CODEX: Journal of Software Engineering
Vol 01(01) 2026 : 11-19

DESIGN AND IMPLEMENTATION OF AN INTERACTIVE WEB-BASED DATA MINING SYSTEM USING KNN, SVM, AND RANDOM FOREST WITH STREAMLIT

***Muhammad Faisal[1], Kholqi Maulana[2]**
Teknik Informatika, Universitas Islam Madura, Pamekasan[1,2]
mhfaisal011@gmail.com[1] , mohkholqimaulana@gmail.com[2]

**Coresponding Author*

ABSTRACT
The rapid growth of digital data requires effective tools to extract meaningful information and support decision-making processes. Data mining and machine learning techniques play an important role in analyzing large datasets and producing accurate classifications. However, implementing machine learning models often requires technical expertise and complex tools. This study aims to design and implement a web-based data mining system using the Streamlit framework integrated with classification algorithms, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest. The research method includes system design, implementation, and evaluation using three datasets: Iris, Wine, and Digit. The system provides an interactive interface that allows users to select datasets, configure algorithm parameters, evaluate classification accuracy, and visualize results. The implementation results show that all algorithms perform effectively, with Random Forest achieving the highest accuracy, followed by SVM and KNN. The developed system successfully integrates machine learning classification methods into a user-friendly web-based platform, enabling efficient data analysis and visualization. This study demonstrates that interactive web-based data mining systems can enhance accessibility and understanding of machine learning applications for academic and practical use.
Keywords : *Classification, Data Mining, Random Forest, Streamlit, Support Vector Machine.*

1. Introduction

The rapid advancement of information technology has led to an exponential increase in data generation across various sectors, including education, business, healthcare, and industry(Rada et al., 2023). This phenomenon has encouraged the development of techniques capable of extracting meaningful knowledge from large datasets. Data mining has become a crucial approach in this context, as it enables the process of discovering patterns, relationships, and useful information from large volumes of data using computational techniques and machine learning methods(Chaurasiya & Pandey, 2025; Guleria et al., 2026). Through data mining, hidden information within datasets can be transformed into valuable knowledge to support decision-making processes and predictive analysis.

Data mining consists of several techniques such as classification, clustering, association rule mining, and regression. Among these techniques, classification is one of the most widely used methods for predicting categorical labels based on input data. Classification works by building a model from training data and using that model to predict the class of new or unseen data. The effectiveness of classification methods depends heavily on the algorithms used and the characteristics of the dataset. Therefore, selecting appropriate algorithms and evaluating their performance are essential steps in the data mining process.

Several machine learning algorithms are commonly applied in classification tasks, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest(Chaurasiya & Pandey, 2025; Guleria et al., 2026). K-Nearest Neighbor is a supervised learning algorithm that classifies data based on similarity measures, typically using distance metrics such as Euclidean distance. It determines the class of a new data point based on the majority class among its nearest neighbors. Support Vector Machine is another supervised learning algorithm that works by finding the optimal hyperplane that separates data into different classes with maximum margin. SVM is known for its high accuracy and effectiveness in handling high-dimensional data. Meanwhile, Random Forest is an ensemble learning method that constructs multiple decision trees

and combines their outputs to improve classification accuracy and reduce overfitting. The combination of these three algorithms provides a comprehensive approach to classification, allowing performance comparison and evaluation across different datasets.

Despite the effectiveness of these algorithms, implementing machine learning models often requires programming skills and technical expertise, which can be challenging for beginners or researchers from non-computing backgrounds. Therefore, the development of user-friendly and interactive platforms for implementing data mining techniques is essential. One of the frameworks that has gained popularity in recent years is Streamlit. Streamlit is an open-source Python-based framework that enables developers and researchers to build interactive web applications for data science and machine learning quickly and efficiently. With Streamlit, users can easily visualize datasets, configure parameters, run machine learning models, and display results in graphical form through a web interface.

Previous studies have demonstrated the effectiveness of classification algorithms such as KNN, SVM, and Random Forest in various domains. Research has shown that KNN performs well in simple classification tasks with smaller datasets, while SVM provides high accuracy for complex and high-dimensional data. Random Forest has also been proven to deliver stable and consistent performance due to its ensemble learning approach(Chaurasiya & Pandey, 2025; Guleria et al., 2026). Additionally, several studies have highlighted the advantages of using Streamlit for developing web-based data science applications, particularly in visualizing data and simplifying the deployment of machine learning models. However, there is still limited research that integrates multiple classification algorithms into a single interactive web-based system that allows users to compare performance, adjust parameters, and visualize results dynamically.

