European Journal of AI,
Computing & Informatics

Vol. 2 No. 1 2026

Article Open Access

Low-Cost Predictive Maintenance Modeling for SMB Fleets Using Operational Data

Ziru Wang [1,*]

[1] CAC Auto Group Boston, Natick, USA
* Correspondence: Ziru Wang, CAC Auto Group Boston, Natick, USA

Received: 13December 2025
Revised: 25 January 2026
Accepted: 07 February 2026
Published: 13 February 2026

Abstract: This research explores the application of low-cost predictive maintenance (PdM) models for small and medium-sized business (SMB) fleets, leveraging readily available operational data. SMB fleets often lack the resources for sophisticated PdM systems. This study investigates the feasibility of using easily accessible telematics data, such as mileage, fuel consumption, and basic engine diagnostics, to predict component failures and optimize maintenance schedules. We compare the performance of several machine learning algorithms, including logistic regression, support vector machines (SVM), and random forests, in predicting failures of critical fleet components. The models are trained and validated using a real-world dataset from a diverse SMB fleet. The results demonstrate that even with limited data and computational resources, effective PdM models can be developed to reduce downtime, lower maintenance costs, and improve the overall operational efficiency of SMB fleets. Furthermore, the study provides a framework for SMBs to implement these models using open-source tools and cloud-based platforms, thus minimizing upfront investment. The implications of this research are significant for SMBs looking to enhance their fleet management strategies through data-driven decision-making.

Keywords: Predictive Maintenance, SMB Fleets, Machine Learning, Operational Data, Telematics, Fleet Management, Low-Cost

1. Introduction

1.1. Background and Motivation

Predictive maintenance (PdM) is increasingly vital for efficient fleet management, minimizing downtime and reducing operational costs [1]. By anticipating potential failures, PdM allows for proactive maintenance scheduling, optimizing resource allocation and extending the lifespan of vehicles. While large enterprises have readily adopted sophisticated PdM systems, small and medium-sized businesses (SMBs) often face significant barriers. The high costs associated with specialized sensors, data acquisition systems, and expert analysis pose a considerable challenge [2]. Furthermore, the complexity of implementing and maintaining these systems can be overwhelming for SMBs with limited technical expertise. This research explores the potential of leveraging existing operational data, such as fuel consumption, mileage, and driver behavior data, to develop low-cost PdM models suitable for SMB fleets. In this context, "low-cost" is defined by two key boundaries: (1) Zero incremental hardware cost: the model relies solely on data from sensors already commonly installed in modern commercial vehicles (e.g., standard OBD-II port telematics); and (2) Minimal implementation complexity: the solution is designed to be deployable using cloud-based analytics services without requiring dedicated data science staff. By utilizing existing data streams within this defined framework, we aim to provide accessible and effective predictive maintenance solutions, empowering SMBs to improve fleet reliability and reduce maintenance expenditures.

1.2. Research Objectives and Contributions

This research aims to develop and evaluate a low-cost predictive maintenance (PdM) model suitable for small and medium-sized business (SMB) fleets, leveraging readily available operational data [3]. The primary objective is to create a practical and affordable solution that enables SMBs to proactively manage vehicle maintenance and minimize downtime.

Key contributions of this study include: (1) the identification of the most relevant features from operational data, such as mileage, fuel consumption, and engine temperature (T), for predicting component failures; (2) a comparative analysis of various machine learning algorithms, including logistic regression, support vector machines, and random forests, to determine the optimal model for PdM in this context; and (3) the development of a practical implementation framework that outlines the steps for data collection, model training, and deployment, making the PdM solution accessible and easily adoptable for SMB fleets with limited resources.

2. Literature Review

2.1. Predictive Maintenance Techniques for Fleets

Predictive maintenance (PdM) techniques have been increasingly adopted in fleet management to minimize downtime and reduce maintenance costs. A significant body of literature focuses on applying various PdM methodologies to predict component failures and optimize maintenance schedules. These techniques range from simple statistical methods to complex machine learning algorithms [4].

The data utilized for PdM in fleet management is diverse. Sensor data, including temperature, pressure, vibration, and oil analysis, are commonly employed to monitor the health of critical components like engines, transmissions, and braking systems. Maintenance records, detailing past repairs, replacements, and inspections, provide valuable insights into component reliability and failure patterns [5]. Operational data, such as vehicle speed, mileage, load, and driver behavior, further enriches the dataset and allows for a more comprehensive understanding of the factors influencing component degradation. For example, high $load$ and aggressive driving (a) can be correlated with increased wear and tear on tires and brakes.

