Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 8.9
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.8.9

Reliability and Performance Improvement Through AI: A Case Study of Sleeper Train Fleet Critical Systems

B. Alkali and C. F. Nworah

Department of Mechanical Engineering, Glasgow Caledonian University
Glasgow, United Kingdom

Abstract

This paper explores the application of machine learning techniques to enhance the reliability of sleeper train critical systems contributing to service operation disruptions. The primary objective of the paper is to develop predictive models capable of identifying and predicting faults to facilitate proactive maintenance. An attempt is made to develop and use two machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), to analyse condition monitoring data and predict system failures. An exploratory data analysis was conducted, and some limitations and imbalance in the dataset were observed. The Synthetic Minority Oversampling Technique (SMOTE) was applied to effectively balance the class distribution and improve the model performance. The proposed models were evaluated using precision, recall, F1-score, and the overall accuracy metrics. The results demonstrated that the Random Forest (RF) model significantly outperformed the Support Vector Machine (SVM) model, thereby achieving a well-balanced trade-off between precision and recall. After addressing data imbalance, the RF model achieved an overall accuracy of 75%, compared to 65% accuracy with imbalanced data. The precision and recall scores for the RF model indicated reliable performance in both fault detection and prediction. In contrast, the SVM model exhibited lower performance metrics, especially in identifying faulty incidents, it achieved perfect recall but low precision for one class, also indicating many false positives. The SVM model on the other hand achieved an overall accuracy of 48% before addressing data imbalance, which improved to 70% with balanced data. This paper contribution

emphasis on the importance of data quality, feature selection, and model ability in handling data imbalance to support decision making.

Keywords: random forest, artificial intelligence, rolling stock, machine learning, predictive model, reliability.

1 Introduction

Railway capacity represents a critical aspect of transport infrastructure management in Great Britain. The capacity utilisation aspect delineates the upper threshold of operability for railway service operations management within a given corridor, station, or network [1]. In the most recent quarter, Great Britain saw a bustling rail network with a total of 417 million passenger journeys, spanning a significant distance of 15.2 billion passenger, and contributing substantially to the economy with a total passenger revenue of £2.6 billion [2]. The safety standards are inherent in railway operations and extend beyond passengers to encompass the well-being of personnel involved in maintenance, logistics, and management, highlighting the multifaceted nature of ensuring safe and reliable rail service operations.

The rising appeal of sleeper trains for long-distance travel, offer travellers the convenience of overnight accommodations and the luxury of a sleeping berth throughout the journey, operating predominantly during nighttime hours with a distinctively comfortable and immersive travel experience is highlighted by Papa in [3]. The MK5 Caledonian Sleeper, a prominent example of such services, operates in Scotland, linking London and Scotland destinations. The Scottish Government's strategic move to establish the Caledonian Sleeper as an independent franchise working towards enhancing the standards and overall experience of overnight rail travel and hence delivery passenger satisfaction. However, within the sleeper fleet, critical systems such as the water closet system is prone to various defects and failures, leading to operational disruptions, passenger inconvenience, and heightened maintenance costs. Present maintenance practices are often reactive, resulting in suboptimal system performance and downtime. There is need to develop proactive maintenance strategies to identify faults early and predict potential failures pre-emptively.

In railway maintenance, apart from corrective and periodic maintenance strategies, predictive maintenance (PdM) is receiving increasing attention because of its ability to predict failures, which minimises service interruptions and lowers the number of unnecessary inspections [4]. Predictive maintenance has the potential to significantly mitigate railway maintenance costs, enhance reliability, and bolster asset availability. Sensors play a pivotal role in monitoring process conditions, however, despite the implementation of monitoring and control mechanisms, processes can deviate from their safe operating parameters due to faults in the overall system. Mou and Zhao [5] believe that by using advanced diagnostic techniques such as machine learning algorithms and condition monitoring systems, one can proactively identify potential faults, minimise downtime, and effectively control costs, thereby fostering a safer and more sustainable operational environment. However, the paper in [6] noticed that this trajectory presents considerable challenges for ensuring the stability of trains and maintaining the safety performance of railway infrastructure.

Machine learning enables the discovery of precise degradation patterns, facilitates the development of predictive models, supports decision-making processes, and facilitates the generation of optimized maintenance plans [7]. A notable challenge in the railway sector persists wherein a vast volume of data is collected without always being effectively transformed into actionable insights. Despite the exponential growth in data accumulation, the industry continues to grapple with the absence of automated solutions and underutilization of machine learning techniques to address operational challenges [8].

