Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/GC03.20250259
Published: February 2, 2026

Predictive Modelling of Vehicle Downtime Risk Using Machine Learning

B. B. S. Makhanya and JHC Pretorius
Postgraduate School of Engineering Management
University of Johannesburg
Johannesburg, South Africa
Bmakhanya@uj.ac.za, jhcpretorius@uj.ac.za

Abstract

This study introduces a predictive modelling framework aimed at assessing vehicle downtime risk through the application of advanced machine learning techniques. Minimising unplanned downtime is crucial for enhancing operational efficiency and reducing costs within transportation and logistics systems. Utilizing maintenance records from a South African coal transportation enterprise, we developed and evaluated several machine learning algorithms, namely Random Forest, Logistic Regression, Support Vector Machine, and XGBoost, to predict high- and low-risk downtime events. The dataset, which consists of 201 records collected over a three-month period, was preprocessed and balanced using the Synthetic Minority Over-sampling Technique (SMOTE) to improve classification accuracy. Among the models assessed, XGBoost delivered the highest predictive performance, achieving an AUC of 0.97, a macro F1-score of 0.87, and a weighted F1-score of 0.88, demonstrating a strong ability to differentiate between risk categories. Feature importance analysis revealed that Failure Mode and Depot were the most influential predictors of downtime, providing valuable insights for targeted maintenance planning. Additionally, the study outlines a deployment strategy for integrating the model into operational systems via local and web-based platforms, facilitating data-driven maintenance scheduling. These findings underscore the transformative potential of machine learning in predictive maintenance and establish a framework for leveraging AI-driven analytics to enhance reliability, resource allocation, and operational resilience within the vehicle and railway sectors.

Keywords
Predictive Modelling, Machine Learning, Downtime Risk, Maintenance Strategies, Operational Efficiency.

1. Introduction
Predictive modelling of vehicle downtime risk, utilising machine learning techniques, presents a complex, multidisciplinary challenge. This approach necessitates a thoughtful integration of data, choice of modelling techniques, and evaluation criteria focused on operational reliability and operational efficiency. Recent advancements in predictive maintenance and downtime reduction underscore the importance of addressing these aspects cohesively. By synthesising findings from contemporary literature, we aim to clarify the essential modelling choices, data requirements, and evaluation metrics necessary for effective predictive modelling in the vehicle sector.

The risk of downtime in vehicle operation can be understood as a time-to-failure or time-to-downtime event, with the primary objective being the prediction of imminent failures or abnormal operational states that can lead to unexpected downtimes. Such preemptive insights enable timely maintenance interventions and operational adjustments(Liang et al., 2023). The existing literature consistently portrays predictive maintenance as a crucial strategy to mitigate downtime, augment availability, and enhance both operational efficiency and overall reliability in industrial applications (Priyanga et al., 2023). Within the vehicle context, downtime is directly linked to operational efficiency, routing efficiencies, and prevailing line conditions, indicating the need for a unified approach that simultaneously predicts failure risks while considering the associated operational implications (Liang et al., 2023).

A practical focus of this research highlights the urgency to minimise unplanned downtimes through data-driven strategies encompassing maintenance scheduling, condition monitoring, and rigorous root-cause analyses of downtime events. It is equally vital to evaluate the operational efficiency ramifications of predictive maintenance strategies (Liang et al., 2023). Aligned with broader literature on predictive maintenance, our findings suggest that employing machine learning-driven predictions can lead to significant reductions in both downtime incidents and associated operational costs (Allahloh et al., 2023).

The paper will proceed to explore the various types of data and sources critical for predicting vehicle downtime, including sensor data, maintenance logs, and operational contexts that collectively enhance the predictive modelling process. We will also discuss suitable modelling approaches and algorithms, addressing common challenges such as data imbalance and ensuring robust predictions. Finally, we will establish evaluation metrics that capture operational impacts and provide a comprehensive methodological blueprint aimed at effectively managing vehicle downtime risks. Through this integrated framework, we aspire to present a solid foundation for future research and application in predictive maintenance within the logistics industry.

1.1 Objectives

The study is aimed at achieving two research objectives:

1. To identify and evaluate the most effective machine learning algorithms for predicting vehicle downtime risks based on diverse data sources, including sensor data, maintenance logs, and operational factors, in order to enhance predictive accuracy and operational efficiency.
2. To develop a comprehensive framework that integrates predictive maintenance strategies with operational planning considerations, thereby enabling railway operators to optimise maintenance schedules and minimise unplanned downtimes while simultaneously mitigating operational disruptions.