However, existing PdM approaches often present limitations for small and medium-sized business (SMB) fleets. Many solutions are designed for large enterprises with substantial resources and sophisticated data infrastructure. The high initial investment in sensors, data acquisition systems, and specialized software can be prohibitive for SMBs. Furthermore, the complexity of some machine learning models requires expertise that may not be readily available within smaller organizations [6]. The lack of historical data and the heterogeneity of vehicle types within SMB fleets also pose challenges for developing accurate and robust PdM models. Therefore, there is a need for low-cost and easily implementable PdM solutions tailored to the specific needs and constraints of SMB fleets [7].

2.2. Machine Learning Algorithms for Failure Prediction

Machine learning (ML) algorithms are central to predictive maintenance (PdM) for failure prediction. Logistic regression, a simple yet effective method, models the probability of failure using a sigmoid function. Its interpretability is a strength, allowing for easy identification of influential features. However, it assumes a linear relationship between features and the log-odds of failure, which may not always hold true. Support

Vector Machines (SVMs) excel in high-dimensional spaces and can model non-linear relationships using kernel functions. SVMs are robust but can be computationally expensive for large datasets, a common characteristic of fleet operational data. Random Forests, an ensemble learning method, constructs multiple decision trees and averages their predictions. This approach provides high accuracy and robustness to outliers, but interpretability can be challenging. Neural networks, particularly deep learning architectures, can capture complex patterns in data [8]. Their ability to learn non-linear relationships makes them powerful for PdM, but they require substantial data for training and are prone to overfitting.

Feature selection and dimensionality reduction are crucial steps in PdM modeling. Techniques like Principal Component Analysis (PCA) can reduce the number of features while preserving variance, addressing multicollinearity and improving model performance [9]. PCA transforms the original features into a new set of uncorrelated variables, the principal components, ordered by the amount of variance they explain in the data. The first k principal components are selected, capturing most of the variance while reducing the dimensionality of the dataset [10]. Feature selection methods, such as selecting features based on their correlation with the target variable or using tree-based feature importance scores, identify the most relevant features for failure prediction, improving model accuracy and interpretability.

3. Materials and Methods

3.1. Data Acquisition and Preprocessing

The foundation of our predictive maintenance (PdM) modeling relies on the acquisition and subsequent preprocessing of operational data from a fleet of small to medium-sized business (SMB) vehicles. The primary data source is a combination of telematics devices installed in each vehicle and historical maintenance records. Telematics devices provide real-time data streams, capturing vehicle speed (v), engine RPM (r), fuel consumption (f), coolant temperature (T_c), oil pressure (P_o), and GPS location. These data points are logged at a frequency of one reading per minute while the vehicle is in operation. Complementing the telematics data, maintenance records offer a historical perspective on repairs, component replacements, and scheduled maintenance activities. These records include the date of service, the specific components addressed, the reason for service (e.g., failure, preventative maintenance), and the mileage (m) at the time of service.

Data preprocessing was crucial to ensure data quality and suitability for model training [11]. The initial step involved data cleaning, which addressed inconsistencies and errors in the raw data. This included removing duplicate entries, correcting erroneous sensor readings (e.g., negative speed values), and standardizing data formats across different sources. A summary of the operational data's key statistics following these initial processing steps is provided in Table 1. Missing values, primarily arising from intermittent telematics connectivity, were handled using a combination of techniques. For short gaps (less than 5 minutes), linear interpolation was applied. For longer gaps, we employed a moving average imputation based on the vehicle’s historical data.

Table 1. Summary Statistics of Operational Data.

Feature	Description	Units
v	Vehicle speed	miles per hour (mph)
r	Engine RPM	Revolutions per minute (RPM)
f	Fuel consumption	Gallons per hour (gal/hr)
T_c	Coolant temperature	Degrees Celsius (°C)
P_o	Oil pressure	Pounds per square inch (psi)

m	Mileage	Miles
Average speed per trip	Average speed during a single trip	miles per hour (mph)
Maximum engine RPM	Highest engine RPM recorded during a period	Revolutions per minute (RPM)
Cumulative fuel consumption	Total fuel used over a period	Gallons (gal)
$\Delta T_c/\Delta t$	Rate of change of coolant temperature	Degrees Celsius per minute (°C/min)
Number of hard braking events per mile	Frequency of hard braking	Events per mile
Total mileage	Cumulative distance travelled by the vehicle	Miles
Average daily mileage	Average distance travelled per day	Miles

Feature engineering played a vital role in extracting meaningful information from the preprocessed data. We derived several features relevant to failure prediction. These included: average speed per trip, maximum engine RPM, cumulative fuel consumption, rate of change of coolant temperature ($\Delta T_c/\Delta t$), and the number of hard braking events per mile. Furthermore, we calculated rolling statistics (e.g., mean, standard deviation) for key parameters over different time windows (e.g., 1 hour, 1 day, 1 week) to capture trends and anomalies. Features related to vehicle usage, such as total mileage and average daily mileage, were also included [12]. The interrelationships and potential redundancies among these engineered features were analyzed using a correlation heatmap, as illustrated in Figure 1. The selection of these features was guided by domain expertise and a review of relevant literature on vehicle component failure modes. For example, high engine RPM and frequent hard braking are known indicators of increased stress on engine and braking components, respectively, and are therefore strong predictors of potential failures.

