Cendana International Conference on Social and Technology (CICoST)
CISoST - 2025
E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

# Decision Support System for Determining Senior High School Students' Majors Using K-Nearest Neighbors and Decision Tree Methods

**Dewi Yohana br Ginting*, Eka Feby Ronauli Lubis, Romulo Aritonang, Rosita br Sianturi, Lastri Marsely br Ginting**
Faculty of Science and Technology, Information Technology Study Program, Indonesian Institute of Technology and Business, Medan, Indonesia
Email: [1]dewiginting052@gmail.com, [2]ekafeby.dosen@itbi.ac.id, [3]gintinglastri1@gmail.com
(* : coressponding author)

**Abstrak**

Determining Student Majors in Senior High School (SMA) is an important process that requires consideration based on academic data. This study aims to develop a web-based Decision Support System (DSS) for student major selection using the K-Nearest Neighbor (KNN) and Decision Tree Classifier algorithms. The system is developed using Python programming language with an interactive interface built on Streamlit. Five core subjects are used as input features, namely Mathematics, English, Biology, Physics, and Chemistry. The research method applied is software engineering with a prototyping approach, as this approach allows direct interaction between users and the system during the design and interface testing stages. The system is equipped with role-based login features, student grade input and management, criteria weight configuration, Excel data import/export, decision tree visualization, and KNN algorithm accuracy testing. The implementation results show that the system can predict student majors (Science or Social Studies) with an accuracy level above 90% on a limited test dataset. In addition, the Decision Tree visualization feature helps users understand the classification logic performed by the system.

**Keyword:** Decision Support System, Student Major Selection, K-Nearest Neighbor, Decision Tree, Streamlit, Python

## 1. INTRODUCTION

Determining student majors in Senior High School (SMA) into Science (IPA) or Social Studies (IPS) is a crucial step that can significantly influence their academic and career trajectories. However, this process is often still based on subjective assumptions or teacher perceptions alone, without support from an objective data analysis system. To address this issue, a system capable of assisting in decision-making in an objective and data-driven manner is needed.

A Decision Support System (DSS) serves as a solution to facilitate a more structured student major selection process. One method that can be used within a DSS is the K-Nearest Neighbors (KNN) algorithm, which classifies data based on the similarity to existing training data. To enhance the reliability of the system, this study also incorporates a Decision Tree visualization feature, providing a logical overview of the decision-making process carried out by the system. Additionally, the system is equipped with an admin panel, user login, Excel data import/export, and accuracy testing to ensure optimal system performance.

According to Nuraeni et al. (2023), the KNN method is suitable in this context because it classifies new data based on similarity with previously available data. The method is also advantageous due to its simplicity of implementation and effectiveness on small to medium-sized datasets. In this system, KNN is used to group students into IPA or IPS majors based on seven major criteria.

Sakti and Daulay (2024) explain that KNN is a popular classification algorithm widely used in machine learning. KNN classifies objects based on the closest training data, considering only the nearest neighbor as the basis for classification.

According to Alkhussayid and Ferdiansyah (2022), the purpose of the KNN algorithm is to classify new objects based on their attributes, where the classification outcome of a new sample is determined by the majority category of its K-nearest neighbors.

Meanwhile, Ari Ardiansyah Sinta, Tezar Maula, and Agus Budiyantara (2023) state that Decision Tree is an effective algorithm for handling multiclass classification problems, where more than two classes must be predicted. Furthermore, the C4.5 algorithm can manage imbalanced classification where the class distribution in the dataset is uneven.

To address these challenges, a technology-based system capable of supporting more accurate and efficient decision-making is necessary. In this study, two classification methods, KNN and Decision Tree, will be implemented and compared to assist in the student major selection process.

According to Widiastuti, Azhar, and Mulyo (2023), the selection of these two methods is based on their effectiveness in classifying data accurately and interpretably. KNN is recognized for its simple yet reasonably accurate approach to classification, while Decision Tree excels in producing results that are easy to interpret due to its tree-based structure.

