CHAPTER 3
RESEARCH METHODOLOGY

3.1.   Literature Study

The first one using this comparing method is finding a dataset from the internet, in this project I am using keggle.com website to find my dataset, and chose Paris housing prices as my dataset there are a lot of attributes in this dataset the one that I use for comparing methods are "prices" attribute.

3.2.   Data and Variable

The main data source in this study is Paris housing prices data having total 10000 data and 17 attributes, the attribute dataset use in this form are squareMeters, numberOfRooms, hasYard, hasPool ,floors - number of floors, cityCode - zip code, cityPartRange - the higher the range, the more exclusive the neighbourhood is, numPrevOwners - number of prevoious owners, made - year, isNewBuilt, hasStormProtector, basement - basement square meters, attic - attic square meteres, garage - garage size, hasStorageRoom, hasGuestRoom - number of guest rooms, price - price of a house, category - Luxury or Basic. From existing data and attributes, the author will compare the accuracy and less error of linear regression and random forest method, which one of those two is better for predicting price.

3.3.   Orange Data Mining Tool

Data Mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various databases. Orange data mining is an open-source machine learning and data visualization, building data to analyze workflow visually, with a large and diverse toolbox. In my research I am using orange to make a comparison using data mining algorithms.