EV BATTERY DEGRADATION PREDICTION

Using Machine Learning

Vaishak VT
Techolas Technologies Calicut

INTRODUCTION

IMPORTANCE OF EV BATTERY HEALTH

Battery health plays a crucial role in determining the performance, range, and lifespan of electric vehicles.

CHALLENGES IN MANUAL ESTIMATION

Battery degradation cannot be accurately estimated manually as it depends on multiple factors like usage, temperature, and charging behavior.

ROLE OF MACHINE LEARNING

Machine learning models can analyze complex patterns in EV usage data and accurately predict battery degradation.

PROJECT GOAL

The goal of this project is to predict EV battery degradation percentage using usage and environmental factors to help improve battery management and planning.

DATASET OVERVIEW

KEY FEATURES

Vehicle Information

These features describe the vehicle's physical and age-related characteristics.

- Vehicle Age (Years)
- Battery Capacity (kWh)

Usage & Charging Patterns

These features represent how the vehicle is used daily, similar to lifestyle habits in humans.

- Daily Usage (km)
- Total Distance Driven (km)
- Charge Cycles
- Fast Charging Percentage (%)

Environmental Conditions

These features represent external conditions affecting battery health.

- Average Operating Temperature (°C)

DATASET SOURCE

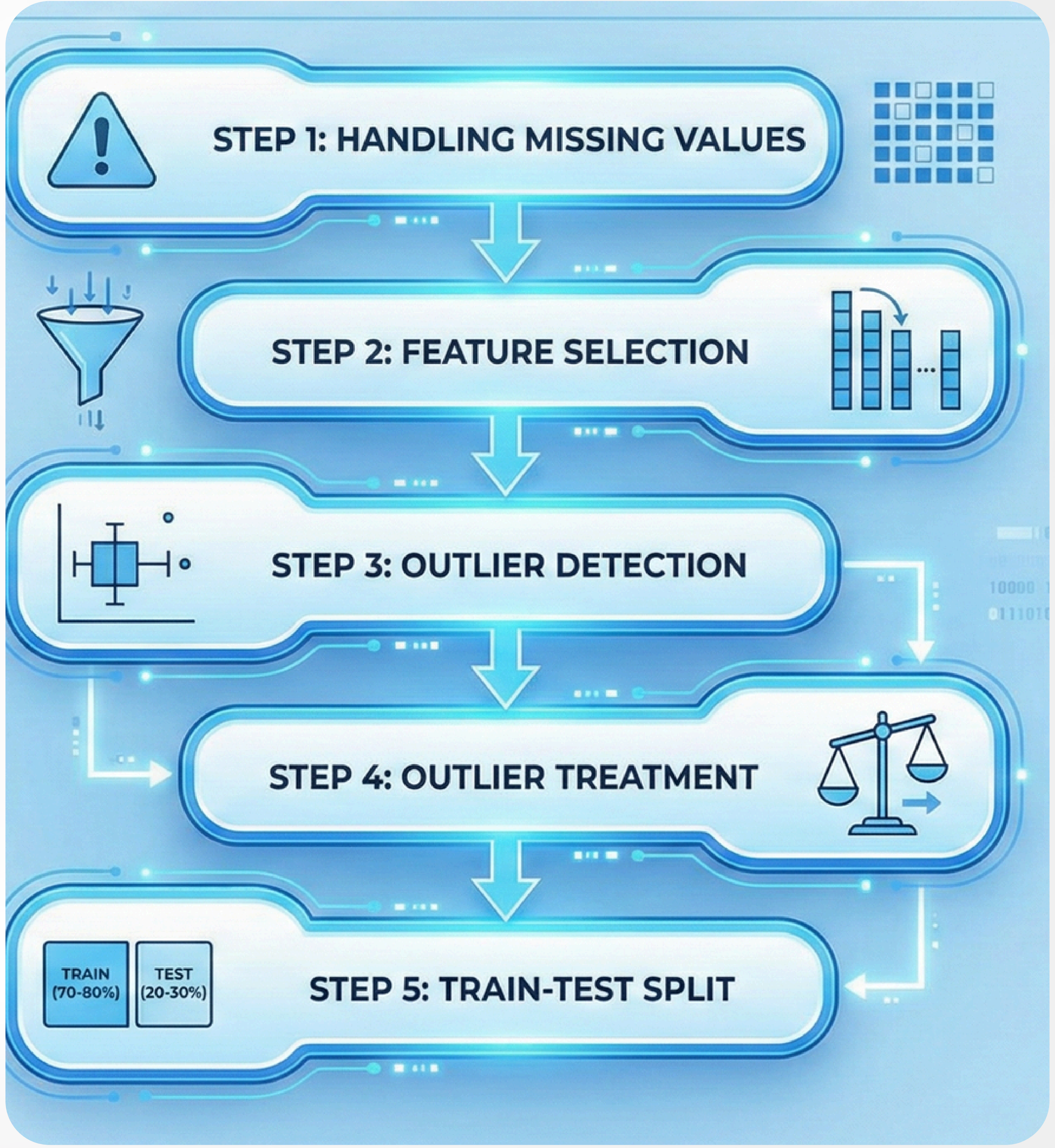
- The dataset is synthetically generated but logically designed to simulate real-world EV usage scenarios.
- Feature relationships are based on known EV battery aging principles.