This study will be applied to SMA Swasta Santo Paulus as a real case study to develop a practical system that is directly useful in education. By utilizing a DSS based on KNN and Decision Tree methods, it is expected that the student major selection process can be conducted more accurately and measurably. This approach

E-ISSN: 3047-2490
Oktober 2025
Page: 010 - 018

leverages student academic scores as the main parameter in interest classification, applying decision tree algorithms to divide student scores into decision branches that represent suitable major pathways.

## 2. RESEARCH METHODE

### 2.1 Decision Support System

According to Prasetyo and Heryanto (2022), a Decision Support System (DSS) is specifically designed to assist decision-makers in solving semi-structured or unstructured problems by utilizing data, models, and user interfaces ("Decision Support System: Theory and Implementation").

Apriyansyah (2022) explains that DSS has the capability to integrate data, analytical models, and interactive user interfaces, thereby facilitating users in understanding and selecting the best solutions from multiple available options ("Literature Review of: Decision Support System: Organization, Human Resources and Knowledge Management").

Ramadhan (2022) states that a Decision Support System consists of a series of processes and mechanisms to acquire and process validated data, which serves as a basis for guidance, useful in identifying problems and supporting systematic decision-making processes ("Journal of Decision Support Systems: Evaluation of Problems in Online Learning Assistance Using the C4.5 Algorithm").

Decision Support Systems (DSS) are designed to assist decision-makers in addressing semi-structured or unstructured problems by integrating data, analytical models, and interactive user interfaces. According to Prasetyo and Heryanto (2022), DSS utilizes data, models, and user interfaces to provide guidance in solving complex decision problems ("Decision Support System: Theory and Implementation"). Apriyansyah (2022) highlights that DSS facilitates users in understanding and selecting the best solution from multiple available alternatives by providing a coherent integration of data, analytical tools, and user-friendly interfaces ("Literature Review of: Decision Support System: Organization, Human Resources and Knowledge Management"). Similarly, Ramadhan (2022) emphasizes that DSS encompasses processes and mechanisms for acquiring and processing validated data, serving as a structured basis for identifying problems and supporting systematic decision-making ("Journal of Decision Support Systems: Evaluation of Problems in Online Learning Assistance Using the C4.5 Algorithm"). Collectively, these perspectives underscore that an effective DSS not only processes information but also enhances the decision-making capabilities of users by providing timely, accurate, and actionable insights.

### 2.2 Metode Decision Tree

According to Monalisa and Hadi (2020), a decision tree is a classification method structured like a flowchart resembling a tree. The model created by a decision tree is easy to understand, making this method highly popular ("C4.5 Algorithm in Determining New Student Majors," *Ultimatics: Journal of Computer Engineering*, 12(2): 108–113).

A decision tree works by splitting a dataset into subsets based on specific attributes. Each node in the tree represents a test on an attribute, and each branch represents the outcome of that test. The leaves of the tree indicate the final class or decision. Every node corresponds to an attribute used as the basis for classification, while the branches from that node show the possible values of the attribute. The ultimate purpose of the tree is to represent the class or decision derived from combinations of attribute values.

This method is interpretable and resembles human decision-making logic, making it suitable for educational contexts and decision support systems. By following the path from the root to the leaf, users can easily trace how the final decision—such as student major assignment—is produced based on academic scores.

The classification process is performed by dividing the dataset into two main parts: training data and testing data. This split allows the model to learn patterns from the training data and then test its ability to predict previously unseen data through the testing dataset. In this study, the test samples were carefully selected or filtered to align with the research objectives, ensuring diversity in student characteristics and score distributions.

### 2.3 Algoritma K-Nearest Neighbor

In the decision support system for determining high school student majors, one of the classification methods used is K-Nearest Neighbor (KNN). Based on research by Rizqian, Irawan, and Wahyuningsih (2024), the KNN method is chosen because it has a high accuracy rate in various classification studies. KNN is an instance-based learning method, which works by comparing new data with training data whose class is already known. Classification is performed based on the majority class of the closest neighbors to the new data.