2 Aim and Objectives

The aim of this paper is to demonstrate how an advanced AI and ML model algorithms can be utilised to analyse conditioned-based monitoring data to pinpoint defect categories of the sleeper train fleet critical systems and use model predictive capabilities to forecast potential time of failure for proactive maintenance interventions.

2.1 Sleeper Coach Technical Description

The design of the Sleeper Coaches is meticulously crafted to ensure passengers enjoy a comfortable and restful overnight journey. These Sleeper Coaches are equipped with essential amenities to enhance the onboard experience as shown in Figure 1. The sleeper fleet considered in this paper, the water closet system is prone to various defects and failures, leading to operational disruptions, passenger inconvenience, and heightened maintenance costs. The maintenance practices conducted is mostly reactive interventions, resulting in suboptimal system performance and increased downtimes.

Figure 1: Sleeper Coach [9] and Toilet [10].

An attempt is made in this paper to develop AI and ML model to analyse condition monitoring data to identify faults early and predict potential failures. An exploratory analysis of real-time condition monitoring data from the Sleeping Room (SR) and Water Closet (WC) system to identify patterns and anomalies indicative of defects.

2.2 Technical information exploratory data analysis

The analysis detailing the technical intricacies of the water closet system is presented, each of the three toilet types incorporates a Fresh water tank (FWT) and a Controlled Emissions Toilet (CET) tank, engineered with a high level of resistance to blockages stemming from misuse or overfilling. The FWT allows for filling from either side of the car using standard connections and is equipped with a manual emptying device. Externally visible water level indicators near the filling point and on the Train Control and Management System (TCMS) screen enable easy monitoring by train crew. The FWT's level sensor delineates five capacity levels (100%, 75%, 50%, 25%, and 0%), automatically sending status updates to the crew when levels reach 25% and locking the toilet at 5%, signalling its "out-of-use" status. Similarly, the Waste Water Tank (WWT) employs a sludge suction mechanism for emptying and features level indicators at 80% and 95% capacity. At 95% capacity, the system triggers automatic door locking and alerts the train crew, indicating the toilet is "out-of-service."

3 Application of Machine Learning in Railway

The accumulation of extensive datasets within the railway sector via condition monitoring technology remains futile unless leveraged to extract valuable insights and actionable intelligence [11].

There are two distinct (ML) methodologies, namely the Support Vector Machine (SVM) and Random Forest (RF), that are considered in this paper. The efficacy of these prediction models can be assessed utilising four distinct statistical metrics, including recall, accuracy, precision, and F1-score [12].

Support Vector Machines (SVM)

The Support Vector Machine (SVM) stands as a supervised learning algorithm rooted in the principles of statistical learning [13]. Support Vector Machine (SVM) stands as a robust machine learning technique renowned for its efficacy in classifying observations by delineating an optimal hyperplane within higher-dimensional spaces [14]. This technique facilitates the mapping of data to a higher dimensional space, thereby enabling linear separation and enhancing the overall flexibility of the model.

Random Forest

Random Forest is a widely embraced ensemble learning algorithm specifically designed for binary classification tasks within the realm of machine learning [15]. However, it is crucial to acknowledge that the decision trees employed in constructing the random forest may exhibit limitations such as low classification accuracies or high correlations, factors that can impact the overall performance and effectiveness of the random forest model [16].

4 Model Development

The AI and ML models, including supervised learning algorithms, were designed and trained to classify defect categories and predict the time of failure. An overview of the model development framework is given in Figure 2, where the input data, set parameter and input feature for prediction is depicted.

