

**A**

**PROJECT REPORT**

**ON**

**“Electric Vehicle Price Analysis and Prediction”**

**SUBMITTED**

**TO**

**SAVITRIBAI PHULE PUNE UNIVERSITY,**

**PUNE FOR THE AWARD OF**

**MASTER OF**

**COMPUTER**

**APPLICATION (SEM-I I)**

**BY**

**Pranalee Solaskar**

**UNDER THE GUIDANCE**

**OF**

**Prof. Navnaath Choudhary**

**THROUGH**

**THE DIRECTOR**

**SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION**

**(SIMCA), NARHE, PUNE**

**(AY. 2021-2022)**

**SINHGAD TECHNICAL EDUCATION SOCIETY’S**

# SINHGAD INSTITUTE OF MANAGEMENT & COMPUTER APPLICATION

**(Affiliated to SavitribaiPhule Pune University & Approved by AICTE)**

**‘NAAC’ Accredited with ‘A’ Grade**

**S. No. 49/1, Off Westernly Bypass, Pune – Mumbai Expressway, Narhe, Pune – 411041, Tel : (020) 66831896 / 66831908 / 66831907**

**E-mail : director\_mca\_simca@sinhgad.edu Website :** [**www.sinhgad.edu**](http://www.sinhgad.edu/)

|  |  |  |
| --- | --- | --- |
| **Prof. M. N. Navale**  M.E. (Elec), MIE, MBA  FOUNDER PRESIDENT | **Dr. (Mrs.) Sunanda M. Navale**  B.A., P.P.M., Ph.D.  FOUNDER SECRETARY | **Dr. Vilas Nandavadekar**  *M.C.A., M.P.M., M.B.A., Ph. D.*  *DIRECTOR, SIMCA - MCA* |

## CERTIFICATE

This is to certify that, the project entitled “**Electric Vehicle Price Prediction and Analysis”,** being submitted for the Partial fulfilment of **Master of Computer Application** by her/him to **Sinhgad Institute of Management and Computer Application affiliated to Savitribai Phule Pune University, Pune** is the result of the original work completed by ***Pranalee Solaskar*** under the guidance of ***Prof. Navnaath Choudhary***

To the best of our knowledge and belief, this work has not been

previously submitted by the award of any degree or diploma of Savitribai Phule

Pune University or any other University.

|  |  |  |
| --- | --- | --- |
| **Prof. Navnaath Choudhary** | **Dr. Ashwini Brahme** | **Dr. Vilas** **Nandawadekar** |
| **Internal Guide** | **Project Coordinator** | **Director SIMCA** |

PLACE: DATE:

**External Examiner**

### DECLARATION

I, the undersigned hereby declare that the project titled **“Electric Vehicle Price Analysis and Prediction “**being submitted for the Partial fulfilment of **Master of Computer Application** by me to **Shinhgad Institute of Management and Computer Application(SIMCA) affiliated to Savitribai Phule Pune University** is the result of an independent work carried out under the guidance of Prof. Navnaath Choudhary**,** is my original work . Further I declare that this project has not been submitted to this or any Institution for the award of any degree.

PLACE: PUNE Pranalee Solaskar

DATE

### ACKNOWLEDGEMENT

I would like to thank all those people who have direct and indirect contributions in the completion of this thesis. First of all, I would like to thank my research guide Dr. Poonam Sawant, for his esteemed guidance, support and encouragement to work on whatever problems motivated me. I express my deep sense of gratitude to my former guide Prof. Navnaath Choudhary, Manager, Big Data Academy, Cap Gemini, Pune for his guidance and encouragement given to me throughout this work. His enthusiasm and energy transformed my vision of the study into reality.

My sincere gratitude goes to Dr. vilas Nandavadekar Head and Director, Sinhgad Institute of Management and Computer Application (SIMCA)-MCA, Pune. His continuous moral support and motivation in every research phase helped me to complete this work. I also wish to thank the Management and MBA-Director of SIMCA, Pune .

I am grateful to my beloved friends, Ms. Vaishnavi Sawant, Ms. Geta Kangude, Mr.Sheshang Harge for their continuous moral support, motivation encouragement and help. My vocabulary falls short to express my deep sense of gratitude to all of them.

My special thanks to my Grand-Mother Mrs. Vimal Kondiram Solaskar and my mother Mrs. Madhavi Solaskar , their blessings are always with me. I fulfilled my dream of research due to inspiration of my father Mr. Mansing Solaskar. Without whom it was impossible for me to carry out my work. I am very thankful to my siblings Mr. Prathamesh Solaskar for their continuous moral support and unconditional love. I owe my special thanks to my friends for his valuable help, motivation and encouragement throughout this research work.

