FEDERAL URDU UNIVERSITY OF ARTS, SCIENCE AND TECHNOLOGY

Muhammad Umer(19122131)

Muzammil(19122148)

Department : BS(Computer Science)

Section: B

Semester : 6

Course Name: Data Mining

Submitted : Miss Uzma Fazal

Project Report : **Pak Used Car Prices**

OBJECTIVE:

A used car dataset typically includes information about various vehicles, such as make, model, year of manufacture, mileage, fuel type, transmission, and other relevant details. Additionally, it might contain pricing information, owner history, and possibly the condition or maintenance history of the cars. These datasets are valuable for analysis in the automotive industry, helping users make informed decisions when buying or selling used cars.

Dataset Major Columns:

These are the major columns in the dataset :

- Body
- Assembly
- Brand
- Model
- Engine
- Type
- Prices

INTRODUCTION AND BACKGROUND OF THE PROBLEM:

The car price dataset is a comprehensive collection of information that captures the prices of various car items across different regions, markets, and time periods. This dataset is a valuable resource for researchers, policymakers, economists, and other stakeholders interested in understanding and analysing trends, patterns, and factors influencing car prices. Car prices have significant implications for individuals, communities, and economies. The volatility in car prices can be influenced by a multitude of factors, including engine conditions, agricultural practices, geopolitical events, market dynamics, and supply chain disruptions. Understanding and mitigating the impact of these factors is crucial for ensuring car security, economic stability, and the well-being of populations.

DATA COLLECTION:

The source of the dataset is kaggle but the data in the dataset is scraped from zameen.com.

https://www.kaggle.com/datasets/talhabarkaatahmad/pakistan-used-car-prices-2023

PRE-PROCESSING:

Before applying any model on the dataset first it undergoes pre-processing. It handle these steps:

- Data Cleaning
- Handle Missing Values
- Remove Duplicate Values

MODELLING:

Random Forest Regression:

The Random Forest Regression model is a flexible and effective tool for predicting property prices. Unlike a single decision tree, it is an ensemble method that aggregates the predictions of multiple trees, resulting in improved accuracy and resistance to overfitting. This model is capable of capturing both linear and non-linear relationships, making it particularly suitable for situations where property prices are influenced by complex interactions among features. The Random Forest Regressor is well-equipped to handle diverse data structures, offering a reliable and powerful solution for predicting property prices with enhanced generalisation capabilities.
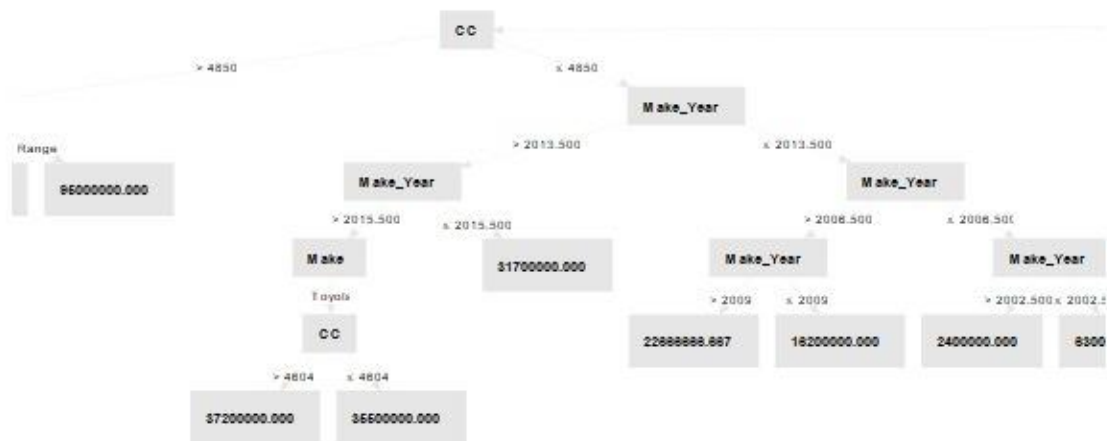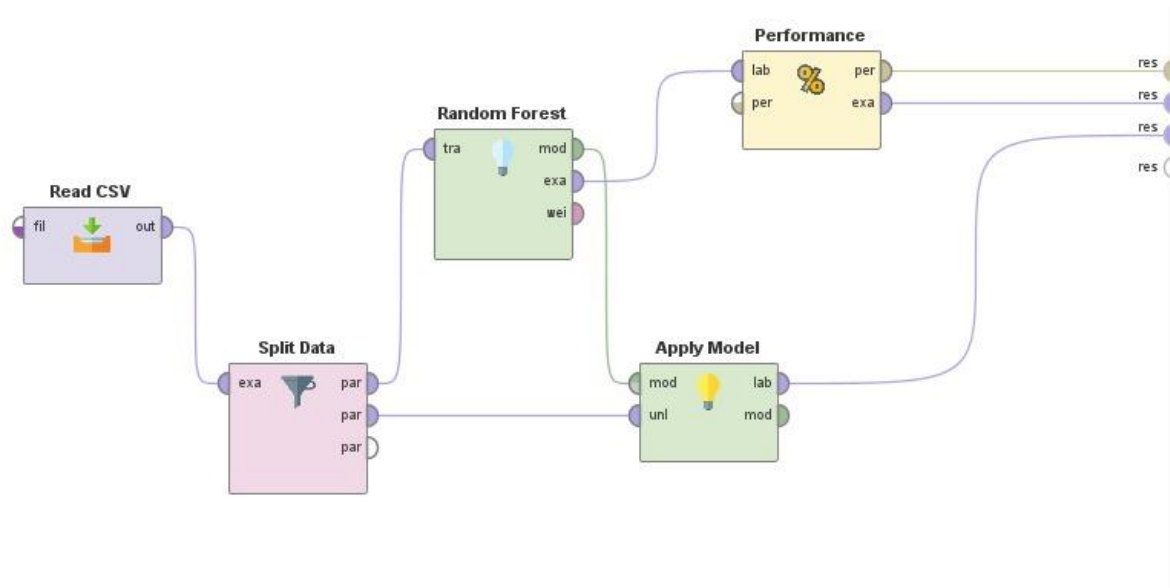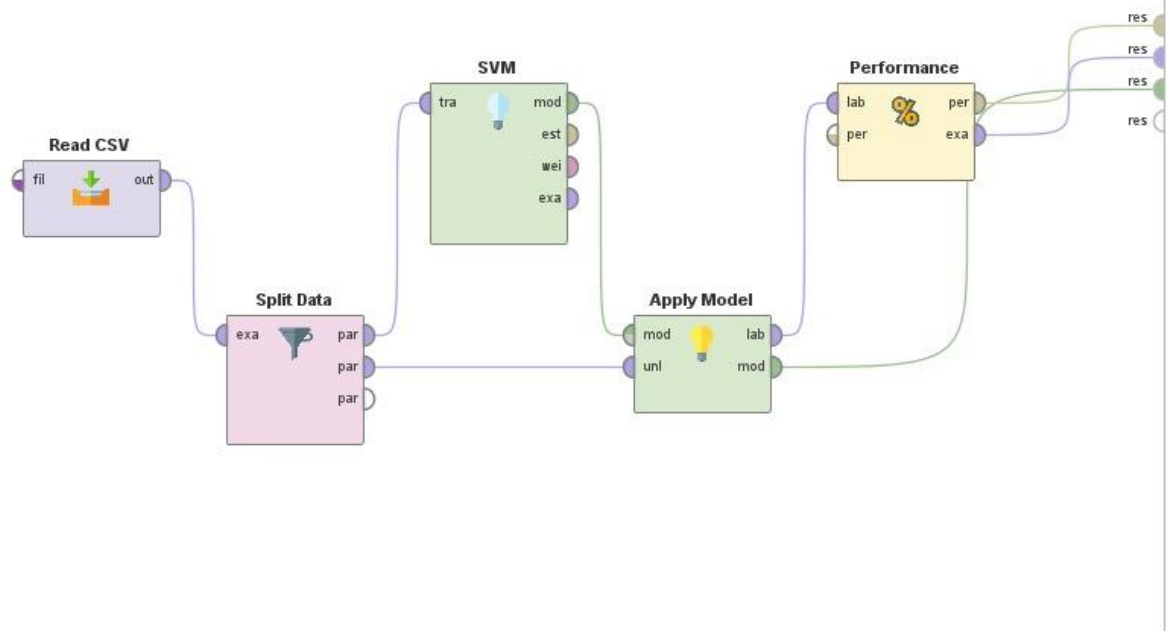
Support Vector Machines (SVM):

SVM is a versatile model capable of handling both linear and non-linear relationships. By using an appropriate kernel, SVM can capture complex interactions among features, making it suitable for scenarios where the relationship between features and property prices is not strictly linear.

np

Row No.	Price	prediction(P...	Make_Year	CC
1	9300000	2274947.135	2015	1800
2	1950000	2274685.167	2017	1000
3	2440000	2274823.054	2017	1300
4	1700000	2274685.167	2017	1000
5	1160000	2274188.679	2005	1300
6	2500000	2274735.677	2011	1800
7	5485000	2275126.436	2021	1500
8	1875000	2275066.669	2019	1600
9	2600000	2275034.512	2021	1300
10	1500000	2274611.595	2013	1300
11	1100000	2274526.573	2014	1000
12	955000	2274579.438	2015	1000
13	2200000	2274632.302	2016	1000
14	2720000	2274855.211	2015	1600

Name			Type	Missing	Statistics			Filter (4 / 4 attributes):	Search for Attributes	
Label **Price**			Integer	0	Min 150000	Max 42500000	Average 3511541.667			
Prediction **prediction(Price)**			Integer	0	Min 2273165.900	Max 2276343.483	Average 2274733.641			
Make_Year			Integer	0	Min 1990	Max 2021	Average 2013.253			
CC			Integer	0	Min 6	Max 4700	Average 1536.397			

CONCLUSION:

Summarise the overall findings and their significance in the context of the car industry. Emphasise how the analysis contributes to informed decision-making, potential improvements in operational efficiency, and strategies for navigating the complexities of the car market. Conclude with a call to action based on the recommendations provided.