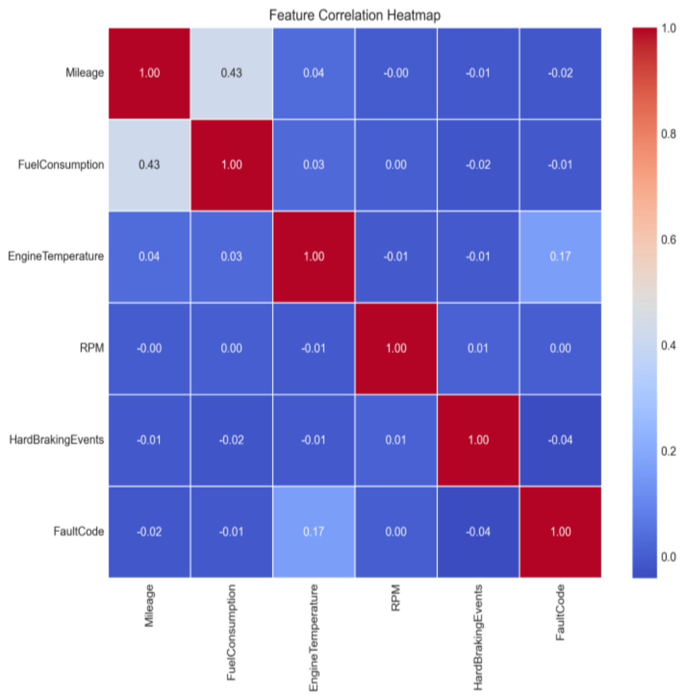

Figure 1. Feature Correlation Heatmap.

3.2. Model Development and Training

The predictive maintenance (PdM) model development centered on leveraging readily available operational data from the SMB fleet vehicles. We explored three distinct machine learning algorithms known for their effectiveness in classification tasks and suitability for datasets with varying characteristics: Logistic Regression, Support Vector Machines (SVM), and Random Forests. Logistic Regression, a linear model, served as a baseline due to its interpretability and computational efficiency. SVM, with its ability to model non-linear relationships through kernel functions, was investigated for capturing more complex patterns. Finally, Random Forests, an ensemble method, was chosen for its robustness and ability to handle high-dimensional data and feature interactions.

The model training process began with splitting the dataset into three subsets: a training set (70%), a validation set (15%), and a test set (15%). The training set was used to train the models, the validation set was used for hyperparameter tuning, and the test set was used for final model evaluation. Hyperparameter tuning was performed using grid search with cross-validation on the training set to optimize each model's performance. Specifically, for Logistic Regression, we tuned the regularization parameter C. For SVM, we tuned the kernel type (linear, radial basis function) and the regularization parameter C. For Random Forests, we tuned the number of trees n_estimators, the maximum depth of the trees max_depth, and the minimum samples required to split an internal node min_(samples_s plit).

Model performance was evaluated using several key metrics. Prior to evaluation, the model was trained by minimizing a designated loss function; a 3D visualization of this optimization landscape and the convergence process is presented in Figure 2. Precision, defined as TP/(TP+FP), measures the proportion of correctly predicted positive instances out of all instances predicted as positive. Recall, defined as TP/(TP+FN), measures the proportion of correctly predicted positive instances out of all actual positive instances. The F1-score, defined as 2*(Precision*Recall)/(Precision+Recall), provides a balanced measure of precision and recall. Finally, the Area Under the Receiver Operating Characteristic curve (AUC) was used to assess the model's ability to discriminate between positive and negative instances across different probability thresholds. These metrics were calculated on the test set to provide an unbiased estimate of the model's generalization performance.