2. Literature Review

Predictive maintenance has emerged as a transformative approach in transportation and logistics, enabling organisations to anticipate equipment failures and reduce unplanned downtimes. Traditional maintenance strategies, reactive and preventive, often fall short in addressing the dynamic and complex nature of modern fleets. Machine learning (ML) and artificial intelligence (AI) offer data-driven alternatives that enhance reliability, optimise maintenance schedules, and reduce operational costs. Mahiyudin et al. (2025) conducted a comprehensive study on vehicle maintenance prediction using ML, applying algorithms such as Decision Trees, Random Forest, and Gradient Boosting on an imbalanced dataset of 50,000 vehicles. Their use of the Synthetic Minority Over-sampling Technique (SMOTE) for data balancing significantly improved model accuracy, achieving up to 99.97% with Decision Trees. Mahale et al. (2025) provided a systematic review of 94 studies on AI-driven predictive maintenance in vehicles. Their work emphasised the importance of explainable AI and highlighted challenges such as data quality, scalability, and integration with existing systems. In the context of autonomous public transit systems, Ejaz and Bakhsh (2023) explored predictive maintenance using real-time diagnostics, deep learning, and digital twin modelling. Their findings demonstrated improved fleet performance and reduced downtimes in smart city environments.

Davari et al. (2021) surveyed data-driven predictive maintenance in the railway industry, categorising ML and deep learning techniques based on tasks, equipment types, and evaluation metrics. They emphasised the need for tailored approaches to address equipment-specific challenges. Krummenacher et al. (2018) investigated ML applications for predicting railway infrastructure defects. Their study highlighted the effectiveness of models like Random Forest, LSTM, and Artificial Neural Networks, while also identifying gaps in data quality and model generalizability. García-Méndez et al. (2025) proposed an explainable ML framework for railway predictive maintenance using real-time data streams. Their pipeline achieved over 98% F-measure and 99% accuracy, demonstrating robustness even under class imbalance and noisy conditions. Aissani et al. (2025) reviewed predictive maintenance technologies for vehicle reliability, focusing on ML, AI, and IoT integration. Their comparative analysis of diagnostic frameworks and digital twin applications underscored the potential of predictive maintenance to reduce costs and improve fleet performance.

These studies collectively affirm the value of machine learning in predictive maintenance across transportation sectors. They also highlight common challenges such as data imbalance, feature engineering, and deployment scalability. Building on these insights, our research aims to develop a practical and interpretable ML-based framework tailored to the vehicle sector, addressing both technical and operational dimensions of downtime risk.

3. Methods

This paper explains the empirical findings of a research team tasked with increasing the operational efficiency of the coal division of a South African enterprise. The organisation, which operates a diverse fleet of Vehicles categorised by various makes and models, is responsible for the transportation of coal from Mpumalanga to Richards Bay. Recent operational challenges have hindered the company's ability to meet its business commitments, primarily due to delays associated with vehicle maintenance. These delays can be traced to several factors, including staff shortages, inadequate availability of essential components, and extended queues of vehicles awaiting servicing.

In an effort to remediate these issues, the company is pursuing the development of a predictive model aimed at forecasting the potential downtime of Vehicles based on varying durations. The proposed classification framework distinguishes between "Low Risk" for downtimes ranging from 0.0 to 5.0 days and "High Risk" for downtimes extending from 6.0 to 242.0 days. The research team has been allocated a three-month timeframe to construct this predictive model.

3. Data Collection and Analysis

The dataset used in this study was sourced from the maintenance records of a South African coal transportation enterprise operating a diverse fleet of vehicles. These vehicles, varying in make and model, are primarily used for hauling coal between Mpumalanga and Richards Bay. The data set was compiled over a three-month period, during which the research team collaborated closely with the company’s maintenance personnel to ensure accurate and consistent data collection.

Initially, the maintenance data was stored in unstructured spreadsheet formats, which posed challenges in terms of accessibility and usability. To address this, the research team developed a temporal SQL database to systematically organise and store the data. This database includes key features relevant to downtime predictions (please see Table 1).

Table 1.Vehicle Downtime Dataset Features

Feature	Description	Purpose in Predictive Modelling
Vehicle Number	Unique identifier assigned to each vehicle.	Used to track individual vehicle history and link maintenance records across time.
Class	Categorisation based on vehicle model or type.	Helps identify patterns in downtime associated with specific vehicle types or configurations.
Depot	The maintenance location is responsible for servicing the vehicle.	Assesses depot-specific performance and resource availability impacting downtime duration.
Failure Mode	Coded representation of the fault type.	Critical for identifying fault patterns and predicting the severity and duration of downtime.
Failure Description	Detailed textual explanation of the fault.	Provides context to the failure mode; useful for feature engineering and understanding root causes.
Stop Date	Date when the vehicle was taken out of service.	Marks the beginning of downtime; used to calculate downtime duration and identify trends.
Planned Release Date	Expected date for the vehicle to return to service.	Indicates maintenance scheduling efficiency and helps compare planned vs. actual downtime.
Actual Release Date	The real date when the vehicle resumed operations.	Used to compute actual downtime and assess deviations from planned maintenance timelines.
Comments	Additional notes from maintenance personnel.	May contain qualitative insights into delays, resource issues, or operational decisions.

The final dataset comprised 201 records, with downtime durations ranging from 0 to 242 days, and a mean downtime of 12.27 days. After preprocessing, including the removal of incomplete records and the standardisation of date formats, the dataset was deemed suitable for binary classification modelling. Downtime risk was categorised into two classes: Low Risk: Downtime between 0.0 and 5.0 days, and High Risk: Downtime between 6.0 and 242.0 days. This structured and cleaned dataset formed the foundation for training and evaluating machine learning models aimed at predicting downtime risk.

