King Saud University
Software Engineering Department
SWE 485 – Selected Topics in Software Engineering

Car Maintenance Tracking system

Sara Alshuwaier
444200957
Reema Almunasser
444201088
Sadeem Alsayari
444201182
Leen Alohali
444200882
Noof Alkhalifa
444200886

1ts Semester | Fall 2025-2026

Problem Statement

Lack of maintenance reminders and tracking causes delayed servicing, leading to:

- Unexpected car breakdowns
- Increased repair costs
- Reduced vehicle lifespan

Objectives

Simplify and automate vehicle maintenance tracking

Predict maintenance needs in advance

Enable proactive rather than reactive maintenance

Use data-driven insights to analyze:

- Maintenance history
- Vehicle condition
- Environmental factors

جامعة
الملك سعود
King Saud University

Business Value

User Value Proposition
Improves user experience through automated tracking and proactive maintenance reminders.

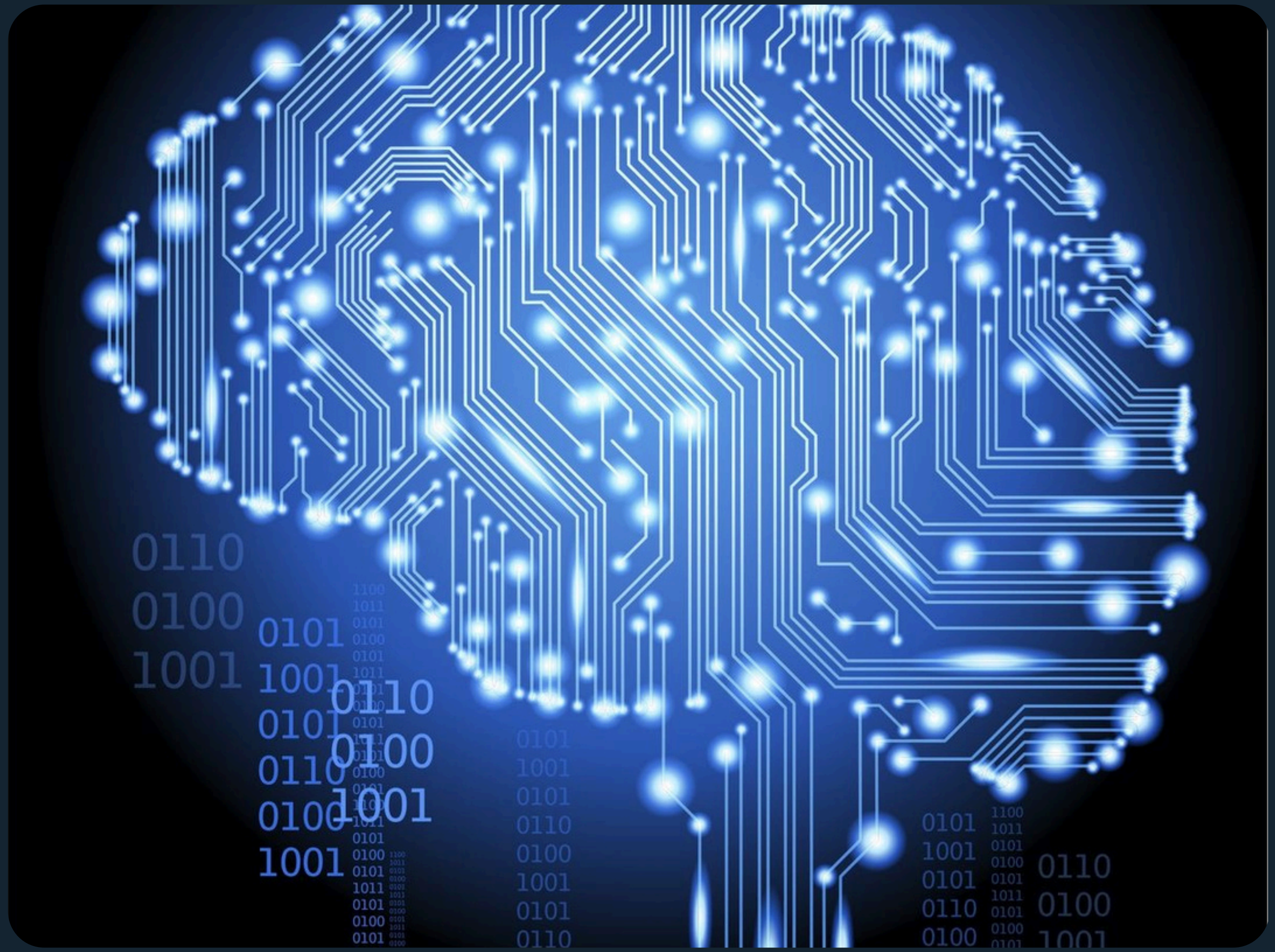
Business & ROI Potential
Supports cost savings, better resource planning, and long-term value through optimized maintenance scheduling.

