**CONTENTS**

**1.Introduction**

**1.1 Introduction**

**1.2 Objectives Of Research**

**1.3 Problem Statement**

**2.Review Of Literature**

**3.Data Collection**

**4.Methodology**

**4.1 Exploratory Data Analysis**

**4.1.1 Figures and Tables**

**4.2 Statistical Methods and Data Visualization**

**4.3 Data Modelling Using Supervised ML Techniques**

**5.References**

**6.Conclusion**

**VEHICLE-RESALE VALUE PREDICTION**

**1.INTRODUCTION:**

**1.1 INTRODUCTION**

Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. AI is accomplished by studying how human brain thinks and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.

Machine Learning may be defined as the field of computer science, more specifically an application of artificial intelligence, which provides computer systems the ability to learn with data and improve from experience without being explicitly programmed.

Basically, the main focus of machine learning is to allow the computers learn automatically without human intervention. It can be started with the observations of data. The data can be some examples, instruction or some direct experiences too. Then on the basis of this input, machine makes better decision by looking for some patterns in data.

**1.2 OBJECTIVES OF RESEARCH**

* The main objective of vehicle-resale prediction model is to determine the resale value of a second-hand vehicle by considering vehicle attributes.
* This Model is very much useful for vehicle makers in second-hand market by predicting the resale value of the vehicle.

**1.3 PROBLEM STATEMET**

Vehicle makers face several challenges in the second-hand market. The depth crisis in the European Union, the general problem of overcapacity, increasing competition from Asian manufacturers, and the trend toward more eco-friendly cars are only a few factors that add to the difficulty of selling used vehicles in the second-hand market and decrease sales margins. Therefore, Vehicle makers require sophisticated decision support systems to sustain the profitability of the used vehicle business. A core component of such systems is a prediction model that estimates resale prices on the basis of vehicle attributes .Although a statistical modelling of resale prices has been considered in previous work (e.g., Purohit, 1992), only very few studies have explicitly attempted to predict resale prices with maximal accuracy to support decision making. As a consequence, we don’t know to which degree are resale prices predictable, what is the relative accuracy of different prediction methods and are some methods particularly effective.

**2.REVIEW OF LITERATURE:**

On doing the literature survey of various methods for vehicle-resale value prediction, we come to the conclusion that to predict the resale value there are multiple approaches like

* Multi-Linear Regression
* Decision Tree Regression
* Random Forest Regression

Multi-Linear Regression:

Multiple Linear Regression is a simple and common way to analyse linear regression. The model is often used for predictive analysis since it defines the relationship between two or more variables. Multiple Linear Regression attempts to model the Relationship between two or more features and a response by fitting a linear equation to observed data.

Decision Tree Regression:

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

Random Forest Regression:

Random forest Regression is an ensemble machine learning algorithm that is used for regression problems. This algorithm is mainly used to forecast values based on some attributes. Random Forest algorithm is easy to use and accurate Machine Learning Algorithm.

**3.DATA COLLECTION :**

Over 370000 used vehicles scraped with scrapy from Ebay-Kleinanzeigen . Those fields are included in autos.csv:

* Date Crawled : when this ad was first crawled, all field-values are taken from this date
* name : name of the vehicle
* seller : private or dealer
* offer Type
* price : the price on the ad to sell the vehicle
* vehicle Type
* year Of Registration : at which year the vehicle was first registered
* gearbox
* power PS : power of the vehicle in PS
* model
* kilo meter : how many kilo meters the car has driven
* month Of Registration : at which month the vehicle was first registered
* fuel Type
* brand
* not Repaired Damage : if the vehicle has a damage which is not repaired yet
* Date Created : the date for which the ad at e-bay was created