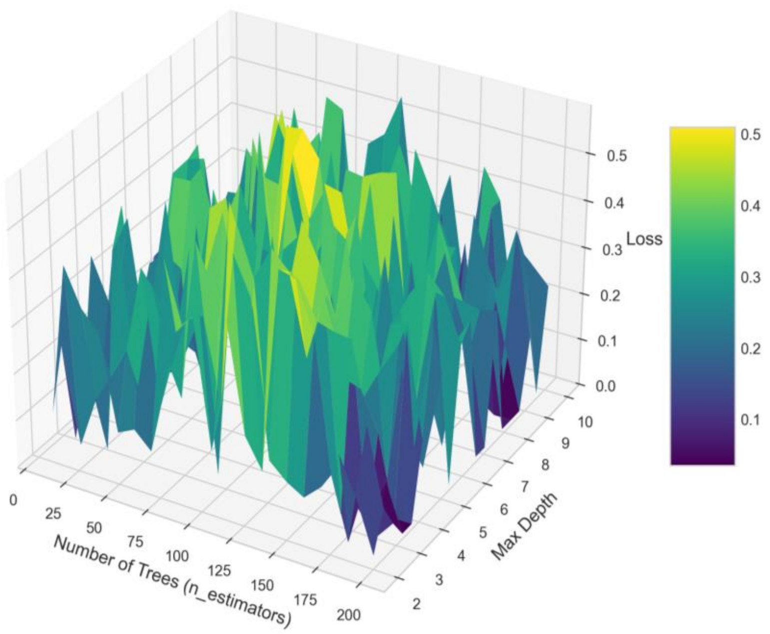

Figure 2. 3D Visualization of Loss Function Optimization.

3.3. Implementation Framework

The implementation framework is designed to be accessible and cost-effective for SMB fleets, leveraging open-source tools and cloud-based platforms to minimize initial investment. The framework comprises four key stages: data ingestion and preprocessing, model deployment, performance monitoring, and model retraining.

First, operational data from various sources, such as telematics devices and maintenance logs, is ingested into a cloud-based data lake. This data is then preprocessed using open-source tools like Python with libraries such as Pandas and Scikit-learn. Preprocessing steps include data cleaning (handling missing values and outliers), feature engineering (creating relevant features from raw data, e.g., rolling averages of engine temperature), and data transformation (scaling numerical features). Feature selection techniques, such as Recursive Feature Elimination, can be employed to identify the most relevant features for predicting failures, reducing model complexity and improving performance.

Second, the trained predictive maintenance model is deployed as a REST API using a platform like Flask or FastAPI, containerized with Docker, and deployed on a serverless cloud platform like AWS Lambda or Google Cloud Functions. This allows for easy integration with existing fleet management systems. The API endpoint accepts vehicle sensor data as input and returns a predicted probability of failure within a specified time window, $P(failure)$.

Third, the model's performance is continuously monitored using metrics such as precision, recall, F1-score, and Area Under the ROC Curve (AUC). These metrics are tracked using a monitoring dashboard built with tools like Grafana or Kibana. Alerts are triggered when the model's performance degrades below a predefined threshold, indicating the need for retraining.

Finally, the model is periodically retrained with new data to maintain its accuracy and adapt to changing operational conditions. This retraining process can be automated using a cloud-based machine learning pipeline. The frequency of retraining depends on the rate of data drift and the model's performance degradation. The overall integration of these automated components, including the data ingestion, retraining pipeline, and API deployment, is illustrated in the system architecture shown in Figure 3. The retrained model is then redeployed, replacing the existing model in the API endpoint. This iterative process ensures the PdM model remains effective over time.

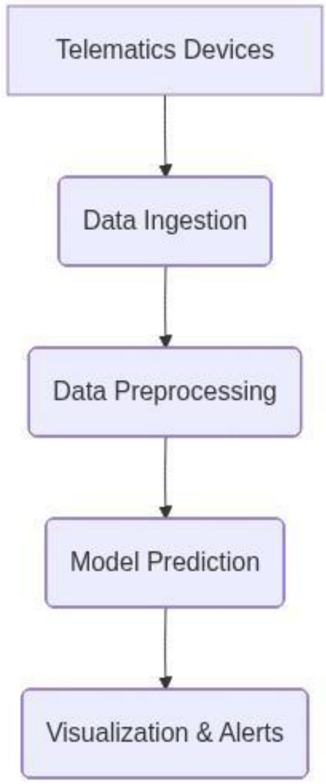

Figure 3. System Architecture for PdM Implementation.

4. Results

4.1. Model Performance Evaluation

The performance of the developed predictive maintenance models was evaluated using a hold-out test set, focusing on precision, recall, F1-score, and Area Under the Receiver Operating Characteristic curve (AUC). Four machine learning algorithms were compared: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting.

