Journal of Computer Science

Original Research Paper

Performance Prediction Model for Predictive Maintenance Based on K-Nearest Neighbor Method for Air Navigation Equipment Facilities

Rachmat Hidayat and Ditdit Nugeraha Utama

Department of Computer Science, Bina Nusantara Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article history
Received: 23-10-2024
Revised: 29-11-2024
Accepted: 02-01-2025

Corresponding Author:
Rachmat Hidayat
Department of Computer Science, Bina Nusantara Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia
Email: rachmat.hidayat002@binus.ac.id

Abstract: The aging of air navigation facilities leads to a decline in equipment performance and reliability, posing significant challenges for aviation organizations in the large-scale replacement of outdated systems, as well as in managing budgets and resources. To address these issues, a performance prediction model for equipment was developed using the K-Nearest Neighbor (KNN) method, aimed at enhancing maintenance planning and budget management. The development process includes collecting damage report data from facilities, preprocessing the data to ensure quality and consistency, and applying the KNN algorithm to generate accurate predictions. The KNN model, with the parameter n_neighbors = 2, achieved a high accuracy of 89.13% on the test data, with the best performance in class 1 classification. These results demonstrate the superiority of KNN over other models, such as Random Forest, which achieved an accuracy of 77%, and Logistic Regression, which only reached 41%. This research not only validates the effectiveness of the KNN model in predicting the performance of air navigation equipment facilities but also contributes significantly to maintenance efficiency. By using the KNN method, aviation organizations can plan maintenance more proactively and efficiently, minimizing the risk of unexpected failures. Moreover, the model aids in more effective budget preparation by adjusting maintenance priorities according to the specific needs and conditions of the facilities. This research focuses on providing a practical and reliable solution for maintenance planning and improved budget management, ultimately enhancing the performance and safety of flight services.

Keywords: Data Mining, Prediction, K-Nearest Neighbour, Machine Learning, Performance

Introduction

The management of air navigation equipment facilities faces significant challenges, particularly with performance degradation as equipment ages. Aging is a common phenomenon in long-operating aviation infrastructures, impacting components such as hardware and supporting systems. This degradation affects the reliability of navigation services, increasing the risk of errors in determining position and direction. Over time, continuous use exacerbates these issues, raising the likelihood of sudden failures. When equipment fails, corrective actions are taken to restore functionality, focusing primarily on repairing or replacing faulty components. While this reactive approach minimizes initial costs, it can result in higher maintenance expenses and prolonged downtime, especially for critical equipment (Zhang *et al.*, 2022).

Budgetary constraints further complicate maintenance prioritization, as simultaneous replacements are often required for multiple facilities. Organizations must strike a balance between managing costs and maintaining operational efficiency to prevent disruptions to flight operations (Kabashkin *et al.*, 2024). Traditional approaches based on the economic life of equipment have been implemented to address these challenges, but their reliance

Science Publications

on mass replacements leads to significant expenditures. This has spurred interest in predictive maintenance strategies, which use data-driven methods to forecast equipment performance and schedule maintenance proactively.

The method that has been implemented to address this issue is using facility aging based on the economic life of the equipment. However, the drawback of this method is that the organization must replace the equipment en masse, which incurs significant costs. One of the methods offered to address this is using machine learning for predictive maintenance, which utilizes damage report data to predict facility performance and plan more effective maintenance. This data-based predictive maintenance has become an important strategy in maintaining operational efficiency and reducing costs (Celikmih *et al*., 2020).

Machine learning has emerged as a pivotal tool for predictive maintenance, with its ability to analyze damage report data and generate actionable insights. Among various machine learning methods, K-Nearest Neighbor (KNN) has proven effective in predicting equipment performance due to its capacity to identify local patterns within historical data (Inyang *et al*., 2023). KNN's simplicity and ability to handle non-linear data make it particularly suitable for navigation facility damage data, which often exhibit complex patterns (Sládek, 2023).

This study applies the KNN method to predict the performance of air navigation facilities, aiming to enhance prediction accuracy, enable timely maintenance, and reduce the risk of service failures. By providing efficient and timely maintenance planning recommendations, this approach addresses the industry's dual priorities of reliability and cost-effectiveness. Previous studies have demonstrated KNN's versatility in diverse domains, such as predicting damage to centrifugal pumps using vibration signals (Chen *et al*., 2021), detecting false alarms in wind turbines (Peco Chacón *et al*., 2023), improving GNSS data accuracy for drone navigation (Aziez *et al*., 2021) and predicting air traffic patterns (Zhuang and Cao, 2023). These applications validate KNN's robustness in handling complex data, making it well-suited for predictive maintenance in dynamic and critical environments like air navigation systems.

However, challenges remain, particularly related to class imbalances in damage report data, which can affect model performance. Addressing these limitations involves incorporating oversampling techniques or exploring alternative algorithms such as Decision Tree or XG Boost for improved accuracy. Future research should also consider integrating real-time sensor data and environmental variables to refine predictive capabilities further.

