# CHAPTER 4
# ANALYSIS AND DESIGN

## 4.1. Analysis

This research was conducted with the help of the Orange data mining application using the Neural Network algorithm model and the K-Nearest Neighbor algorithm. The dataset is taken from Kaggle with 5 manufacturers and years of manufacture from 1939 to 2020. The attributes consist of Id, price, levy, manufacturer, model, production year, category, leather interior, fuel type, engine volume, mileage, cylinders, gearbox type, drive wheels, doors, wheels, colors, and airbags. The price column is the target for predicting the car's price.

Neural networks, which are a subset of machine learning and are at the core of deep learning algorithms, are also known as artificial neural networks (ANNs) or simulated neural networks (SNNs). In order to mirror the way that organic neurons communicate with one another, their name and structure are both derived from the human brain. Node layers make up artificial neural networks (ANNs), which have an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, is interconnected and has a threshold and weight that go with it. A node gets activated and begins transferring data to the network's next layer if its output is greater than the predefined threshold value for that node. No data is otherwise transmitted to the network's next tier.

"K-nearest neighbors (KNN) is a type of supervised learning algorithm used for both regression and classification. KNN tries to predict the correct class for the test data by calculating the distance between the test data and all the training points. Then select the K number of points that are closest to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes of 'K' training data and the class that holds the highest probability will be selected. In the case of regression, the value is the mean of the 'K' selected training points."[1]

[1] https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4

## 4.2. Desain

What data will be used for forecasts is indicated by the project's data. The dataset for this project was collected from Kaggle. Car Price Deep Contractor-owned Dataset 5 months ago. There are 18 columns and 19237 rows in this dataset. The car's ID, price, maker, model, production year, fuel type, and driving wheels are the datasets used in this model. Choose data in the orange program, then choose CSV file import. Then, click Select columns, then Data transform. Target the attribute price in a few columns after that. Next, make a data sampler. Determine the fixed proportion of data in the data sampler. Connect the data sampler to the CSV file import after that choose the columns.

In this project, transform refers to deciding which column and how much data will be used for prediction. Choose data in the Orange data mining applications, then choose CSV file import. After that, select columns and then data transform. Target the attribute price in a few columns after that. Next, make a data sampler. Determine the fixed proportion of data in the data sampler. Connect the data sampler to the CSV file import to choose columns at that point.

The project's model refers to the algorithm that will be utilized to forecast the dataset. Neural Network and K-Nearest Neighbors (KNN) are used here. To make predictions, the algorithm is given sample data. Select Neural Network and k-NN from the orange program. Correlate the datasets from the Data sampler and the training data.

In this project, evaluation and reporting mean that, following the input of the dataset into the model, predictions will be made. Next, compare the output of the two algorithms to discover which yields more accurate predictions. Choose Prediction after choosing Evaluate in Orange application. After that, link the data sampler to the prediction. Change the related link from the data sample to the remaining data in edit connections.

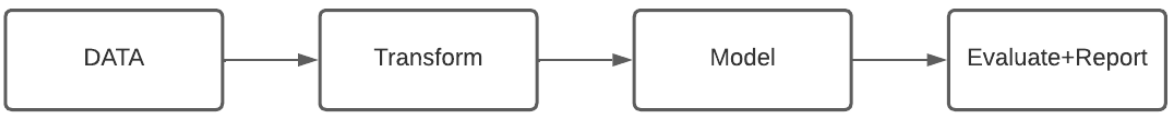


**Figure 4.1** Flowchart Process Predictions