

CHAPTER 1

INTRODUCTION

1.1 Background

Now, cars come in many shapes and use a variety of fuels. It can affect the price of the car. So we need to find the average car price across all car brands.

The author use the K-Nearest Neighbors (KNN) and Random Forest algorithms. K-Nearest Neighbors (KNN) stands for "K-nearest neighbor" algorithm, while Random Forest is an algorithm based on classification and regression techniques that utilize decision trees. It is a way of selecting variables or predictors to build an accurate model to predict the target variable.

The author start by collecting examples of prices for several cars. Each brand's price, model, and year are unique. Then, using orange data mining, we train a K-Nearest Neighbors (KNN) and a random forest on the training data.

The final result of this project will be the RMSE of K-Nearest Neighbors (KNN) and Random Forest. The results of the car price prediction are needed at this time considering the many car models from various brands. then it can be used for marketing and to see the average price of each brand.

1.2 Problem Formulation

Based on the explanation above, questions that can be taken:

1. Can Orange run the car price dataset with the K Nearest Neighbor (KNN) algorithm for predicting?
2. Can Orange run the car price dataset with the Random Forest algorithm for predicting?
3. Which algorithm has better result?

1.3 Scope

This project use orange using algorithm K-Nearest Neighbor (KNN) and Random Forest with data taken from kaggle. Dataset has attributes including ID, price, manufacturer, model, production year, fuel type, and drive wheels.

1.4 Objective

The results of this project will determine whether orange has good performance and which algorithm has the best results for car price datasets using K-Nearest Neighbors (KNN) or Random Forest.