By building on previous work, this research focuses on the unique challenges of air navigation facilities, aiming to enhance the predictive performance of KNN while addressing operational complexities. The study not only seeks to improve prediction accuracy but also provides practical insights for optimizing maintenance budgets and ensuring reliable navigation services, meeting the aviation industry's critical needs.

Materials and Methods

Prediction models are approaches used to estimate future values or behaviors based on available historical data. The main goal of predictive models is to generate accurate and reliable forecasts by identifying patterns and trends observed in past data. Prediction involves analyzing historical data using analytical algorithms to estimate future events or outcomes (Zhao, 2022).

A machine learning approach is applied in this study to predict equipment maintenance using historical data. Machine learning, a field within artificial intelligence, focuses on creating algorithms that enable systems to detect patterns in data (Choi *et al*., 2020). Data mining plays a critical role in this research, as it utilizes historical data to uncover hidden patterns related to equipment failures. This process provides insights into potential equipment damage that may not be visible through manual inspection (Gupta and Chandra, 2020). In this context, data mining is employed to identify factors contributing to the failure of air navigation equipment.

Among the many algorithms used in predictive modeling, KNN stands out for its effectiveness in classification based on data proximity. KNN categorizes new items by evaluating their similarity to existing data points (Arifiansyah, 2023). This research employs the KNN algorithm to forecast the performance of air navigation equipment. This method was chosen due to its ability to categorize data based on feature similarity, making it well-suited to the type of data used, such as facility damage reports. KNN has demonstrated effectiveness in handling non-linear data patterns and enabling real-time monitoring of equipment conditions to predict operational failures. Comprehensive and real-time data collection enhances prediction models (Namoun and Alshanqiti, 2020). This research is designed with a quantitative approach. The process of predicting equipment performance involves facility damage report data that is collected and analyzed using KNN. This model is developed by simulating based on the available historical data to produce accurate predictions. The user defines the number of nearest neighbors, represented as k, which influences the classification process (Isnain *et al*., 2021). Optimizing k is essential for model performance, as the value is data-dependent (Lubis *et al*., 2020). Additionally, the algorithm's effectiveness hinges on selecting an appropriate distance metric (Amrutha and Prabu, 2021). This study determined $k = 2$ as the optimal value through cross-validation, balancing model accuracy, and generalization.

Alternative methods such as Random Forest and Logistic Regression were considered but found less suitable for this dataset. Random Forest achieved only 77% accuracy and Logistic Regression struggled with the non-linear patterns present in the data. The decision to adopt KNN was based on its capacity to handle non-linear patterns and its compatibility with the dataset's characteristics.

To ensure the reliability of the predictions produced by the KNN model, the evaluation process employed the cross-validation technique. This approach allowed the model to be thoroughly tested on different subsets of the dataset. Cross-validation is considered a robust and comprehensive method for evaluating predictive models. This technique demonstrated that the KNN model provided consistent predictions with a high level of accuracy, making it an optimal choice for contexts such as facility maintenance and complex systems. The validation findings confirm that the KNN model utilized in this study offers reliable predictions, making it well-suited for the context of air navigation facility maintenance (Nti *et al.*, 2021).

Predicting the performance of air navigation equipment is critical for enhancing operational reliability and minimizing unexpected failures. The aviation industry demands effective methods to forecast equipment damage and optimize maintenance schedules. For instance, KNN has been applied successfully in other contexts, such as predicting false alarms in wind turbines, where it reduced unnecessary operational costs by identifying 22% of false alarms (Peco Chacón *et al*., 2023).

Applying this method to air navigation facilities introduces additional challenges due to higher operational complexity and more dynamic variations in system conditions. Compared to wind turbines, the complexity and variability of air navigation systems require models that can predict not only physical failures but also operational performance, incorporating variables such as average failure interval and average repair time.

In general, the KNN model has proven effective in various scientific applications. This research expands on previous studies by applying KNN to more complex scenarios in air navigation systems, addressing limitations in data coverage, and offering practical solutions to enhance equipment performance forecasting. As a result, this research contributes to the development of predictive methods in the aviation sector while demonstrating KNN's relevance across various fields of predictive technology.

Addressing the challenges of managing aging navigation facilities requires identifying and evaluating the condition of equipment that needs replacement. This involves analyzing factors such as economic lifespan, actual performance, and their impact on flight operations. Such an approach is essential for prioritizing budget allocation to ensure critical equipment receives proper attention. A mature infrastructure management strategy capable of predicting and identifying damage conditions provides more efficient solutions for managing aging facilities (Assaad and El-Adaway, 2020). The replacement of critical assets must be carefully planned to avoid unexpected financial losses while maintaining operational reliability (Balanta *et al.*, 2023).