According to Zaidah (2025), the number of nearest neighbors is determined by the K parameter, which can be optimized to achieve the best results. Choosing the right K value significantly affects model performance: if K is too small, the classification may be sensitive to outliers; if K is too large, the classification may become too general and blur the boundaries between classes.

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

In the context of student majoring, the academic scores of five main subjects (Mathematics, English, Biology, Physics, and Chemistry) are used as input features. The system then compares the new student's data with other students whose majors have already been classified (Science/ Social Science) and determines the new student's major based on the majority label of the K nearest neighbors.

One of the main advantages of KNN is its simplicity in implementation, especially for small to medium-sized datasets, while also being able to handle unstructured datasets. The method is flexible and easy to interpret, making it suitable for educational contexts.

The use of the K-Nearest Neighbors (KNN) algorithm in this system enables relatively accurate and straightforward classification, particularly for small to medium-sized datasets. By selecting an appropriate value of K, the system can provide more optimal predictions of student majors, closely reflecting their actual abilities. A major advantage of KNN is its capability to handle unstructured datasets and its simplicity in implementation, making it a widely used method in various classification studies. Euclidean distance is employed to measure the closeness between two points in Euclidean space and is used in KNN to identify the nearest neighbors. This distance is calculated by measuring the straight-line distance between two points in two or more dimensions, a method favored in machine learning due to its simplicity and effectiveness in interpreting data proximity.:

1. Determine the value of K, which represents the number of nearest neighbors.
2. Calculate the distance between the test data and each training data point, typically using the Euclidean Distance.

$$d(x) = \sqrt{(X_1 + Y_1)^2 + (X_2 + Y_2)^2 + (X_2 + Y_2)^2 + \cdots + (X_n + Y_n)^2}$$

3. Select the k nearest data points based on the smallest distances.
4. Determine the majority class among these k neighbors.
5. After identifying the k nearest neighbors using the K-Nearest Neighbor (KNN) algorithm, perform the classification process for the test data.

## 3. RESULT AND DISCUSSION

### 3.1 Result

The Major Selection Decision Support System (DSS) developed using Python and Streamlit provides an interactive and user-friendly interface. The main features of the system are presented through the sidebar navigation menu, including. Introduction to the system and explanation of the criteria used to determine majors, namely: Mathematics, English, Biology, Physics, and Chemistry.

