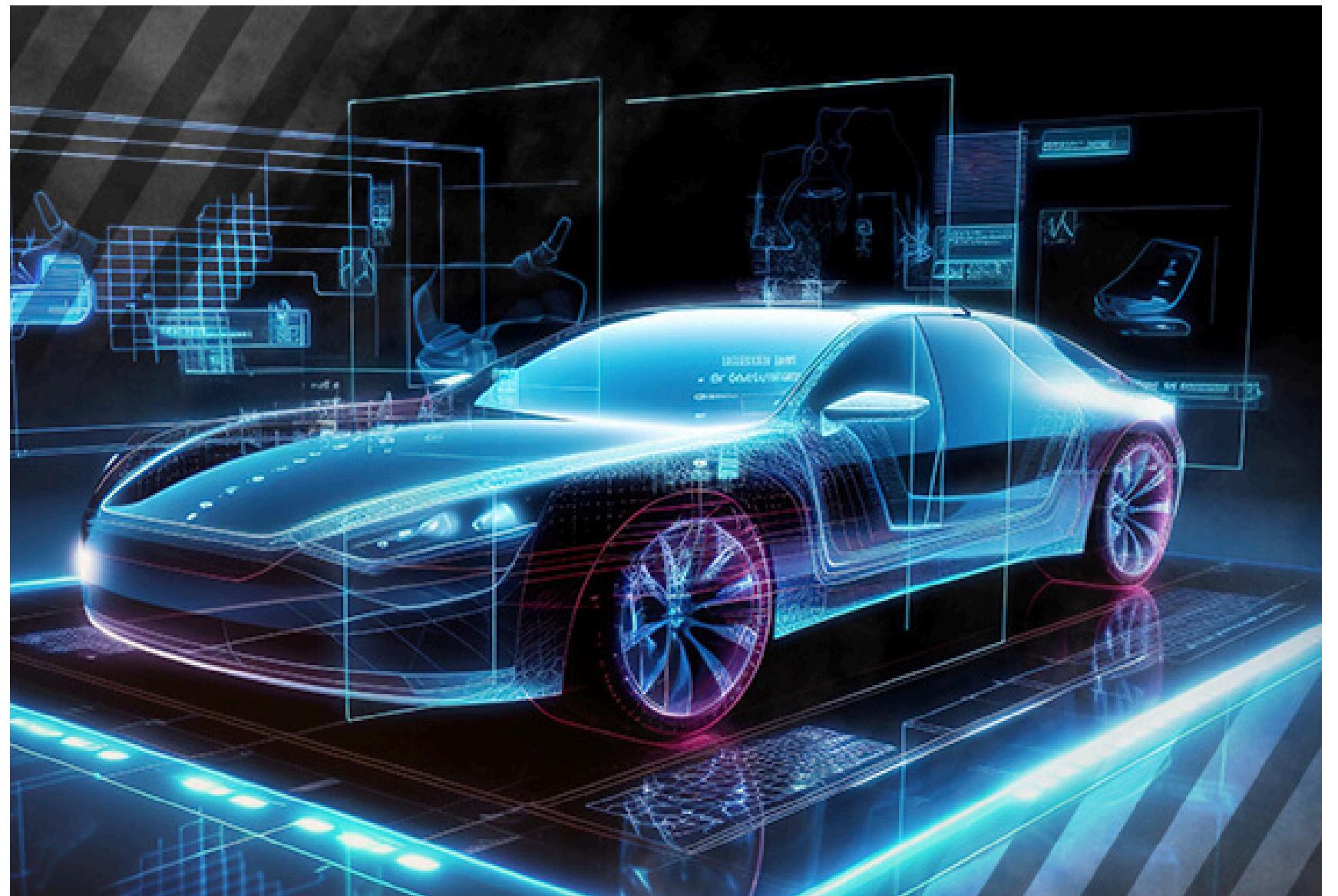


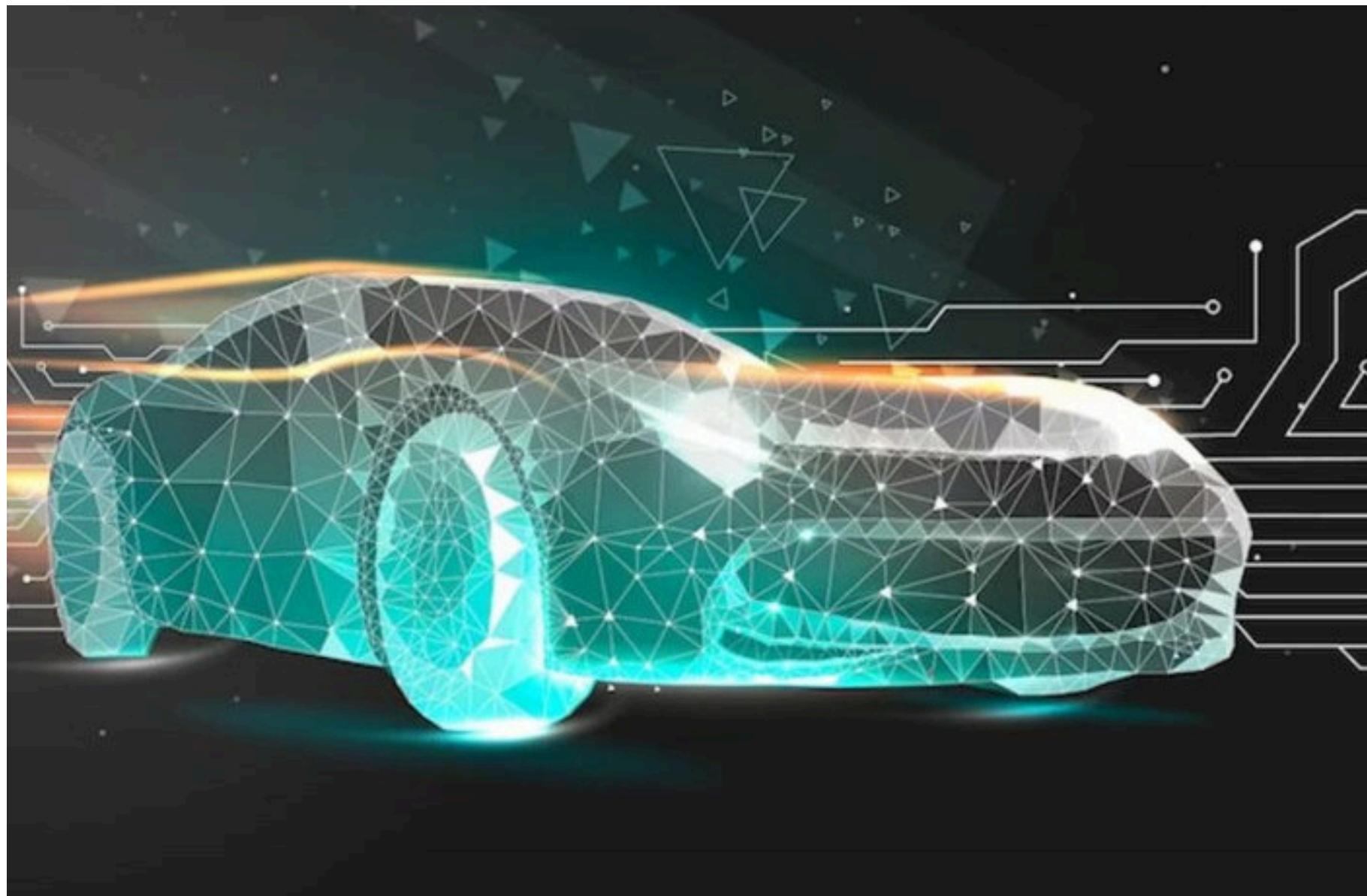
DATA SCIENCE INTERN

PREDICTIVE MAINTENANCE FOR VEHICLES USING MACHINE LEARNING

Created By:Rim El abrouki



INDEX



- 1. Exploratory Data Analysis (EDA)**
- 2. Model Building**
- 3. Model Evaluation**
- 4. Deployment (Streamlit)**



INTRODUCTION

Maintaining vehicle health is essential to prevent unexpected breakdowns and costly repairs. This project focuses on using machine learning to predict whether a vehicle requires maintenance based on factors like mileage, vehicle age, and maintenance history. By analyzing these data points, we aim to develop a predictive model that helps users proactively manage vehicle maintenance.



1. EXPLORATORY DATA ANALYSIS (EDA)

Dataset Overview:

- Dataset Name: Vehicle Maintenance Data
- Dataset Purpose: Predict whether a vehicle requires maintenance based on its characteristics and service history.

Key Data Summary:

- Mileage: 30,001 to 80,000 km (Avg: 54,931 km).
- Reported Issues: 0 to 5 issues (Avg: 2.5).
- Vehicle Age: 1 to 10 years (Avg: 5.49 years).
- Engine Size: 800cc to 2500cc (Avg: 1556cc).