Table 1: Data parameters

No.	Attribute	Valid entries	Data type
0	Location	4599 valid	text
1	Equipment type	4599 valid	text
2	Equipment identifier	4599 valid	int
3	Fault report	4599 valid	int
4	Average failure interval	4599 valid	float
5	Average repair time	4599 valid	float
6	Operational duration	4599 valid	int
7	Actual operation time	4599 valid	float
8	System availability	4599 valid	float

The study analyzed five years of historical data on recurring equipment failures, focusing on key attributes such as Location, Equipment Type, Equipment Identifier, Fault Reports, Average Failure Interval, Average Repair Time, Operational Duration, Actual Operation Time, and System Availability, as summarized in Table (1). Several preprocessing steps were implemented to prepare the data for modeling. Missing values were addressed using mean imputation to prevent gaps from affecting the analysis. The data was normalized with Min-Max scaling to standardize all parameters within a uniform range, ensuring consistency across the dataset. To address class imbalance, oversampling techniques were applied, enhancing the model's performance by providing sufficient representation for underrepresented categories during training.

Next, feature selection was performed to identify the attributes most relevant to equipment failures. This process reduced dimensionality, improved computational efficiency, and increased the model's interpretability. Techniques such as correlation analysis and mutual information were used to remove redundant or irrelevant features, retaining only significant predictors. The dataset was then split into training and testing subsets using a stratified sampling method, which preserved the distribution of failure classes across both sets. This approach reduced the risk of biased performance metrics and enabled robust model evaluation. By integrating comprehensive preprocessing, feature selection, and balanced dataset preparation, the study established a strong foundation for accurately predicting recurring equipment failures.

The data collection stage involved gathering historical data from key systems, including communication, navigation, surveillance, and automation equipment. This dataset spans five years of operational records and includes critical parameters: Location, Equipment Type, Equipment Identifier, Fault Reports, Average Failure Interval, Average Repair Time, Operational Duration, Actual Operation Time, and System Availability. These attributes provide comprehensive insights into equipment performance, enabling the identification of patterns and trends essential for effective predictive maintenance.

The research methodology, as illustrated in Fig. (1), is structured into four stages: Data collection, data preprocessing, model design, and model evaluation. Each stage is designed to systematically develop a reliable predictive model for air navigation equipment performance.

In the data preprocessing stage, the collected dataset was refined to ensure quality and consistency. Missing values were addressed using mean imputation to avoid gaps in the analysis. Data normalization was performed using Min-Max scaling to standardize all parameters within a uniform range, ensuring compatibility with the KNN algorithm. Additionally, to handle class imbalance caused by underrepresented failure categories, oversampling techniques were applied, thereby improving the model's ability to generalize across all classes.

The model design phase focused on the implementation of the KNN algorithm, selected for its simplicity and proven effectiveness in handling non-linear data patterns. The value of k = 2 was determined through cross-validation to achieve optimal accuracy and generalization. This algorithm categorizes equipment performance by evaluating the proximity of data points within the dataset, making it particularly suitable for predicting recurring failures in complex systems.

Finally, in the model evaluation stage, the KNN model's performance was rigorously tested using unseen test data. Evaluation metrics such as accuracy, precision, recall, and f1-score were employed to assess the reliability and robustness of the predictions in real-world scenarios. This comprehensive evaluation ensured that the model could effectively support maintenance planning and operational decision-making for air navigation facilities.

This structured methodology leverages robust data collection and preprocessing techniques combined with an optimized KNN design, providing a reliable framework for predicting equipment performance in communication, navigation, surveillance, and automation systems.

Model evaluation is conducted using metrics such as accuracy, precision, recall, and f1-score to validate the model performance (Vujović, 2021). This modeling strategy aims to predict air navigation equipment failures, enabling proactive maintenance and reducing operational disruptions. By forecasting potential issues, it optimizes maintenance schedules, minimizes downtime, and enhances reliability and safety. This approach ensures critical equipment performs optimally while improving resource allocation for maintenance activities.

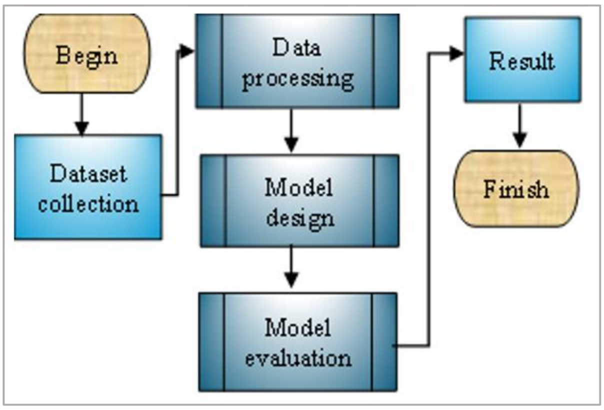

Fig. 1: Research stages

Results

The results of this study are organized to emphasize the effectiveness of the K-Nearest Neighbor (KNN) model in predicting the performance of air navigation facilities. The research process followed a systematic methodology that included data collection, preprocessing, model design, and evaluation, ensuring a robust framework for predictive maintenance. Key parameters used in the analysis include attributes such as location, equipment type, operational time, and availability. Preprocessing steps were undertaken to prepare the data for effective modeling, including imputing missing data using mean values and normalizing parameters to ensure consistency across the dataset. These preprocessing steps were essential for enhancing the quality of the dataset and ensuring its suitability for analysis. Additionally, the study highlights how machine learning methods, including KNN, Random Forest, and Logistic Regression, are integrated into the predictive framework. It focuses on critical attributes for performance prediction, such as average failure interval and availability. This comprehensive approach enhances the understanding of the study's methodological rigor and ensures the replicability of the proposed framework.