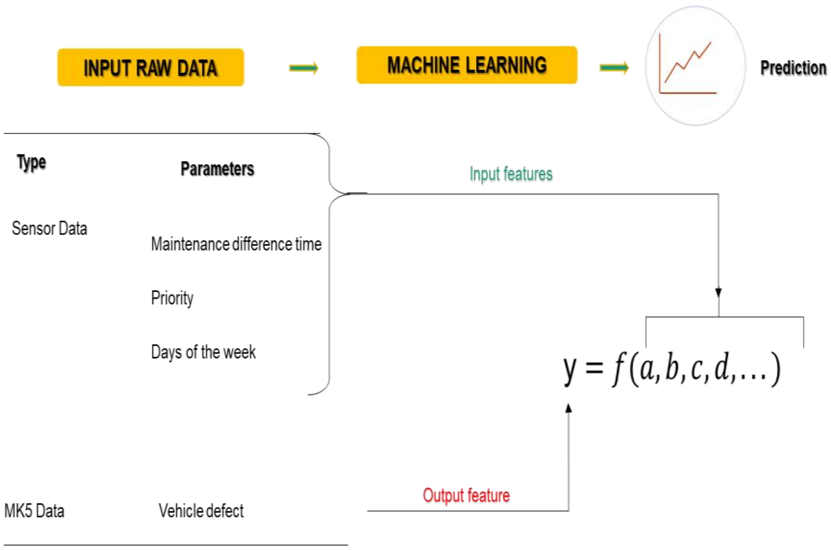

Figure 2: A block diagram of the Model Development.

4.1 Model evaluation and validation

The performance and accuracy of the developed models are evaluated through testing and validation. This validation process ensure that the models are robust and capable of generalising the data and classification. A binary classification approach is used, and a 2x2 confusion matrix in Table 1 that looks like this is considered:

	Predicted Positive (P)	**Predicted Negative (N)**
Actual Positive (P)	True Positive (TP)	False Negative (FN)
Actual Negative (N)	False Positive (FP)	True Negative (TN)

Table 1: A 2x2 Confusion Matrix.

- True Positive (TP): The number of instances correctly predicted as positive.
- True Negative (TN): The number of instances correctly predicted as negative.
- False Positive (FP): The number of instances incorrectly predicted as positive (also known as Type I error).

False Negative (FN): The number of instances incorrectly predicted as negative (also known as Type II error.

4.2 The use of Confusion Matrix

Accuracy Measurement: the accuracy is calculated as the proportion of correct predictions (both TP and TN) out of the total predictions using equation (1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Precision: The proportion of true positive predictions among all positive predictions in (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Recall (Sensitivity): The proportion of true positive predictions among all actual positive instances in (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1-Score: The harmonic means of precision and recall, providing a single metric for model performance.

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Specificity: The proportion of true negative predictions among all actual negative instances.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5}$$

4.3 Data Analysis and Visualization

Analysis based on the notification descriptions, highlighting the most frequent defects using work plot is shown in the data visualisation in Figure 3. A systematic approach is used to analyse condition monitoring data to enhance Sleeping Room (SR), Water Closet (WC) system reliability within the sleeper train fleet.

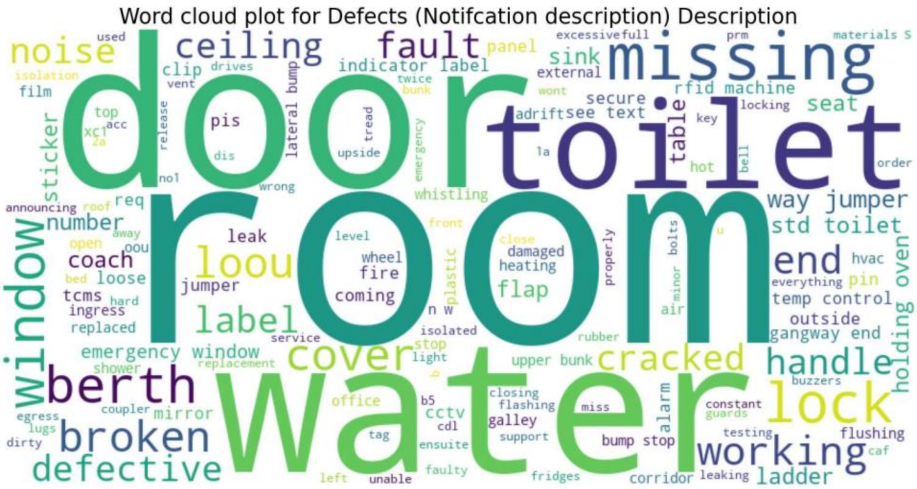

Figure 3: Word plot of various defects

Further investigation and analysis show that the sleeping room-related maintenance constitutes the majority, encompassing 68.4% of all reported issues, and toilet-related defects at 31.6%. These percentage distributions highlight a significant data imbalance, particularly notable in the Toilet category with its minimal ratio.