5. Results and Discussion

5.1 Numerical Results

After setting up the database, the research team needed to provide training to the company's maintenance team and supervisors. This training was meant to show them how to record and store the data correctly. Over a three-month period, the database collected a total of 201 records, as shown in Table 2.

Table 2. Data set descriptions

Statistic	Downtime Stats
count	201
mean	12.273
std	27.194
min	0
25%	2
50%	5
75%	9
max	242

5.2. Data Wrangling

The raw dataset was preprocessed to ensure consistency and reliability. Date fields, including Stop Date, Release Date, and Actual Date, were converted to standard datetime format. Entries with negative downtime values were excluded, and rows with missing values in essential fields Class, Depot, and Status Description were removed (Table 3). This ensured that only complete and valid records were used for modelling.

Table 3. Missing values

Features	Missing Values
Vehicle Number	0
Class	3
Depot	6
Status	1
Status Description	0
Stop Date	0
Release Date	0
Actual Date	0
Down Time	0

5.3. Assumption testing

To validate the suitability of the dataset for classification modelling, several assumptions were tested. Firstly, the analysis revealed a moderate class imbalance in the target variable, with 110 instances classified as low-risk and 85 as high-risk. This indicates a relatively balanced dataset for the purpose of modelling. Secondly, it was confirmed that there were no missing values in the selected features, specifically Class, Depot, and Status Description, which ensures the completeness and reliability of the data. Additionally, an assessment of feature cardinality showed that there were eight unique Vehicle classes, six distinct maintenance depots, and 187 unique fault descriptions. Collectively, these results indicate that the dataset is sufficiently diverse and balanced, making it appropriate for training a binary classification model.

5.3.1 Addressing Class Imbalance in Downtime Risk Prediction

Although the imbalance is not severe, it can still influence model performance, particularly in correctly identifying minority class instances (i.e., high-risk downtimes). To mitigate this, the Synthetic Minority Over-sampling Technique

(SMOTE) was employed. SMOTE generates synthetic examples of the minority class by interpolating between existing samples. This approach was chosen over other techniques, such as ADASYN and cost-sensitive learning, for the following reasons:

- SMOTE vs. ADASYN: While ADASYN also generates synthetic samples, it focuses more on difficult-to-learn examples, which can sometimes introduce noise and reduce model generalisation. SMOTE, on the other hand, provides a more balanced and controlled oversampling strategy, which was deemed suitable given the moderate imbalance in our dataset.

- SMOTE vs. Cost-Sensitive Learning: Cost-sensitive learning adjusts the algorithm's loss function to penalise misclassification of minority classes more heavily. However, this approach often requires careful tuning and may not be compatible with all classifiers. In contrast, SMOTE is model-agnostic and integrates seamlessly with tree-based algorithms like XGBoost.

To ensure that the use of SMOTE did not lead to overfitting, the following precautions were taken:

- Train-Test Split Before SMOTE: SMOTE was applied only to the training set, preserving the integrity of the test set for unbiased evaluation.
- Cross-Validation: Stratified k-fold cross-validation was used to assess model performance across multiple data splits, ensuring robustness.
- Performance Metrics: The model's high AUC score (0.97) and balanced precision-recall values across both classes indicate strong generalisation and minimal overfitting.

Future work may explore hybrid approaches, such as combining SMOTE with ensemble methods or experimenting with ensemble resampling techniques to further enhance predictive performance and fairness across classes.

5.3. Model selection

This section presents a comparative analysis of four machine learning models used to predict high-risk downtime in Vehicles. The models evaluated include Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost. Performance metrics such as Accuracy, Precision, Recall, and F1-Score for the High Risk class are reported. Among the models evaluated, XGBoost and Support Vector Machine demonstrated the highest overall accuracy (0.69) (Table 4). XGBoost achieved the best recall (0.56) and F1-score (0.66) for the High Risk class, indicating its superior ability to identify Vehicles at risk of extended downtime. While Random Forest showed the highest precision (0.92), its low recall (0.31) suggests it may miss many high-risk cases. Logistic Regression offered a balanced performance but was outperformed by XGBoost and SVM in terms of recall and F1-score. Hence, XGBoost was selected for further analysis and to predict the model.

Table 4. Model

Model	**Accuracy**	**Precision (High Risk)**	**Recall (High Risk)**	**F1-Score (High Risk)**
Random Forest	0.62	0.92	0.31	0.46
Logistic Regression	0.65	0.80	0.44	0.57
Support Vector Machine	0.69	0.89	0.47	0.62
XGBoost	0.69	0.80	0.56	0.66

The dataset was partitioned into training (70%) and testing (30%) sets. To further evaluate the model's robustness, Stratified K-Fold Cross-Validation with 5 folds was implemented on the training data. This method preserves the original class distribution across the folds, which is crucial given the moderate class imbalance within the dataset. During cross-validation, the XGBoost model achieved average performance metrics of 57% ± 7% for accuracy, 45% ± 10% for the F1 Score, and 52% ± 9% for ROC AUC, indicating a moderate level of predictive capability. In contrast, the final evaluation on the held-out test set revealed significantly enhanced performance, with an accuracy of 88%, a macro F1 Score of 0.87, a weighted F1 Score of 0.88, and an AUC of 0.97. This improvement can be attributed to the full utilisation of the training set and the effectiveness of SMOTE in balancing the classes. The primary goal of the model was to predict Downtime Risk, categorizing vehicles as either High or Low risk based on features such as Class, Depot, and Failure Mode (Table 5). The final model demonstrated strong generalisation across both risk categories, underscoring its potential for deployment in operational environments.

