Title: "Analysing and Predicting Prices of Used Vehicles"

# Introduction:

The proposed capstone project aims to address the critical issue of understanding and predicting car pricing in the used vehicles market. The used car market is a big part of the car industry, and it affects both individuals as well as businesses. “There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features.” (Gokce, 2020).

This project is important because it deals with real-world problems like car price inflation, how some people have more information than others, and the need for accurate pricing. This project will help potential car buyers, sellers and dealers make informed decisions, ultimately contributing to a more transparent and efficient marketplace.

# Objectives:

Development of a Predictive Pricing Model:

Create an advanced pricing model that considers various factors influencing used car prices, Manufacture, Model, Production year, Category, Fuel type, Mileage, and Engine size.

Data collection and Analysis:

Collect and analyze extensive historical data to train and validate the pricing model, ensuring its accuracy and reliability. (Simplilearn, 2021)

Market insights:

Studying and understanding how the used car market works. To understand why people, prefer certain cars, why prices change in different places, and how factors like the economy or the time of year affect car prices. By doing this, we can make sure our pricing model is smart and can adapt to all these changes in the market.

Compare the predicted price against the available price.

To be able to make a better informed decision by comparing the available price of a car vs the predicted price of a similar model to decide whether the buyer is getting a bargain or the car is overpriced.

User-friendly tool:

Develop a user-friendly interface or application that allows consumers and dealers to estimate car prices easily based on the model's predictions.

# Problem Definition:

The used car market suffers from pricing uncertainty, often leaving buyers and sellers unsure of the fair market value of vehicles especially when prices of vehicles have inflated about %77.5 since the start of the pandemic according to Dondeal. This has led to overpricing, affecting consumers' purchasing power and businesses' profitability. Therefore, this project aims to develop an accurate pricing model that addresses the issues of inflation and overpricing of vehicles and provides users with reliable pricing estimates, that can leave the buyer and the seller with a better informed decision.

# Scope:

## Business Understanding:

In this phase, the project objectives are defined, and data mining goals are established. This involves understanding the problem, its relevance, and how it can add value to the business.

## Data Understanding:

This phase focuses on data collection, exploration, and initial data analysis. It is crucial for understanding the available data, its quality, and its suitability for the project.the team will collect and clean a large dataset of historical car sales data. This dataset should include information about used car prices, Manufacture, Model, Production year ,Category, Fuel type, Mileage, and Cylinders.

Step that we are going to implements:

“Collect initial data, Describe data, Explore data, Verify data quality” (Hotz, 2022).

Data Preparation:

Data preprocessing and cleaning are performed in this phase. Data is transformed and made ready for analysis and modeling. That helps ensure that our machine learning model has clean, relevant and representative data for can lead with the most accurate price prediction.

Step that we are going to implements:

“Select data, Clean data, Construct data, Integrate data and Format data” (Hotz, 2022).

Model Development:

Machine learning models will be developed to predict car prices. Various regression algorithms such as linear regression, decision trees, random forests, and K-Nearest Neighbors (KNN) will be explored and tested.

Team Roles:

Business Understanding: (Alaa Shammary , Federico Ariton)

Data Understanding: (Alaa Shammary , Federico Ariton)

Data Preparation: (Alaa Shammary , Federico Ariton)

Model Development: (Alaa Shammary , Federico Ariton)

Evaluation: (NAME OF THE MEMBER OF THE TEAM)

Deployment: (NAME OF THE MEMBER OF THE TEAM)

Documentation: (NAME OF THE MEMBER OF THE TEAM)

# Ethical Considerations:

Ethical considerations include safeguarding user privacy and ensuring that data collected is used solely for the purpose of the project. Data permissions will be sought from all relevant sources, and any personally identifiable information will be anonymized. The project will adhere to all data protection and privacy regulations. There are no medical aspects involved in this capstone project.

# References:

Gokce, E. (2020). Predicting Used Car Prices with Machine Learning Techniques. [online] Medium. Available at: <https://towardsdatascience.com/predicting-used-car-prices-with-machine-learning-techniques-8a9d8313952>.

Simplilearn (2021). *What is Collection of Data? Methods, Types & Everything You Should Know*. [online] Simplilearn.com. Available at: <https://www.simplilearn.com/what-is-data-collection-article#:~:text=Data%20collection%20is%20the%20process%20of%20collecting%20and%20analyzing%20information>