TARGET VARIABLE

- Battery Degradation (%) – Represents the percentage of battery health loss over time.

DATA PREPROCESSING

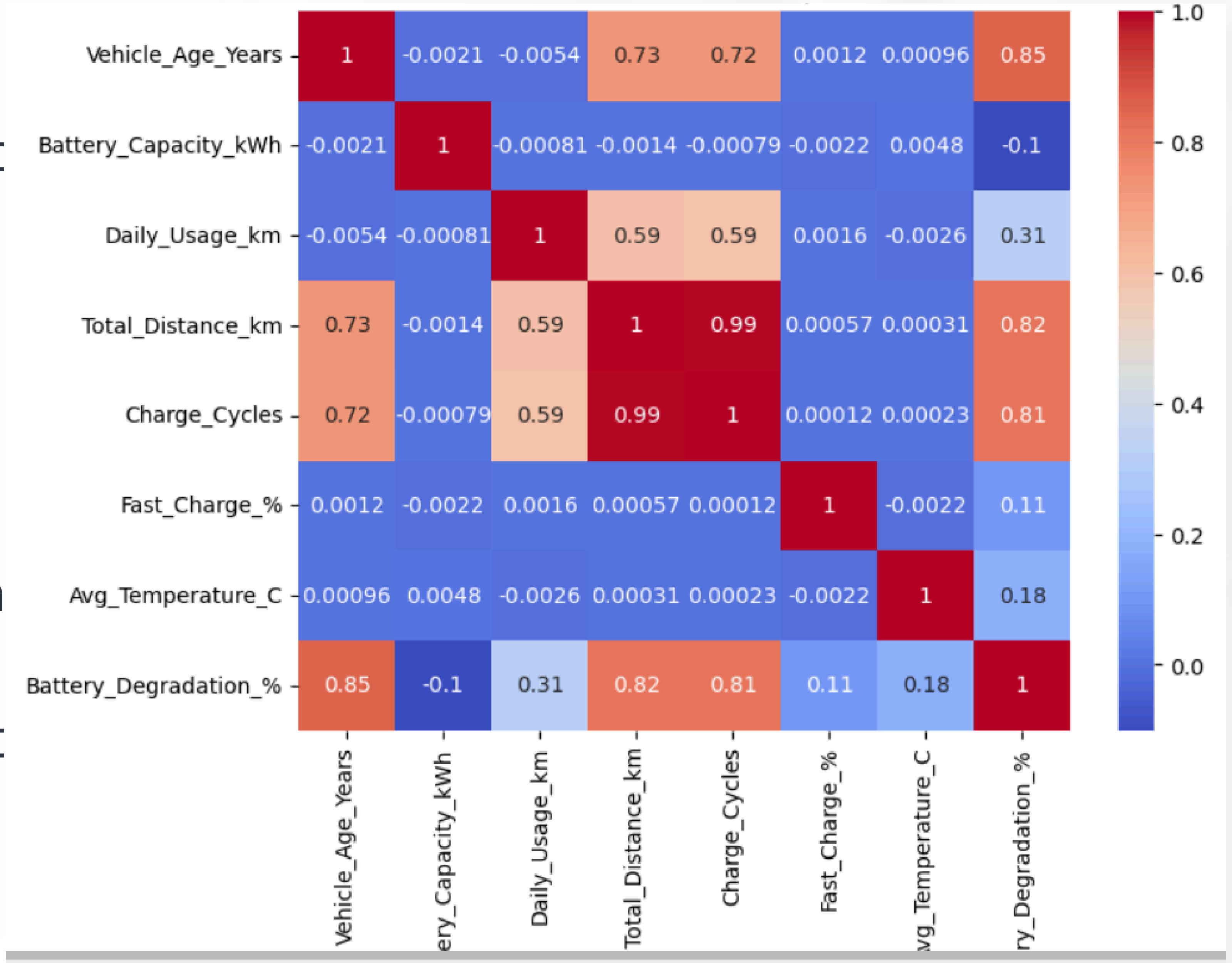
1. Handling Missing Values
   - Checked and handled missing values to ensure data completeness.
2. Feature Selection
   - Relevant features related to EV usage and battery stress were selected.
3. Outlier Detection
   - Boxplots were used to identify extreme values in the dataset.
4. Outlier Treatment
   - Outliers were detected using boxplots and treated using the IQR method to improve model robustness and reduce noise.
5. Train-Test Split
   - Dataset was split into 80% training data and 20% testing data.