Table 5. XGBoost Model Performance

Risk Level	Precision	Recall	F1-Score	Support
Low Risk (0)	0.92	0.87	0.89	39
High Risk (1)	0.81	0.88	0.85	25
Macro Avg	0.87	0.88	0.87	64
Weighted Avg	0.88	0.88	0.88	64

5.2 Graphical Results

This section provides a detailed analysis of Vehicle downtime categorized by both Vehicle class and depot, highlighting the varying risk levels associated with each. The findings reveal significant disparities in maintenance performance, with a clear distinction between low-risk and high-risk units, emphasising the need for targeted interventions and improvements in repair processes (Figure 1- Figure 4).

Data Expiration

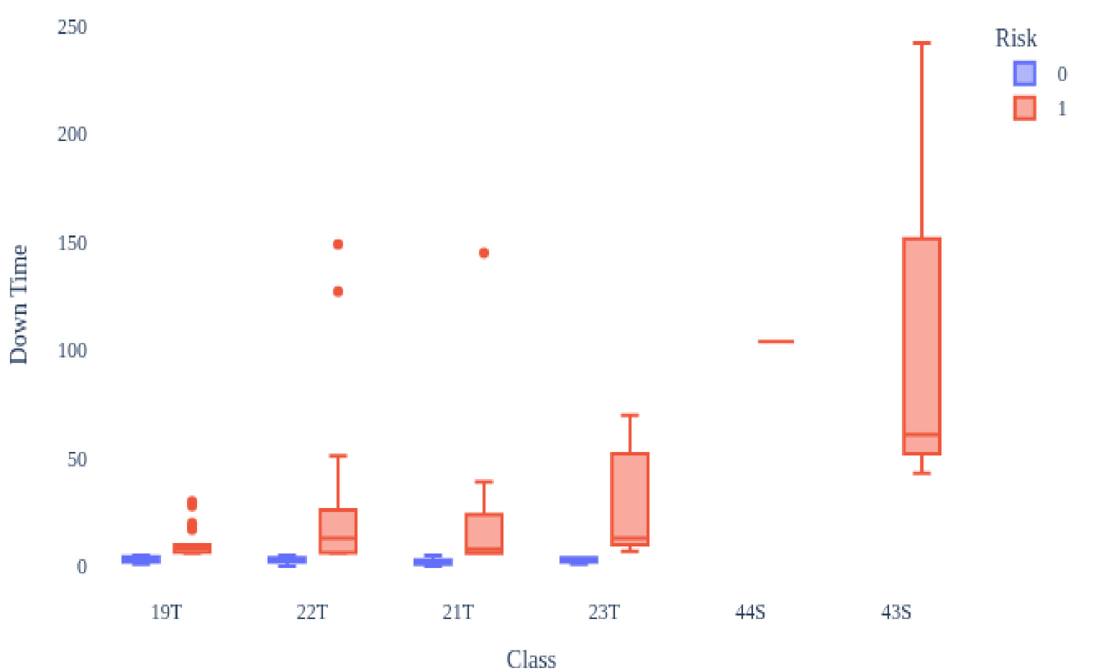

Figure 1. Downtime by Vehicle Class and Risk

In the analysis of Vehicle classes, Class 19T demonstrates a predominance of low-risk Vehicles, with 55 units experiencing an average downtime of 4.67 days, ranging from 1 to 10 days. In contrast, five Vehicles in this class are categorised as high risk, with an average downtime of 22.6 days and a range between 17 and 30 days. This suggests that most 19T Vehicles maintain relatively stable performance and manageable maintenance needs. Class 21T presents a mixed scenario; while 43 of its Vehicles are low risk, averaging 3.91 days of downtime within a range of 0 to 9 days, the nine high-risk Vehicles exhibit a troubling average downtime of 41.78 days, reaching up to 145 days. This discrepancy indicates that, despite the majority being low risk, the high-risk group is affected by significant faults or delays in repair processes.

Similarly, Class 22T has a notable number of high-risk Vehicles, with 13 units averaging 40.38 days of downtime, ranging from 13 to 149 days. In contrast, 51 low-risk Vehicles show an average of 3.94 days, with downtime ranging from 0 to 12 days. The substantial downtime among high-risk Vehicles points to potential reliability concerns or maintenance bottlenecks in this class. Class 23T features three high-risk Vehicles, which have a striking average downtime of 45.33 days, ranging from 14 to 70 days. In comparison, the 13 low-risk Vehicles exhibit a mere 4.08 days of downtime, ranging from 1 to 12 days. The elevated downtime in the high-risk group suggests it may be prudent to conduct closer inspections of fault types or depot handling procedures.