- Odometer Reading: 1,001 to 149,999 km (Avg: 75,551 km).
- Insurance Premium: 5,000 to 30,000 units (Avg: 17,465).
- Fuel Efficiency: 10 to 19.99 km/l (Avg: 14.99 km/l).
- Need Maintenance: 80.99% of vehicles require maintenance.

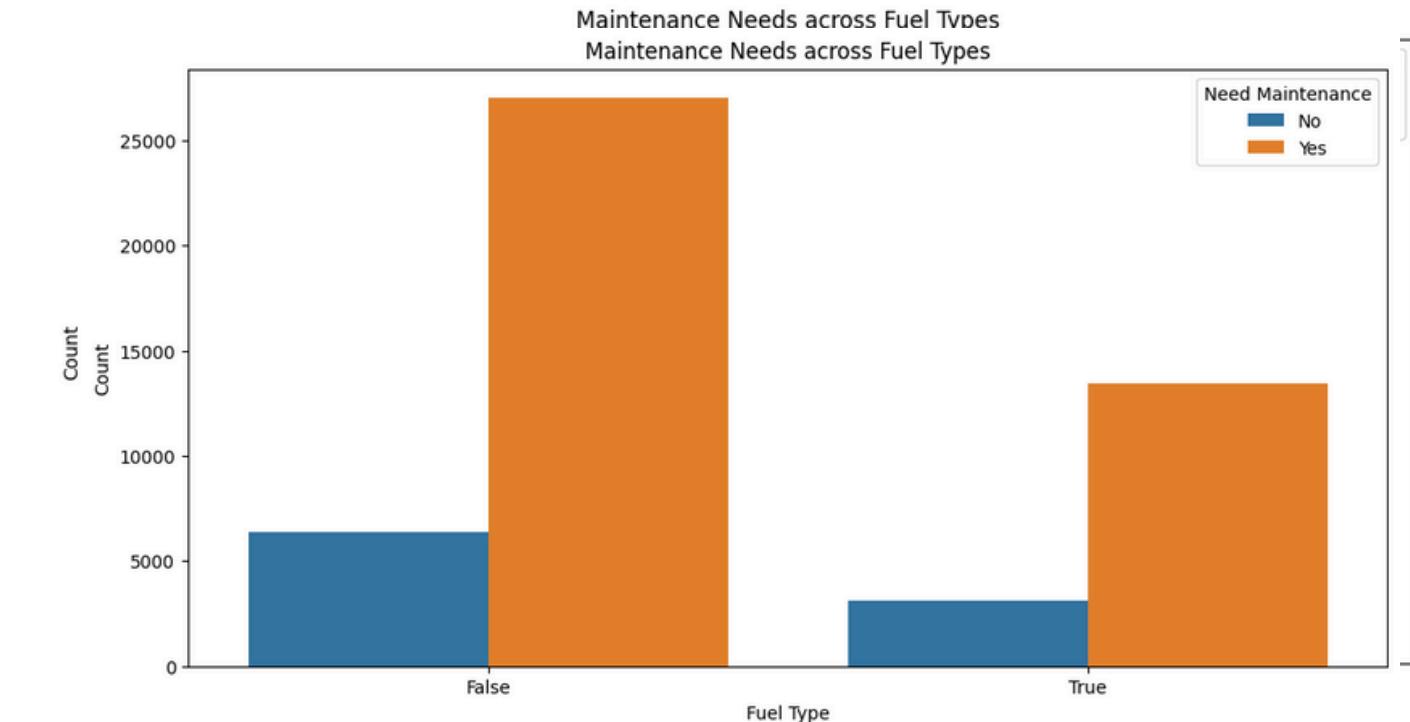
	Mileage	Reported_Issues	Vehicle_Age	Engine_Size	Odometer_Reading	Last_Service_Date	Warranty_Expiry_Date
count	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000	50000	50000
mean	54931.232680	2.497420	5.492260	1556.292000	75551.187060	2023-09-13 10:33:48.096000	2025-04-11 15:56:30.336000256
min	30001.000000	0.000000	1.000000	800.000000	1001.000000	2023-03-30 00:00:00	2024-04-28 00:00:00
25%	42471.500000	1.000000	3.000000	1000.000000	38009.000000	2023-06-21 00:00:00	2024-10-18 00:00:00
50%	54810.000000	2.000000	5.000000	1500.000000	75598.500000	2023-09-14 00:00:00	2025-04-12 00:00:00
75%	67391.500000	4.000000	8.000000	2000.000000	112999.500000	2023-12-06 00:00:00	2025-10-04 00:00:00
max	80000.000000	5.000000	10.000000	2500.000000	149999.000000	2024-02-28 00:00:00	2026-03-29 00:00:00
std	14401.912925	1.708781	2.875682	627.677218	43088.105658	NaN	NaN