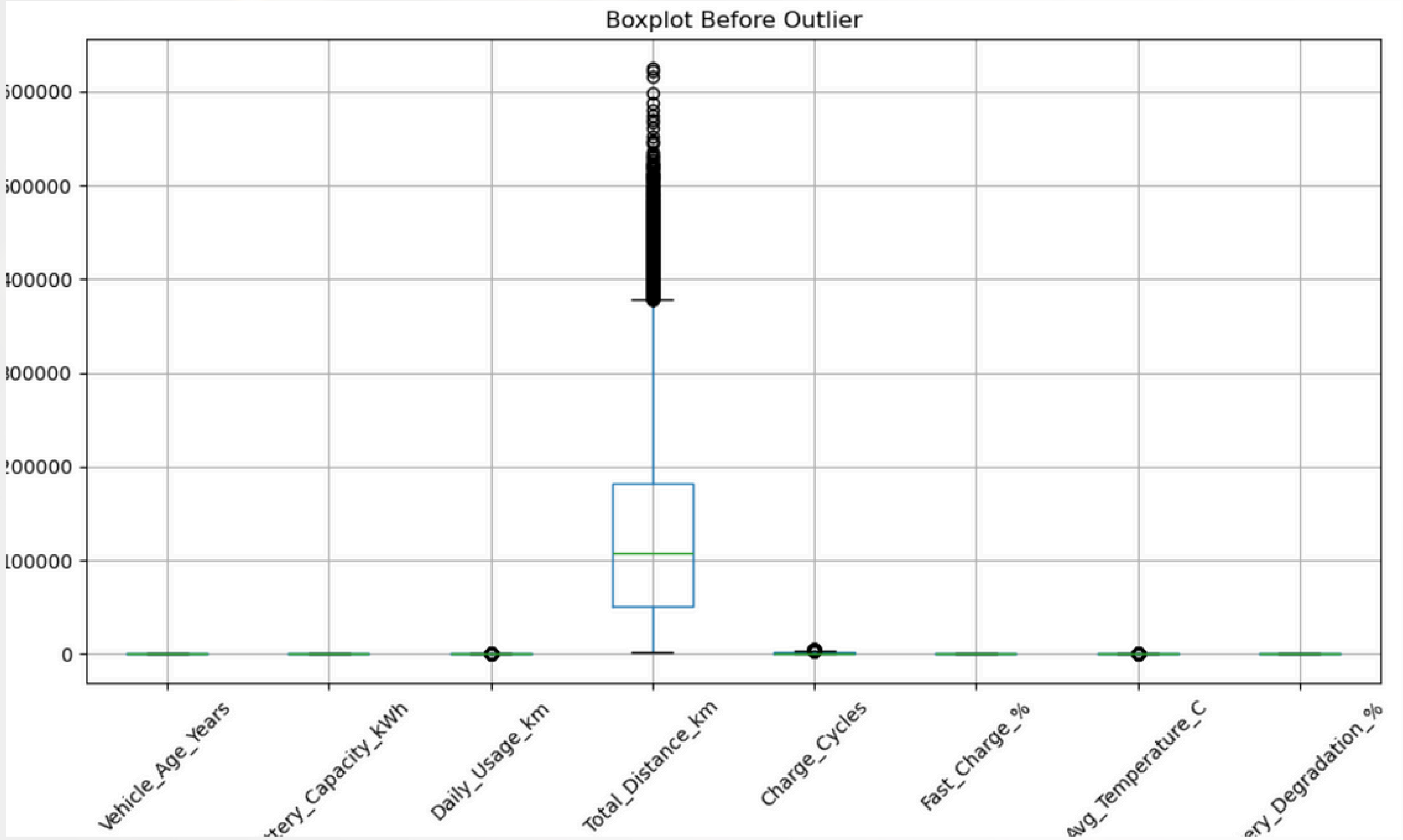
VISUALIZATIONS & INSIGHTS

CORELATION HEAT MAP

- Used to analyze the relationship between input features and battery degradation.
- Strong positive correlation observed with:
  - Vehicle Age
  - Total Distance Driven
  - Charge Cycles
- Battery Capacity shows a weak negative correlation with degradation.
- Correlation analysis helped identify the most influential features before model training.