Class diagrams are used because they effectively visualize the structure and relationships between system elements. This visualization simplifies the understanding of the roles of each component and their interactions, contributing to a more organized and efficient system design. UML Class Diagrams are particularly well-suited for representing object-oriented systems and facilitating software design (Gosala *et al*., 2021). Neighbor, Random Forest, and Logistic Regression used to analyze historical equipment data.

The interconnection between system elements is depicted in the class diagram shown in Fig. (2), which

DOI: 10.3844/jcssp.2025.800.809

illustrates the structure and relationships among system components used in predicting the performance of air navigation equipment. The diagram highlights three key machine-learning algorithms: KNN, random forest, and logistic regression. These algorithms are employed to predict equipment performance using historical data. Each piece of equipment, such as communication, navigation, surveillance, and automation tools, is represented by the facility equipment class, which includes critical attributes like location, average time between failures, and availability.

The performance prediction process begins with data collection and preprocessing. The prepared data is then analyzed using each machine learning algorithm to generate predictions and evaluate model performance. This evaluation involves key metrics, including accuracy, precision, recall, and f1-score. The generalization relationship between the facility equipment class and specific systems (e.g., communication, navigation, surveillance, automation) ensures that the algorithms can be applied broadly to various types of equipment in the air navigation environment. This approach supports enhanced predictive maintenance efficiency and improved equipment reliability.

Thus, the Class Diagram is essential for developing structured object-oriented systems, enabling the seamless integration of machine-learning methods to enhance the efficiency and accuracy of equipment maintenance. By clearly defining the relationships between system components and their attributes, the Class Diagram ensures that all relevant data is organized in a way that supports effective predictive modeling. This structured approach not only facilitates the implementation of machine learning algorithms but also ensures that the system can be easily modified or expanded to accommodate future requirements or technological advancements.

Additionally, the use of class diagrams provides a robust framework for aligning system design with operational goals. By incorporating critical attributes such as average time between failures, repair times, and system availability, the diagram ensures that the data required for accurate performance predictions is readily accessible and systematically analyzed. This clarity in design allows for more reliable predictions, enabling maintenance teams to make informed decisions that minimize downtime and optimize resource allocation. The Class Diagram thus acts as a foundation for creating intelligent maintenance systems that can adapt to the evolving needs of complex operational environments.

The training procedure for the KNN algorithm to predict the performance of air navigation facilities begins with the collection of data from diverse sources, including parameters such as location, equipment type, average failure interval, average repair duration, and operational duration. Once the data is collected, it undergoes preprocessing steps, such as cleaning, normalization, and feature selection, to ensure its suitability for analysis. The dataset is then divided into training and testing subsets to facilitate model development and evaluation.

The KNN model is developed by determining the optimal value of *k* and calculating distances using metrics such as Euclidean distance. Based on these calculated distances, the model identifies the nearest neighbors to make predictions. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and f1-score, ensuring its reliability and effectiveness in predicting equipment performance.

After evaluation, the KNN model supports decision-making processes related to equipment maintenance or replacement. This predictive capability enables technicians to make accurate, data-driven decisions, thereby improving operational efficiency and extending the lifespan of air navigation equipment. The workflow of the KNN algorithm is visually represented in the form of an activity diagram, as shown in Fig. (3).