Based on these considerations, this research focuses on the design and implementation of a Streamlit-based web data mining system that integrates K-Nearest Neighbor, Support Vector Machine, and Random Forest algorithms. The system is designed to provide an interactive platform where users can select datasets, configure algorithm parameters, evaluate model accuracy, and visualize classification results. By integrating multiple algorithms into a single web-based platform, this research aims to provide a practical and user-friendly tool for learning, experimentation, and implementation of data mining techniques. The objective of this study is to design and develop a Streamlit-based web data mining system that integrates classification algorithms, including K-Nearest Neighbor, Support Vector Machine, and Random Forest, while providing an interactive interface for dataset selection, parameter configuration, and visualization of classification results, as well as evaluating and comparing the performance of these implemented algorithms based on accuracy and visualization outputs.

Table 1. Comparison of Classification Algorithms Used in the Study

Algorithm	Method Type	Working Principle	Advantages	Limitations
K-Nearest Neighbor (KNN)	Supervised Learning	Classifies data based on similarity and nearest distance to neighboring data points(Rada et al., 2023)	Simple implementation, effective for small datasets, no training phase required	Sensitive to noise and large datasets, computationally expensive
Support Vector Machine (SVM)	Supervised Learning	Finds optimal hyperplane to separate classes with maximum margin(Rada et al., 2023)	High accuracy, effective for high-dimensional data, robust to overfitting	Requires parameter tuning, less efficient for very large datasets
Random Forest	Ensemble Learning	Combines multiple decision trees and uses voting for classification(Rada et al., 2023)	High accuracy, reduces overfitting, handles large datasets well	Requires more computational resources, less interpretable

Streamlit Framework	Web-based Data Science Framework	Provides interactive web interface for data visualization and machine learning	Easy to use, fast development, interactive visualization	Limited for complex backend systems

Table 1 summarizes the characteristics of the classification algorithms and framework used in this study, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, and Streamlit. KNN is a simple supervised learning algorithm that classifies data based on similarity to its nearest neighbors and works well for small datasets, although it is sensitive to noise and large data volumes. SVM is a supervised learning method that finds the optimal hyperplane to separate classes, offering high accuracy and strong performance for high-dimensional data but requiring careful parameter tuning. Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve accuracy and reduce overfitting, though it requires higher computational resources. Meanwhile, Streamlit functions as a web-based framework that enables the development of interactive data mining applications with easy visualization and user-friendly interfaces, supporting the implementation and evaluation of the three classification algorithms in this research.

Table 2. Summary of Related Research

Author/Year	Research Focus	Method Used	Result
Previous Study 1	Implementation of KNN for classification	KNN	Achieved good accuracy for small datasets
Previous Study 2	SVM-based classification system	SVM	High accuracy in binary classification
Previous Study 3	Random Forest classification	Random Forest	Stable and consistent performance
Previous Study 4	Streamlit-based data science web	Streamlit	Interactive and user-friendly visualization
This Research	Web-based data mining system	KNN, SVM, RF + Streamlit	Integrated interactive classification system

Table 2 presents a summary of related studies relevant to this research. Previous studies have shown that classification algorithms such as K-Nearest Neighbor, Support Vector Machine, and Random Forest are widely used in data mining due to their ability to produce accurate classification results across various datasets. Research focusing on KNN demonstrates its effectiveness for simple classification tasks and smaller datasets. Studies using SVM indicate that the algorithm provides high accuracy, especially in binary classification and high-dimensional data. Meanwhile, research on Random Forest shows stable and consistent performance because it combines multiple decision trees to improve prediction accuracy and reduce overfitting.

In addition, several studies have explored the use of Streamlit as a framework for developing interactive web-based data science applications. These studies highlight the advantages of Streamlit in providing interactive visualization, ease of development, and rapid deployment of machine learning models into web interfaces. However, most previous research focuses on individual algorithms or standalone implementations rather than integrating multiple classification methods within a single interactive web-based system.

Therefore, this research differs from previous studies by integrating KNN, SVM, and Random Forest algorithms into a single Streamlit-based web data mining application. The system enables users to compare algorithm performance, adjust parameters, and visualize results interactively. Table 2 strengthens the research position by showing that this study combines methods and technologies that have been proven effective in previous research while offering a more integrated and interactive implementation.