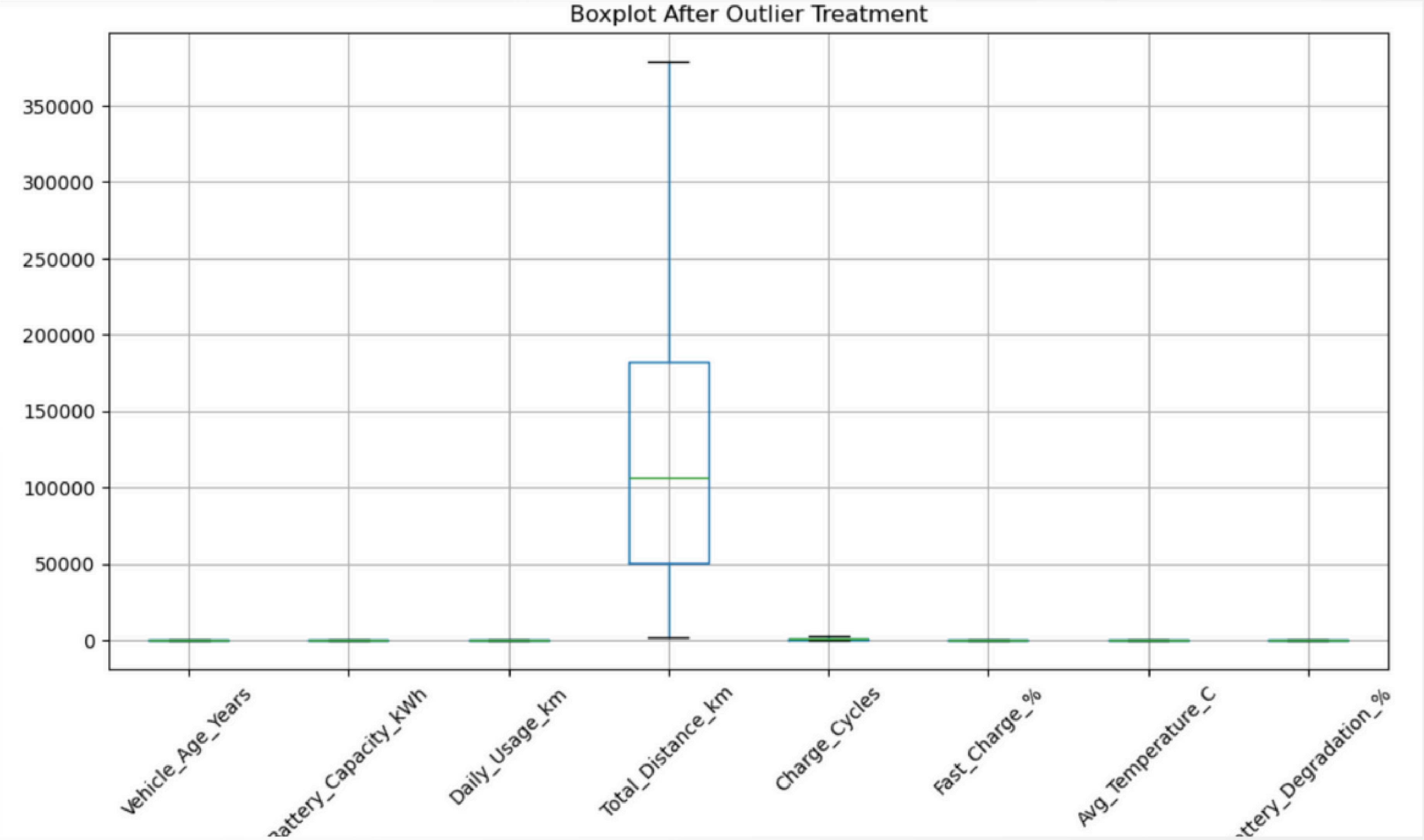


BOXPLOT – BEFORE OUTLIER TREATMENT



- Identified extreme values in:
  Total Distance
  Charge Cycles
  Battery Degradation
- These outliers represent heavy EV usage patterns.

BOXPLOT – AFTER OUTLIER TREATMENT



- Outliers were treated using the IQR method.
- Data distribution became more stable and realistic.
- Important usage patterns were retained without removing valid data.