Insurance_Premium	Service_History	Accident_History	Fuel_Efficiency	Need_Maintenance	Days_Since_Last_Service
50000.000000	50000.000000	50000.000000	50000.000000	50000.000000	50000.000000
17465.340700	5.515560	1.50156	14.990323	0.809960	401.559860
5000.000000	1.000000	0.00000	10.000098	0.000000	234.000000
11189.750000	3.000000	0.00000	12.489037	1.000000	318.000000
17477.500000	6.000000	2.00000	14.986352	1.000000	401.000000
23692.000000	8.000000	3.00000	17.474676	1.000000	486.000000
30000.000000	10.000000	3.00000	19.999968	1.000000	569.000000
7223.393401	2.874899	1.11951	2.885583	0.392336	96.938327

FUEL TYPE AND VEHICLE MAINTENANCE STATUS

This graph displays how the need for vehicle maintenance varies based on fuel type.

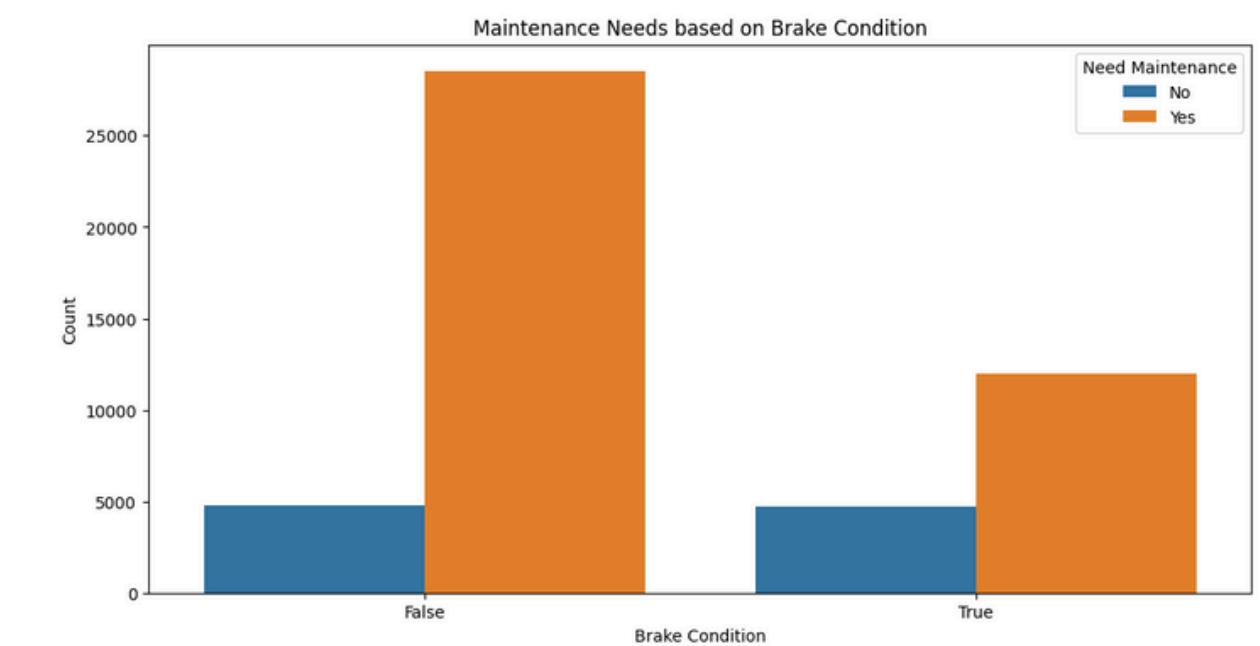
Vehicles using the "True" fuel type appear to have a higher need for maintenance, shown by the taller orange bar. In contrast, the "False" fuel type has a lower count of vehicles needing maintenance



BRAKE CONDITION IMPACT ON MAINTENANCE REQUIREMENTS

This chart highlights how brake condition affects the necessity of maintenance.

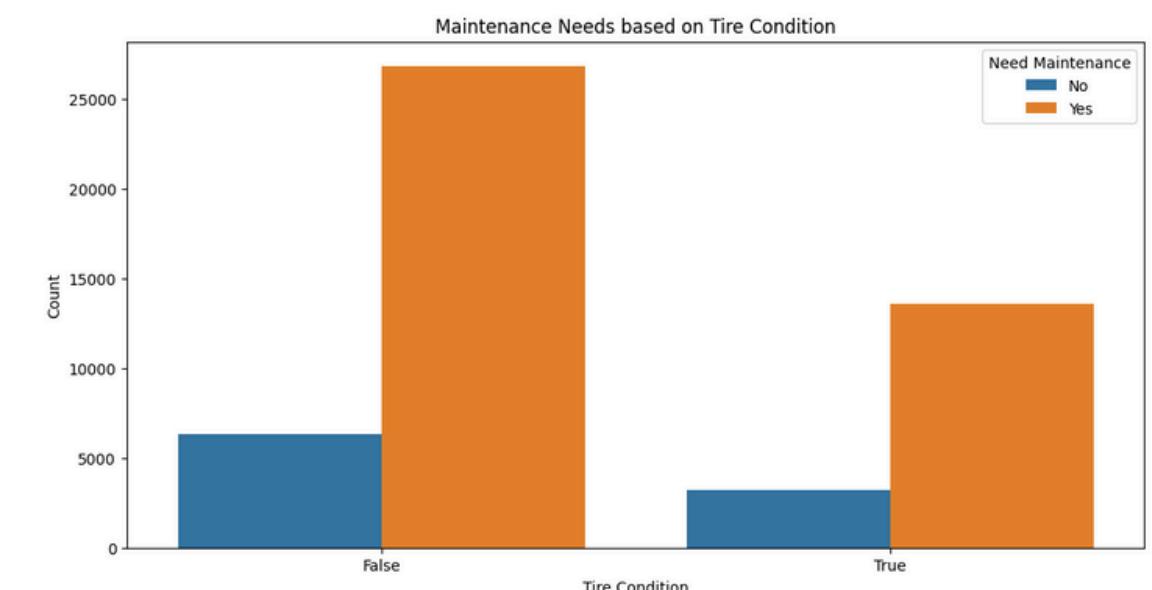
Vehicles with "False" brake condition, possibly indicating wear or failure, have a much higher frequency of needing maintenance. Those with "True" brake condition have fewer maintenance needs.