Operational Impact
Reduces unexpected breakdowns by enabling early identification of maintenance needs through predictive insights.

Cost Efficiency
Helps lower maintenance and repair costs by preventing major failures and reducing emergency repairs.

Asset Value
Extends vehicle lifespan by supporting timely and data-driven maintenance decisions.

Data Source	Key Features	Collection Methods
• Dataset obtained from Kaggle	• Maintenance frequency	• Historical maintenance logs
• Logistics Vehicle Maintenance History Dataset	• Type of maintenance performed	• Recorded service and repair events
• Represents historical records of vehicle maintenance activities	• Vehicle condition indicators	• Time-stamped vehicle records
	• Time-based features (dates, intervals)	• Collected from operational logistics data



Data Overview

Process/Approach

Data Cleaning
Handle missing values, remove duplicates, normalize data

EDA (Exploratory Data Analysis)
Visualize distributions, statistical summaries, and class balance

Supervised Learning
Train models to predict advice using Gradient Boosting Decision Trees (GBDT) algorithm and Random Forest (RF)

Unsupervised Learning
Apply clustering to find patterns and improve recommendations

Generative AI (GAI)
integrate TinyLlama and Microsoft Phi-2 models to generate detailed advice

King Saud University
جامعة الملك سعود

Tools Used

python
The main programming language used for data analysis, machine learning modeling, and Generative AI integration.

Pandas & NumPy
Python libraries used for data cleaning, manipulation, and numerical computations

Matplotlib & Seaborn
visualization python libraries used for creating charts, graphs, and statistical plots

Scikit-learn
Machine learning python libraries Used to build and evaluate supervised & unsupervised models

Transformers
Python library used to access and integrate Generative AI models (TinyLlama and Microsoft Phi-2 models)

Google Colab & GitHub
Platforms used for Version control & collaboration

💡 Data Exploration Findings

Main Preprocessing Steps:

- ⚙ Cleaned & standardized text/categorical features (labels unified, rare categories → “other”)
- ⚙ Scaled main numeric features to [0, 1] and kept meaningful outliers
- ⚙ Extracted time features from Last_Maintenance_Date (service year/month/day, days since last service, recent service flag).
- ⚙ Engineered Load_Utilization = Actual_Load / Load_Capacity

Key Results:

- ⚙ Final cleaned file

💡 Supervised Learning Findings

Models Tested:

- Gradient Boosting (GBDT)
- Random Forest

Key Results:

- Both models achieved near perfect performance on all metrics
- GBDT selected as the final predictive model due to flawless classification and strong generalization

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
0	RandomForest	0.9996	0.9996	0.9996	0.9996	1.0
1	GBDT	1.0000	1.0000	1.0000	1.0000	1.0

💡Unsupervised Learning Findings

Method Used:

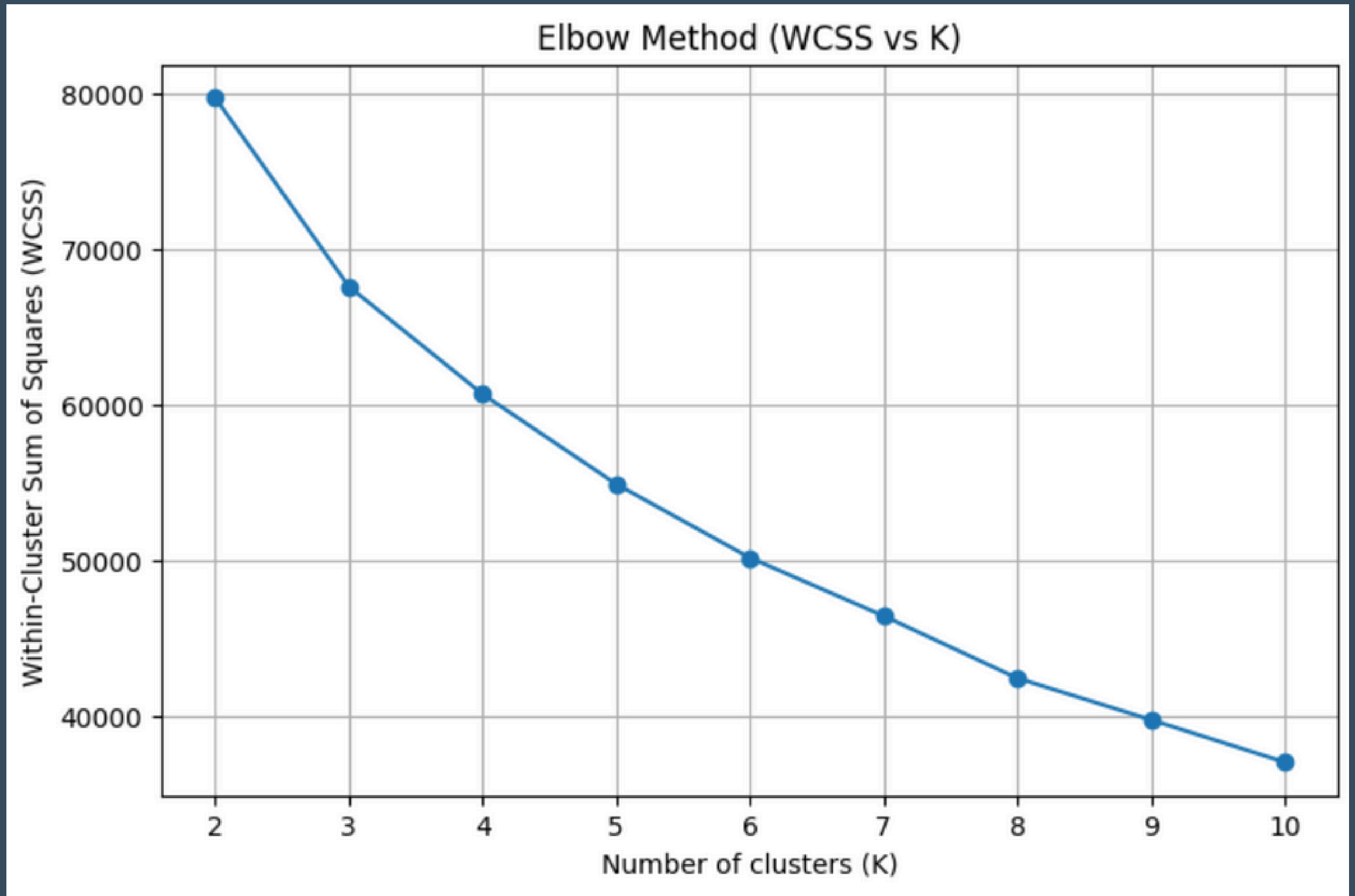
- K-Means Clustering

Key Results:

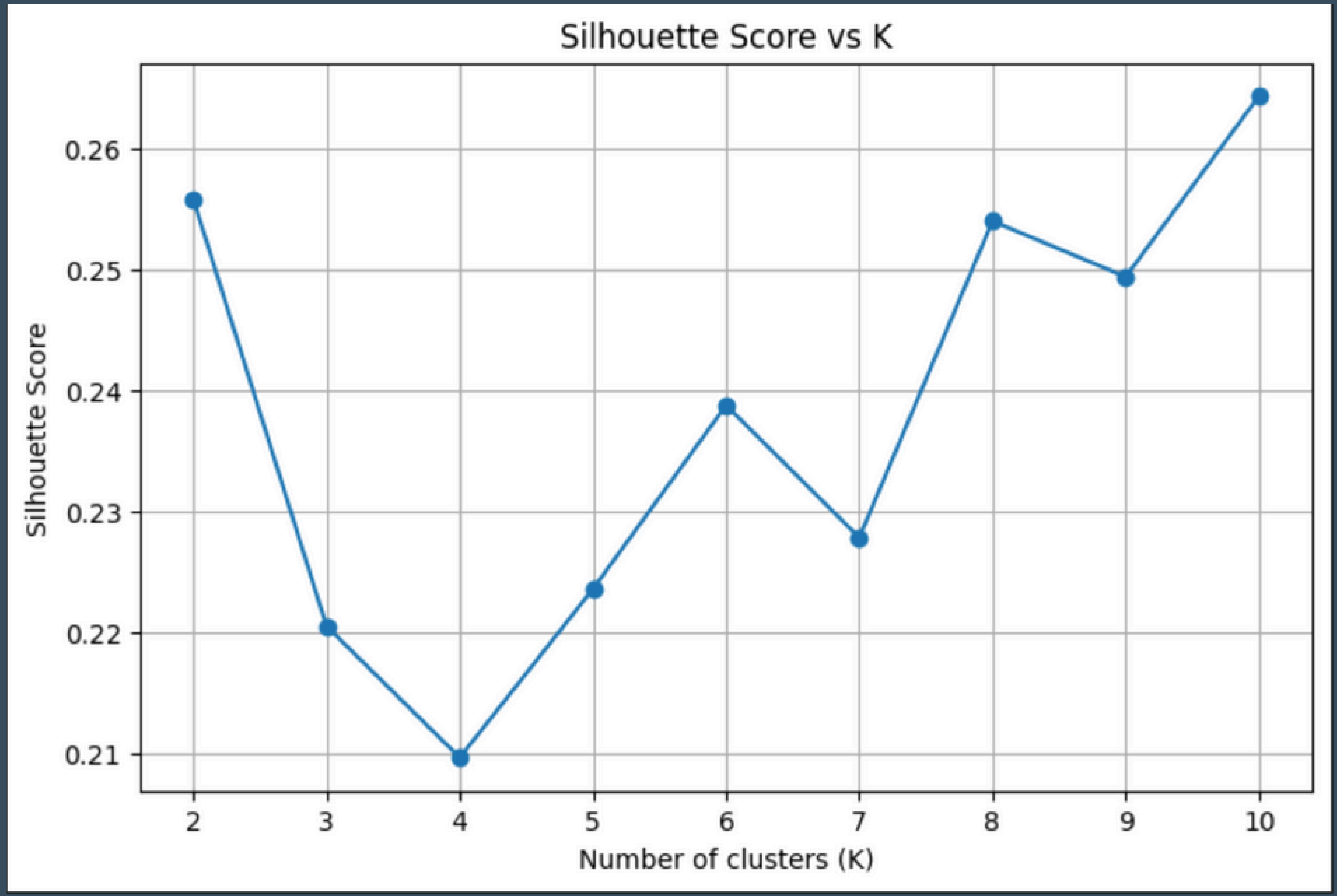
- K-Means was applied after removing the class label to explore natural group patterns in the fleet.
- K = 10 was selected based on the Elbow and Silhouette evaluation trends.
- The resulting clusters reflect different operational and maintenance usage behaviors across vehicles.

Key Results:

Elbow Method



Silhouette Score VS K



King Saud University جامعة الملك سعود

💡 Generative AI Integration Findings

Models Used: TinyLlama-1.1B - Microsoft Phi-2

Templates:

Template A – Concise Advice
Generates short and quick maintenance advice in 1–2 lines

Template B – Detailed Advice
Provides structured, expert-level guidance with multiple sections

Template Hybrid
Balanced Prompt:

Template	Strengths	Weaknesses	Metrics (Quality / Detail / Relevance)
Template A (Concise)	- Clear and brief - Easy to scan quickly	- Lacks depth - Missing actionable steps	5 / 0 / 5
Template B (Detailed)	- Well-structured sections (Immediate, Preventive, Risk, Next Service) - Rich in detail	- Verbose - May include redundant checks	5 / 5 / 5
Final Prompt (Hybrid)	- Balanced: structured yet concise - Practical and deployment-ready - Actionable advice	- Slightly less depth than Template B	5 / 5 / 5

Conclusion & Recommendations

• Key Insights

- Data consistency and preprocessing proved essential: cleaning, standardizing, and handling missing values significantly improved model reliability and performance.
- Integration of supervised and unsupervised methods provided complementary perspectives: predictive models delivered accuracy, while clustering uncovered hidden operational patterns that validated and enriched the analysis.

• Actionable Recommendations

- Deploy the GBDT model as the core predictive engine, with continuous monitoring of performance.
- Use clustering results to categorize vehicles into risk groups and tailor maintenance schedules accordingly.
- Integrate explainability tools (e.g., SHAP, cluster reports) into dashboards to build trust among technicians and managers.

Conclusion & Recommendations

- **Future Work**

Optimize predictive thresholds by balancing breakdown costs against maintenance costs.

Conduct field experiments to measure the real-world impact of the model on reducing breakdowns and operational costs

Enrich the dataset with additional indicators such as driving style and environmental conditions.

Lessons Learned

Signal strength matters more than feature quantity; a few strong indicators proved more impactful than dozens of weaker ones.

Addressing class imbalance is critical to ensure fairness and prevent bias toward one category.

Data consistency is essential for the success of any model; cleaning and standardization directly improve results.