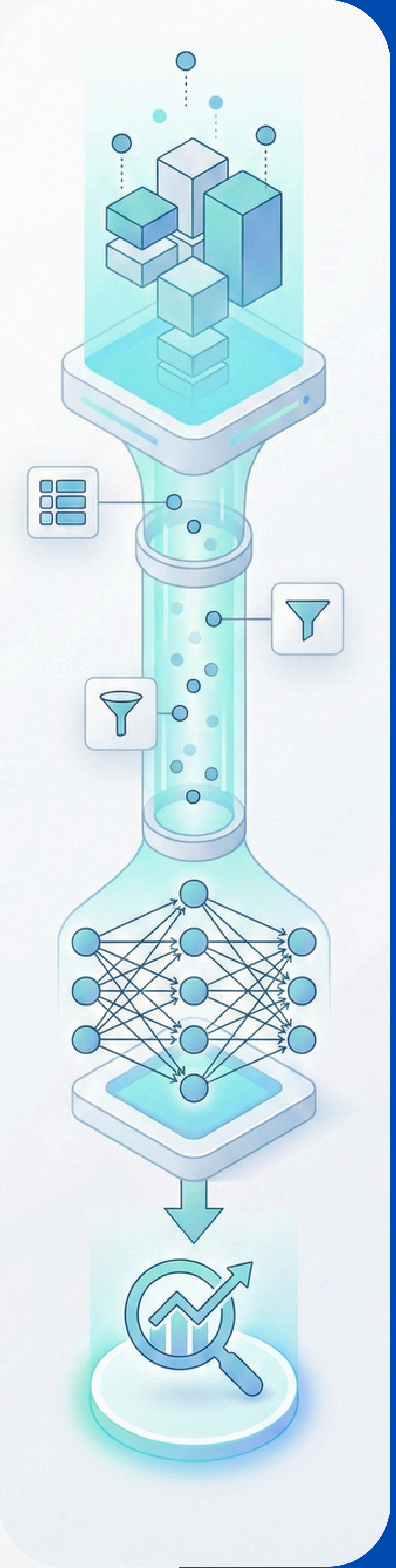
MODEL **SELECTION**

1. Multiple Regression Models Tested

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor
- KNN Regressor
- AdaBoost Regressor
- Gradient Boosting Regressor

2.Train–Test Split

- Used to evaluate initial model performance on unseen data.
- Dataset split into 80% training and 20% testing.
- Helps evaluate model performance and avoid overfitting.



3.K-Fold Cross Validation

- 5-Fold Cross Validation was applied to ensure model stability and reduce dependency on a single data split.

4. Hyperparameter Tuning

- GridSearchCV was used to optimize Random Forest parameters such as number of trees and tree depth to prevent overfitting.

5.Best Performing Model

- Random Forest Regressor achieved the highest and most consistent R^2 score across both test data and cross-validation.

MODEL RESULTS & PERFORMANCE EVALUATION

EVALUATION METRICS USED

- R^2 Score – Measures how well the model explains variance in battery degradation
- MSE (Mean Squared Error) – Measures average squared prediction error
- RMSE (Root Mean Squared Error) – Indicates prediction stability and penalizes large errors

	Model	R2 Score
2	Random Forest	0.985775
5	Gradient Boosting	0.982644
1	Decision Tree	0.968117
4	AdaBoost	0.933380
0	Linear Regression	0.885688
3	KNN	0.875885

K-FOLD CROSS VALIDATION

- 5-Fold Cross Validation applied
- Ensures model stability and avoids bias due to single train-test split
- Performance evaluated across multiple data partitions

BEST MODEL SELECTION

- Random Forest Regressor
- Achieved:
  - Highest Mean R^2 Score
  - Lowest RMSE
- Selected as the final model for prediction and deployment

PREDICTION EXAMPLE &

MODEL DEPLOYMENT