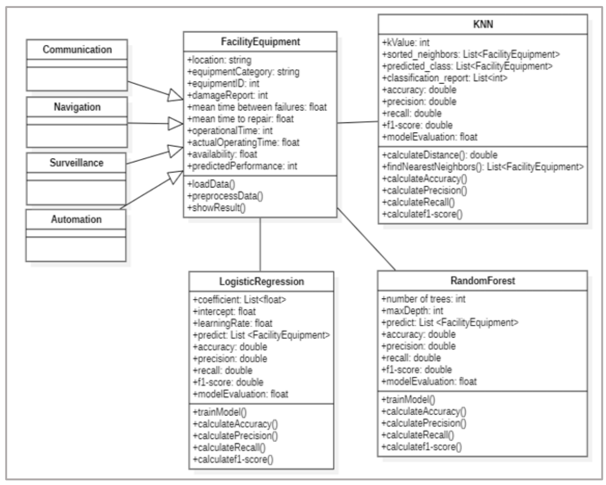

Fig. 2: Class diagram of entities

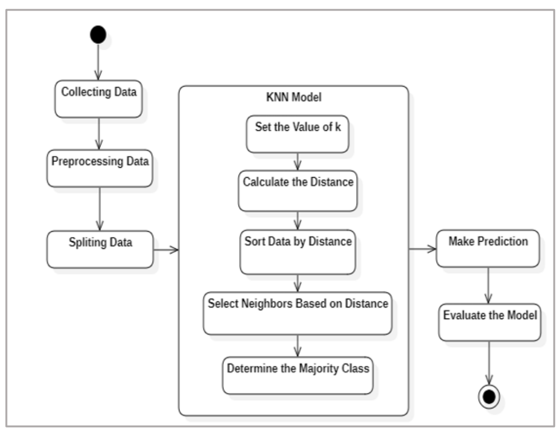

Fig. 3: Activity diagram algorithm model

The Activity Diagram illustrates the systematic workflow of the KNN algorithm for predicting equipment performance. The process begins with the data collection stage, where relevant information is gathered from various sources. Next, the data undergoes preprocessing, which involves cleaning and normalization to ensure the dataset is prepared for analysis. Once the preprocessing is complete, the data is split into training and testing sets to enable model training and evaluation.

Within the KNN model, several key steps are performed. First, the value of k is set, which determines the number of nearest neighbors to consider during the prediction process. The algorithm then calculates the distances between data points using a suitable metric, such as Euclidean distance. These distances are subsequently sorted and the closest neighbors are selected based on the calculated values. From these selected neighbors, the algorithm identifies the majority class, which forms the basis for making predictions.

Once predictions are made, the model's performance is evaluated using various metrics, such as accuracy, precision, recall, and f1-score. This workflow ensures a structured and efficient approach to leveraging the KNN algorithm for predictive maintenance, enabling accurate predictions and informed decision-making for equipment reliability.

The interpretation of the results is summarized in Table (2). The evaluation of the K-Nearest Neighbors (KNN) model was conducted using a classification report that includes precision, recall, and f1-score for each class. Class 1 exhibited the best performance, achieving a precision of 0.92, a recall of 0.97, and an f1-score of 0.95. Classes 0 and 2 also demonstrated strong performance, with f1-scores of 0.90 and 0.84, respectively. However, classes 3 and 4 showed weaknesses, particularly in recall.

Overall, the KNN model demonstrated strong performance, achieving a macro-average f1-score of 0.86 and a weighted-average f1-score of 0.89. The macro-average f1-score reflects consistent performance across all classes, regardless of sample size. In contrast, the weighted-average f1-score highlights the algorithm's overall effectiveness, accounting for the sample distribution and yielding more accurate predictions for the more prevalent classes. While the KNN model performs well in classifying data and managing class imbalance, there remains room for improvement in classes with lower recall. Further tuning of hyperparameters, such as the value of k and distance metrics, may enhance model performance. Incorporating additional data preprocessing techniques could also address class imbalance and improve recall for weaker classes. Future work could explore ensemble methods to combine KNN with other algorithms for enhanced classification accuracy.

The dataset used in this study exhibited class imbalance, which can negatively affect the performance of machine learning models, particularly for underrepresented classes. To address this issue, oversampling techniques were applied during the preprocessing stage to balance the class distribution. By artificially increasing the representation of minority classes, the model was better able to learn patterns across all classes. This approach enhanced the KNN model's ability to handle imbalanced data, as demonstrated by the improved f1-scores for minority classes in Table (2). Despite these improvements, some challenges remain, particularly in classes with lower recall values, indicating that further optimization may be necessary.

After making predictions, the next step is to evaluate the KNN model through verification and validation procedures. Model verification ensures that the model is constructed according to the algorithm's specifications, while validation assesses its performance on unseen data.

Several methods, such as the ROC curve and AUC, are used to evaluate the model's classification accuracy. Fig. (4) presents the ROC curve and AUC results for each class. The ROC curve and AUC are essential metrics for assessing classification models, particularly in evaluating the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR).

Table 2: KNN model predictions

Class	Recall	Precision	F1-Score	Support
0	0.91	0.89	0.90	232
1	0.97	0.92	0.95	340
2	0.86	0.83	0.84	181
3	0.70	0.91	0.80	105
4	0.77	0.86	0.81	62
Accuracy				0.89