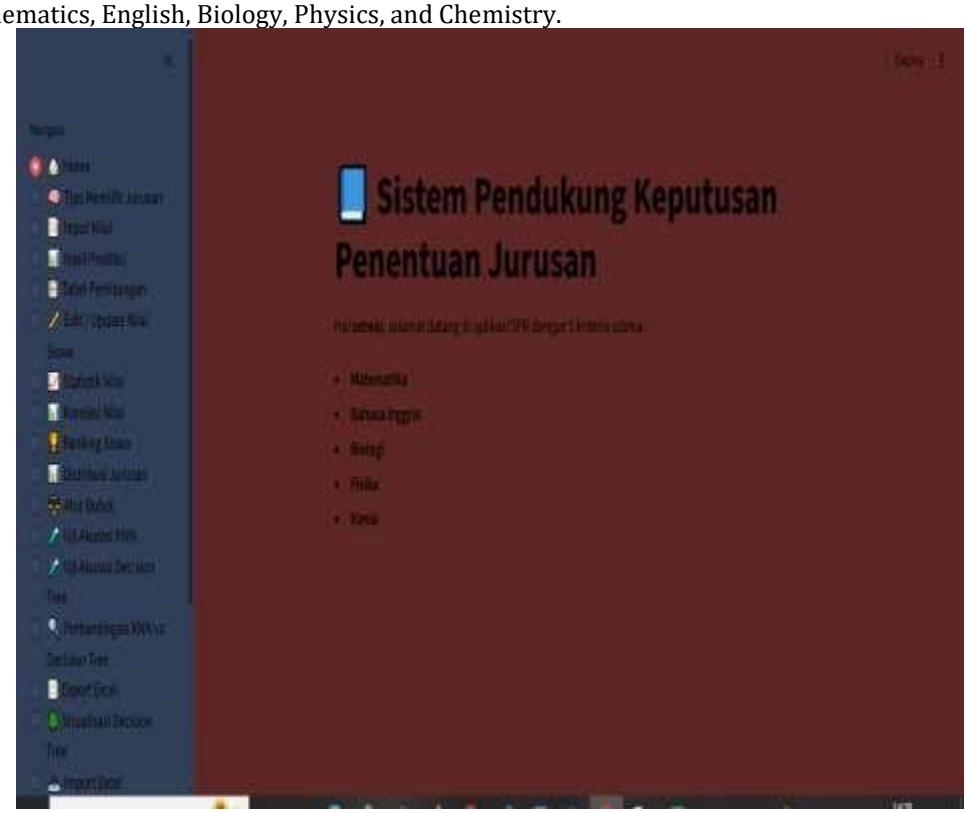


**Figure 1.** DSS Home Page

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

With this guidance, students can better recognize their potential, make accurate decisions, and reduce the risk of choosing the wrong major, which could lead to regret in the future.


**Figure 2** Tips for Choosing a Major

Student Data Input Form: This feature allows teachers, admins, or general users to input student data that will be processed using the KNN or Decision Tree algorithms.

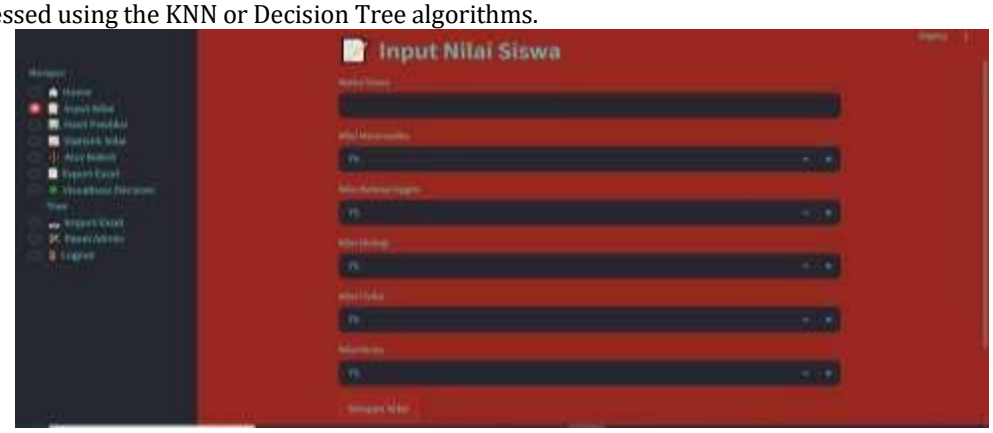


**Figure 3** Student Grade Input

**Prediction Results:** Displays the predicted major (Science/IPA or Social Studies/IPS) based on final scores calculated from the weighted subjects.