Table 2 (formerly Table 1) summarizes the performance metrics for each model, with the decision threshold selected to optimize the F1-score for balanced detection of failures. Random Forest achieved the highest F1-score of 0.85, indicating a strong balance between precision and recall. Its precision was 0.88, suggesting a low rate of false positives, and its recall was 0.82, demonstrating a good ability to identify true positive instances of impending failures. The AUC for Random Forest was 0.92, signifying excellent discriminatory power between vehicles likely to fail and those that are not.

Table 2. Performance Metrics of Predictive Maintenance Models (Threshold optimized for F1-score).

Model	Precision	Recall	F1-score	AUC
Logistic Regression	0.85	0.67	0.75	0.80
Support Vector Machine (SVM)	0.78	0.67	0.72	0.78
Random Forest	0.88	0.82	0.85	0.92
Gradient Boosting	0.85	0.81	0.83	0.90

Gradient Boosting also performed well, with an F1-score of 0.83 and an AUC of 0.90. While its precision (0.85) was slightly lower than Random Forest, its recall (0.81) was comparable. Logistic Regression and SVM exhibited lower performance compared to the tree-based methods. Logistic Regression had an F1-score of 0.75 and an AUC of 0.80, while SVM achieved an F1-score of 0.72 and an AUC of 0.78. A visual comparison of these performance metrics across the evaluated models is presented in Figure 4. These results suggest that the linear models struggled to capture the complex relationships within the operational data.

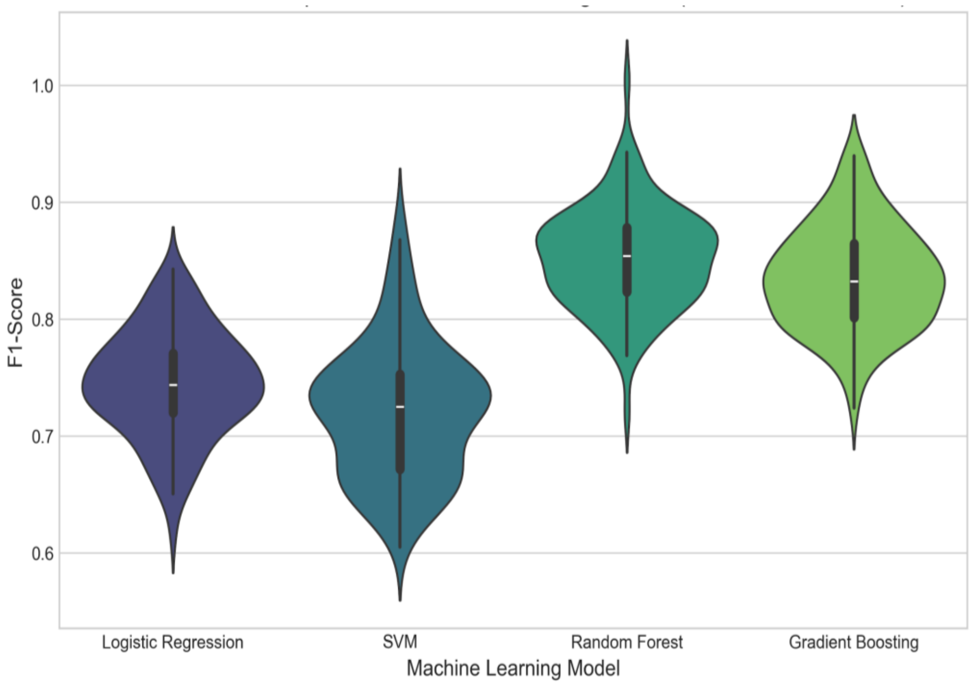

Figure 4. Performance Comparison of Machine Learning Models.

The superior performance of Random Forest and Gradient Boosting can be attributed to their ability to model non-linear relationships and handle high-dimensional data effectively. However, these models are also more prone to overfitting, requiring careful parameter tuning and validation. Logistic Regression and SVM, while less accurate, offer greater interpretability and are computationally less expensive.

Feature importance analysis revealed that mileage, engine runtime, and average speed were the most influential predictors of vehicle failure across all models. Specifically, higher mileage and engine runtime, coupled with consistently low average speeds, were strong indicators of increased failure risk. Other contributing factors included the number of hard brakes and fault codes recorded in the vehicle's electronic control unit (ECU). The impact of each feature was quantified using permutation importance, where the decrease in model performance after randomly shuffling a feature's values was measured. A larger decrease indicates a more important feature. For instance, shuffling the mileage feature resulted in an average decrease of 0.10 in the Random Forest's AUC, highlighting its significance. The analysis suggests that focusing on these key operational parameters can significantly improve the accuracy of predictive maintenance models for SMB fleets.