Class 43S is exclusively high risk, with four Vehicles averaging a formidable downtime of 101.75 days and ranging from 43 to 242 days. This class's performance indicates serious reliability issues or insufficient repair capacity, raising concerns for operational efficiency. Finally, Class 44S also solely comprises high-risk Vehicles, with two units both experiencing 104 days of downtime. Like Class 43S, this class exhibits consistently high downtime, which may imply systemic issues such as ageing components or limited availability of spare parts.

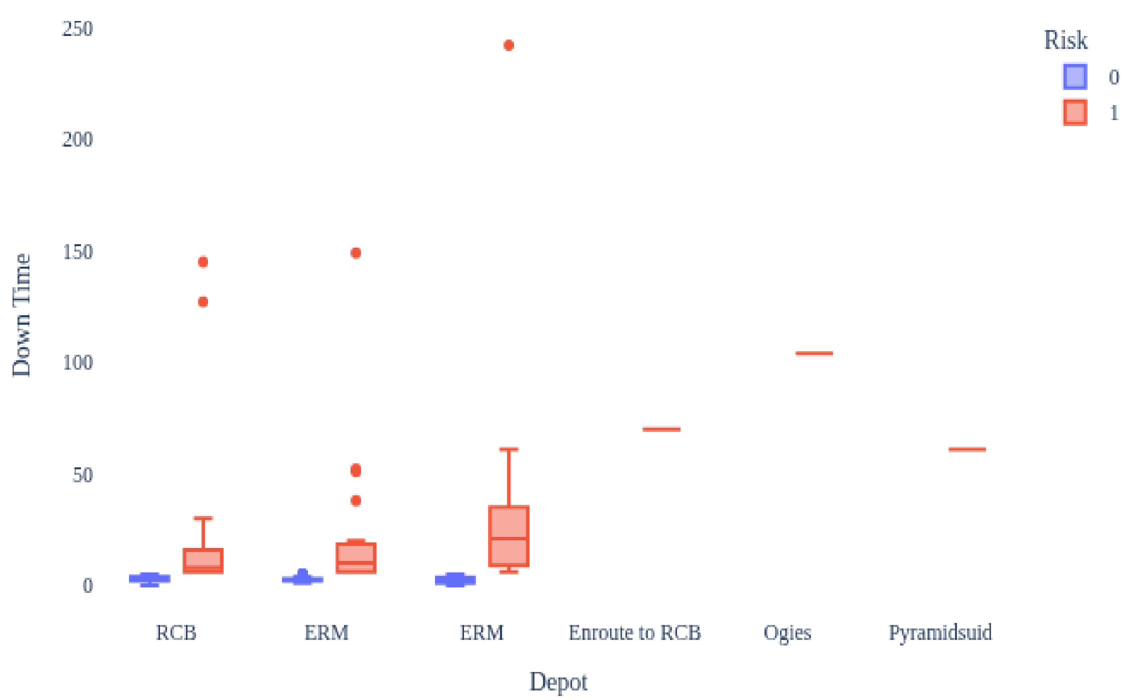

Figure 2.Downtime by Depot and Risk

At the RCB Depot, there are 13 Vehicles categorised as high risk, with an average downtime of 38.4 days, which ranges from 14 to 145 days. In contrast, the low-risk category consists of 88 Vehicles, averaging only 4.2 days of downtime, with periods spanning from 0 to 12 days. This indicates that while the RCB Depot manages a significant number of Vehicles efficiently, a notable proportion experiences substantial delays, potentially due to complex faults or resource constraints. Similarly, the ERM Depot reports 18 high-risk Vehicles, split across two entries, with an average downtime of 50.1 days and a maximum of 242 days, the longest downtime recorded among all depots. This depot also has 72 low-risk Vehicles, averaging about 4 days of downtime. The persistent pattern here resembles that of RCB, indicating a predominance of low-risk cases alongside some extreme outliers. The maximum downtime of 242 days suggests potential challenges in addressing specific fault types or limitations in available resources.

At the Ogies Depot, there are only 2 Vehicles classified as high risk, both having substantial downtime of 104 days. Although the sample size is limited, these consistently high downtime points point to possible systemic issues at the depot, likely stemming from restricted repair capabilities or a shortage of spare parts. The Pyramidsuid Depot has a single high-risk Vehicle with 61 days of downtime, mirroring the situation at Ogies. This isolated case may indicate specific issues or inadequate capacity to address complex repairs effectively.

Finally, there is one high-risk Vehicle enroute to RCB that experienced 70 days of downtime. This situation most likely reflects delays related to transit or scheduling, contributing to extended downtime even prior to the commencement of repairs. Before building a predictive model, it's important to check whether the data meets certain assumptions. Assumption testing included class imbalance assessment (110 low-risk vs. 85 high-risk), missing value checks (none found), and cardinality analysis (8 classes, 6 depots, 187 fault types). The classes are somewhat imbalanced, but not severely. However, the model might still favour the majority class (low-risk), so techniques like resampling or adjusting class weights could be considered to improve fairness and accuracy.

