**BIKE BUYER PREDICTION**

**Team Name-** KING CODERS

**Team Members-** 1) Rohith Somisetty

2) V S S Karthik

**Submitted to-** Ms.Ramya Guntamukkala

Topic Page No. • Title of the project 2 ♣ Introduction 2 ♣ Objectives of research 2 ♣ Problem Statement 2 ♣ Industry Profile 3 • Review of literature 4 • Data Collection 5 • Methodology 5 ♣ Exploratory Data Analysis 7 o Figures and tables 8 ♣ Statistical Techniques and Data Visualization 9 ♣ Data Modeling using Supervised ML techniques 9 • Conclusion 13

BIKE BUYER PREDICTION

**INTRODUCTION-**

In the present day scenario, we come across that the usage of vehicles for transport has increased at a tremendous rate and hence the business of the automobile industry has witnessed an enormous growth. But it is also important to keep in mind who all prefer a 4 wheeler or a 2 wheeler depending upon the family income, the number of children they have, marital status of a person, the distance he travels everyday and many other factors influence the fact whether which type of vehicle a person opts for.

**OBJECTIVES OF RESEARCH-**

1) To visualize and understand the given database by using graphs, diagrams and images.

2) To test the dataset using different machine learning algorithms and find the best one.

3) To fit the dataset into the best model.

4) To predict which classification the input given by the user belongs to.

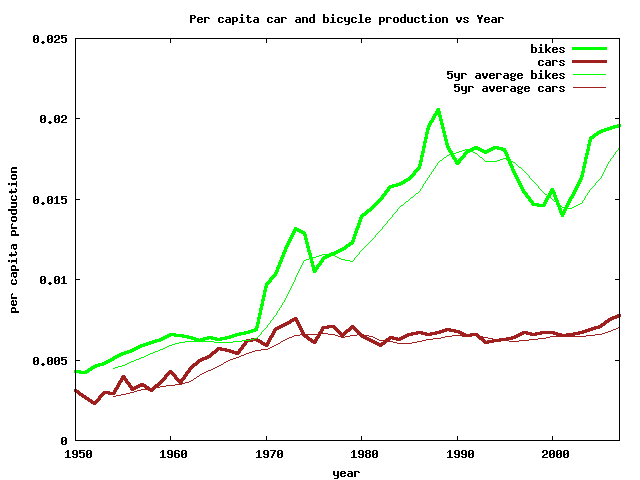
5) Finally to make the User Interface for easy use of the application.

**PROBLEM STATEMENT-**

As the demand for vehicles is increasing it is also important to know for the industry to have a prediction whether a person will go for a bike or not from his date base to have a figure of how many vehicles he has to manufacture. This saves a lot of raw material, labor, capital as well as time. To make the work for the industries easier, the code helps in predicting if a person buys a bike or not based on the data set given to it containing the information of people of the past years who bought or did not buy a bike based on few parameters.

**INDUSTRY PROFILE-**

Based on a recent census report, India stands in the first position in bike manufacturing by manufacturing a total of 48,000 units everyday which sums up to a total of 17.7 million units per annum overtaking China which produces for about 16.8 million units per annum.



As per the above report the number of bikes being manufactured has increased at an overwhelming rate and is ruling the automobile industry by being manufactured the most.

**REVIEW OF LITERATURE-**

* **Existing System-** Presently there is no such company which uses the method of predicting whether a person buys a bike or not until the code we wrote is of that purpose. As it is an important factor for the industries for efficient manufacturing, the prediction helps a lot in doing so for the industries.
* **Scope Of Extension-** Since there is no such system which is used to predict the buying of bike considering all the factors, the Machine Learning algorithm we used for doing the same is a very helpful and a key aspect for an efficient production of motorcycles in the coming future.
* **Technologies-**

1. Operating System: Machine Learning is platform independent, it can be accessed on Android, Windows, Linux or any other OS the user wishes to open.

2. Programming Languages: Only Python is used in the project.

3. Tools: The tools used in the Machine Learning are:

♣ IBM Cloud

♣ IBM Watson studio

♣ Node – Red

♣ Jupyter Notebook

**Data Collection**

In the following project, here we are using data-set named ‘bike buyer.csv’ which is a CSV (Comma Separated Values) file containing all the different contents of wine which differs and categorises quality of wine. Using the mentioned programming language “Python” we import the data-set in the Jupyter Notebook using one of the Python libraries named “Pandas”.