Addressing such data discrepancies is pivotal, with the addition of more relevant data being a recommended approach. Alternatively, leveraging data augmentation techniques could serve as a viable alternative in mitigating the challenges posed by this imbalance.

4.4 Handling Data Imbalance

To address the issue of data imbalance within our dataset, we employed the Synthetic Minority Oversampling Technique (SMOTE). The SMOTE generates new synthetic instances by interpolating between existing examples, thereby providing a more sophisticated and effective solution and the implementation of SMOTE in the study is detailed in Figure 4.

```
from imblearn.over_sampling import SMOTE

# Count the class distribution before SMOTE
print("Class distribution before SMOTE:", y.value_counts())

# Applying SMOTE
smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)

# Count the class distribution after SMOTE
print("Class distribution after SMOTE:", y_resampled.value_counts())
```

```
Class distribution before SMOTE: target
0         78
1         36
Name: count, dtype: int64
Class distribution after SMOTE: target
0         78
1         78
Name: count, dtype: int64
```

Figure 4: Display showing code and solving the data imbalance.

Following the data observation, water defects appeared to be the least prevalent among reported issues. However, this observation warrants caution as it may be influenced by the limited availability of data. The frequency of repairs based on the notification dates is given in Figure 5.

It is observed that maintenance notifications occurred during the period spanning from 2019-05-27 to 2019-06-06, primarily attributed to sleeping room and toilet-related defects. This makes sense since these are new trains that was produce around 2018 according to technical data information. Visualising this data days of the week provide further insight into the temporal distribution of maintenance notifications, offering a deeper understanding of the patterns and trends within the dataset.

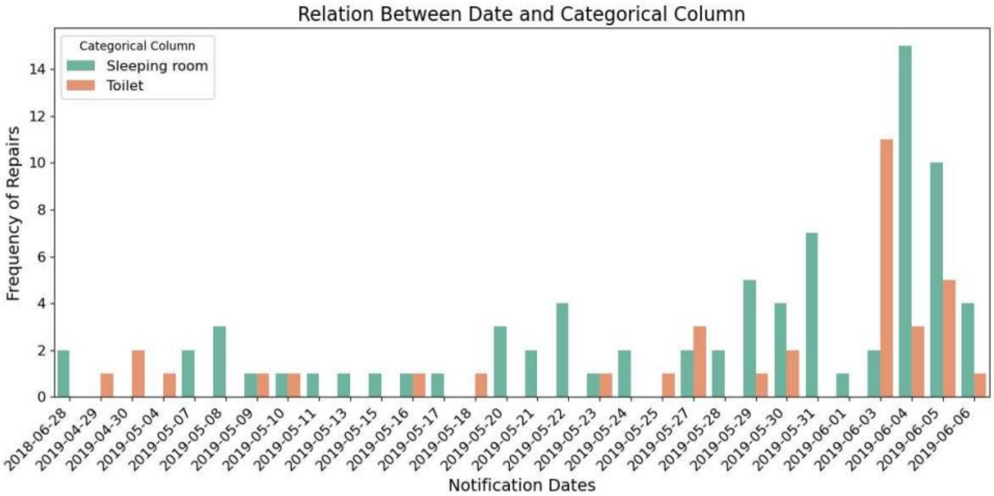

Figure 5: Frequency of repair by dates.

The frequency of repair data in days in Figure 4 show that major maintenances are carried out on sleeping rooms related defects for all maintenance days. While the next top maintained vehicle defects relate to the toilet system, however, it can be observed that there were no records of maintenance on Sunday which is reasonable as most maintenance companies are shut on Sundays.