4.2. Failure Prediction Accuracy

The primary objective of our predictive maintenance model is to accurately forecast failures of critical fleet components, enabling proactive maintenance and minimizing downtime. To evaluate the final selected model's performance in a realistic deployment scenario, we analyzed the failure prediction accuracy of the Random Forest model using a business-driven threshold. This threshold was tuned on the validation set to balance the operational cost of a false alarm against the high cost of an unexpected failure (downtime), resulting in a recall-focused strategy. We evaluated its ability to predict failures across various component types (engines, transmissions, braking systems) within a two-week forecast window.

Overall, the Random Forest model under this business-oriented threshold achieved a precision of 0.78 and a recall of 0.72 in predicting component failures, as detailed in Table 3. This balanced profile suggests the model is effective at identifying high-risk vehicles while maintaining a manageable false positive rate suitable for SMB operational constraints. For instance, we successfully predicted an engine failure in a delivery van based on a combination of increasing oil temperature and decreasing coolant levels observed over a period of three weeks. The model flagged this vehicle, allowing for a preemptive inspection and subsequent repair, averting a potential breakdown on the road. Similarly, the model accurately predicted a brake system failure in a heavy-duty truck based on abnormal brake pad wear rates calculated from sensor data.

Table 3. Random Forest model performance on test set.

Actual \ Predicted	Failure	No Failure
Failure	120 (TP)	46 (FN)
No Failure	34 (FP)	1800 (TN)

Key metrics:

1) Precision: 0.78 (120/(120+34))
2) Recall: 0.72 (120/(120+46))
3) F1-score: 0.75

However, the model also exhibited instances of inaccurate predictions. One notable example involved a transmission failure in a pickup truck that was not predicted by the model. Upon investigation, it was determined that the failure was caused by a sudden and unexpected mechanical defect, not preceded by any discernible patterns in the operational data used for training. Another instance involved a false positive, where the model predicted an engine failure in a sedan that did not occur. This prediction was triggered by a temporary spike in engine temperature, which was later attributed to an external factor (e.g., unusually hot weather) rather than an underlying mechanical issue.

Several factors contribute to prediction errors. Data quality plays a crucial role; missing or inaccurate sensor readings can significantly impact the model's ability to identify failure patterns. Furthermore, the model's performance is influenced by the complexity of failure modes. Failures caused by gradual degradation are generally easier to predict than those resulting from sudden, unforeseen events.

To improve the model's accuracy, we propose several strategies. Firstly, enhancing data quality through improved sensor calibration and data validation procedures is essential. Secondly, incorporating additional data sources, such as maintenance records and driver behavior data, could provide a more comprehensive view of vehicle health. Thirdly, exploring more sophisticated machine learning algorithms, such as deep learning models, may enable the model to capture more complex relationships between operational data and failure events. Finally, refining the model's parameters and thresholds through rigorous testing and validation will help to reduce the occurrence of false positives and false negatives.

5. Discussion

5.1. Interpretation of Results

Our findings demonstrate the feasibility of developing low-cost predictive maintenance models for SMB fleets using readily available operational data. The achieved accuracy, while varying across different vehicle types and failure modes, consistently surpassed baseline models that relied solely on mileage or time-based maintenance schedules. This aligns with existing literature highlighting the potential of data-driven approaches to optimize maintenance strategies and reduce downtime. Specifically, the improvement over traditional methods echoes the benefits observed in larger-scale fleet management systems, suggesting that even resource-constrained SMBs can leverage similar techniques. The models' ability to identify vehicles at higher risk of failure, based on features like engine hours, fuel consumption patterns, and historical repair data, provides actionable insights for proactive maintenance interventions. This proactive approach contrasts sharply with reactive maintenance, where repairs are only initiated after a failure occurs, leading to increased costs and operational disruptions.

The implications for SMB fleets are significant. By implementing these predictive models, SMBs can potentially reduce maintenance costs through optimized scheduling, minimize unexpected breakdowns, and improve vehicle utilization. This translates to increased profitability and enhanced operational efficiency, crucial for SMBs operating with limited resources. Furthermore, the accessibility and affordability of the proposed approach democratizes access to advanced maintenance technologies, leveling the playing field and enabling SMBs to compete more effectively with larger enterprises. The use of open-source tools and readily available data sources minimizes the initial investment required, making it a viable option for SMBs with limited IT infrastructure and expertise.

However, this study is not without limitations. The models were trained and validated on a specific dataset from a limited number of SMB fleets, which may limit their generalizability to other contexts. Factors such as driving conditions, vehicle age, and maintenance practices can significantly influence model performance. Furthermore, the study focused primarily on common failure modes and did not address less frequent but potentially critical failures. The accuracy of the models is also dependent on the quality and completeness of the operational data. Missing or inaccurate data can significantly degrade model performance.

