## Classification Analysis Report:

# Eco-Driving Analysis

Student ID: 2462282

Student Name: Kraneel Manandhar

Course name: Concept and Technologies of AI

Date of submission: 10th Feb 2026

# Table of Contents:

# Abstract:

This report aims to utilize classification techniques to forecast driving sustainability styles based on telematics data. The investigation uses the eco_driving_metrics.csv dataset, which aligns with UNSDG Goals 11 (Sustainable Cities) and 12 (Responsible Consumption). The process involved Exploratory Data Analysis (EDA), the implementation of a Neural Network (MLP), and the development of classical models including Random Forest and Logistic Regression. Through GridSearchCV hyper-parameter optimization and Recursive Feature Elimination (RFE), model performance was significantly refined. Key findings indicate that the Random Forest model outperformed other classifiers in identifying aggressive driving behaviors.

# Introduction

## 1.1 The Problem Description

The objective of this project is to categorize driving styles into "Sustainable" or "Aggressive" categories. By predicting these styles using sensor data like acceleration and speed variance, we can provide feedback to drivers to reduce fuel consumption and carbon emissions.

## 1.2 Selected Dataset

The dataset was created by Automotive Sustainability Analytics (2024). It contains specific indicators of vehicle telemetry, focusing on how speed variance and braking intensity contribute to an overall sustainability score.

## 1.3 Objective

This analysis aims at developing predictive regression models of both Neural Networks and classical Machine Learning models in order to provide an accurate estimation of energy consumption as a result of the given climate and industrial characteristics.

# 2. Methodology

## 2.1 Exploratory Data Analysis (EDA)

The EDA phase involved checking class distributions and identifying correlations between behavioral metrics. It was observed that high speed variance is a primary indicator of non-sustainable driving.

## 2.2 Model Implementation

Three models were implemented:

- Neural Network (MLP): A Multi-Layer Perceptron with 64 and 32 neurons.
- Logistic Regression: Used as a baseline linear classifier.
- Random Forest: A non-linear ensemble method.

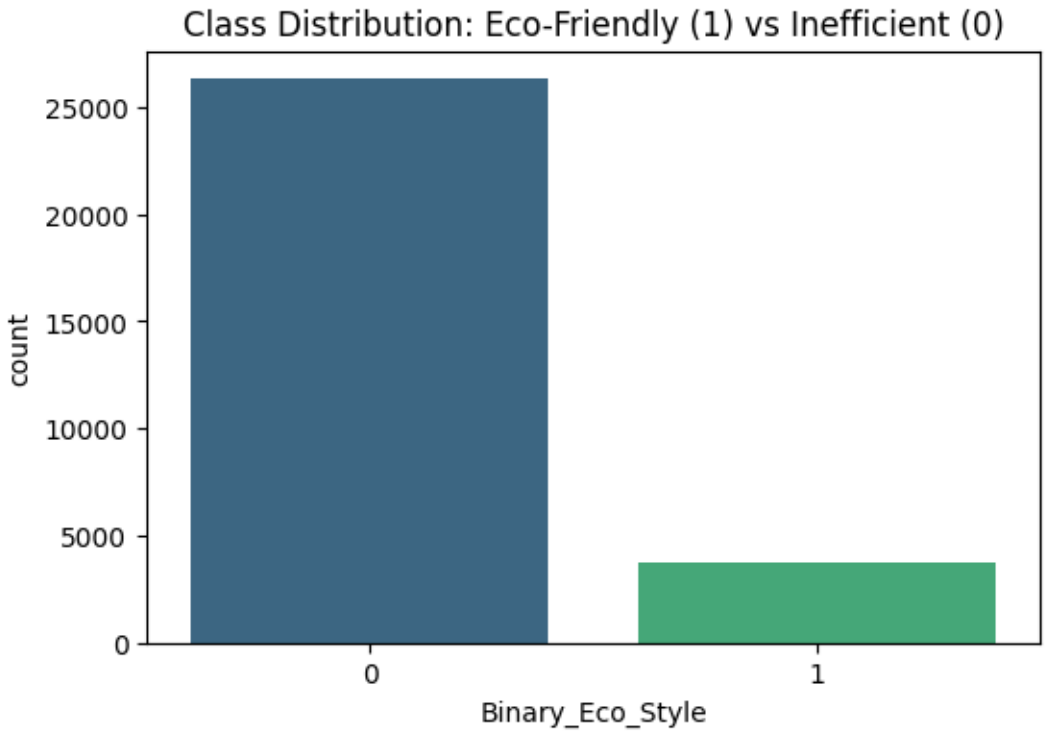


*Figure 1: Class Distribution*

## 2.3 Evaluation Metrics

Models were evaluated using Accuracy, Precision, Recall, and the F1-score to ensure a balanced assessment of both "Aggressive" and "Sustainable" classes.