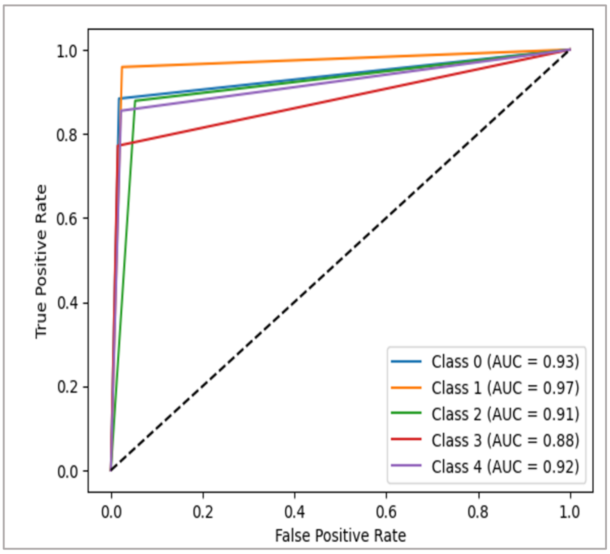

Fig. 4: ROC Curve and AUC

The analysis revealed that class 1 achieved the highest performance with an AUC of 0.97, followed by classes 0 and 4, which had AUCs of 0.93 and 0.92, respectively. Classes 2 and 3 also performed well, with AUCs of 0.91 and 0.88. The high AUC values (ranging from 0.88 to 0.97) confirm the model's effectiveness in distinguishing instances. Consistent performance across most classes indicates that the KNN model handles variations in the dataset effectively. These results highlight the robustness of the KNN algorithm in classification tasks, even when faced with imbalanced data. The findings also underscore the importance of preprocessing steps, such as oversampling, in achieving reliable model performance. Future improvements could involve testing additional metrics or tuning hyper parameters to further refine the model's predictive capabilities.

Additionally, the confusion matrix provides a detailed summary of classification errors. By analyzing the confusion matrix, the strengths and weaknesses of the classification system can be identified. Classes 0 and 1 demonstrated strong performance, with high True Positive rates and minimal errors. However, classes 2, 3, and 4 faced greater challenges, exhibiting higher False positive and false negative rates. The overall accuracy of the KNN model is 80.39% but the confusion matrix analysis shows room for improvement, especially in reducing prediction errors in certain classes. As illustrated in Fig. (5), The confusion matrix provides an extensive overview of the classification system's outcomes, emphasizing both the accurate and inaccurate predictions made for each class.

To further validate the model, its performance was compared with two additional models: Random forest and logistic regression. The comparison aimed to identify the best model for the dataset based on metrics such as accuracy, precision, recall, and f1-score.

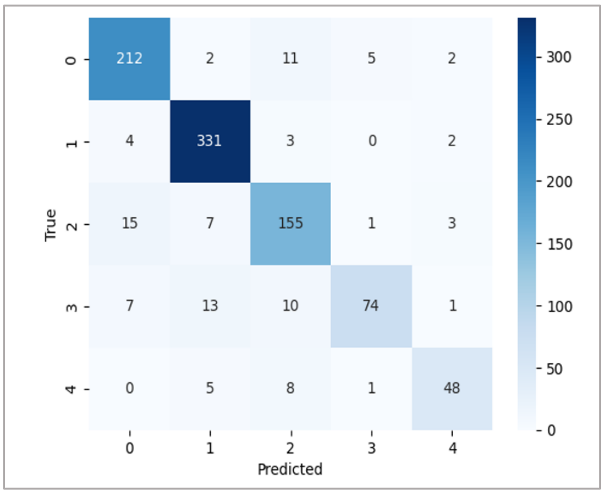

Fig. 5: Confusion matrix of KNN

Table 3: Model comparison evaluation results

Model	Precision	Recall	F1-Score	Accuracy
KNN	0.89	0.89	0.89	0.89
Random Forest	0.77	0.77	0.77	0.77
Logistic regression	0.35	0.41	0.32	0.41

The comparison results indicate that the KNN model outperforms the others, achieving 89% accuracy along with strong precision, recall, and f1-score of 0.89. This demonstrates that KNN is capable of providing accurate and balanced predictions for data classification. Meanwhile, Random Forest achieves an accuracy of 77%, which, although quite good, still falls short compared to KNN. This model may require further tuning to improve its performance. Conversely, Logistic Regression shows the lowest performance, with an accuracy of only 41%, along with low precision, recall, and f1-score values. This suggests that Logistic Regression is not suitable for this classification problem or may require better parameter adjustments. Overall, KNN emerged as the best model for this dataset, while Random Forest served as a fairly good alternative, and Logistic Regression was less effective in this context. Table (3) presents the findings from the evaluation of the machine learning model comparison.

Discussion

The results indicate that the KNN model provides accurate predictions for air navigation facility performance, particularly for frequently occurring classes. The superior classification performance, as reflected in the f1-score and AUC values, demonstrates the effectiveness of KNN in predictive maintenance. The structured representation highlights the comprehensive design of the predictive framework, ensuring that relevant attributes are incorporated into the modelling process.

The training and evaluation workflow illustrates the systematic approach undertaken to optimize KNN's predictive accuracy. The confusion matrix shows that misclassification occurred primarily in minority classes, which can be attributed to class imbalance. Although oversampling improved performance, further enhancements could be achieved through advanced resampling techniques or hyper parameter tuning.

The comparison with Random Forest and Logistic Regression confirms that KNN is the most suitable model for this dataset. The poor performance of Logistic Regression suggests that linear decision boundaries are insufficient for complex patterns in air navigation equipment data.