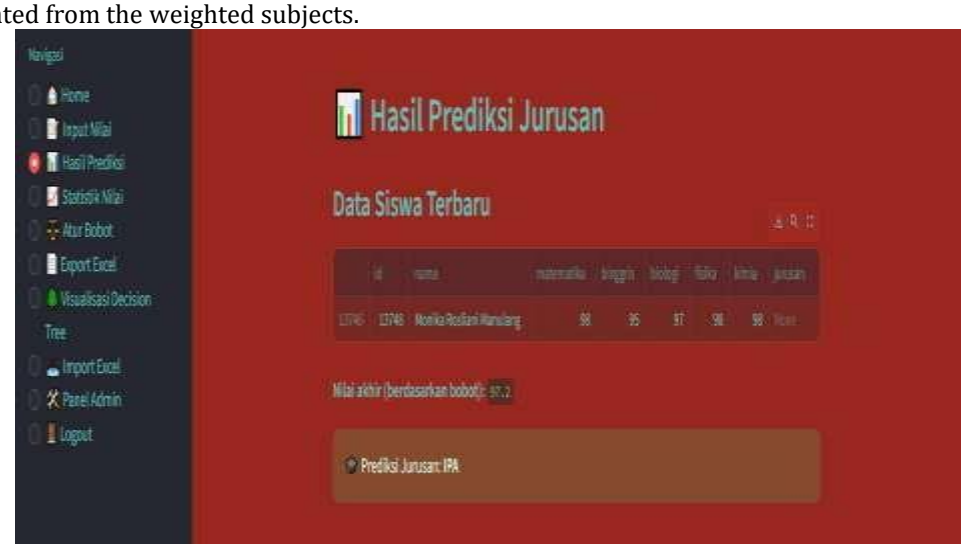


**Figure 4** Prediction Results

**Student Score Statistics:** Displays a bar chart of the average scores of all students for each subject.

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

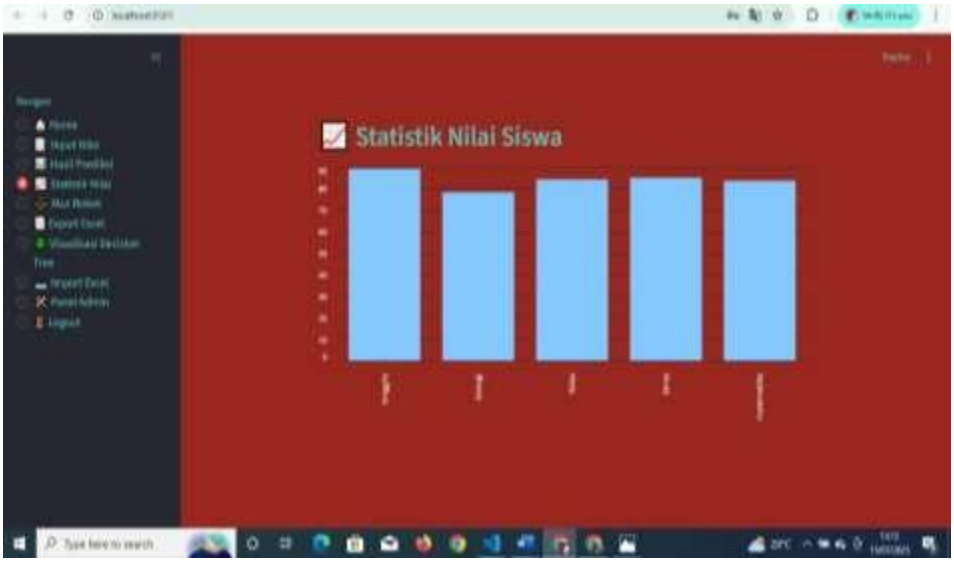


**Figure 5** Student Score Statistics

**Subject Weight Adjustment:** An interactive feature to set weights for each subject. The system provides a warning if the total weight does not equal 1.0.


**Figure 6** Adjust Subject Weights

**xport Student Data:** Provides a button to download all student data in Excel (.xlsx) format.

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

**Figure 6** Export Student Data

**Decision Tree Visualization:** Displays the structure of the decision tree used in classifying majors using the Decision Tree Classifier algorithm.