**4.METHODOLOGY :**

**4.1 EXPLORATORY DATA ANALYSIS**

**4.1.1 Figures and Tables**

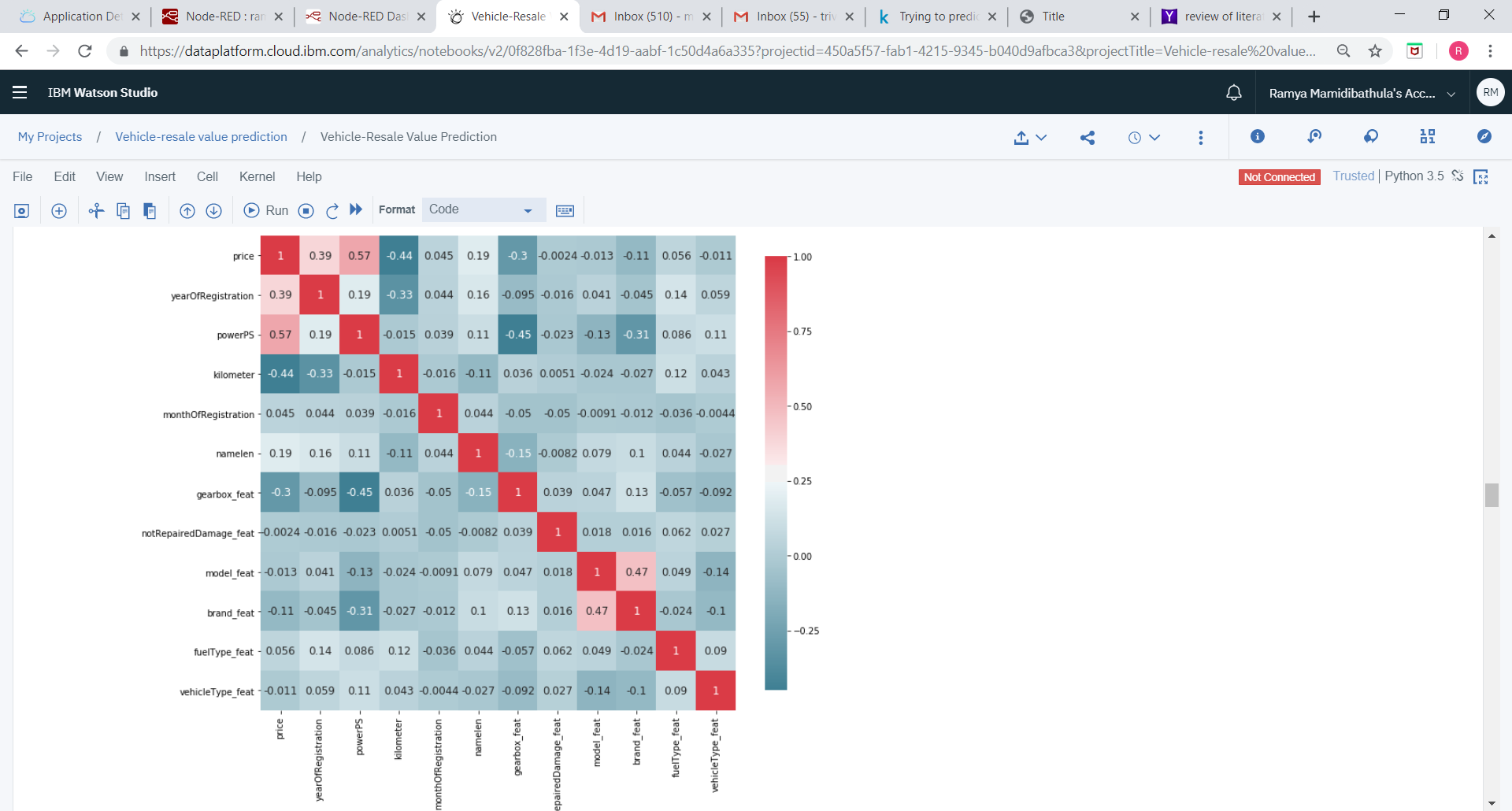


Fig 1 : Heat-Map

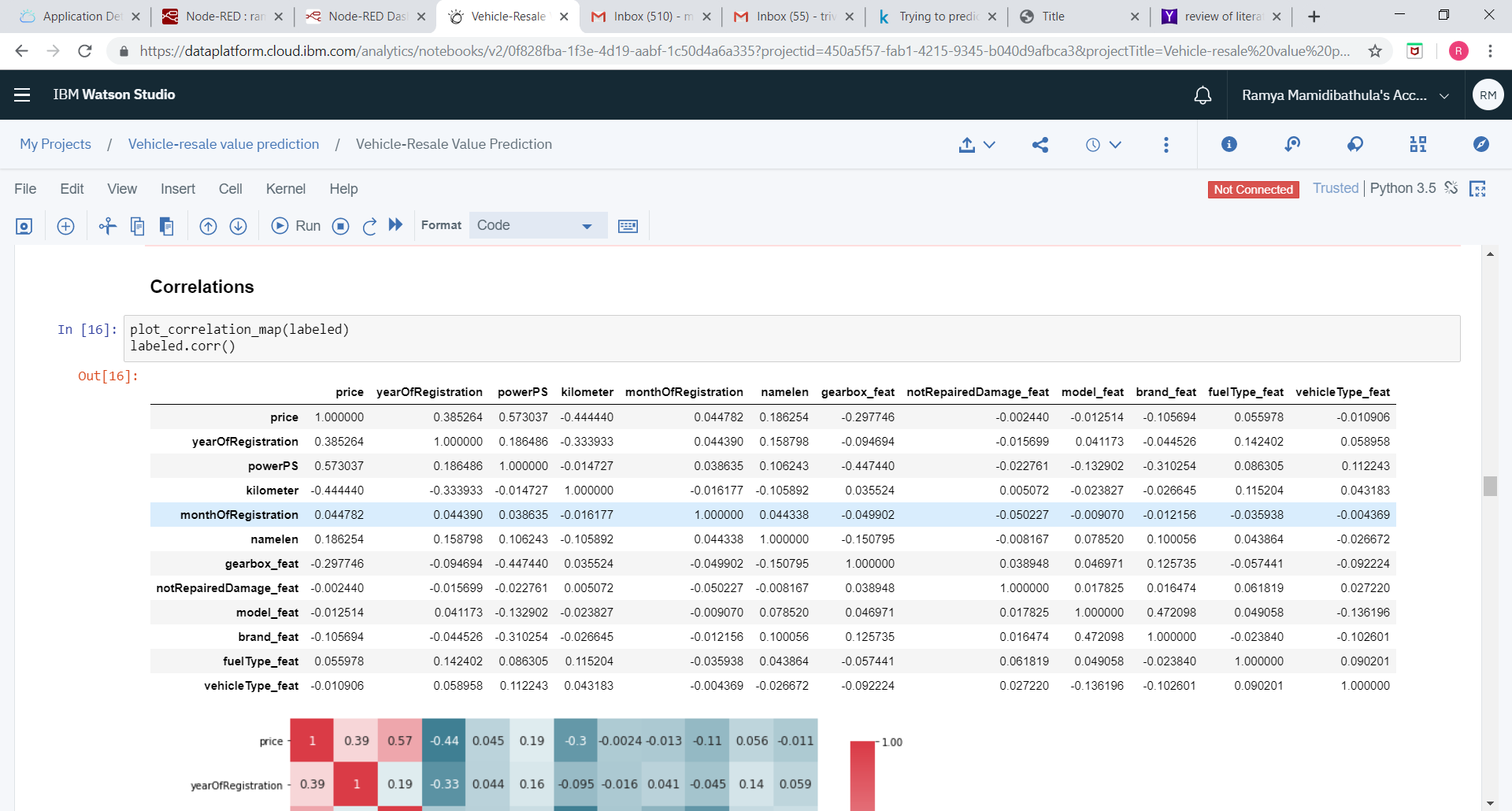


Fig 2 : Correlation Table

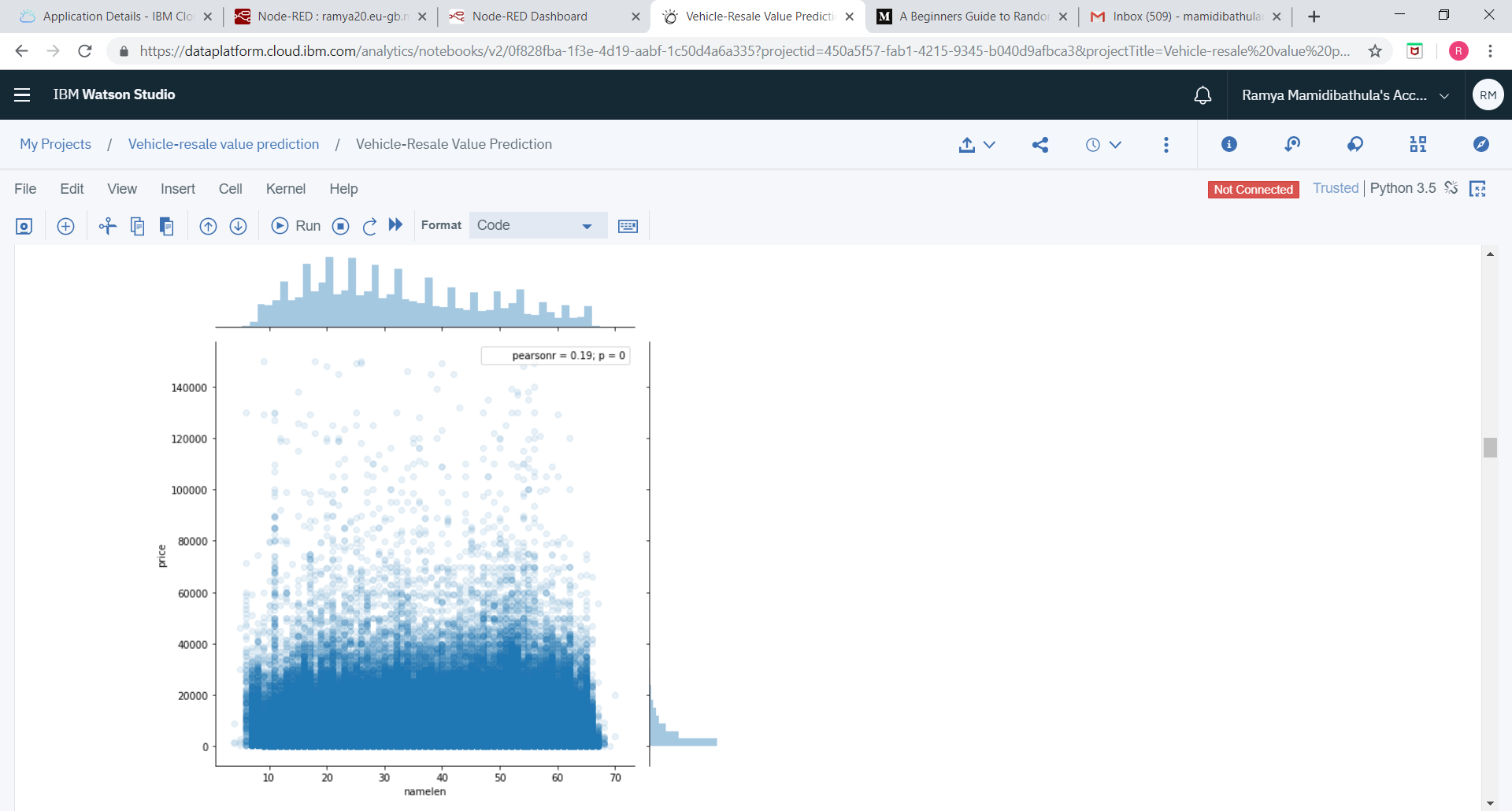


Fig 3 : Joint Plot of name length and price

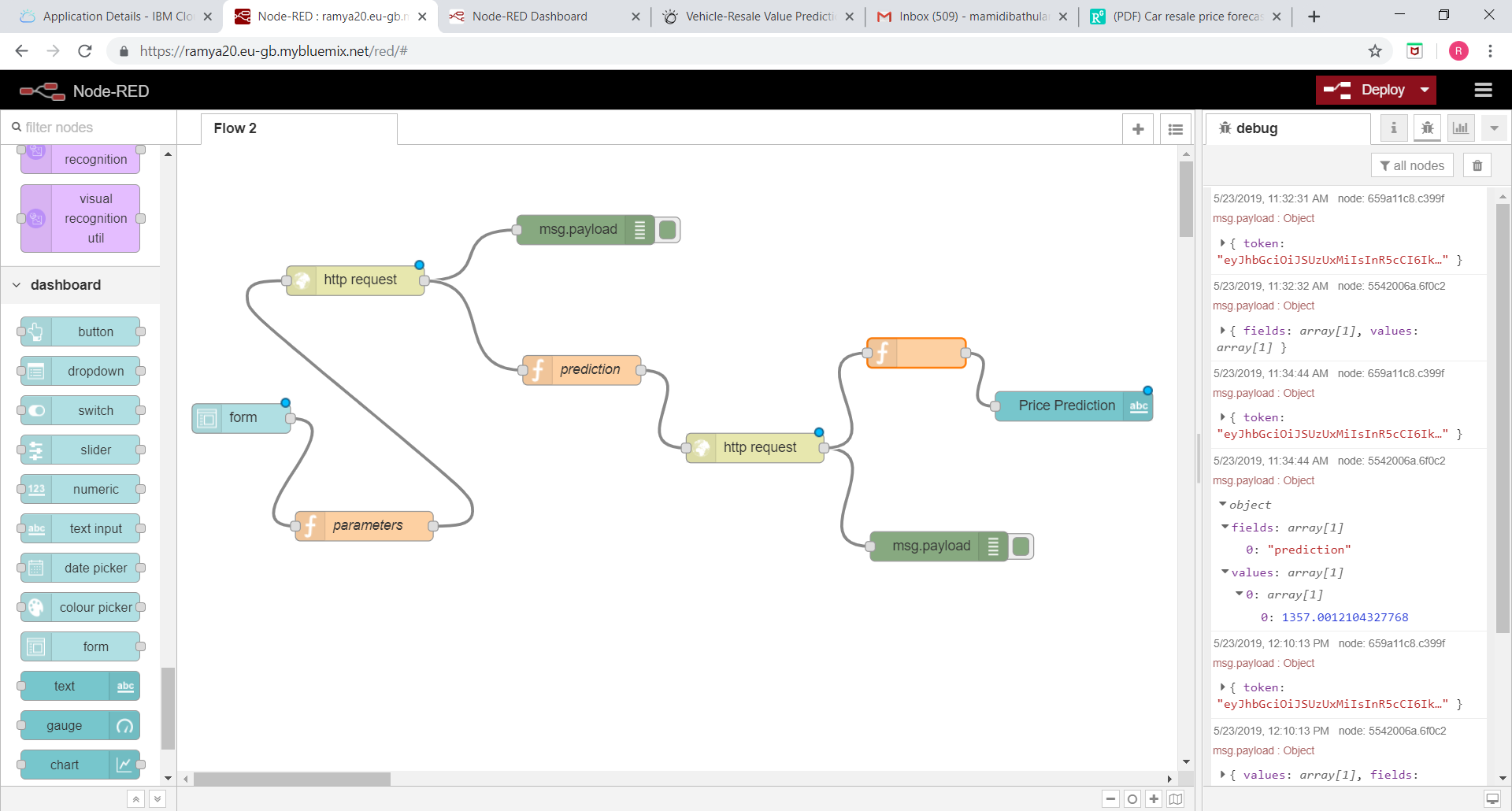


Fig 4 : Flow of User Interface

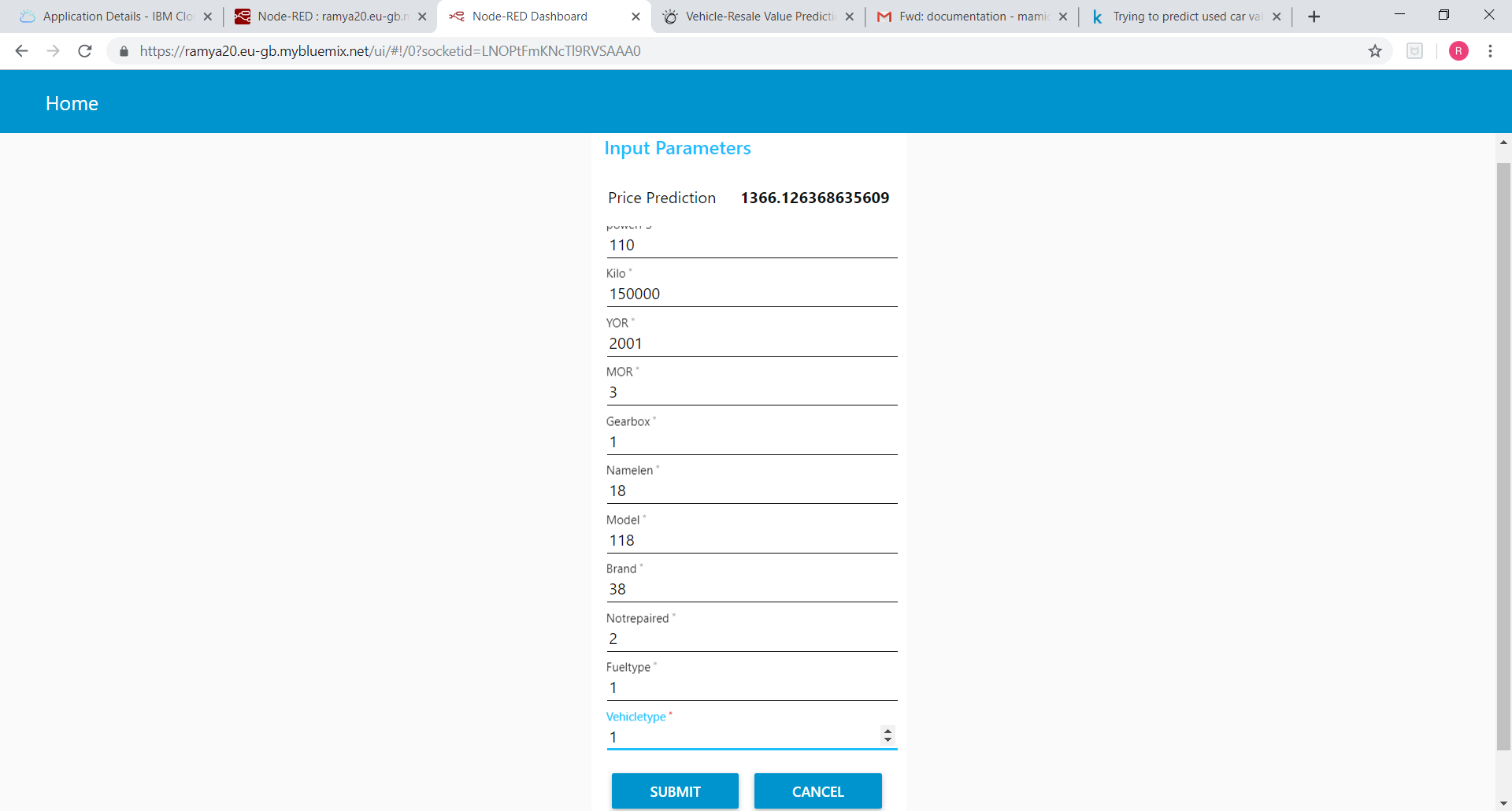


Fig 5 : Predicted Value

**4.2 STATISTICAL TECHNIQUES AND DATA VISUALIZATION**

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral to yourself and stakeholders than measures of association or significance.

The following list of visualization techniques are used in the Predicting vehicle-resale value are:

**Bar charts:** Rectangles (bars) are used to represent different entities, where the height or the width encodes quantitative values. For example, we have an example of stacked bar charts that is being utilized for ranking. This visualization technique is well suited to represent relative differences. There are different types of bar charts such as horizontal bar charts, stacked bar charts, and range bar charts that were classified equally in this category.

**Scatter plots.** These are graphs where each sample is represented by a point or symbol. Each point or symbol position is defined according to two dimensions, or two generated features of these samples. Those graphs are useful to illustrate trends and correlations.

**Heat Maps.** This graphical representation represents values by colors. In fraud detection this technique is usually used to visually query for patterns or outliers in a large amount of data.

**4.3 DATA MODELLING USING SUPERVISED ML TECHNIQUES**

Random forest is a type of ensemble  supervised machine learning algorithm. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

To implement a Random Forest Regression Model using Python. we need to import a fews libraries:

**from sklearn.ensemble import RandomForestRegressor**

An object is created for RandomForestRegressor

**forest = RandomForestRegressor()**

we may also consider some parameters like criterion, random\_state.

Then we have to fit the model by using training data by using fit method.

**forest.fit(X\_train, y\_train)**

At last, we have to predict the values by using predict method and predicted values are stored in a variable.

**y\_predict=forest.predict(X\_test)**

If we want to check the accuracy of a model, r2\_score is used which is to be imported from sklearn.metrics

The parameters for r2\_score is y\_test which contains independent

**from sklearn.metrics import r2\_score**

**r2\_score(y\_test,y\_predict)**

**5.REFERENCES:**

<https://www.kaggle.com/ddmngml/trying-to-predict-used-car-value/data>

<https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/>

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

<https://www.datascience.com/blog/classification-random-forests-in-python>

**6.CONCLUSION:**

In this project, three different machine learning techniques are used to forecast the price of used vehicles. The accuracy obtained for Multi-Linear Regression is 0.5 , for Decision Tree Regression is 0.7 and for Random Forest Regression is 0.8. So, we prefer Random Forest Regression as it has more accuracy. Finally, we are building a machine learning model by using Random Forest Regression to predict the resale value of used vehicles by considering some attributes of a vehicle.