5.4. Feature importance for downtime prediction

The analysis revealed that the "Failure Mode" is the most influential feature in predicting whether a Vehicle is categorised as high or low downtime risk. This finding aligns with the understanding that the nature of the fault directly impacts the time required for repairs. Additionally, the "Depot" variable also plays a significant role in the prediction model, likely due to variations in resource availability, levels of technical expertise, and overall operational efficiency across different locations. While the "Class" feature has the least impact of the three, it still contributes to the prediction; this may be attributed to differences in factors such as the age of the Vehicles, their technological specifications, or the complexity of their maintenance needs.

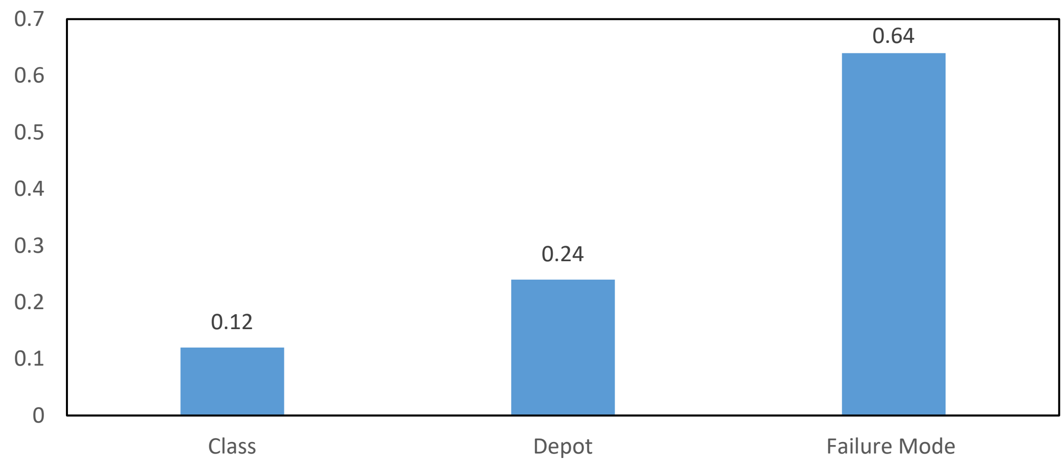

Figure 3. Feature Importance for Downtime Risk Prediction

Figure 4. The curve illustrates the trade-off between the True Positive Rate, also known as Recall, and the False Positive Rate, providing insight into the model's performance. The AUC, or Area Under the Curve, is a critical metric in this context, with a value of approximately 0.97, signifying a strong discriminative ability. This high AUC indicates that the model is particularly effective at distinguishing between High Risk and Low Risk Vehicles, demonstrating its potential utility in predictive maintenance applications.

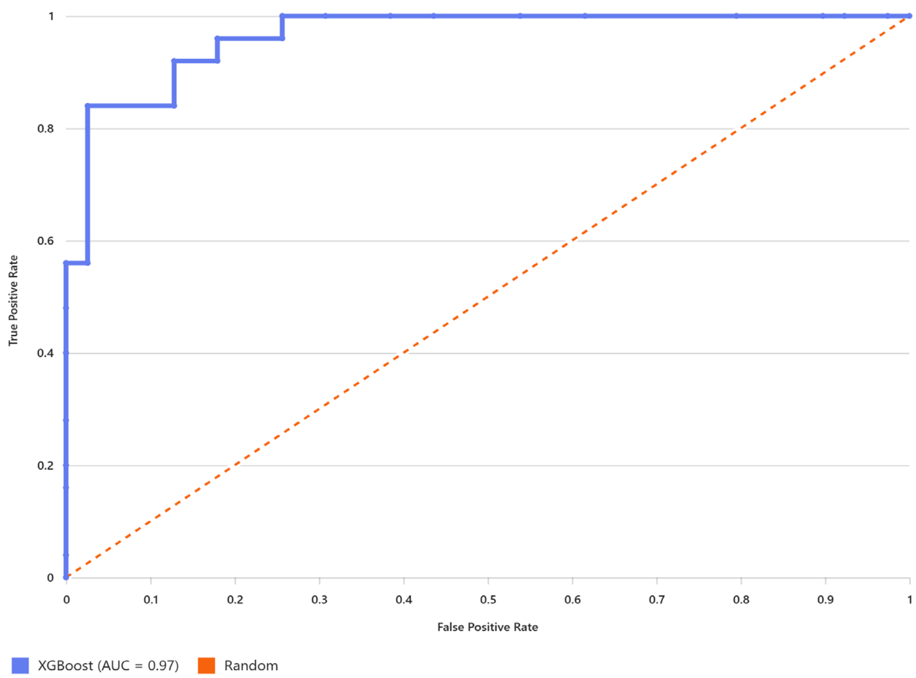

Figure 4. ROC Curve: XGBoost Downtime Risk Model

5.3 Proposed Improvements

This section outlines the essential steps for deploying the XGBoost Downtime Risk Model, designed to enhance Vehicle operations through predictive analytics. It begins with local deployment, tailored for professionals like planners and engineers, enabling them to utilise historical data for accurate downtime forecasts. The narrative then progresses to web-based deployment, emphasising accessibility through cloud platforms and the development of user-friendly interfaces. Integration with existing systems, technical tools required, and a sample workflow are also discussed to provide a comprehensive guide for effectively implementing the model and optimising railway performance.