The findings of this study demonstrate the effectiveness of the KNN algorithm in predictive maintenance and align with several prior research efforts. For instance, previous studies such as those by Zhang *et al.* (2022) on maritime vessel trajectory prediction and (Peco Chacón *et al.*, 2023) on wind turbine maintenance have highlighted the capability of KNN to handle non-linear datasets and improve prediction accuracy. Zhang *et al.* (2022) achieved an accuracy of 99.2%, whereas (Peco Chacón *et al.*, 2023) reached 98% accuracy in identifying false alarms.

In the context of air navigation, this study contributes to the field by focusing specifically on equipment performance prediction and addressing class imbalance through oversampling, achieving an overall accuracy of 89%. Unlike earlier works, such as (Jang, 2023), which applied KNN for frost prediction on bridges with 95% accuracy, this study incorporates multiple evaluation metrics (precision, recall, f1-score, accuracy) and compares the KNN model with Random Forest and Logistic Regression to validate its robustness.

Additionally, while prior research like that of (Manyol *et al.*, 2022) utilized KNN for data imputations in large-scale datasets, this study extends its application to predictive maintenance for critical air navigation infrastructure, highlighting its scalability and adaptability. These comparisons emphasize the novelty of integrating KNN into predictive maintenance frameworks for aviation systems, contributing to enhanced operational reliability and maintenance planning.

This study corroborates earlier research that employed KNN for equipment failure prediction, showing high accuracy. However, the inclusion of more comprehensive data has yielded even better results.

Conclusion

This study aims to predict the performance of air navigation equipment facilities using the KNN algorithm. The key findings demonstrate that the KNN model can achieve a high accuracy of 89%, with an average precision of 0.87 and a recall of 0.83. This indicates that KNN is effective in predicting equipment failures, especially in classes with a high frequency of failures. Further verification and evaluation using cross-validation, the ROC curve, AUC, and the Confusion Matrix show very good performance.

The primary contribution of this study is the deployment of an enhanced KNN model for predicting the maintenance of air navigation equipment, providing more accurate predictive solutions compared to models used in earlier research. Thus, this research contributes to the development of more reliable predictive maintenance methods in the field of air navigation, reducing the likelihood of system failures and improving operational efficiency.

From a practical perspective, the results of this research can be utilized by air navigation facility maintenance management to be more proactive in detecting equipment malfunctions and taking preventive actions before a larger failure occurs. Additionally, the theoretical implications of this research enhance the understanding of utilizing machine learning algorithms, such as KNN, for critical equipment maintenance within the aviation sector.

This research shows that KNN is an effective model for predicting air navigation equipment failures and can provide practical benefits in maintenance management. The use of this method not only aids in better decision-making related to maintenance but also promotes higher efficiency in the operations of the aviation industry.

One limitation of the study is that it relies solely on damage reports. Integrating real-time sensor data or additional variables, such as environmental conditions, could improve prediction accuracy. Future research should explore models like Decision Tree or XG Boost and incorporate more variables to enhance prediction accuracy.

Acknowledgment

We would like to thank the Research Interest Group on Quantitative and Decision Sciences (RIG Q&DS) and Bina Nusantara University for their helpful support and facilitation in fostering collaboration and stimulating insightful discussions throughout the research work.

Funding Information

This study was made possible through the generous funding provided by Bina Nusantara University. The authors express their heartfelt gratitude for the financial support from the University's Research Fund, which was instrumental in facilitating the research and enabling the dissemination of the findings presented in this journal paper.

Author’s Contributions

Rachmat Hidayat: Carrying out all the research, performing the analysis, and written the paper.

Ditdit Nugeraha Utama: Supervising the research and reviewed the paper.

Ethics

This manuscript represents original work by the authors and has not been published elsewhere. The authors have thoroughly reviewed and approved the content, confirming its accuracy and adherence to academic standards. The research and publication processes were conducted with a strong commitment to integrity and ethical principles. No ethical issues or conflicts of interest arose during this study. Furthermore, we adhered to the ethical guidelines established by Bina Nusantara University to ensure responsible conduct throughout the research.

References

Amrutha, K., & Prabu, P. (2021). ML-Based Sign Language Recognition System. *2021 International Conference on Innovative Trends in Information Technology (ICITIIT)*, 1-6. https://doi.org/10.1109/icitiit51526.2021.9399594

Arifiansyah, M. N. (2023). Prediction of Life Expectancy of Lung Cancer Patients Post Thoracic Surgery using K-Nearest Neighbors and Bat Algorithm. *Journal of Advances in Information Systems and Technology*, *4*(2), 170-179. https://doi.org/10.15294/jaist.v4i2.60846

Assaad, R., & El-adaway, I. H. (2020). Bridge Infrastructure Asset Management System: Comparative Computational Machine Learning Approach for Evaluating and Predicting Deck Deterioration Conditions. *Journal of Infrastructure Systems*, *26*(3). https://doi.org/10.1061/(asce)is.1943-555x.0000572

Aziez, S. A., Al-Hemeary, N., Reja, A. H., Zsedrovits, T., & Cserey, G. (2021). Using KNN Algorithm Predictor for Data Synchronization of Ultra-Tight GNSS/INS Integration. *Electronics*, *10*(13), 1513. https://doi.org/10.3390/electronics10131513