2. Literature Review

Data mining has become a fundamental discipline in computer science due to its capability to extract meaningful patterns and knowledge from large volumes of data.(Chaurasiya & Pandey,

2025; Gupta et al., 2024) It refers to the process of discovering hidden patterns, correlations, and useful information from datasets using statistical, computational, and machine learning techniques. The rapid growth of digital data across various sectors has increased the importance of data mining as a tool to support decision-making, prediction, and knowledge discovery. As organizations continue to generate massive amounts of data, effective data mining techniques are required to transform raw data into actionable insights.

Classification is one of the most widely used techniques in data mining and machine learning(Halder et al., 2023; Mishra et al., 2022). It is a supervised learning process that involves training a model using labeled data and then using that model to classify new data into predefined categories. Classification techniques are commonly applied in various domains such as education, healthcare, finance, and business analytics. The performance of classification models depends on the selection of appropriate algorithms and parameter configurations. Among the most widely used classification algorithms are K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest, which have been extensively studied and applied in various research contexts.

K-Nearest Neighbor (KNN) is a simple yet effective supervised learning algorithm that classifies data based on similarity or distance measures(Pan et al., 2021). The algorithm determines the class of new data by examining the majority class among its nearest neighbors in the training dataset. Due to its simplicity, KNN is easy to implement and does not require a complex training process. However, its performance can be affected by large datasets, irrelevant features, and noise, as it relies heavily on distance calculations. Despite these limitations, KNN remains widely used in classification problems due to its effectiveness in handling simple and structured datasets.

Support Vector Machine (SVM) is another powerful supervised learning algorithm commonly used for classification and regression tasks(Jabardi, 2025; Mohammed & Mahdi, 2024). SVM works by finding the optimal hyperplane that separates data points into different classes with maximum margin. This algorithm is known for its high accuracy and strong generalization capability, particularly in high-dimensional data. SVM also applies the structural risk minimization principle, which helps reduce overfitting and improve prediction performance. However, SVM requires careful parameter tuning and appropriate kernel selection to achieve optimal results, and its computational complexity may increase when applied to large-scale datasets.

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification performance and stability(Zhao et al., 2024). By constructing several decision trees using random subsets of data and features, Random Forest generates predictions based on a voting mechanism among the trees. This ensemble approach helps reduce overfitting and increases accuracy compared to single decision tree models. Random Forest is capable of handling large datasets and complex feature interactions, making it suitable for various data mining applications. Nevertheless, the algorithm requires greater computational resources and may produce models that are less interpretable due to the large number of trees involved.

In addition to algorithmic development, the implementation of machine learning models through interactive and accessible platforms has become an important area of research. Web-based applications provide a convenient environment for deploying data mining models and enabling user interaction. Streamlit has emerged as one of the most popular frameworks for developing interactive data science and machine learning web applications. It is an open-source Python-based framework that allows rapid development of web applications for data visualization, analysis, and model evaluation. Streamlit enables users to interact with datasets, adjust parameters, and visualize outputs dynamically, making it suitable for educational, research, and practical implementations of data mining.

Previous studies have demonstrated the effectiveness of KNN, SVM, and Random Forest algorithms in various classification tasks. Research has shown that KNN performs well for simple classification problems, while SVM provides high accuracy for complex and high-dimensional data. Random Forest has been widely recognized for its stability and ability to improve prediction accuracy through ensemble learning. Meanwhile, several studies have explored the use of Streamlit in building interactive data science applications, highlighting its advantages in usability, visualization, and rapid deployment. However, most existing studies focus on individual

algorithm implementation or standalone applications rather than integrating multiple classification algorithms into a single interactive web-based system.

Based on the reviewed literature, it can be observed that there is a need for an integrated platform that combines multiple classification algorithms with an interactive web interface to facilitate comparison, experimentation, and visualization. Therefore, this study aims to develop a Streamlit-based web data mining system that integrates K-Nearest Neighbor, Support Vector Machine, and Random Forest algorithms. The proposed system is expected to provide an interactive environment that enables users to explore datasets, configure algorithm parameters, evaluate classification accuracy, and visualize results effectively. This integration not only enhances usability but also contributes to the development of accessible and practical data mining applications for academic and research purposes.

3. Research Methods

This research employs a system development and experimental approach to design and implement a web-based data mining application using the Streamlit framework integrated with machine learning classification algorithms(Nantasenamat et al., 2023; Pandey, 2024; Raghavendra, 2022). The research method consists of several stages, including problem identification, literature study, system design, implementation, testing, and evaluation. Each step is conducted systematically to ensure that the developed system functions effectively and meets the research objectives.