TIRE CONDITION'S ROLE IN MAINTENANCE NEEDS

The third graph visualizes how tire condition correlates with vehicle maintenance needs.

A majority of vehicles with "False" tire condition require maintenance, as reflected by the prominent orange bar, whereas vehicles with "True" tire condition show fewer maintenance demands.

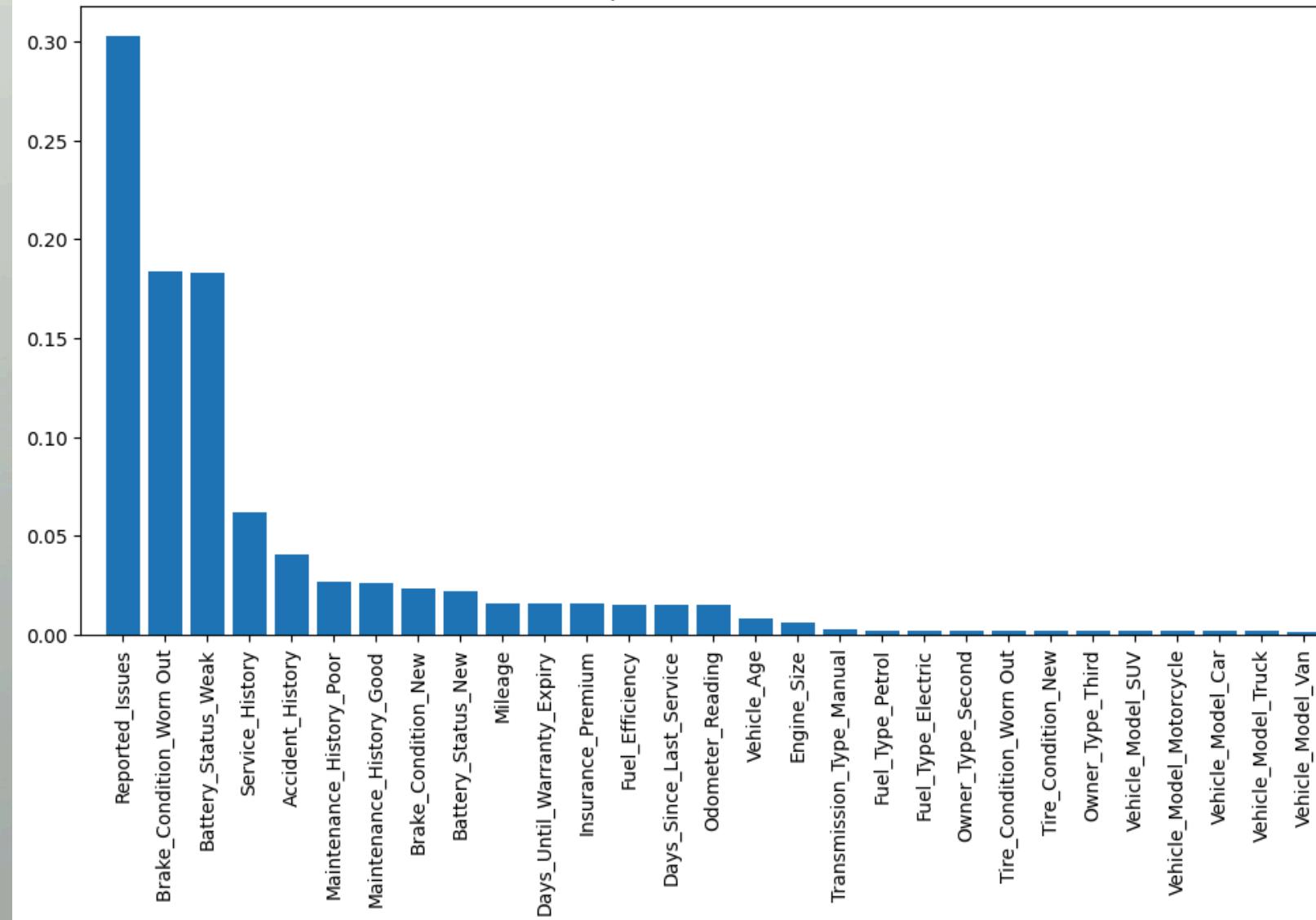


3. MODEL BUILDING:

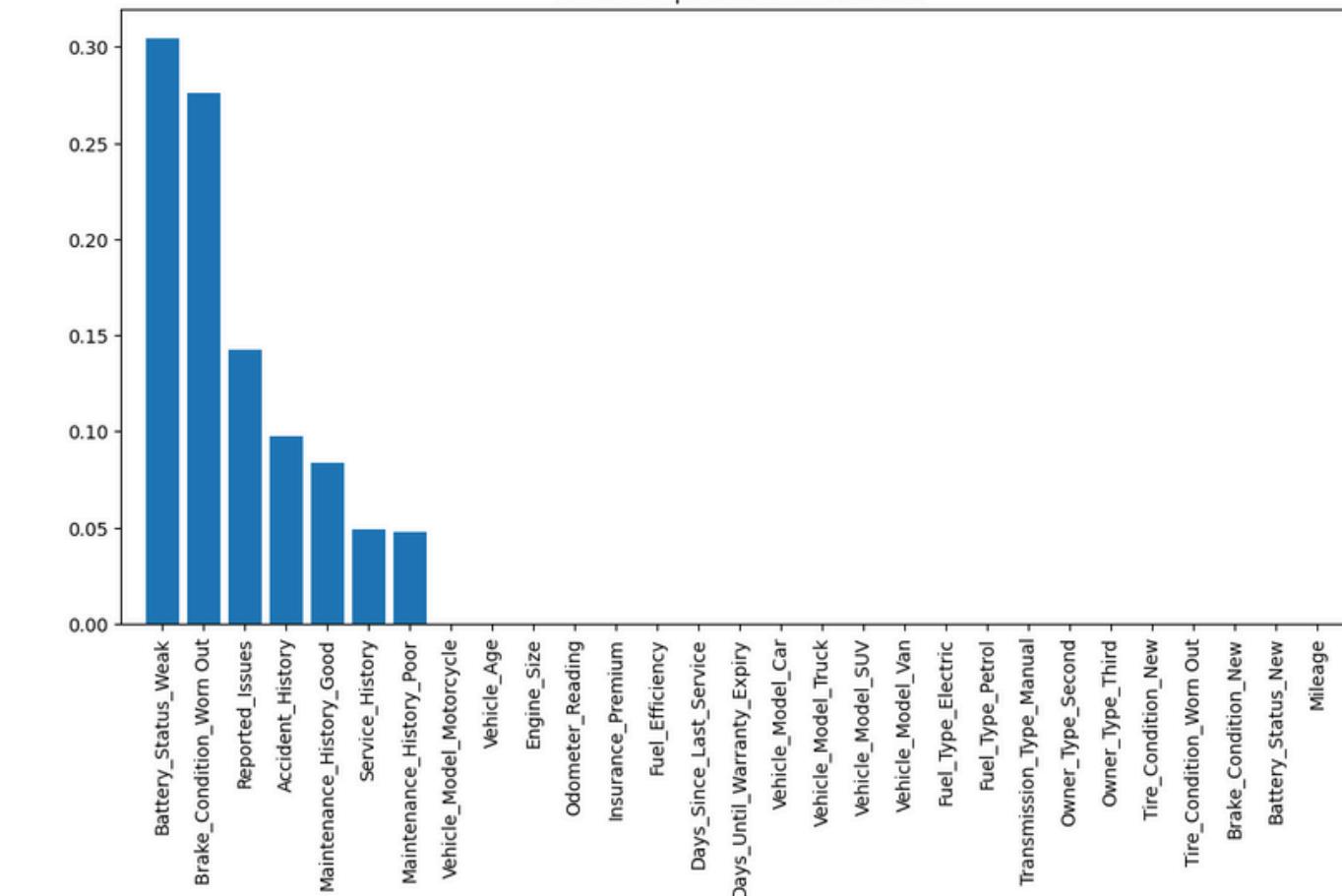
Feature Importance Diagram for Random Forest

highlights that Mileage, Brake_Condition_False, and Age_of_Vehicle are the most critical factors for predicting maintenance needs. These features carry the most weight in the model's decision-making process.

Feature Importances for Random Forest