Balanta, J., Rivera, S., Romero, A., & Coria, G. (2023). Planning and Optimizing the Replacement Strategies of Power Transformers: Literature Review. *Energies*, *16*(11), 4448. https://doi.org/10.3390/en16114448

Celikmih, K., Inan, O., & Uguz, H. (2020). Failure Prediction of Aircraft Equipment Using Machine Learning with a Hybrid Data Preparation Method. *Scientific Programming*, *2020*, 1-10. https://doi.org/10.1155/2020/8616039

Chen, Y., Yuan, J., Luo, Y., & Zhang, W. (2021). Fault Prediction of Centrifugal Pump Based on Improved KNN. *Shock and Vibration*, *2021*(1). https://doi.org/10.1155/2021/7306131

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Peter Campbell, J. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science and Technology*, *9*(2). https://doi.org/https://doi.org/10.1167/tvst.9.2.14

Gosala, B., Chowdhuri, S. R., Singh, J., Gupta, M., & Mishra, A. (2021). Automatic Classification of UML Class Diagrams Using Deep Learning Technique: Convolutional Neural Network. *Applied Sciences*, *11*(9), 4267. https://doi.org/10.3390/app11094267

Gupta, M. K., & Chandra, P. (2020). A Comprehensive Survey of Data Mining. *International Journal of Information Technology*, *12*(4), 1243-1257. https://doi.org/10.1007/s41870-020-00427-7

Inyang, U. G., Ijebu, F. F., Osang, F. B., Afoluronsho, A. A., Udoh, S. S., & Eyoh, I. J. (2023). A Dataset-Driven Parameter Tuning Approach for Enhanced K-Nearest Neighbour Algorithm Performance. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(1), 380-391. https://doi.org/10.18517/ijaseit.13.1.16706

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm for Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121. https://doi.org/10.22146/ijccs.65176

Jang, J. (2023). Bridge Frost Prediction Using K-nearest Neighbor Classifier. *The Open Transportation Journal*, *17*(1). https://doi.org/10.2174/0126671212263648230921115634

Kabashkin, I., Perekrestov, V., Tyncherov, T., Shoshin, L., & Susanin, V. (2024). Framework for Integration of Health Monitoring Systems in Life Cycle Management for Aviation Sustainability and Cost Efficiency. *Sustainability*, *16*(14), 6154. https://doi.org/10.3390/su16146154

Lubis, Z., Sihombing, P., & Mawengkang, H. (2020). Optimization of K Value at the K-NN algorithm in clustering using the expectation maximization algorithm. *IOP Conference Series: Materials Science and Engineering*, *725*(1), 012133. https://doi.org/10.1088/1757-899x/725/1/012133

Manyol, M., Eke, S., Massoma, A. J. M., Biboum, A., & Mouangue, R. (2022). Preprocessing Approach for Power Transformer Maintenance Data Mining Based on k-Nearest Neighbor Completion and Principal Component Analysis. *International Transactions on Electrical Energy Systems*, *2022*, 1–10. https://doi.org/10.1155/2022/8546588

Namoun, A., & Alshanqiti, A. (2020). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Applied Sciences*, *11*(1), 237. https://doi.org/10.3390/app11010237

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation. In *International Journal of Information Technology and Computer Science* (Vol. 13, Issue 6, pp. 61-71). https://doi.org/10.5815/ijitcs.2021.06.05

Peco Chacón, A. M., Segovia Ramírez, I., & García Márquez, F. P. (2023). K-Nearest Neighbour and K-Fold Cross-Validation Used in Wind Turbines for False Alarm Detection. *Sustainable Futures*, *6*, 100132. https://doi.org/10.1016/j.sftr.2023.100132

DOI: 10.3844/jcssp.2025.800.809

Sládek, D. (2023). Application of K-Nearest Neighbor Classification for Static Webcams Visibility Observation. *Advances in Meteorology*, *2023*, 1-19. https://doi.org/10.1155/2023/6285569

Vujović, Ž. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, *12*(6), 599-606. https://doi.org/https://doi.org/10.14569/IJACSA.2021.0120670

Zhao, L. (2022). Event Prediction in the Big Data Era: A Systematic Survey. *ACM Computing Surveys*, *54*(5), 1-37. https://doi.org/10.1145/3450287

Zhang, P., Gao, Z., Cao, L., Dong, F., Zou, Y., Wang, K., Zhang, Y., & Sun, P. (2022). Marine Systems and Equipment Prognostics and Health Management: A Systematic Review from Health Condition Monitoring to Maintenance Strategy. *Machines*, *10*(2), 72. https://doi.org/10.3390/machines10020072

Zhuang, W., & Cao, Y. (2023). Short-Term Traffic Flow Prediction Based on a K-Nearest Neighbor and Bidirectional Long Short-Term Memory Model. *Applied Sciences*, *13*(4), 2681. https://doi.org/10.3390/app13042681