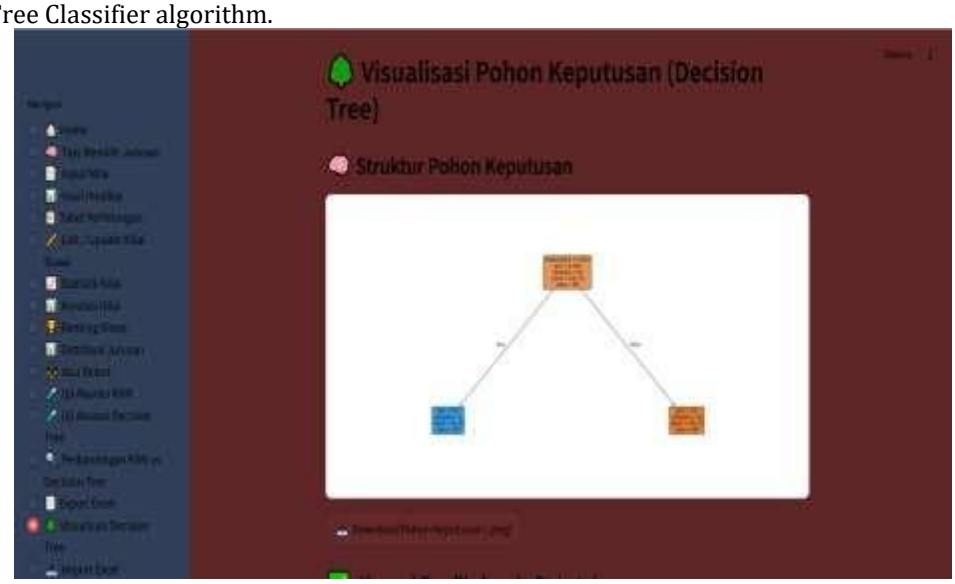


**Figure 7** Decision Tree Visualization


**Figure 8** Import Excel

**Admin Panel:** Displays all student data from the database. Admins can also delete data based on ID

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**


**Figure 9** Admin Panel

**KNN Accuracy Test:** Implements the KNN algorithm to measure prediction accuracy. Accuracy is displayed as a percentage and accompanied by a confusion matrix for performance evaluation.

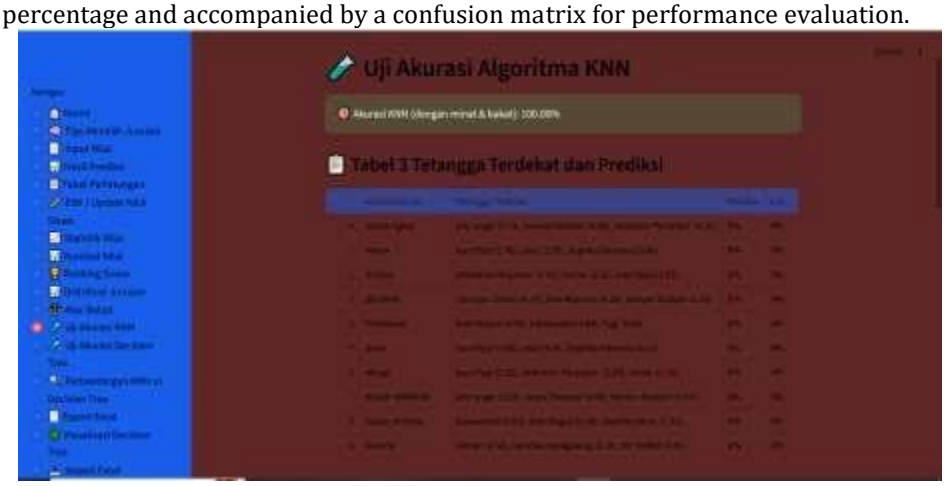


**Figure 10** KNN Accuracy Test

**Logout Function:** Ends the active user session. After pressing the Logout button, the system clears the login status from session_state and automatically redirects the user back to the login page


**Figure 11** Logout and Login Page

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

### 3.2 Discussion

This system works by inputting student grades and then calculating the final score based on predetermined weights. Next, the system predicts the appropriate major using a simple rule-based approach: if the final score ≥ 80, the recommended major is Science (IPA); otherwise, it is Social Studies (IPS). Additionally, the implementation of a Decision Tree as a simple form of machine learning is used to illustrate the classification process based on historical data. The decision tree visualization provides users with additional insight into the logic behind the decision-making process derived from the available data.

