CHAPTER 1
INTRODUCTION

1.1. Background

Regression is a data mining function that predicts a number. Profit, sales, mortgage rates, house values, square footage, temperature, or distance could all be predicted using regression techniques, and also using methods and algorithms to compare which one of them has a better prediction for a specific dataset.

And this project is to find a better regression method for predicting prices by using data on Paris housing prices, and get the best prices for a house in Paris by using two methods, two methods used in this calculation Linear regression and Random Forest which are processed in the Orange application.

Paris housing prices data will be processed using Linear regression and Random Forest methods and then processed in the Orange application to predict prices. After calculating the data that has been obtained from each of these methods, the data is compared with each algorithm to see which prediction is better.

1.2. Problem Formulation

In this project, only a few issues will be explored.

1. Can the Linear Regression method predict prices with less error accurately?
2. Can the Random Forest method predict less error accurately?

1.3. Scope

This project is based on the Kaggle dataset in 2021 and is located in Paris, this prediction is carried out using the help of the orange data mining tool and also using the orange data mining widget by utilizing linear regression and random forest algorithm, also comparing both algorithms which one of them is suitable for this dataset, this study only comparing prices of the houses.

1

1.4. Objective

This project will be discussing the use of the Orange Data Mining application to make predictions. The result of this project will prove that Orange is a great tool for making predictions because it has both high accuracy and low variance.