Research Approach

This study adopts an experimental and development-based research approach. The experimental approach is used to test the performance of classification algorithms, while the development approach is applied to design and implement a web-based data mining system. The system integrates three classification algorithms: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest, implemented through a Streamlit-based interactive web interface.

Research Steps

The research procedure consists of the following stages:

a. Problem Identification

The initial stage involves identifying the need for an interactive and user-friendly platform that can implement and evaluate classification algorithms efficiently. Many machine learning implementations require programming expertise and complex tools, making them less accessible to beginners. Therefore, this research aims to develop a web-based system that simplifies the implementation and evaluation of classification algorithms.

b. Literature Study

A literature review is conducted to collect theoretical references related to data mining, classification algorithms, and web-based machine learning implementation. Sources include books, scientific journals, and previous research studies related to KNN, SVM, Random Forest, and Streamlit-based applications. This stage provides theoretical support and identifies research gaps.

c. System Design

At this stage, the system architecture and interface are designed. The system is designed as an interactive web application using the Streamlit framework. The design includes:

1. Dataset selection interface
2. Algorithm selection menu (KNN, SVM, Random Forest)
3. Parameter configuration panel
4. Accuracy evaluation and visualization display

The design also includes modeling of classification processes and system workflow.

d. Data Preparation

The datasets used in this research consist of standard datasets available in machine learning libraries, including:

1. Wine dataset
2. Iris dataset
3. Digit dataset

These datasets are used to test and evaluate the performance of classification algorithms.

e. System Implementation

The implementation stage involves developing the web-based data mining system using Python programming language and Streamlit framework. Supporting libraries used include:

1. NumPy and Pandas for data processing
2. Scikit-learn for machine learning modeling
3. Matplotlib for data visualization

Each classification algorithm is implemented and integrated into the system. Users can select datasets, configure parameters, and run classification processes through the web interface.

f. Testing and Evaluation

System testing is conducted to ensure that all features function correctly. The evaluation focuses on:

1. Accuracy of classification results
2. Visualization output
3. System usability and interaction

Performance comparison is conducted to evaluate the effectiveness of KNN, SVM, and Random Forest algorithms across different datasets and parameter settings.

g. Result Analysis

The results of classification accuracy and visualization are analyzed to determine the effectiveness of each algorithm. The analysis also evaluates how the interactive web interface supports user understanding of classification performance.

Algorithm and System Design

The system design integrates three classification algorithms into a single web-based application. Each algorithm follows its respective computational process:

1. **K-Nearest Neighbor (KNN):** Calculates distance between data points and classifies based on nearest neighbors.
2. **Support Vector Machine (SVM):** Determines optimal hyperplane for classification.
3. **Random Forest:** Builds multiple decision trees and determines classification through voting.

Users can interactively select datasets, adjust algorithm parameters, and view classification accuracy and visualization results. This integrated design ensures that the system not only performs classification tasks but also serves as an interactive learning and evaluation platform for data mining and machine learning.

4. Results and Discussions

Results

This study successfully developed a web-based data mining system using the Streamlit framework integrated with three classification algorithms: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest. The system was designed to provide an interactive interface that allows users to select datasets, configure algorithm parameters, run classification processes, and visualize the results. The datasets used in this study include the Wine dataset, Iris dataset, and Digit dataset obtained from the scikit-learn library.

The developed system consists of several main components, including dataset selection, algorithm selection, parameter configuration, accuracy calculation, and visualization output. Users can interact with the system through a sidebar menu that enables the selection of classification algorithms and adjustment of parameters. The system then processes the selected dataset using the chosen algorithm and displays classification accuracy along with graphical visualization.

System Implementation Results

The implementation results show that the Streamlit-based web application runs successfully and is able to perform classification tasks for all three datasets. Each algorithm was tested using default and adjustable parameters to evaluate its classification performance. The system displays the number of samples, number of features, and number of classes for each dataset, followed by the classification accuracy and visualization results.

Classification Accuracy Results

The classification accuracy obtained from the implementation of KNN, SVM, and Random Forest algorithms on each dataset is presented in Table 3

Table 3. Classification Accuracy Results.

Dataset	KNN Accuracy	SVM Accuracy	Random Forest Accuracy
Iris Dataset	0.97	0.98	0.99
Wine Dataset	0.96	0.98	0.99
Digit Dataset	0.94	0.97	0.98

The results indicate that all three algorithms achieve high accuracy across the datasets. Random Forest shows the highest accuracy in most cases, followed by SVM and KNN. The accuracy values vary depending on the dataset and parameter settings selected by the user.