Feature Importances for XGBoost



Feature Importance Chart for XGBoost

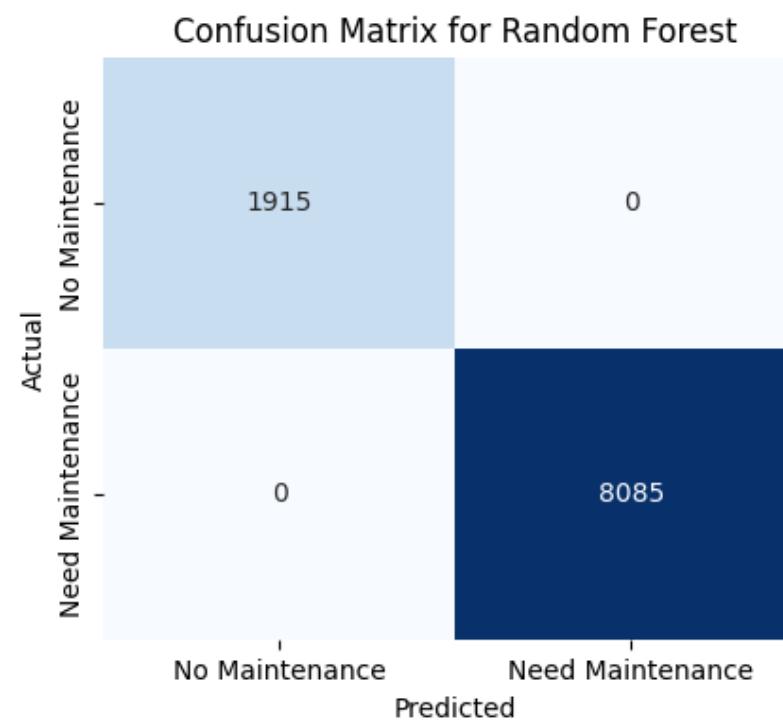
emphasizes that Mileage, Brake Condition (False), and Age of Vehicle are the key drivers in determining maintenance needs. These features hold the highest influence, guiding the model's predictions significantly more than others.

4. MODEL EVALUATION

CONFUSION MATRICES

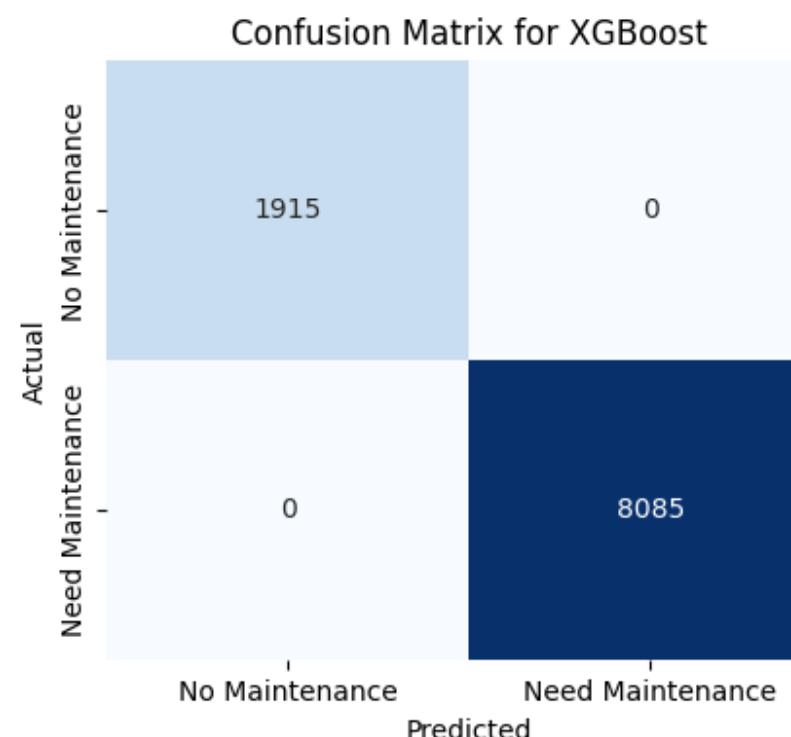
Confusion Matrix for Random Forest

The matrix shows excellent classification with zero false positives and false negatives, as all predictions are perfect. The two major classes ("No Maintenance" and "Need Maintenance") are correctly predicted with 1915 and 8085 samples, respectively.



