**CIDM 6355 Project Proposal**- **Used Car Price Prediction**

**Team Name:** Team E

**Team** **Number:** 5

**Project Propos[[1]](#endnote-2)al:** Used Car Price Prediction

**Team Members:**

|  |  |
| --- | --- |
| 1. Joshua Tarrant | 2. Farzaneh Noroozi |
| 3. Pradnya Bhere | 4. Joshua John |
| 5. Brian Doss | 6. Mallika Alan |
| 7. Trevor Hofmann |  |

**Introduction**

The used car market is a complex and ever-changing industry where pricing plays a pivotal role. Accurate price predictions are not only beneficial for sellers but also essential for buyers looking to make informed decisions. Considering this, our project aims to employ data mining techniques to predict the prices of used cars. This proposal outlines the problem statement, data source, and project plan for this endeavor.

**Problem Statement**

The primary goal of this project is to predict the price range of used cars based on various attributes. We aim to provide a solution that empowers both buyers and sellers in the used car market. Buyers will benefit from understanding how different car attributes affect prices, helping them make well-informed purchase decisions. Sellers, on the other hand, will be able to set competitive and realistic prices, ultimately increasing their chances of a successful sale.

**Data Description**

Our dataset is sourced from Kaggle and consists of detailed information on used automobiles. It comprises 4009 records and 12 attributes, including categorical variables such as Brand, Model, Model Year, Mileage, Engine Type, Transmission, Exterior, Interior Colors, Accident History (Yes/No), and Clean Title (Yes/No) The target attribute, "Price range prediction," will serve as the basis for our prediction model.

**Data Source**

The dataset can be accessed and downloaded from Kaggle using the following link: [Used Car Price Prediction Dataset](https://www.kaggle.com/datasets/taeefnajib/used-car-price-prediction-dataset?resource=download). Kaggle is a reputable platform for data science and machine learning datasets, ensuring data quality and reliability.

**Project Plan**

Our project plan encompasses the following key steps:

1. **Data Preprocessing:** To ensure the quality and reliability of our dataset, we will begin with a thorough data preprocessing phase. This will include handling missing values, detecting, and addressing outliers, and ensuring data consistency.
2. **Feature Selection:** We will identify the attributes that have the most significant impact on used car prices. Feature selection is critical for building an accurate prediction model. Understanding which attributes influence price the most will provide valuable insights.
3. **Model Building:** We will employ classification techniques to construct a robust prediction model. The selected features will serve as input variables, and the target attribute will be "Price range prediction."
4. **Model Evaluation:** The performance of our prediction model will be rigorously evaluated using appropriate metrics, such as accuracy, precision, recall, and F1-score. This evaluation will help us determine the model's effectiveness in predicting used car prices.
5. **Interpretation:** We will analyze the results to gain insights into the attributes that have the most significant influence on used car prices. These insights will be presented in an understandable format, making them accessible to both buyers and sellers.

**Conclusion**

In conclusion, our project endeavors to provide a data-driven solution to the challenge of predicting used car prices. By leveraging data mining techniques, we aim to empower individuals involved in the used car market with valuable insights. Our solution will facilitate better decision-making, improve pricing strategies, and contribute to a more transparent and efficient marketplace.

We anticipate that the results of this project will not only have practical applications but also serve as a valuable reference for future research in the automotive industry.

1. [↑](#endnote-ref-2)