Table 3 presents the classification accuracy obtained from the implementation of three machine learning algorithms, namely K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest, across three different datasets: Iris, Wine, and Digit. The table shows the quantitative results of the classification process in terms of accuracy values produced by each algorithm on each dataset.

Based on the results shown in Table 3, all three algorithms demonstrate high classification performance on the tested datasets. For the Iris dataset, Random Forest achieved the highest accuracy, followed by SVM and KNN. Similarly, for the Wine dataset, Random Forest and SVM produced higher accuracy values compared to KNN, indicating that ensemble and margin-based classification methods perform effectively on structured datasets with clear class separation. On the Digit dataset, which contains more complex and high-dimensional data, Random Forest again produced the highest accuracy, while SVM also showed strong performance. KNN achieved slightly lower accuracy compared to the other algorithms but still maintained a high level of classification performance.

These results indicate that Random Forest consistently provides the most stable and highest accuracy across different datasets due to its ensemble learning mechanism, which combines multiple decision trees to improve prediction accuracy and reduce overfitting. SVM also demonstrates strong performance, particularly in datasets with complex patterns and higher dimensions, due to its ability to find optimal separating hyperplanes. Meanwhile, KNN performs effectively but tends to produce slightly lower accuracy, as its performance depends heavily on distance calculations and the distribution of data points.

Overall, Table 3 confirms that the implemented classification algorithms function properly within the developed Streamlit-based web data mining system. The accuracy results demonstrate that the system can successfully perform classification tasks and provide measurable performance outputs for comparison. These findings support the effectiveness of integrating multiple machine learning algorithms into a single interactive web-based platform for data mining and classification purposes.

Visualization Results

The system also provides visualization of classification results using matplotlib. The visualization displays the distribution of data and classification outcomes in graphical form. This feature allows users to observe patterns and relationships within the data more clearly. Each dataset produces different visualization outputs based on its characteristics and classification results.

Overall, the system successfully demonstrates the integration of classification algorithms into a web-based interactive platform. The results confirm that the implemented system can perform classification, display accuracy values, and provide visualization outputs effectively.

Discussion

The results of this study demonstrate that the integration of KNN, SVM, and Random Forest algorithms into a Streamlit-based web application can effectively perform data classification tasks and present results interactively. The successful implementation of the system indicates that web-based platforms can be used as accessible tools for implementing machine learning and data mining techniques.

The high accuracy values obtained from the classification process indicate that the selected algorithms perform well on the tested datasets. Random Forest achieved the highest accuracy in most cases due to its ensemble learning approach, which combines multiple decision trees to improve prediction performance and reduce overfitting. This finding is consistent with previous studies that highlight the effectiveness of Random Forest in achieving stable and accurate classification results. SVM also demonstrated strong performance, particularly in handling high-dimensional data, which aligns with its theoretical capability to find optimal hyperplanes for classification. Meanwhile, KNN produced slightly lower accuracy compared to the other algorithms but remained effective due to its simplicity and ability to classify data based on similarity.

The implementation of Streamlit as the system development framework proved to be effective in providing an interactive and user-friendly interface. Users can easily select datasets, adjust algorithm parameters, and visualize classification results in real time. This supports previous research that emphasizes the advantages of Streamlit in developing interactive data science and machine learning applications. The integration of visualization features further enhances user understanding of classification results and data patterns.

When compared with previous studies, the findings of this research confirm that KNN, SVM, and Random Forest are effective classification algorithms with high accuracy levels. However, this study contributes by integrating these algorithms into a single web-based system that allows interactive experimentation and comparison. This integration provides added value by enabling users to evaluate algorithm performance directly within a unified platform.

Scientifically, this study contributes to the development of practical and accessible data mining systems by demonstrating how machine learning algorithms can be implemented within a web-based interactive environment. The developed system can be used as a learning tool, experimental platform, and practical application for classification tasks. Furthermore, the research supports the idea that interactive web technologies can enhance the usability and accessibility of machine learning applications for students, researchers, and practitioners.