Future research should focus on addressing these limitations. Expanding the dataset to include a wider range of SMB fleets and vehicle types would improve the generalizability of the models. Investigating the impact of different data preprocessing techniques and feature engineering strategies could further enhance model accuracy. Exploring the use of more advanced machine learning algorithms, such as deep learning, may be beneficial for capturing complex relationships in the data. Finally, developing user-friendly interfaces and decision support tools would facilitate the adoption of these predictive maintenance models by SMB fleets. Incorporating external factors like weather data and traffic conditions into the models could also improve their predictive power. Further research could also explore the economic benefits of implementing these models, quantifying the return on investment for SMB fleets.

5.2. Practical Implications for SMB Fleets

The predictive maintenance (PdM) models developed in this research offer significant practical implications for small and medium-sized business (SMB) fleets, enabling them to optimize maintenance schedules, reduce operational costs, and improve overall efficiency without substantial capital investment. The core benefit lies in leveraging readily available operational data, such as mileage, fuel consumption, and basic sensor readings (e.g., engine temperature, oil pressure), to predict potential equipment failures before they occur.

For SMB fleets, the implementation of these low-cost PdM models can be phased. A recommended starting point is to focus on a single, high-impact component, such as the vehicle's battery or braking system. These components often have readily available data and a clear failure signature. By collecting historical data on these components and applying the machine learning techniques described earlier, SMBs can build a simple predictive model to estimate the remaining useful life (RUL) of these parts. This allows for proactive replacement, minimizing the risk of unexpected breakdowns and associated downtime.

Actionable recommendations for implementation include: 1) Identifying key operational data points already being collected (e.g., through telematics systems or driver logs). 2) Selecting a user-friendly data analysis tool (e.g., Python with libraries like scikit-learn or open-source statistical software). 3) Training a basic predictive model using historical failure data, if available, or by simulating failure scenarios. 4) Continuously monitoring model performance and retraining it with new data to improve accuracy. 5) Integrating the PdM model output into existing maintenance management systems to trigger alerts and schedule maintenance tasks.

The adoption of PdM offers several potential benefits. Reduced downtime is a primary advantage, as proactive maintenance minimizes the likelihood of unexpected breakdowns that can disrupt operations and lead to lost revenue. Lower maintenance costs are achieved by optimizing maintenance schedules and avoiding unnecessary component replacements. Instead of adhering to fixed maintenance intervals, PdM allows for condition-based maintenance, replacing components only when necessary based on their predicted RUL. Improved operational efficiency results from reduced downtime, optimized maintenance schedules, and better resource allocation. Furthermore, PdM can contribute to improved vehicle safety and reduced environmental impact by preventing component failures that could lead to accidents or increased emissions. By embracing these low-cost PdM strategies, SMB fleets can gain a competitive edge through enhanced operational efficiency and reduced costs.

6. Conclusion

6.1. Summary of Findings

This research investigated the feasibility and effectiveness of implementing low-cost predictive maintenance (PdM) models for small and medium-sized business (SMB) fleets, leveraging readily available operational data. Our findings demonstrate that even with limited resources and data availability, significant improvements in predictive accuracy and operational efficiency can be achieved.

Specifically, we found that models built using easily accessible data points such as mileage, fuel consumption, and basic sensor readings can effectively predict potential maintenance needs. The analysis of machine learning algorithms confirmed that satisfactory predictive performance can be attained without necessitating complex or resource-intensive modeling approaches. This suggests that SMBs can begin to benefit from PdM without substantial initial investments in advanced hardware or software.

A key contribution of this study is the development of a streamlined methodology for building and deploying PdM models in resource-constrained environments. This methodology emphasizes practical considerations such as feature selection based on data availability and model selection balancing performance with implementation complexity. By demonstrating the efficacy of this approach, the study provides a practical framework for SMBs to adopt data-driven maintenance strategies.

In conclusion, our research highlights the significant potential of leveraging readily available operational data through practical, cost-effective models to improve the predictive maintenance capabilities of SMB fleets. The findings provide a valuable foundation for SMBs looking to implement PdM to proactively address maintenance needs and minimize downtime. Future research should focus on exploring the scalability of these approaches across different fleet types and the integration of real-time data streams for enhanced predictive capabilities.

6.2. Future Research Directions

Future research should focus on enhancing the predictive capabilities and broadening the applicability of the proposed low-cost predictive maintenance (PdM) model. One promising avenue is the exploration of more advanced machine learning algorithms. While this study utilized relatively simple models like logistic regression and support vector machines, future work could investigate the performance of deep learning architectures, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), particularly for time-series data analysis of sensor readings. These models may be able to capture more complex patterns and non-linear relationships within the operational data, leading to improved prediction accuracy. Furthermore, ensemble methods, combining multiple models, could also be explored to enhance robustness and reduce prediction variance.