## 4. CONCLUTION

Based on the design and implementation results of the student majoring decision support system using the K-Nearest Neighbor (KNN) and Decision Tree methods, it can be concluded that the system successfully assists the decision-making process in determining students' majors (Science/IPA or Social Studies/IPS) objectively, based on subjects such as Mathematics, English, Biology, Physics, and Chemistry. The KNN algorithm provides classification results based on data similarity, achieving a relatively high accuracy in the tests conducted on the dataset. The Decision Tree supports visualization of the major selection logic, making it easier for teachers to understand the system's decision patterns. The interactive and responsive Streamlit-based interface facilitates users (teachers and admins) in inputting data, viewing prediction results, and exporting outputs to Excel. Additional features, such as subject weight settings, an admin panel, data import-export, and statistical visualization, enhance the system's functionality, making it suitable and practical for use in educational environments.

## REFERENCES

Alkhussayid, M. Daffa, and Ferdiansyah Ferdiansyah. 2022. "Implementasi Algoritma K-Nearest Neighbors Pada Penentuan Jurusan Siswa." Jurnal Sistem Komputer Dan Informatika (JSON) 4 (1): 25. https://doi.org/10.30865/json.v4i1.4759.

Amanda, Kiki, Darjat Saripurna, and Mhd. Zulfansyuri Siambaton. 2024. "Penerapan Algoritma Cart Dalam Penentuan Jurusan Siswa Di SMA." Hello World Jurnal Ilmu Komputer 2 (4): 169–77. https://doi.org/10.56211/helloworld.v2i4.404.

Apriyansyah, Hijrah. 2022. "Literature Review of: Decision Support System: Organization, Human Resources and Knowledge Management." Dinasti International Journal of Economics, Finance & Accounting 3 (2): 136–48. https://dinastipub.org/DIJEFA/article/view/1246.

Apung, Hidroponik Rakit. 2020. "Jurnal Pendidikan Informatika Dan Sains" 9 (1): 46–58. https://doi.org/10.31571/saintek.v9i1.

Ari ardiansyah sinta, Tezar Maula, and Agus Budiyantara. 2023. "Perbandingan Algoritma K-Nearest Neighbor (K-Nn) Dengan Decision Tree (C4.5) Untuk Menentukan Judul Buku Berdasarkan Peminatan Pada Perpustakaan Kantor Pusat Bpk Ri." Jurnal Teknik Informatika Dan Komputer 2 (2): 50–55. https://doi.org/10.22236/jutikom.v2i2.12654.

Fashihullisan, Dodi Vionanda, Yenni Kurniawati, and Fadhilah Fitri. 2023. "Comparing Classification and Regression Tree and Logistic Regression Algorithms Using 5×2cv Combined F-Test on Diabetes Mellitus Dataset." UNP Journal of Statistics and Data Science 1 (4): 344–52. https://doi.org/10.24036/ujsds/vol1-iss4/84.

Hasanah, Faujatun, Tati Suprapti, Nining Rahaningsih, and Irfan Ali. 2022. "Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Buku Berdasarkan Peminatan." Jurnal Accounting Information System (AIMS) 5 (1): 102–11. https://doi.org/10.32627/aims.v5i1.467.

Hidayati, Nur, and Arief Hermawan. 2021. "K-Nearest Neighbor (K-NN) Algorithm with Euclidean and Manhattan in Classification of Student Graduation." Journal of Engineering and Applied Technology 2 (2): 86–91. https://doi.org/10.21831/jeatech.v2i2.42777.

Imam, Komarul, and Zaehol Fatah. 2024. "Gudang Jurnal Multidisiplin Ilmu Implementasi Metode K-Nearest Neighbors ( KNN ) Untuk MenentukanJurusan Siswa Di SMK Sumber Bunga" 2 (November): 289–94.