5. Conclusion

This study successfully designed and implemented a Streamlit-based web data mining system that integrates K-Nearest Neighbor, Support Vector Machine, and Random Forest algorithms for classification tasks. The system provides an interactive platform for dataset selection, parameter configuration, accuracy evaluation, and visualization of classification results. The findings indicate that all implemented algorithms perform effectively, with Random Forest achieving the highest and most consistent accuracy across datasets, followed by SVM and KNN. The main advantage of the developed system lies in its user-friendly interface and ability to integrate multiple classification methods into a single interactive web-based platform, making it accessible for learning and experimentation. However, the system still has limitations, particularly in handling larger datasets and more complex real-time data processing. Therefore, future research is recommended to enhance system performance by integrating additional algorithms, optimizing computational efficiency, and expanding the system with larger and more diverse datasets to improve its applicability and effectiveness.

Acknowledgement

The authors would like to express their sincere gratitude to Universitas Nurul Jadid for providing academic support and facilities that enabled the completion of this research. Appreciation is also extended to the lecturers of the Informatics Engineering Study Program for their valuable guidance, suggestions, and encouragement throughout the research process. The authors are deeply thankful to all parties who contributed directly or indirectly to the development of this study, especially those who provided insights and support during the design and implementation of the system. Their contributions have been invaluable in completing this research successfully.

References

Chaurasiya, H., & Pandey, A. K. (2025). Data mining applications used in medical healthcare. In *Progressive Computational Intelligence, Information Technology and Networking* (pp. 236–240). https://doi.org/10.1201/9781003650010-38

Guleria, M., Sharma, L., Malik, U., Singh, N., & Nigam, A. (2026). Role of Emerging Techniques in Data Mining. In *Studies in Computational Intelligence* (Vol. 1240, pp. 437–448). https://doi.org/10.1007/978-3-032-06732-6_23

Gupta, S., Sharma, N., Kaur, G., & Sardana, A. (2024). Prediction of diabetes using voting classification algorithms. In *Medical Imaging Informatics: Machine learning, deep learning and big data analytics* (pp. 89–114). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85186820901&partnerID=40&md5=8a5c8664a3a359a3e3b955d766596a15

Halder, M., Shopnil, S., Arafat, Y., Chowdhury, M. M., Hossain Jobayer, S., & Farid, D. M. (2023). Clustering as a Catalyst for Big Data Classification (CC-BC). *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*. https://doi.org/10.1109/ICCIT60459.2023.10441188

Jabardi, M. (2025). Support Vector Machines: Theory, Algorithms, and Applications. *Infocommunications Journal*, *17*(1), 66–75. https://doi.org/10.36244/ICJ.2025.1.8

Mishra, A., Khan, M. H., Khan, W., Khan, M. Z., & Srivastava, N. K. (2022). A Comparative Study on Data Mining Approach Using Machine Learning Techniques: Prediction Perspective. In *EAI/Springer Innovations in Communication and Computing* (pp. 153–165). https://doi.org/10.1007/978-3-030-77746-3_11

Mohammed, S. F., & Mahdi, G. J. M. (2024). Non-linear support vector machine classification models using kernel tricks with applications. *AIP Conference Proceedings*, *3036*(1). https://doi.org/10.1063/5.0196147

Nantasenamat, C., Biswas, A., Nápoles-Duarte, J. M., Parker, M. I., & Dunbrack, R. L. (2023). Building bioinformatics web applications with Streamlit. In *Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development* (pp. 679–699). https://doi.org/10.1016/B978-0-443-18638-7.00001-3

Pan, Z., Pan, Y., Wang, Y., & Wang, W. (2021). A new globally adaptive k-nearest neighbor classifier based on local mean optimization. *Soft Computing*, *25*(3), 2417–2431. https://doi.org/10.1007/s00500-020-05311-x

Pandey, S. (2024). Streamlit Essentials: From basics to advanced data app development. In *Streamlit Essentials: From Basics to Advanced Data App Development*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105022971937&partnerID=40&md5=04546039d9c45110949cad11d0f99cee

Rada, R., Bedalli, E., Shurdhi, S., & Cico, B. (2023). A comparative analysis on prototype-based clustering methods. *12th Mediterranean Conference on Embedded Computing, MECO 2023*. https://doi.org/10.1109/MECO58584.2023.10154917

Raghavendra, S. (2022). Beginner's guide to streamlit with python: Build web-based data and machine learning applications. In *Beginner's Guide to Streamlit with Python: Build Web-Based Data and Machine Learning Applications*. https://doi.org/10.1007/978-1-4842-8983-9

Zhao, J., Lee, C.-D., Chen, G., & Zhang, J. (2024). Research on the Prediction Application of Multiple Classification Datasets Based on Random Forest Model. *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems, ICPICS 2024*, 156–161. https://doi.org/10.1109/ICPICS62053.2024.10795875