Last but not least, I would like to thanks all my friends for their best wishes. Above all I am grateful to the almighty lord Ganesha, lord Shiva and Sai baba, by the grace of whom this thesis is completed.

**Pranalee Solaskar**

**PREFACE**

Predicting the price of electric vehicles has been an area of ​​great research interest, as it requires considerable effort and industry expert knowledge. For a reliable and accurate prediction, a large number of different attributes are analysed. To build an electric vehicle price prediction model in Europe and the United States, we are applying three machine learning techniques (Data cleaning, data preprocessing, prediction).However, the techniques mentioned here apply to the function altogether.We are using Python for forecasting data. We are applying and comparing different algorithms to data sets for better forecasting.

Early in the 21st century, new companies were formed to take advantage of the absence of the large car manufacturers in the EV market, such as Tesla Motors in the U.S, Think in Norway, BYD in China and REVA in India. Globally all these new manufacturers had released one or more EV models and changed the equation of Auto Industry. Looking into Indian Scenario, still, long way to go, the paper presents the current Indian EV market, market players in two and four-wheeler with recent developments along with the current status of Indian road transportation. Policies and initiatives of government are discussed. A case study is presented with consumer’s perspective to understand ground reality. Tariff for EV charging is discussed. Challenges for Indian market growth, policies, and promotions required are discussed with feasible options along with global scenario.

Transport systems are expected to widely shift towards electric propulsion in the next decade. The diffusion of Electrical Vehicles (EVs) however creates great challenges; among the others, EV charging patterns are non-controllable,thus it is mandatory to have at disposal high quality EV load forecasts in order to reach the operational excellence of networks with a wide EV penetration.

As a key pillar of smart transportation in smart city applications, electric vehicles (EVs) are becoming increasingly popular for their contribution in reducing greenhouse gas emissions. One of the key challenges, however, is the strain on power grid infrastructure that comes with large-scale EV deployment. The solution to this lies in utilization of smart scheduling algorithms to manage the growing public charging demand. Using data-driven tools and machine learning algorithms to learn the EV charging behavior can improve scheduling algorithms.

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Title** | **Page No.** |
|  | Certificate | i |
|  | Declaration | ii |
|  | Acknowledgement | iii |
|  | Preface | iv |
|  | Contents | v |
|  | Keywords | vi |
|  | Abbreviation | vii |
|  | List of Figures | viii |
|  | List of Tables | ix |
| **1** | **Introduction** | **1** |
| **2** | **Review of Literature** | **8** |
| **3** | **Theoretical Framework** | **21** |
| **4** | **Design and Implementation** | **25** |
| **5** | **Experimental Findings and Suggestions** | **34** |
| 6 | **Bibliography** | **38** |
| 7 | **Annexure** | **40** |
|  | Annexure A : Data Collection Sites |  |
|  | Annexure B : Sample Data |  |
|  | Annexure C : Program Code |  |

**Keywords**

* Encoding
* Replace
* astype
* Rename
* Value
* Map
* Reduce
* Countplot
* Value\_Count
* Figure
* hue
* heatmap
* figsize
* replot
* LinearRegression
* Predict
* Fit
* Scatter
* Sklearn

**ABBREVIATION**

* GHG – Greenhouse Gases
* ICE - Internal Combustion Engine
* MPGE- [Miles per gallon of gasoline equivalent](https://afdc.energy.gov/glossary.html#MilesperGallonofGasolineEquivalent)
* PHEVs – Plug in Hybrid Hybrid Electric Vehicle

**List of Figures**

Figure -1.1: Problem Identification

Figure -1.2: Different Stages of Research Work

Figure -3.1: Predictive Analytics Process

Figure -3.2:ML Algorithms

Figure -4.1:Environmental Concern

Figure -4.2: Interested people in electric vehicle

Figure -4.3: Evaluation Parameters

Figure -4.4: Types Of Drive Graphical

Figure -4.5: Number Of Seats In vehicle

Figure -4.6: Number of seats and drive

Figure -4.7: Acceleration

Figure -4.8: Heatmap of DataFrame Column

Figure -4.9: Acceleration and kilowatt and drive

Figure -4.10: Fast Charge Speed and Efficiency

**List of Tables**

Table -4.1: Evaluation Parameters and Concluding Remarks

Table -4.2: Types of Drive

Table -4.3: Number Of Seats In vehicle

Table -4.4: Drive and No of seats

***Chapter 1***

**Introduction**

**1.1 Introduction**

In today’s world, there is a huge amount of climate change we are observing. Different committees are established on a world level for getting the solution. The electric vehicle is one of the solutions for tackling Climate change. Eclectic vehicles are the next-generation cars that everyone will utilise in their daily lives. Which is good for our health and environment. As EV will be in use more so mass production of these vehicles are in progress. Many automobile companies are launching their EV in the market. Many different features are given by them which decide the cost of the vehicle. Battery life, Speed, Acceleration, performance, safety features, and many factors are there to calculate the cost of the EV. We decided to refer to different research papers focusing on specific features of EV and on the basis of this we can predict the cost of the EV using Machine learning.