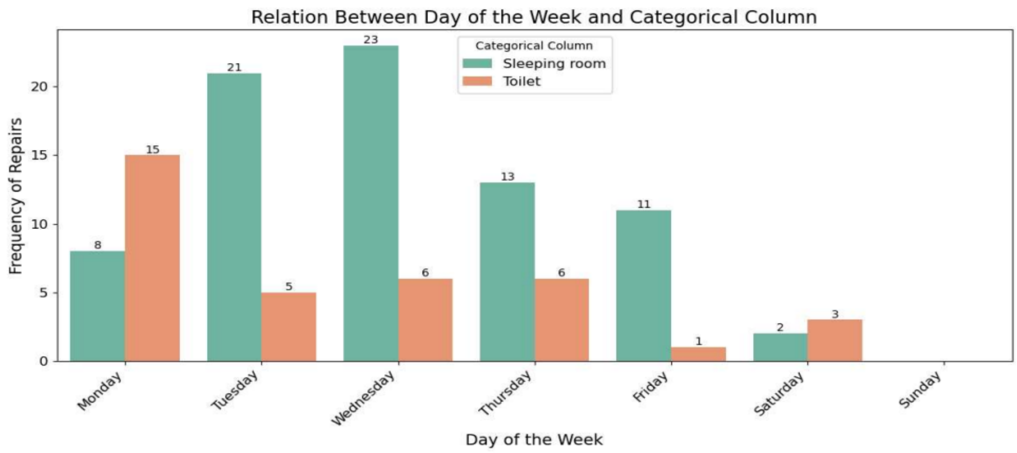

Figure 6: Frequency of repair by days.

4.5 Random Forest Model Implementation - Without Optimization

The Random Forest model was trained on a dataset split into 80% training and 20% testing sets. The classification in Table 3 and Figure 7 show the model performance.

	precision	recall	f1-score	support
0	0.59	0.91	0.71	11
1	0.83	0.42	0.56	12
accuracy			0.65	23
Macro average	0.71	0.66	0.63	23
Weighted avg.	0.72	0.65	0.63	23

Table 3: Random Forest Evaluation (Before handling data Imbalance).

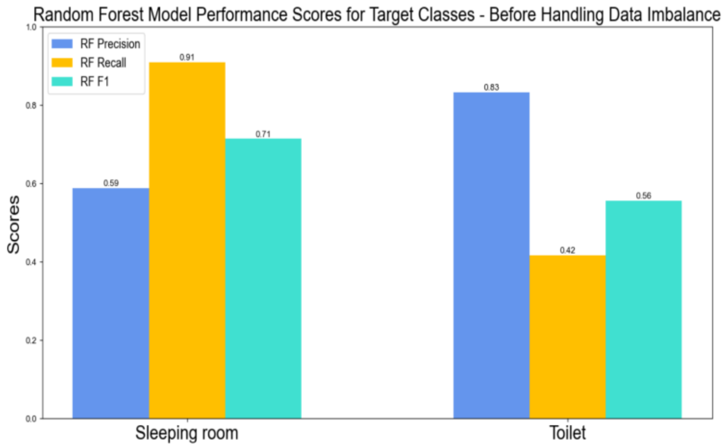

Figure 7: Visualising random forest result.

5 Analysis and Results

The Random Forest model, without any optimization, demonstrates a significant difference in its performance between the two classes. Class 0 achieves a high recall (0.91), indicating that most positive instances of this class are correctly identified. However, its precision is relatively lower (0.59), suggesting a fair number of false positives. Conversely, Class 1 exhibits high precision (0.83) but considerably lower recall (0.42), meaning the model misses a substantial number of actual positives for this class. The overall accuracy of 0.65 reflects these imbalances, which could be problematic in scenarios where both classes are equally important.

5.1 Confusion Matrix – Random Forest

The model correctly identified 10 instances of "Sleeping room" and 5 instances of "Toilet" as shown in the graph in Figure 8. It misclassified 7 "Toilet" instances as "Sleeping room" and 1 "Sleeping room" instance as "Toilet". This indicates a reasonably balanced performance with some misclassification, particularly higher false positives for "Toilet".

