2025

Fleet report

OGBECHIE CHRISTOPHER, IBRAHIM LUKMAN, MATTHEW ASARE

AI-Based Fleet Performance and Passenger Demand Prediction System

MetroMove Transit Services

Author: Christopher F. Ogbechie
Course: ANLT202
Date: December 2025

---

1. Introduction & Problem Statement

MetroMove Transit Services operates a regional vehicle fleet that handles thousands of logistics and transportation trips daily. The organization faces recurring operational challenges related to delayed deliveries, unsafe driving behavior, and unpredictable vehicle maintenance costs. These challenges result in increased operating expenses, inefficient fleet allocation, and reduced overall service quality.

The objective of this project is to design and implement an AI-driven Decision Support System using machine learning and deep learning to:

- Identify high-risk trips that are likely to experience delays,
- Predict maintenance costs for vehicles,
- Improve operational decision-making through data-driven analytics,
- Deploy a cloud-accessible prediction system for operational staff.

This system supports MetroMove's goal of improving reliability, reducing operational risks, and enabling proactive fleet management.

---

2. Dataset Description

Dataset Source

The dataset used in this project was obtained from Kaggle:

Fleet Management Dataset
Source: https://www.kaggle.com/datasets/nhmishuk/fleet-dataset
File: fleet_dummy_5000.csv

Dataset Overview

- Records: 5,000

- Features: 25
- Data Types:
  - Numeric: 16
  - Categorical: 9

Key Variables

Category	Examples
Location	gps_start_lat, gps_end_lon
Costs	fuel_cost, maintenance_cost, toll_cost
Behavior	violation_count, speeding_incidents
Time	pickup_time, delivery_time
Label	status

Data Quality

- No missing values.
- No corrupted records.
- Suitable for classification and regression modeling.

---

3. Machine Learning Problem Definition

3.1 Classification Task — Trip Risk Prediction

Objective: Predict whether a trip is **High Risk** or **Normal**.

The risk label is derived from trip status:

Status	Label
Delayed	1 (High Risk)
Delivered / In Transit / Scheduled	0 (Normal)

Class Distribution:

- Normal Trips: 4,499
- Delayed Trips: 501

Purpose:

This classification allows MetroMove to:

- Identify risky trips early,
- Prioritize operational attention,
- Reduce future delays.

3.2 Regression Task — Maintenance Cost Prediction

Objective: Predict the **maintenance cost** per trip using vehicle and trip attributes.

Target variable:

- maintenance_cost

Features:

- Distance traveled
- Fuel cost
- Toll charges
- Load value
- Speeding incidents
- GPS coordinates
- Pickup hour
- Day of week

---

4. Data Preprocessing

Feature Engineering

The following features were added:

- pickup_hour

- pickup_dayofweek
- delivery_delay_min (derived time difference)

Data Leakage Prevention

An initial regression model achieved unrealistic accuracy (R^2 = 1.0) due to the inclusion of profit_margin, which directly depends on maintenance_cost.

Therefore, the following column was removed:

profit_margin

This ensured valid model performance evaluation.

Data Splitting

- 80% Training
- 20% Test
- Stratified split for classification

Scaling

StandardScaler was applied for Logistic Regression and deep learning models.

---

5. Classification Models & Results

Models Implemented

1. Logistic Regression (Baseline)
2. Random Forest Classifier
3. Balanced Logistic Regression
4. Deep Learning Model (Dense Neural Network)

Baseline Models Performance

Both Logistic Regression and Random Forest predicted only the majority class.

Model	Delayed Recall
Logistic Regression	0%
Random Forest	0%

Balanced Logistic Regression (Final Model)

Class weighting was applied to handle imbalance:

Metric (Delayed Trips)	Score
Recall	~47%
Precision	~9%
Accuracy	50%

Confusion Matrix

[[454 446]

[54 46]]

Deep Learning Performance

The neural network model predicted only the majority class, failing to identify delayed trips.

Conclusion:

Deep learning did not outperform classical machine learning for this dataset.

---

6. Regression Models & Results

Models Implemented

1. Linear Regression
2. Random Forest Regressor

After removing leakage:

Model	MAE	R^2
Linear Regression	62.9	0.00
Random Forest	64.1	Negative

Conclusion:

The dataset does not contain strong maintenance cost predictors. Data quality limits regression accuracy.

---

7. Hyperparameter Tuning Results

Classification (Logistic Regression)

Grid search tuning resulted in:

Parameter	Value
Solver	liblinear
Regularization (C)	0.01

Best Recall (Delayed): **46.8%**

Regression (Random Forest)

Best configuration:

Parameter	Value
max_depth	10
n_estimators	50

MAE improved slightly, but regression remained weak.

---

8. Cloud Deployment

A live web application was deployed using Streamlit Cloud.

Application Link

Public URL:
https://ai-fleet-demand-metromove-o8qov4sdpgtreftqbj6ysu.streamlit.app/

Features

- User input form
- Risk prediction
- Maintenance cost estimation
- Interactive interface
- Cloud-accessible design

Technology Stack

- Python
- Streamlit
- Scikit-learn
- Pandas

Screenshots

(Insert screenshots here from browser once captured.)

---

9. Conclusion

This project demonstrates:

- How machine learning improves fleet monitoring
- The importance of avoiding data leakage
- The value of recall over accuracy in risk detection
- Why model complexity alone does not guarantee performance

Balanced Logistic Regression achieved the best risk detection accuracy for delayed trips.

---

10. Future Improvements

Future development could include:

- Real-time telemetry integration
- Deep learning using time-series data
- GPS-based trajectory modeling
- Cost forecasting using external conditions
- Interactive operational dashboards

---

11. References

1. Kaggle — Fleet Dataset
   https://www.kaggle.com/datasets/nhmishuk/fleet-dataset
2. Scikit-learn Documentation
   https://scikit-learn.org
3. TensorFlow Documentation
   https://www.tensorflow.org