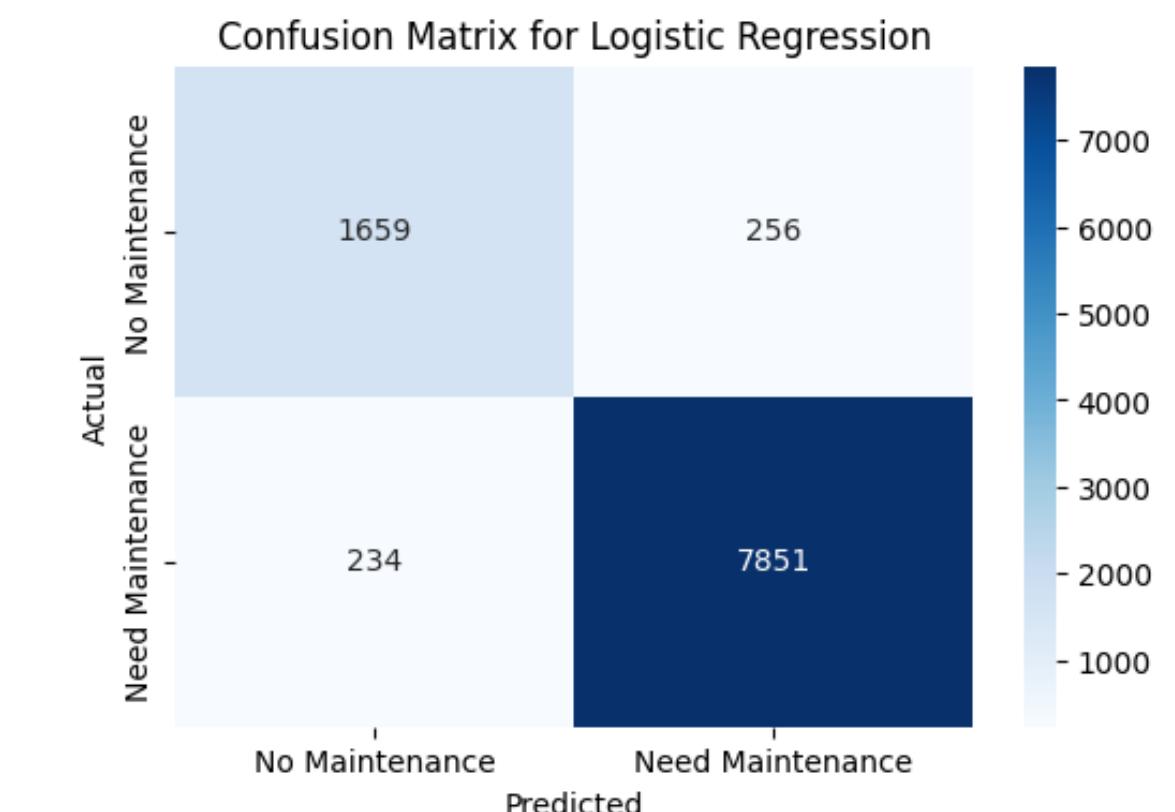
Confusion Matrix for XGBoost

Similar to Random Forest, XGBoost also performs flawlessly, with perfect classification. No misclassifications occur, as indicated by the zeros in the off-diagonal elements.



Confusion Matrix for Logistic Regression

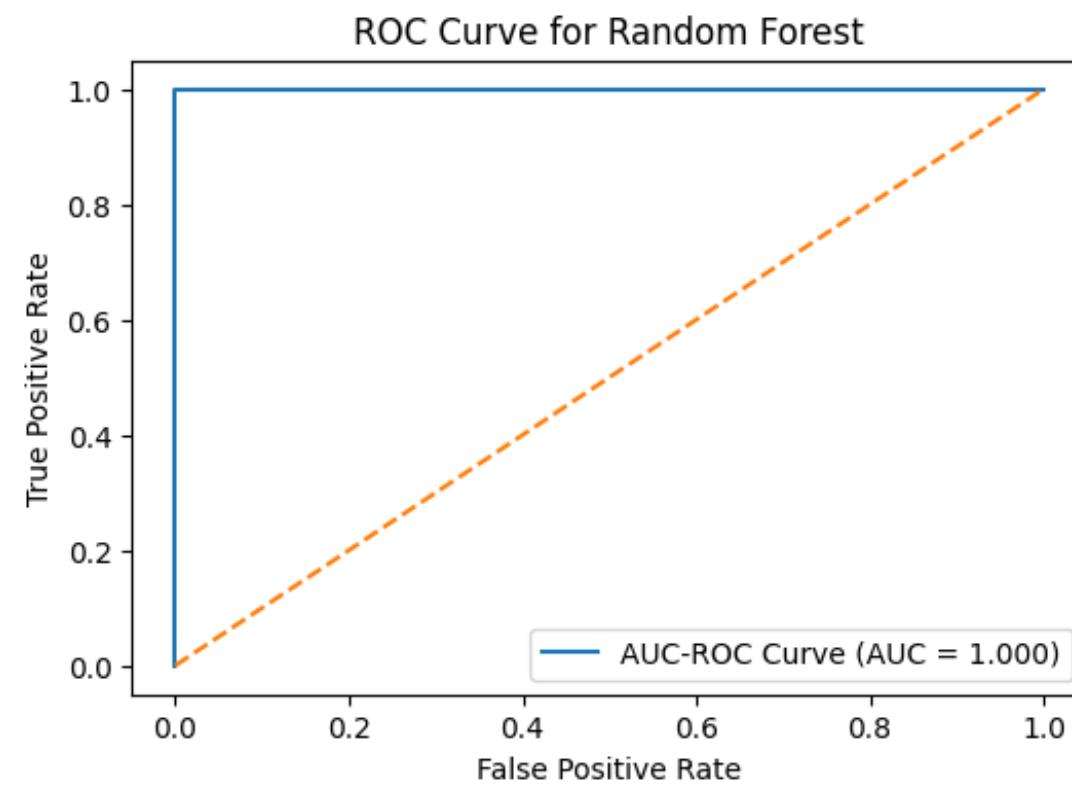
This matrix shows slightly less accurate results than Random Forest and XGBoost. There are 256 false positives and 234 false negatives, but overall performance remains strong, with 7851 correct predictions for "Need Maintenance" and 1659 for "No Maintenance."