5.3.1. Local Deployment

This method is specifically designed for internal use by professionals such as planners, engineers, and depot managers who are tasked with managing Vehicle operations and minimising downtime. The predictive model developed for assessing Downtime Risk leverages historical data and advanced algorithms to provide accurate forecasts. To ensure smooth integration into existing workflows, the model is exported using popular serialisation libraries like joblib or pickle. This allows for efficient storage and retrieval of the model, making it readily available for subsequent analysis and predictions. To facilitate user interaction, a dedicated Python application or web dashboard is created using frameworks such as Streamlit or Flask. This user-friendly interface allows authorised personnel to input relevant Vehicle data, such as operational history, maintenance schedules, and performance metrics. Once the data is entered, the application processes it through the predictive model and generates Downtime Risk predictions. These predictions empower frontline managers to make informed decisions regarding maintenance and operations, ultimately leading to optimised Vehicle performance and reduced downtime. As a result, this method enhances operational efficiency, supporting the overall goal of maintaining reliable and effective railway services.

5.3.2. Web-Based Deployment

To enhance accessibility to the application, consider deploying the model on a cloud computing platform such as Microsoft Azure or Amazon Web Services (AWS). You can use a web framework like Flask or FastAPI to establish an Application Programming Interface (API) endpoint. Flask is a lightweight web framework for Python that allows developers to create web applications easily. FastAPI, on the other hand, is a modern web framework for building APIs with Python 3.6 and above, known for its high performance and support for standard Python type hints.

For broader access, host the model on a cloud platform (e.g., Azure or AWS). Utilise a web framework, such as Flask or FastAPI, to create the API endpoint. Additionally, develop a front-end user interface that interacts with the API and implement robust security measures, including user authentication, to protect the application.

5.3.3. Integration with Existing Systems
If the company uses systems like SAP or Maximo, convert the model into a microservice or API. Integrate it via REST API or middleware to automate predictions based on real-time data from maintenance logs or sensors.

5.3.4. Technical Tools Required
- Python (for model and API)
- XGBoost, scikit-learn, joblib
- Flask or FastAPI (for API)
- Streamlit or Dash (for dashboards)
- Docker (for containerization)
- Cloud platform (optional for scalability)

5.4 Validation
To ensure the reliability and robustness of the predictive modelling framework developed in this study, a multi-layered validation strategy was employed, encompassing data integrity checks, statistical hypothesis testing, and model performance evaluation.

5.4.1 Data Integrity and Assumption Testing
Prior to model development, the dataset underwent rigorous preprocessing to ensure quality and consistency. Key steps included the conversion of date fields to standard datetime formats, the removal of entries with negative downtime values, and the elimination of records with missing values in critical fields such as Class, Depot, and Failure Mode. Assumption testing confirmed a moderately imbalanced target variable with 110 low-risk versus 85 high-risk instances, as well as adequate feature diversity with 8 unique Vehicle classes, 6 depots, and 187 fault types. Furthermore, there were no missing values in the selected features used for modelling. These results validated the dataset's suitability for binary classification modelling.

5.4.2 Statistical Hypothesis Testing
To further validate the classification framework and feature relevance, a two-sample independent t-test was conducted to compare the mean downtime between low-risk and high-risk categories. The test yielded a statistically significant result (t-statistic = −20.34, $p < 0.0001$), confirming that the two groups differ meaningfully in downtime duration, thereby justifying the chosen risk thresholds. Additionally, Chi-square tests were performed, revealing the following results: for Depot vs. Risk, $\chi^2 = 3.97$ with a p-value of 0.2647; and for Class vs. Risk, $\chi^2 = 4.26$ with a p-value of 0.5127. These results indicate no statistically significant association between depot or class and downtime risk at the 5% significance level, suggesting that other features, particularly Failure Mode, play a more critical role in prediction.

5.4.3 Model Evaluation and Cross-Validation
The dataset was split into training (70%) and testing (30%) sets to evaluate the model's generalisation capability. To further assess robustness, Stratified K-Fold Cross-Validation (5 folds) was performed on the training data. This method maintains the original class distribution across folds, which is essential given the moderate class imbalance. During cross-validation, the XGBoost model achieved average performance metrics of 0.57 ± 0.07 for accuracy, 0.45 ± 0.10 for the F1 Score, and 0.52 ± 0.09 for ROC AUC. While these results indicate moderate predictive performance during cross-validation, the final evaluation on the held-out test set demonstrated significantly improved metrics, with an accuracy of 88%, a macro F1 Score of 0.87, a weighted F1 Score of 0.88, and an AUC of 0.97. This improvement suggests that the model benefits from the full training set and the application of SMOTE for balancing, though further tuning and feature engineering may be necessary to enhance consistency across unseen data splits.

6. Conclusion
This research underscores the transformative potential of predictive modelling in managing Vehicle downtime risks through the application of machine learning techniques. By integrating structured data from maintenance logs, sensor readings, and operational contexts, the study demonstrates how predictive analytics can significantly improve operational reliability, reduce unplanned downtimes, and enhance operational efficiency in the railway sector.