Prediction Example

Input Values Provided to the Model:

- Vehicle Age: 4.3 years
- Battery Capacity: 30 kWh
- Daily Usage: 60 km
- Total Distance Driven: 82,000 km
- Charge Cycles: 720
- Fast Charging Usage: 30%
- Average Temperature: 32°C

Predicted Output

Predicted Battery Degradation:

- ≈ 28.5%

This indicates moderate battery wear due to frequent charging and usage.

Model Deployment

- The trained Random Forest model was saved using Joblib.
- A Streamlit web application was developed for deployment.
- Users can input EV usage details and get real-time battery degradation predictions.

The deployed model enables users to estimate EV battery health and make informed decisions regarding battery maintenance and replacement.

CONCLUSION & FUTURE SCOPE

Conclusion

- The project successfully predicts EV battery degradation using machine learning.
- Exploratory Data Analysis identified key factors affecting battery health.
- Outlier treatment improved data quality and model stability.
- Multiple regression models were evaluated.
- Random Forest Regressor performed best with high R^2 score and low error.
- The final model was deployed using Streamlit for real-time prediction.

Key Takeaways

- Battery degradation increases with:
  - Vehicle age
  - Charge cycles
  - Frequent fast charging
  - High operating temperature
- Ensemble models outperform simple regression models for this problem.

Future Scope

- Use real-world EV data instead of synthetic data.
- Apply advanced hyperparameter tuning techniques.
- Extend the model to time-series battery degradation prediction.
- Deploy as a mobile or cloud-based application.
- Integrate with EV battery management systems.


Thank you