**Data set description:**

**ID-** The identity number of a person.

**MARITAL STATUS** - Whether or not a person is married.

**GENDER**- Defines the sex of a person.

**YEARLY INCOME**- The earnings of a person per year.

**CHILDREN**- The number of children a person has.

**EDUCATION**- Upto what level has a person studied.

**OCCUPATION**- What the person does for a living.

**HOME OWNER**- If a person owns a house or not.

**CARS**- The number of cars a person owns.

**COMMUTE DISTANCE**- The distance a person travels everyday.

**REGION**- The region in which the person lives.

**AGE**- The age of that person.

**BIKE BUYER**- Whether that person bought a bike or not.

**Methodology**

The following application was made to find if a person buys a bike or not. The given datasets had 13 properties where the final property is the category in which that a person buys a bike or not. The category of bike buyer can be figured out by using the other 12 features present in the dataset. The following application was made in Python for its back end and Node red for the front end. All the coding as well as the execution of the program was done in Watson Studio. First the dataset was uploaded and inserted in the Jupyter Notebook using the following code.

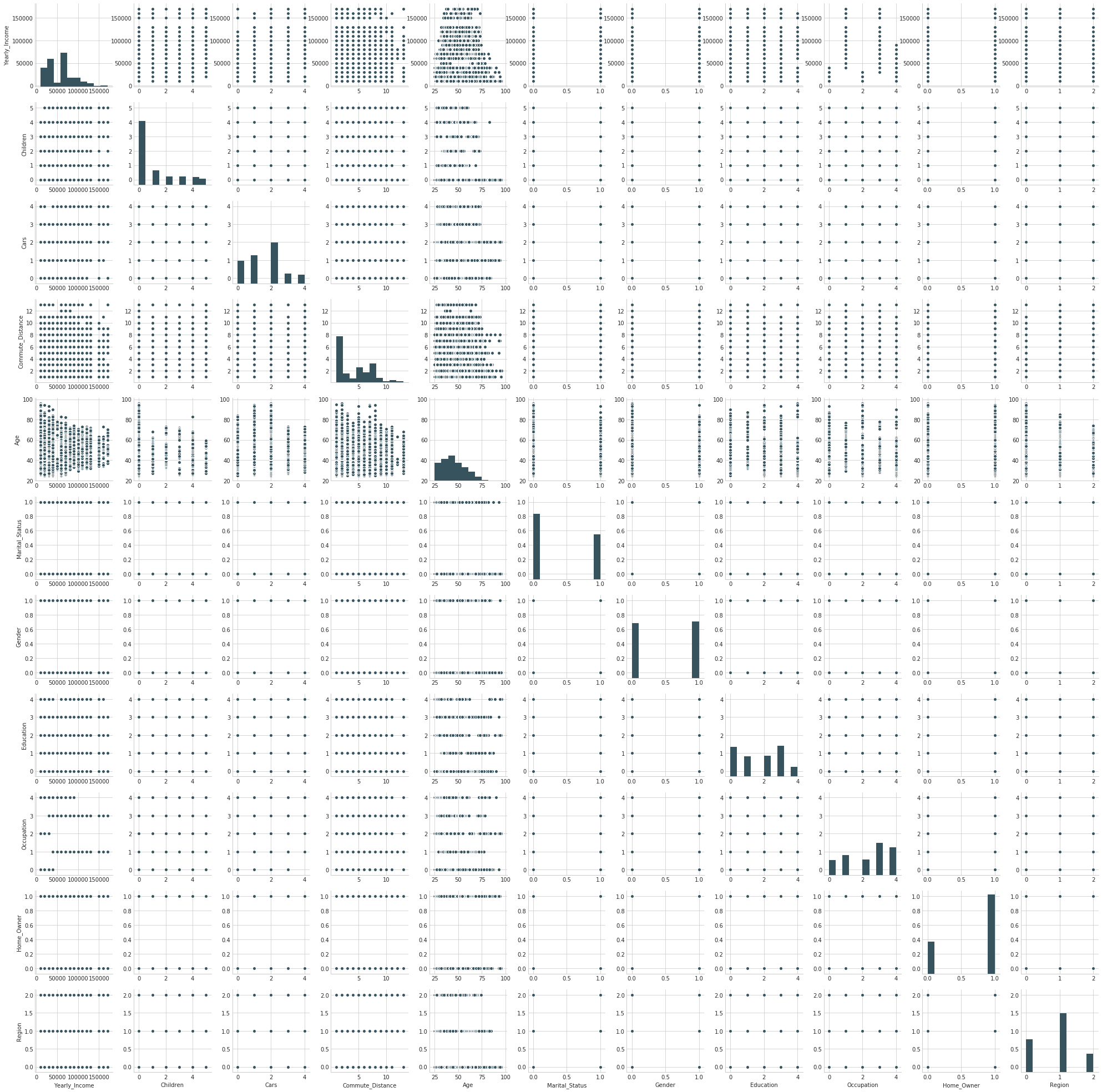
* First the dataset was uploaded and inserted in the Jupyter Notebook using the following code.