The comparative analysis of machine learning models revealed that XGBoost offers superior performance in identifying high-risk downtime cases, especially when combined with data balancing techniques like SMOTE. The model achieved a high AUC score of 0.97, indicating strong discriminative power and reliability in real-world applications. Beyond model performance, the study highlights the importance of feature selection, with "Failure Mode" and "Depot" emerging as key predictors of downtime risk. These insights can guide targeted interventions, resource allocation, and strategic planning across maintenance depots.

To ensure practical utility, the paper proposes both local and cloud-based deployment strategies, enabling planners, engineers, and depot managers to integrate predictive tools into their daily operations. The outlined technical roadmap, including API development, dashboard creation, and system integration, provides a scalable framework for implementation.

Looking ahead, future research should explore the integration of real-time sensor data, advanced deep learning models, and economic impact analysis to further refine predictive capabilities. Additionally, longitudinal studies could assess the long-term benefits of predictive maintenance on fleet performance, cost savings, and sustainability. By fostering a data-driven culture and leveraging AI-powered tools, railway operators can transition from reactive to proactive maintenance strategies, ultimately driving efficiency, reliability, and resilience in transportation systems.

References

Allahloh, A. S., Sarfraz, M., Ghaleb, A. M., Al-Shamma'a, A. A., Hussein Farh, H. M., and Al-Shaalan, A. M., "Revolutionizing IC Genset Operations with IIoT and AI: A Study on Fuel Savings and Predictive Maintenance," *Sustainability (Switzerland)*, vol. 15, no. 11, 2023. doi: 10.3390/su15118808.

Davari, N., Veloso, B., Costa, G. de A., Pereira, P. M., Ribeiro, R. P., and Gama, J., "A Survey on Data-Driven Predictive Maintenance for the Railway Industry," *Sensors*, vol. 21, no. 17, 2021. doi: 10.3390/s21175739.

Liang, H., Zhang, Y., Yang, P., Wang, L., and Gao, C., "Comparison and Analysis of Prediction Models for Locomotive Traction Energy Consumption Based on Machine Learning," *IEEE Access*, vol. 11, 2023. doi: 10.1109/ACCESS.2023.3268531.

Mahale, Y., Kolhar, S., and More, A. S., "A Comprehensive Review on Artificial Intelligence-Driven Predictive Maintenance in Vehicles: Technologies, Challenges and Future Research Directions," *Discover Applied Sciences*, vol. 7, no. 4, 2025. doi: 10.1007/s42452-025-06681-3.

Mahiyudin, G., Hussain, M., and Dewi, D. D., "A Comprehensive Study on Predicting the Need for Vehicle Maintenance Using Machine Learning," in *The 7th International Global Conference Series on ICT Integration in Technical Education & Smart Society*, vol. 89, 2025. doi: 10.3390/engproc2025107089.

Priyanga, P., Ashwini, K., Deepa, S. R., and Sridevi, S., "The Smart Factory of Tomorrow: Artificial Intelligence and Machine Learning Reshaping Manufacturing Processes," in *2023 2nd International Conference on Smart Technologies for Smart Nation (SmartTechCon 2023)*, 2023. doi: 10.1109/SmartTechCon57526.2023.10391663.

Ejaz, U., and Bakhsh, M. M., "AI-Driven Predictive Maintenance in Autonomous Public Transit Systems for Smart Cities," *World Journal of Advanced Research and Reviews*, vol. 17, no. 1, pp. 1400–1415, 2023. doi: 10.30574/wjarr.2023.17.1.0129.

Biographies

Dr Bheki B. S. Makhanya is a Senior Researcher at the Postgraduate School of Engineering Management, University of Johannesburg. He is also the Technical Fleet Manager at Transnet Freight Rail. Dr Makhanya holds a PhD in engineering management from the University of Johannesburg and has more than a decade of experience in asset management and reliability improvement in the railway industry. He has been a co-author of numerous peer-reviewed research papers since 2017 and has supervised over 30 master's students in Engineering Management from the University of Johannesburg, South Africa. Dr Makhanya's research interests include cost of quality, total quality management, reliability improvement, and risk management.

Prof Jan Harm Christiaan Pretorius obtained his BSc Hons (Electrotechnics) (1980), MIng (1982) and DIng (1997) degrees in Electrical and Electronic Engineering at the Rand Afrikaans University and an MSc (Laser Engineering and Pulse Power) at the University of St Andrews in Scotland (1989). He worked at the South African Atomic Energy Corporation as a Senior Consulting Engineer for 15 years. He also worked as the Technology Manager at the Satellite Applications Centre. He is currently a professor at the Postgraduate School of Engineering Management in the Faculty of Engineering and the Built Environment where he has worked since 1998. He has co-authored over 240 research papers and supervised 61 PhDs and over 270 Master's students. He is a registered professional engineer, professional Measurement and Verification practitioner, senior member of the Institute of the IEEE, a fellow of the SAIEE and a fellow of the South African Academy of Engineering.