W. (2012). Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika, Statistika, & Komputasi, 9*(1), 57-68.