Capstone Project

Diagnostic Scantool Communication Analysis Final Report

1. **Define the Problem Statement:** The project addresses communication issues faced by the ALLDATA automotive diagnostic scan tool, specifically instances when it cannot successfully connect to certain vehicles. The primary goals are to clearly identify these connection problems, classify the causes, and recommend solutions to enhance diagnostic tool reliability and improve customer support efficiency.
2. **Model Outcomes or Predictions:** The project utilizes a classification approach, which is a type of supervised machine learning. The expected outcome from the model is to predict whether an attempt to connect the diagnostic tool to a specific vehicle will be successful or unsuccessful based on vehicle attributes and operational details.
3. **Data Acquisition:** Telemetry data was acquired from the ALLDATA mobile diagnostics application, stored in Azure cloud storage in JSON format. Using KQL (Kusto Query Language), this raw telemetry data was converted into a structured, tabular format suitable for analysis. The data covered details of vehicle connectivity attempts, vehicle attributes, shop details, and the results of communication attempts.

**Data Preprocessing/Preparation:** The following steps ensured the data was cleaned and ready for analysis:

* Unnecessary features (such as billing information) were removed.
* Missing or invalid battery voltage data was managed (missing voltages replaced with zero, invalid values removed).
* Vehicle makes were standardized by combining multi-word entries.
* Invalid or demo mode scans were removed.
* Vehicle system names were normalized into consistent categories.
* Data was split into training (70%) and testing sets (30%) to validate model performance effectively.
* Feature engineering was used to simplify the 'vehicle model' and 'system' fields, focusing on frequently occurring items to reduce complexity.

1. **Modeling:** Baseline logistic regression models were developed to predict connection success or failure based on attributes like vehicle make, model year, system category, and scanning method. Battery voltage was also evaluated as a potential influencing factor. However, analysis indicated battery voltage had a minimal impact on the results.
2. **Model Evaluation:** Models were evaluated using several metrics including Accuracy, Recall (Sensitivity), Precision, Specificity, and F1 Score. All baseline models demonstrated approximately 89% accuracy, with excellent recall (approximately 99%) indicating a strong ability to identify successful connections. However, specificity (correctly identifying failed connections) was low, indicating the need for further model refinement to improve this aspect.
3. **Significance of the Project:** This project provides actionable insights, identifying which vehicles and conditions lead to communication failures. This information allows suppliers to prioritize improvements on actual tool-related issues rather than investigating all failures. Moreover, customer support teams can quickly discern real hardware issues and vehicle-specific problems, significantly enhancing customer experience and increasing overall efficiency.