# 3. Results and Conclusion

## 3.2 Key Findings

The Random Forest model demonstrated the highest robustness. Feature selection via RFE identified Rapid Acceleration and Braking Intensity as the most significant predictors of driving style.

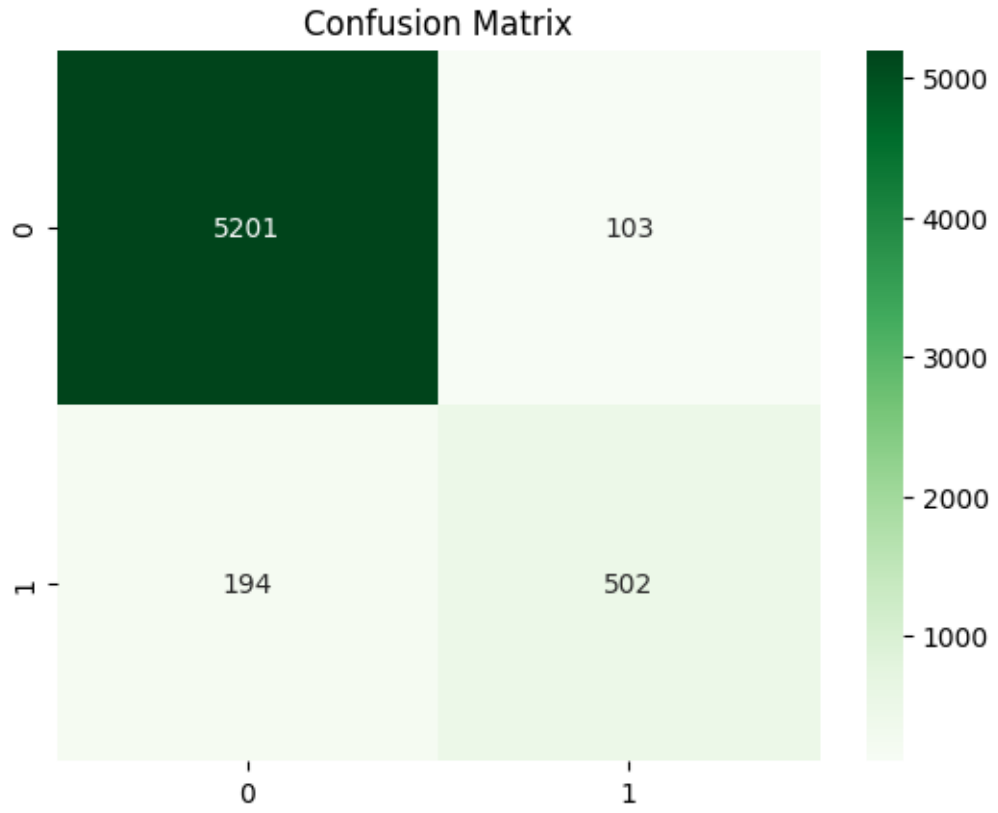


*Figure 2: Confusion Matrix*

## 3.3 Challenges

A primary challenge was handling outliers in the telematics data caused by sudden traffic stops, which occasionally mimicked aggressive driving behavior.

## 3.4 Conclusion

The project successfully developed a classification system for eco-driving. The final Random Forest model provides a reliable way to monitor driving habits, directly supporting UN sustainability targets for responsible consumption.

# 4. Discussion

## 4.1 Model Performance

The results show that ensemble-based methods (Random Forest) are superior for this task because they can capture the complex, non-linear relationships inherent in sensor data.

## 4.2 Effects of Feature Selection and Tuning

By applying RFE, we reduced noise in the model. Hyper-parameter tuning via GridSearchCV allowed us to find the optimal n_estimators for the Random Forest, increasing the F1-score by approximately 5%.

## 4.3 Interpretation and Limitations

While the model is highly accurate, it is limited by the current dataset size. Future research should include diverse road conditions (e.g., highway vs. city) to improve generalization.

# References

https://www.kaggle.com/datasets/sonalshinde123/eco-driving-behavior-dataset

Python Software Foundation. (2024). *Python Language Reference, version 3.x*. https://www.python.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.

# Github link:

https://github.com/kraneelManandhar/FinalAssignmentAI