

## Problem

### Introduction

A car company has the data for all the cars that are present in the market. They are planning to introduce some new ones of their own, but first, they want to find out what would be the popularity of the new cars in the market based on each car's attributes. We will provide you a dataset of cars along with the attributes of each car along with its popularity. Your task is to train a model that can predict the popularity of new cars based on the given attributes.

### Dataset

You are given a training dataset, `train.csv`. The file is a comma separated file with useful information for this task:

- `train.csv` contains the information about a car along with its popularity level. Each row provides information on each car. Information such as `buying_price`, `maintenance_cost`, `number_of_doors`, `number_of_seats`, etc. The definition of each attribute is as follows:
  1. `buying_price`: The `buying_price` denotes the buying price of the car, and it ranges from `[1...4]`, where `buying_price` equal to 1 represents the lowest price while `buying_price` equal to 4 represents the highest price.
  2. `maintenance_cost`: The `maintenance_cost` denotes the maintenance cost of the car, and it ranges from `[1...4]`, where `maintenance_cost` equal to 1 represents the lowest cost while `maintenance_cost` equal to 4 represents the highest cost.
  3. `number_of_doors`: The `number_of_doors` denotes the number of doors in the car, and it ranges from `[2...5]`, where each value of `number_of_doors` represents the number of doors in the car.
  4. `number_of_seats`: The `number_of_seats` denotes the number of seats in the car, and it consists of `[2, 4, 5]`, where each value of `number_of_seats` represents the number of seats in the car.
  5. `luggage_boot_size`: The `luggage_boot_size` denotes the luggage boot size, and it ranges from `[1...3]`, where `luggage_boot_size` equal to 1 represents smallest luggage boot size while `luggage_boot_size` equal to 3 represents largest luggage boot size.
  6. `safety_rating`: The `safety_rating` denotes the safety rating of the car, and it ranges from `[1...3]`, where `safety_rating` equal to 1 represents low safety while `safety_rating` equal to 3 represents high safety.
  7. `popularity`: The `popularity` denotes the popularity of the car, and it ranges from `[1...4]`, where `popularity` equal to 1 represents an unacceptable car, `popularity` equal to 2 represents an acceptable car, `popularity` equal to 3 represents a good car, and `popularity` equal to 4 represents the best car.

We also provide a test set of car along with the above attributes excluding popularity, in `test.csv`. The goal is to predict the popularity of the car based on its attributes.

**Tool used:** Jupyter notebook

**Language/Libraries:** Python machine learning libraries

train.csv

- No preprocessing.
  - Train-test split the data
- train the data on various classifier

1. svm
  2. logistic regression
  3. kneighbors classifier
  4. naive bayes
  5. decision tree
  6. random forest classifier
  7. gradient boosting regression
  8. mlp classifier
  9. extra tree classifier
  10. bagging classifier
- choosing random forest classifier given best

test.csv

- column name manually defined( they just only 6, for large number of feature surely done with programming).
- Making prediction using RandomForestClassifier
- saving the prediction result to prediction.csv

## Result:

Problem	Submissions	Leaderboard	Discussions	
Problem	Language	Time	Result	Score
<a href="#">Car Popularity Prediction</a>	Text	a month ago	Processed	909.05
				<a href="#">View Results</a>