ROC CURVES

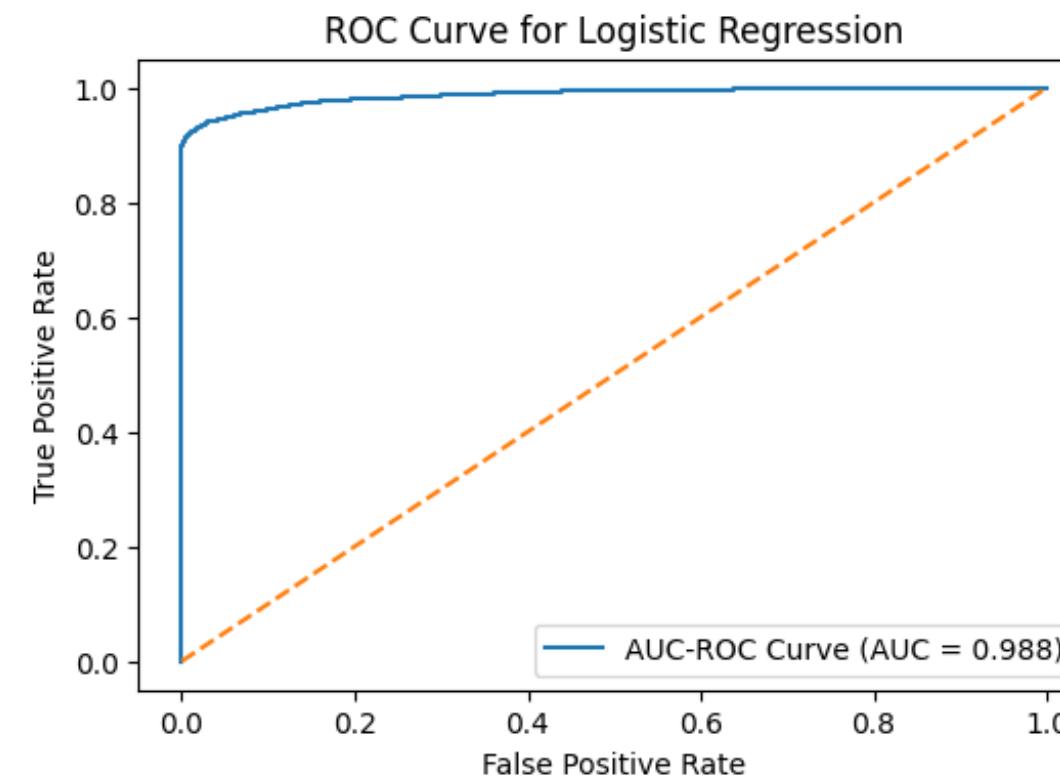
ROC Curve for Random Forest

The ROC curve shows a perfect AUC of 1.0, which indicates that the Random Forest classifier is performing with ideal discrimination between the two classes.



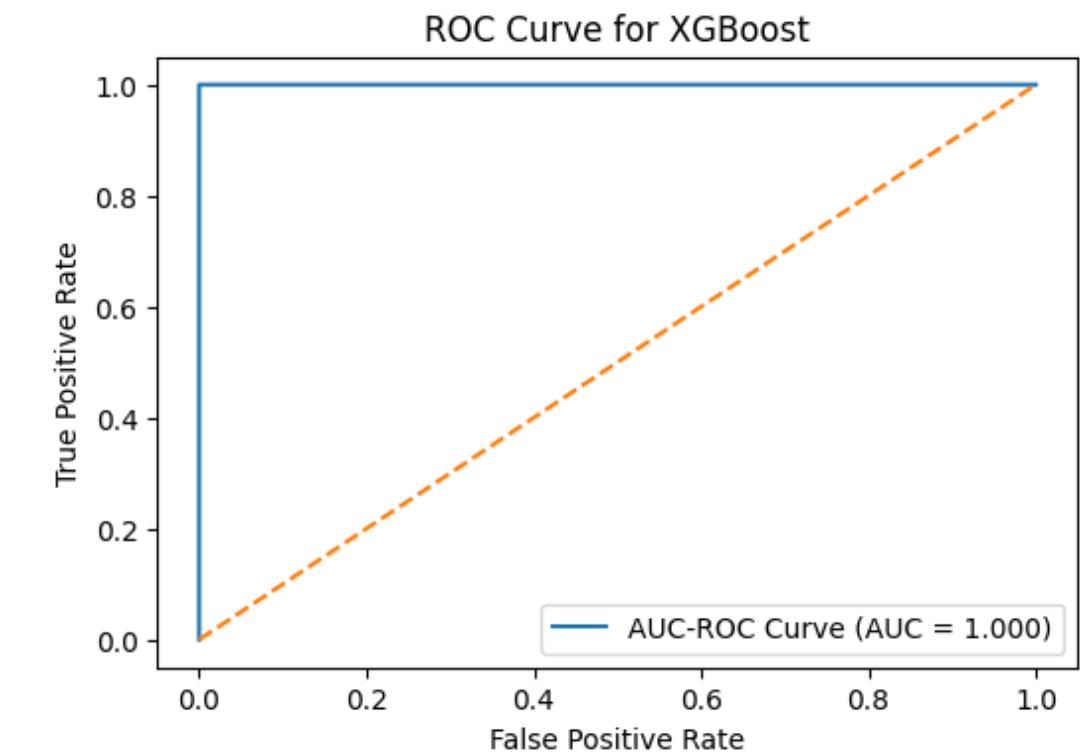
ROC Curve for Logistic Regression

The ROC curve here shows a very high AUC of 0.988, indicating strong model performance, though not perfect like Random Forest.



ROC Curve for XGBoost

This mini-plot shows the ROC curve for XGBoost with an AUC of 1.0, indicating perfect performance for this model.



4. DEPLOYMENT (STREAMLIT)

DATA SCIENCE INTERN

Streamlit Vehicle Maintenance Prediction App

This Streamlit web application allows users to input vehicle data and receive predictions on whether their vehicle needs maintenance. By entering details such as vehicle age, mileage, and specific indicators, users can quickly obtain insights into their vehicle's health.