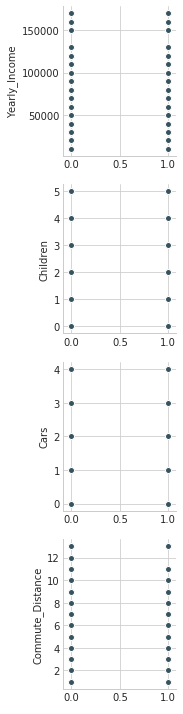
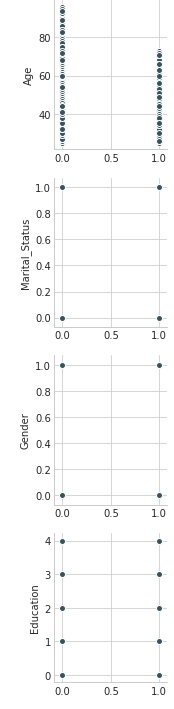
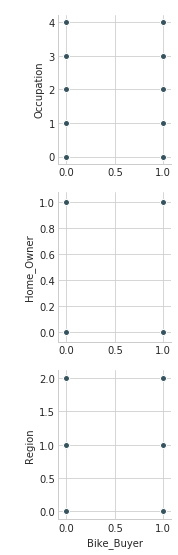
* Defining the independent variable ‘x’ and the dependent variable ‘y’. And Deleting the ‘NaN’ Values.Then Label Encoding The Data



* Graphs
* Pair Plots



* Input Vs Output Graph

* Splitting the data

Then the data is spilt into test and train data. This is done so there is certain unseen data that the predicted values can be tested on. The train data is the data that the model is trained on and the test data the data on which the predicted value will be tested

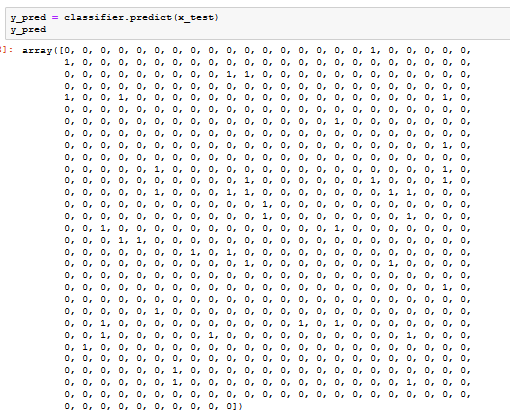


* Data Modelling using Supervised ML techniques:

Finally the model is fit. We have used the Random Forest algorithm to create the classifier needed to fit the model. Then the model is fitted with x\_train and y\_train variables using the classifier.fit function.



* The .predict function is then called to predict the values of x\_test



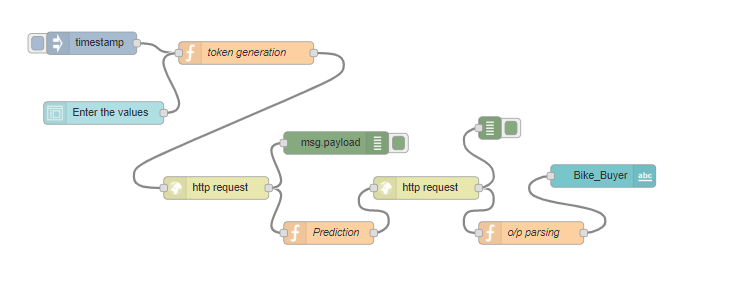
* The values for X\_test is collected now we compare the predicted values with the real values i.e. y\_test .We then score the predicted value according to its accuracy.



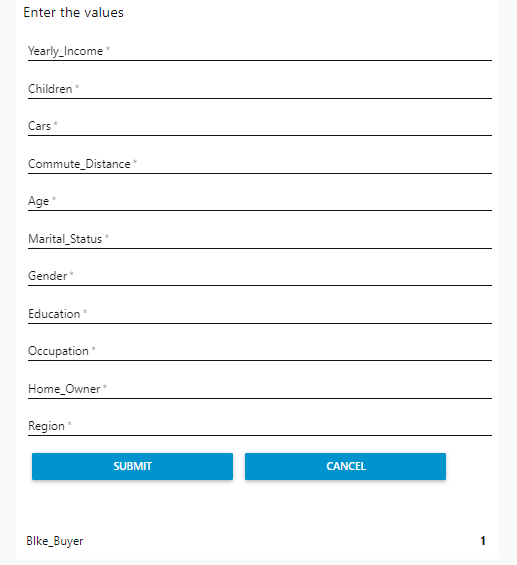


We got a accuracy of 91.5% So the model is properly trained.

* The Node Red Flow



* Web Application



**CONCLUSION-**

The algorithm we used is efficient enough to predict whether a person buys a bike or not and since there is no algorithm which can predict with a 100% accuracy the above algorithm is sufficient enough to be reliable upon.