Jean Cross Sihombing, Denny. 2023. "Development of an Application for Document Management in Construction Companies Using Prototyping Method." Jurnal Sains Komputer & Informatika (J-SAKTI 7 (2): 915–28.

Miftahuddin, Yusup, Sofia Umaroh, and Fahmi Rabiul Karim. 2020. "Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan." Jurnal Tekno Insentif 14 (2): 69–77. https://doi.org/10.36787/jti.v14i2.270.

Monalisa, Siti Monalisa, and Fakhri Hadi. 2020. "Algoritma C4.5 Dalam Penentuan Jurusan Siswa Baru." Ultimatics : Jurnal Teknik Informatika 12 (2): 108–13. https://doi.org/10.31937/ti.v12i2.1838.

Nuraeni, Septiya, Sri Putri Aulia Syam, Muhammad Farid Wajdi, Bachtiar Firmansyah, and Muhammad Malkan. 2023. "Implementasi Metode K-NN Untuk Menentukan Jurusan Siswa Di SMAN 02 Manokwari." G-Tech: Jurnal Teknologi Terapan 7 (1): 89–95. https://doi.org/10.33379/gtech.v7i1.1905.

Ramadhan, Abiyyu. 2022. "Sistem Pendukung Keputusan Evaluasi Problematika Pendampingan Pembelajaran Daring Dengan Algoritma C4.5." Jurnal Sistim Informasi Dan Teknologi 4:58–63. https://doi.org/10.37034/jsisfotek.v4i2.124.

Ray, Swati, Joyati Das, Ranjana Pande, and A Nithya. 2024. "Swati Ray 1 , Joyati Das 2* , Ranjana Pande 3 , and A. Nithya 2" 7482:195–222. https://doi.org/10.1201/9781032622408-13.

Ridlwanul Fitra, Ibnu Rusdi. 2022. "Penerapan Metode Algoritma K-Nearest Neighbor Menggunakan Rapidminer Studio Pada Klasifikasi Status Sosial Ekonomi." Smart Comp: Jurnalnya Orang Pintar Komputer 11 (4): 4524–30.

Rizqian, Mohamad Arjun, Bambang Irawan, and Puji Wahyuningsih. 2024. "Implementasi Algoritma KNN Untuk Klasifikasi Jurusan Siswa Kelas XI Di SMA Al Irsyad Tegal." Journal of Citizen Research and Development 1 (2): 233–41. https://doi.org/10.57235/jcrd.v1i2.3309.

E-ISSN: **3047-2490**
Oktober 2025
Page: **010 - 018**

Sakti, Raja, and Arief Daulay. 2024. "Analisis Kritis Dan Pengembangan Algoritma K-Nearest Neighbor ( KNN ): Sebuah Tinjauan Literatur" 4 (2): 131–41.

Siti hidayatul Khoirun Nisa', and R. Yuniardi Rusdianto. 2024. "Pemanfaatan Visualisasi Data Dalam Meningkatkan Pengambilan Keputusan Bisnis." Jurnal Informasi, Sains Dan Teknologi 7 (2): 200–208. https://doi.org/10.55606/isaintek.v7i2.290.

Syafarina, Gita Ayu, and Zaenuddin Zaenuddin. 2023. "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi." Metik Jurnal 7 (2): 121–25. https://doi.org/10.47002/metik.v7i2.608.

Widiastuti, Nur Aeni, Maulana Azhar, and Harminto Mulyo. 2023. "Implementasi Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Pada Peserta Didik Baru." Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer 14 (2): 195–208. https://doi.org/10.24176/simet.v14i2.10092.

Zaidah, Farkhatun. 2025. "Implementasi Metode K-Nearest Neighbor Dalam Menentukan Klasifikasi Strata Posyandu Di Kabupaten Brebes" 5 (1).