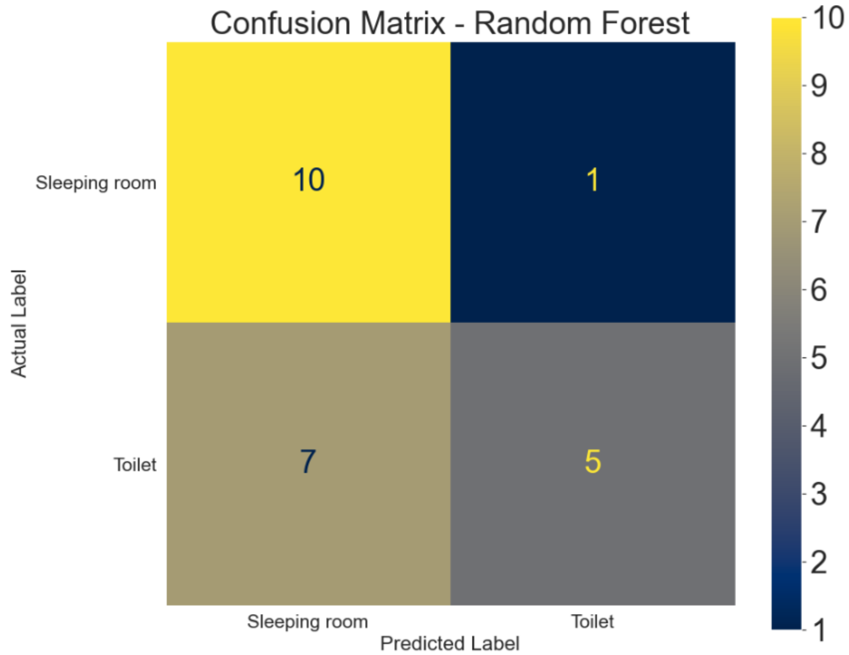

Figure 8: Confusion Matrix–RF.

5.2 SVM Model Implementation – Without Optimization

The SVM model was trained on a dataset split into 80% training and 20% testing sets. The classification report for the model's performance on the testing in Table 4

	precision	recall	f1-score	support
0	0.48	1.00	0.65	11
1	0.00	0.00	0.00	12
accuracy			0.48	23
Macro average	0.24	0.50	0.32	23
Weighted avg.	0.23	0.48	0.31	23

Table 4: SVM Evaluation (Before handling data Imbalance).

The SVC model's performance is notably poorer compared to the Random Forest model. Class 0 achieves perfect recall (1.00) but low precision (0.48), indicating a high number of false positives. Class 1's metrics are particularly concerning, with both precision and recall at 0.00, indicating the model fails to identify any positive instances of this class. The overall accuracy of 0.48 highlights the model's inability to generalize well across both classes. This result suggests that the SVC, without optimization, is unsuitable for this imbalanced dataset as shown in Figure 9.

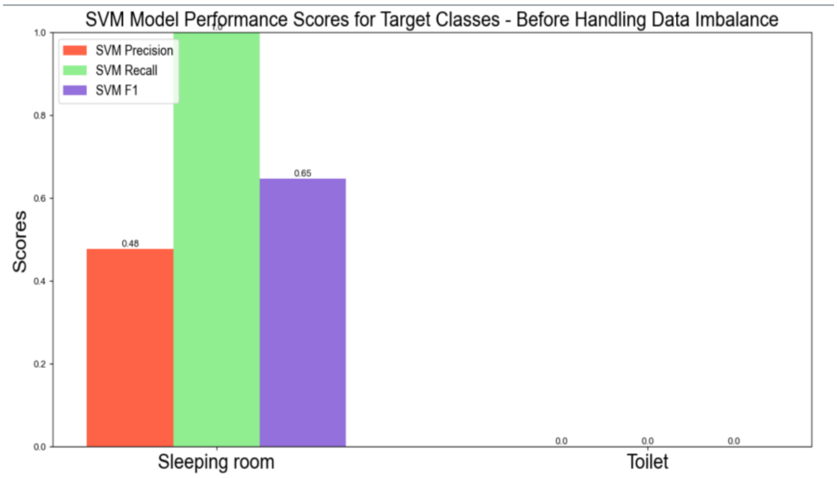

Figure 9: Visualising SVM result.

5.3 Confusion Matrix – SVM

The SVM model correctly identified 11 instances of "Sleeping room" but failed to identify any instances of "Toilet" as shown in Figure 10. All "Toilet" instances (12) were misclassified as "Sleeping room". This shows a strong bias towards predicting "Sleeping room" and a failure to correctly classify "Toilet".

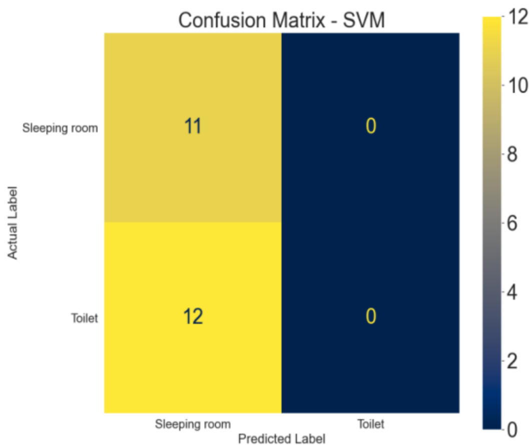

Figure 10: Confusion Matrix–SVM.