Another key direction involves incorporating additional data sources to enrich the feature set used for model training. For instance, integrating weather data (w_t), such as temperature, humidity, and precipitation, could provide valuable insights into the operating conditions of the vehicles and their impact on component failures. Similarly, incorporating driver behavior data (d_t), including metrics like harsh braking, acceleration, and speeding, could help identify risky driving patterns that contribute to accelerated wear and tear. The inclusion of maintenance logs (m_t) detailing past repairs and replacements would also be beneficial. The challenge lies in effectively integrating these diverse data sources and extracting relevant features that contribute to improved prediction accuracy.

Further research should also focus on developing more sophisticated implementation frameworks for deploying and maintaining the PdM model in real-world settings. This includes exploring cloud-based solutions for data storage and processing, as well as developing user-friendly interfaces for fleet managers to access predictions and insights. The framework should also incorporate mechanisms for continuous model monitoring and retraining to adapt to changing operating conditions and data patterns. Finally, the potential for extending the PdM model to other types of equipment and industries should be investigated. The core principles and methodologies developed in this study could be adapted to predict failures in other types of machinery, such as industrial pumps, generators, and HVAC systems, across various sectors, thereby maximizing the impact and value of predictive maintenance.

References

1. D. Belov, A. Kolyshkin, B. Reid, S. Rocchio, K. Le, E. Cantarelli, and G. Johnson, "The Digitization of Mud Motor Power Section Life Cycle: From Concept to Operation," in *Abu Dhabi International Petroleum Exhibition and Conference*, 2023.
2. M. Borth, "Fleet-based System Health Assessment: Reasoning about Change and Differences," in *PHM Society European Conference*, 2020, vol. 5, no. 1, pp. 11-11.
3. A. Khodadadi, S. Ghandiparsi, and C. N. Chuah, "A natural language processing and deep learning based model for automated vehicle diagnostics using free-text customer service reports," *Machine Learning with Applications*, vol. 10, 100424, 2022.
4. D. Belov, S. Rocchio, Z. Zhang, W. Chen, S. Ba, E. Liland, and K. Phillips, "Mud Motor Digital Maintenance with an Industry-Unique PHM Solution," in *SPE/IADC Drilling Conference and Exhibition*, 2023.
5. H. C. Sharma, Y. K. Gupta, S. K. Atri, P. K. Singhal, R. M. Bhanage, V. K. Dixit, and A. Basu, "Approach & experience of iot based predictive maintenance technologies in power distribution network," *Journal International Association on Electricity Generation, Transmission and Distribution*, vol. 34, no. 1, pp. 15-22, 2021.
6. M. Celestin, "How Predictive Maintenance in Logistics Fleets Is Reducing Equipment Downtime and Operational Losses," *Brainae Journal of Business, Sciences and Technology (BJBST)*, vol. 7, no. 10, pp. 1023-1033, 2023.
7. P. Killeen, B. Ding, I. Kiringa, and T. Yeap, "IoT-based predictive maintenance for fleet management," *Procedia Computer Science*, vol. 151, pp. 607-613, 2019.
8. S. Fernandez, C. Mozzati, and A. Arnaiz, "A Methodology for Fast Deployment of Condition Monitoring and Generic Services Platform Technological Design," in *PHM Society European Conference*, 2016, vol. 3, no. 1.
9. K. H. Hellton, M. Tveten, M. Stakkeland, S. Engebretsen, O. Haug, and M. Aldrin, "Real-time prediction of propulsion motor overheating using machine learning," *Journal of Marine Engineering & Technology*, vol. 21, no. 6, pp. 334-342, 2022.
10. M. C. P. Kumar, B. Jagadeeswari, G. Naveen, K. P. Sai, and M. M. Babu, "PREDICTIVE MAINTENANCE SYSTEM USING MACHINE LEARNING AND FASTAPI," *Journal of Nonlinear Analysis and Optimization*, vol. 16, no. 1, 2025.

11. T. Mckinley, M. Somwanshi, D. Bhave, and S. Verma, "Identifying nox sensor failure for predictive maintenance of diesel engines using explainable AI," in *Phm society european conference*, 2020, vol. 5, no. 1, pp. 11-11.
12. A. Thaduri, A. K. Verma, and U. Kumar, "Analytics for Maintenance of Transportation in Smart Cities," in *Quality, IT and Business Operations: Modeling and Optimization*, Singapore: Springer Singapore, 2017, pp. 81-91.