BIKE BUYER PREDICTION

**PROJECT REPORT**

**Submitted by**

**UNBEATABLES**

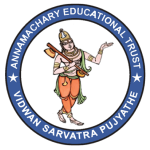
**TEAM MEMBER NAMES:**

1. **Pasupuleti Utkarsh**
2. **Thavva Virajitha**
3. **Gali Yamini**

***In partial fulfilment for the award of the Certificate***

**of**

**SUMMER INTERNSHIP PROGRAM**

**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village , Renigunta Mandal , Tirupati , Andhra Pradesh 517520**

**July 2019.**

### BONAFIDE CERTIFICATE

This is to certify that the project entitled ”**PROJECT TITLE**” submitted by **Team Members Names** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### Signature of Supervisor                                       Signature of Head of the Department

 Akshay Kumar Kothuri, M.Tech       Mrs.B.Rupa Devi M.Tech,Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS,Tirupati.

BIKE BUYER PREDICTION

Abstract :

In this project, I am working on prediction of bike buyers. Transportation has been part of our life as long as humanity exists. The most common road vehicle for transportation is automobile. Automobile includes cars, motorbikes etc. Bikes are the affordable vehicles for all kinds of people. As the need of transportation increases the usage of bikes also increased which results in the sudden rise of automobile industries. The automotive industry continues to face set of challenges. Most manufacturing operations in automotive industries are still largely dependent on experiences-based human decisions. I am highly interested to apply Machine Learning in the automotive industry to make a remarkable ability to bring out hidden relationships among datasets and make predictions.

1 . INTRODUCTION

The problem with every manufacturing company is about the amount of production of the product. If they couldn’t understand the requirement of the product before manufacturing it results in the incurring of losses to the companies. In this project we focused on automobile industry in which bikes are key roles. Our project helps the industry to predict the correct requirement of the product based on customers behavior in the previous deals.

1.1 Python :

Python, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks. The simple syntax rules of the programming language further makes it easier for you to keep the code base readable and application maintainable. Main reasons to use python language is:

1. Readable and Maintainable Code

2. Multiple Programming Paradigms

3. Compatible with Major Platforms and Systems

4.  Robust Standard Library

5. Many Open Source Frameworks and Tools

6. Simplify Complex Software Development

7. Adopt Test Driven Development.

The Libraries in python which can be used by developers to implement Machine Learning in their existing applications are :

* TensorFlow
* Scikit-Learn
* Numpy
* Keras
* Pandas

1.2 Machine Learning :

Machine learning is an subset of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Basically Machine Learning is of three types:

1.Supervised Learning: Learning from characterized data.

2.Unsupervised Learning: Learning from raw data.

3.Reinforcement Learning: Learning from self mistakes (or) self learning data.

Predictive modeling is the way of building a model that is capable of making predictions. The process includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions. Predictive modeling can be divided further into two areas: Regression and pattern classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables. In contrast to regression models,the task of pattern classification is to assign discrete class labels to particular data value as output of a prediction Here we used multiple linear regression algorithm to predict the future crime rate.

Steps involved in Machine Learning :

Steps involved in the machine learning is shown in the below figure

1. Data Collection

2. Data Wrangling

3. Analyse the data

4. Train the algorithm

5. Test Algorithm

6. Deployment

2 . LITERATURE REVIEW

2.1 Project Profile :

In this project, we are working on prediction of bike buyers. Transportation has been part of our life as long as humanity exists. The most common road vehicle for transportation is automobile. Automobile includes cars, motorbikes etc. Bikes are the affordable vehicles for all kinds of people. As the need of transportation increases the usage of bikes also increased which results in the sudden rise of automobile industries. The automotive industry continues to face set of challenges. Most manufacturing operations in automotive industries are still largely dependent on experiences-based human decisions. I am highly interested to apply Machine Learning in the automotive industry to make a remarkable ability to bring out hidden relationships among datasets and make predictions.

2.2 Search Blog :

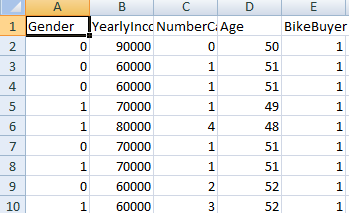
The Search Blog we have studied to fetch the info regarding project is :

<http://ijraset.com/fileserve.php?FID=21941>

3. DATA COLLECTION

3.1 Data Description :

Our dataset consists of 5 attributes and 547 rows which includes entities like Gender ,Yearly income , age ,number of cars owned , bike buyer.



3.2 Data Preprocessing :

In this step of data pre-processing we will pre-process the data. We will read the data from dataset and replace the null values. We will know the information about all the data types. We will know the mean standard deviation and various metrics regarding to the numerical data. To know the correlation between the numerical attributes we will plot the graphs to visualize the data (this context we will use pair plot and heatmap). We will now remove the true outliers. We will consider the data which is three standard deviations away from the mean as the outliers and remove them. Now we need to perform a One Hot Encoding OneHotEncoder from the sklearn.preprocessing module. Normalising to rescale the features to a standard range of values using MinMaxScaler. After that the whole dataset is divided into training and testing data using train\_test\_split from sklearn.model\_selection.

We have collected our dataset from the following link :

https://github.com/xoraus/ML-BikeBuyersPrediction-/blob/master/BBC%20Data%20Set/BBC.csv

4. METHODOLOGY

4.1 Model Selection :

As we need to predict whether a person is a bike buyer or not we do not deal with continuous values in our dependent variable, we have opted classification model for our project. Three supervised learning approaches are selected for this problem. Care is taken that all these approaches are fundamentally different from each other, so that we can cover as wide an umbrella as possible in term of possible approaches. For example- We will not select Random Forest and RandomForest together as they come from the same family of ‘ensemble’ approaches. For each algorithm, we will try out different values of a few hyper parameters to arrive at the best possible classifier. This will be carried out with the help of grid search cross validation technique.

We have tried all the classification models and finally opted decision tree classifier method for our project because we observed a high accuracy value that is 0.69 in this method.

Decision Tree Classifier The decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a treelike model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, it’s also widely used in machine learning.

1) Advantages

a) Able to handle categorical and numerical data. b) Doesn’t require much data pre-processing, and can handle data which hasn’t been normalized, or encoded for Machine Learning Suitability. c) Simple to understand ,visualize and interpret.

2) Disadvantages

a) Complex Decision Trees do not generalize well to the data and can result in over fitting. b) Unstable, as small variations in the data can result in a different decision tree.

5. FINDINGS AND SUGGESTIONS

5.1 Findings :

From among the purchase of alternatives the consumer makes the solution. It may be to buy or not to buy. If the decision is to buy the result displayed will be 1 and if the decision is not to buy the result displayed will be 0.

5.2 Suggestions :

* + When the buyers are busy we cant get accurate data from them.
  + According to the time limit of our project we can cover only the some area.
  + During survey some respondents may not give answer in a proper manner.
  + Due to lack of time for our project we didn’t show the ROC curves for our Machine Learning Models.

6. CONCLUSION

In this project, we have proposed methods for prediction of bike buyers using machine learning techniques. The four machine learning techniques that were used include SVM, Random Forest, Decision tree, KNN. The system was implemented using all the models and their performance was evaluated. Performance evaluation was based on certain performance metric. Decision Tree techniques resulted highest Accuracy of 0.69. From the above results Decision Tree plays a key role in shaping improved classification accuracy of a dataset.