Summary – Without Optimization

The Random Forest classifier outperformed the SVM in terms of overall accuracy, precision, recall, and F1-score. This indicates that the Random Forest model is more robust and reliable for this application, particularly in handling the imbalance between classes. The better performance of the Random Forest model can be attributed to its ability to handle large datasets with higher dimensionality and its ensemble nature, which reduces the risk of overfitting. On the other hand, the SVM model exhibits significant bias and fails to correctly identify any "Toilet" instances, making it unsuitable for this classification task without further optimization.

To achieve optimal results, it is crucial to address the data imbalance issue within the dataset.

5.4 Comparison of model performance

To ensure a fair and accurate comparison between the models. The confusion matrices for the Random Forest and Support Vector Machine (SVM) models in Figure 11 provide a visual representation of the classification performance of each model on the test dataset. The confusion matrices allow us to understand how well the models distinguish between the two classes: "Sleeping room" and "Toilet".

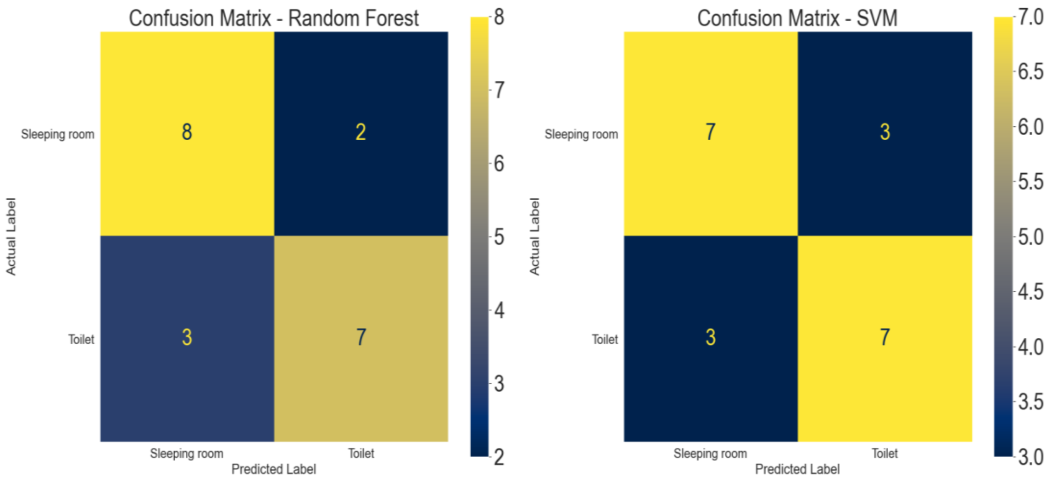

Figure 11: Confusion Matrix–After handling Data imbalance.

Model Performance Comparison show the following:

1. Accuracy
Random Forest: The model correctly classified 75% (15 out of 20) of the test samples.
SVM: The model correctly classified 70% (14 out of 20) of the test samples.
2. Precision
Random Forest: Higher precision in identifying "Sleeping room" due to fewer false positives compared to SVM.
SVM: Slightly lower precision for "Sleeping room" as it misclassified more Toilets as Sleeping rooms.
3. Recall
Random Forest: Better recall for "Sleeping room" with more true positives.
SVM: Equal recall for "Toilet" but slightly lower for "Sleeping room".
4. F1-Score
Random Forest: Higher F1-score for "Sleeping room" indicating a better balance between precision and recall.
SVM: Lower F1-score for "Sleeping room" reflecting a trade-off between precision and recall.

6 Conclusions and Contributions

The results demonstrate that the Random Forest classifier is more effective than the SVM classifier for this particular problem. The RF model's higher precision and recall rates suggest that it is better suited for predicting faults in the WC system. The use of SMOTE was crucial in addressing the class imbalance, allowing both models to perform better than they would have with the original imbalanced data. The improved data handling and model performance indicate that the WC system's reliability can be significantly enhanced through the application of machine learning techniques. By accurately predicting failures, maintenance can be performed proactively, reducing downtime and improving passenger experience.

The study successfully demonstrates the application of AI and machine learning techniques to enhance the reliability of the WC system in the MK5 sleeper fleet. The Random Forest classifier outperformed the SVM in terms of overall accuracy, precision, recall, and F1-score, particularly in handling imbalanced data. This superiority can be attributed to its ensemble nature and ability to manage large datasets with high dimensionality, which mitigates the risk of overfitting.

