USED CARS PREDICTION

Saudi Arabia Used Cars Prediction

By: Nourah Alsaadan

SADAIA | T5

Abstract

Table of Contents

List of Figures

List of Tables

No table of figures entries found.

Ch1 | Introduction

Introduction

This chapter represents a simple introduction to the project and a discussion of the problem statement. Also, investigates the dataset and explains the attributes.

1.1 Problem Statement

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. I am going to implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models across cities in Saudi Arabia.

1.2 Dataset

For this project, we are using the dataset on used car sales from all over Saudi Arabia, available on Kaggle [1].

The dataset contains 8248 records of used cars collected from syarah.com. Each row represents a used car with a link to its webpage. Other information regarding each car is the brand name, model, manufacturing year, origin, the color of the car, options, capacity of the engine, type of fuel, transmission type, the mileage that the car covered, region price, and negotiable.

1.3 Tools

In order to predict the cars prices, we are going to use Python and some of the packages that will help us to achieve our goal such as: NumPy, Pandas, Scikit-learn, Seaborn and matplotlib.

And since it's a regression problem we are going to compare several regression models to find the best model.

Further, we are going to make the visualizations using Tableau.

1.4 MVP

Minimum Viable Product

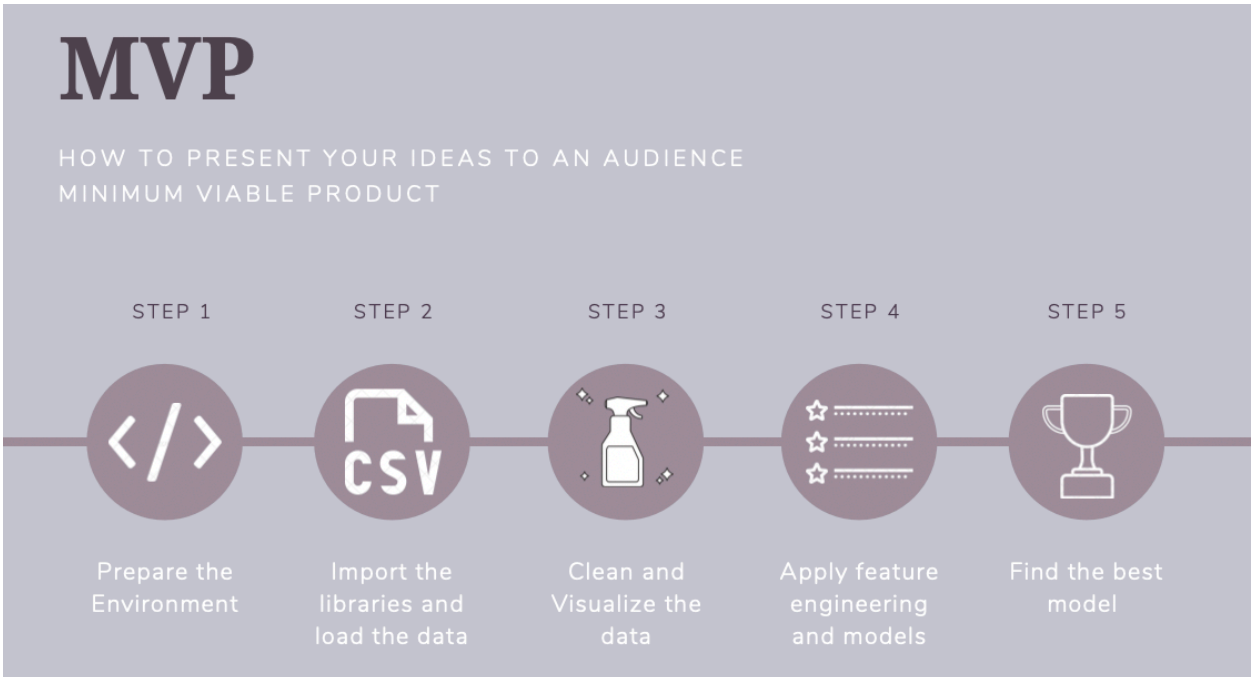

Figure 1

References

[1] https://www.kaggle.com/turkibintalib/saudi-arabia-used-cars-dataset?select=UsedCarsSA_Unclean_EN.csv