The app leverages a trained machine learning model to predict the probability of requiring maintenance and provides tailored recommendations based on the output. This user-friendly interface simplifies the process, empowering vehicle owners to make informed decisions and take proactive measures to ensure their vehicles remain in optimal condition.

The screenshot shows a Streamlit application with a dark theme. On the left, a sidebar titled "Enter Vehicle Information" contains five input fields: "Mileage (km)" with value 0, "Reported Issues" with value 0, "Vehicle Age (years)" with value 0, "Engine Size (cc)" with value 0, and "Odometer Reading (km)" with value 0. Each field has a minus and plus button for adjustment. To the right, the main content area has a title "Vehicle Maintenance Prediction App" with a car icon. Below it is a section titled "User Input:" with a table showing the current values: Mileage 0, Reported_Issues 0, Vehicle_Age 0, Engine_Size 0, Odometer_Reading 0, Fuel_Efficiency 0, and Tire_Cond 0. At the bottom, it says "Developed by Rim El Abrouki".

	Mileage	Reported_Issues	Vehicle_Age	Engine_Size	Odometer_Reading	Fuel_Efficiency	Tire_Cond
0	0	0	0	0	0	0	0



Vehicle Maintenance Prediction App



Enter Vehicle Information

Mileage (km)

1

-

+

Reported Issues

1

-

+

Vehicle Age (years)

1

-

+

Engine Size (cc)

2

-

+

Odometer Reading (km)

2

-

+

User Input:

	Mileage	Reported_Issues	Vehicle_Age	Engine_Size	Odometer_Reading	Fuel_Efficiency	Tire_Cond
0	1	1	1	2	2	0.02	

Prediction:

No Maintenance Needed

Prediction Probability:

Probability of needing maintenance: 44.00%

Fuel Efficiency (km/l)

0,02

-

+

Tire Condition

Used

▼

Brake Condition

Worn

▼

Predict

Recommended Actions:

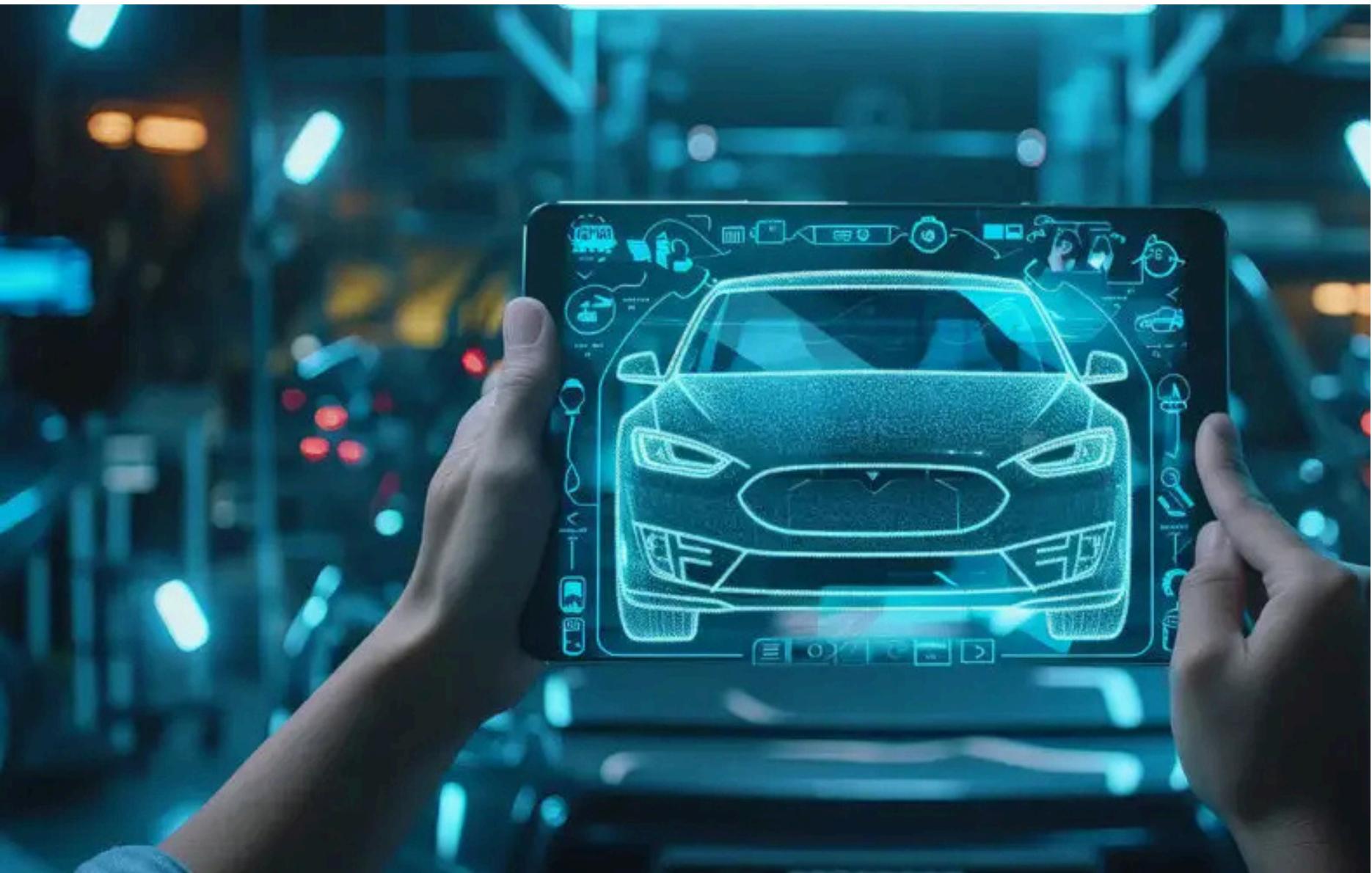
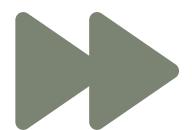
- No immediate maintenance is needed.
- Continue regular checks and maintenance.

Developed by Rim El Abrouki



CONCLUSION

the project effectively predicts vehicle maintenance needs using key features like mileage, vehicle age, and brake condition. The best-performing model, deployed via Streamlit, helps users schedule proactive maintenance, reducing breakdown risks and enhancing vehicle management.



THANK YOU