Acknowledgements
Thanks to Caledonian Sleeper for providing access to data to support the research and results presented in this paper.

References

[1] Szymula, C., Nikola Bešinović and Nachtigall, K. (2024). Quantifying periodic railway network capacity using petri nets and macroscopic fundamental diagram. *Transportation research. Part C, Emerging technologies*, 158, pp.104436–104436. doi:https://doi.org/10.1016/j.trc.2023.104436.

[2] Office of Rail and Road (2023). *Passenger Rail Usage | ORR Data Portal*. [online] Orr.gov.uk. Available at: https://dataportal.orr.gov.uk/statistics/usage/passenger-rail-usage/.

[3] Papa, E. (2020). Could sleeper trains replace international air travel? *theconversation.com*. [online] Available at: https://westminsterresearch.westminster.ac.uk/item/qy39v/could-sleeper-trains-replace-international-air-travel [Accessed 25 Feb. 2024].

[4] Minh-Huong Le-Nguyen, Fabien Turgis, Pierre-Emmanuel Fayemi and Bifet, A. (2023). Real-time learning for real-time data: online machine learning for predictive maintenance of railway systems. *Transportation research procedia*, 72, pp.171–178. doi:https://doi.org/10.1016/j.trpro.2023.11.391.

[5] Mou, M. and Zhao, X. (2024). Industrial process fault diagnosis based on domain adaptive broad echo network. *Journal of the Taiwan Institute of Chemical Engineers (Print)*, 159, pp.105453–105453. doi:https://doi.org/10.1016/j.jtice.2024.105453.

[6] Liu, H., Song, L., Xu, L., Hou, J. and Yu, Z. (2023). Stochastic analysis and

dynamic reliability of the heavy-haul train–bridge system based on direct probability integral method. *Probabilistic engineering mechanics*, 74, pp.103510–103510. doi:https://doi.org/10.1016/j.probengmech.2023.103510.

[7] Liu, H., Song, L., Xu, L., Hou, J. and Yu, Z. (2023). Stochastic analysis and dynamic reliability of the heavy-haul train–bridge system based on direct probability integral method. *Probabilistic engineering mechanics*, 74, pp.103510–103510. doi:https://doi.org/10.1016/j.probengmech.2023.103510.

[8] Attoh-Okine, N. (2014). *Big data challenges in railway engineering*. [online] IEEE Xplore. doi:https://doi.org/10.1109/BigData.2014.7004424.

[9] Lannie's Food & Travel (2023). *Comfy overnight train to Scotland: Caledonian Sleeper Review*. [online] Available at: https://lanniesfoodandtravel.com/2023/03/14/overnight-train-to-scotland/ [Accessed 11 May 2024].

[10] BBC Scotland (2017). New Caledonian Sleeper trains to offer 'hotel' experience. *BBC News*. [online] 30 Aug. Available at: https://www.bbc.co.uk/news/uk-scotland-41095412 [Accessed 11 May 2024].

[11] M. Bezuidenhout, J.L. Jooste, Lucke, D. and Fourie, C.J. (2023). Leveraging digitilisation and machine learning for improved railway operations and maintenance. *Procedia CIRP*, 120, pp.702–707. doi:https://doi.org/10.1016/j.procir.2023.09.062.

[12] Scheu, M.N., Tremps, L., Smolka, U., Kolios, A. and Brennan, F. (2019). A systematic Failure Mode Effects and Criticality Analysis for offshore wind turbine systems towards integrated condition-based maintenance strategies. *Ocean Engineering*, 176, pp.118–133. doi:https://doi.org/10.1016/j.oceaneng.2019.02.048.

[13] Vapnik, Vladimir Naumovich and Vapnik, V. (1998). Statistical learning theory.

[14] Jonatha Sousa Pimentel, Ospina, R. and Ara, A. (2024). A novel fusion Support Vector Machine integrating weak and sphere models for classification challenges with massive data. *Decision analytics journal*, pp.100457–100457.

[15] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), pp.5–32. doi:https://doi.org/10.1023/a:1010933404324.

[16] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M. and Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, [online] 237, p.121549. doi:https://doi.org/10.1016/j.eswa.2023.121549.