* 1. **Motivation**

The increase in fuel prices encourages the development and the integration of electric Hence Electric Vehicles (EV) are being presented as a sustainable solution to reduce Greenhouse Gases (GHG) emissions and oil dependency.

The EV presents many advantages when compared to vehicles with Internal Combustion Engine (ICE). It has lower energy consumption and emission rates, it is rather silent and it has lower operating costs. However, the EV also presents disadvantages since it have limited autonomy. The long charging times, the limited charging stations and the undeveloped smart grid infrastructures demands for a hard planning of the daily use of the EV. Based on the available energy sources. currently, the EVs cannot compete with the conventional vehicles in terms of driving range and initial cost.

The main goal of the presented work. based on Information and Communication Teclmologies (C) is the interaction between the driver and the IV. This is done aiming planning the top of the FV and then wage optimization, including the energy use and the battery charging process. Through the developed solution, the EV driver can interact with the EV battery charging system using the information of the charging points. This information is related with the charging point position, the availability, the energy price, and the reservation with the associated parking place. The integration of this information associated with driver interaction devices and applications to be developed, will contribute for a better trip planning and energy usage. Consequently, it is reduce the range anxiety of the EV driver. In this research work we describe a personalized range prediction in a mobile application.

**Justification of the Problem**

There are, however, many challenges associated with electric vehicles. They have a limited range. Charging the vehicle takes time and can be a hassle sometimes. The availability of charging stations is also a big issue. Incompatibility of charging stations can also be a problem. Despite many challenges and issues, switching to electric vehicles is good for the environment and is more economically viable in the long term. Many have predicted that, by 2040, most of the vehicles will be electric. Rising fossil fuel costs and high maintenance costs of petrol and diesel vehicles coupled with environmental concerns are the main reasons. Many developed countries have given incentives for purchasing electric vehicles. Automobile manufacturers are already manufacturing some impressive electric vehicles.

The energy cost of manufacturing an electric vehicle is also very high, but considering everything and the fact that charging electric vehicles is very cheap, EVs are a great option. Manufacturing batteries is an important task in the production of Electric vehicles.

**Significance of the Study**

Using more energy efficient vehicles like hybrid and electric vehicles **supports the Indian economy and helps diversify the U.S. transportation fleet**. The multiple fuel sources used to generate electricity results in a more secure energy source for the electrified portion of the transportation sector.

Electric vehicles can reduce fuel costs dramatically because of the high efficiency of electric-drive components. Because all-electric vehicles and PHEVs reply in whole or part on electric power, their fuel economy is measured differently than that of conventional vehicles. [Miles per gallon of gasoline equivalent](https://afdc.energy.gov/glossary.html#MilesperGallonofGasolineEquivalent) (MPGe) and kilowatt-hours (kWh) per 100 miles are common metrics. Depending on how they are driven, today's light-duty all-electric vehicles (or PHEVs in electric mode) can exceed 130 MPGe and can drive 100 miles consuming only 25–40 kWh.

* The study is needful for Electric Vehicle Company to understand what their customers are saying about their products and services to create more personalized Travelling experiences. This can allows EV Company to categorize customers dynamically and experiment with unique, more personally targeted offers in order to ensure customer satisfaction.
* The present study is more helpful to society and government to know best product and Electric Vehicle , so that they can take proper decision while selecting Vehicles for investments.
  1. **Scope of the Problem**

The popularity of electric vehicles (EVs) has significantly increased over the last few years, causing changes not only in the transportation industry but generally in business and society. This paper covers one possible angle to the (r) evolution instigated by EVs, i.e., it provides the data science perspective review of the interdisciplinary area at the intersection of green transportation, energy informatics, and economics. Namely, the review summarizes data-driven research in EVs by identifying two main research streams: (i) socio–economic, and (ii) socio–technical. The socio–economic stream includes research in: (i) acceptance of green transportation in countries and among different populations, (ii) current trends in the EV market, and (iii) forecasting future sales for the green transportation. The socio–technical stream includes research in: (i) electric vehicle battery price and capacity and (ii) charging station management. This kind of study is especially important now when the question is no longer whether the transition from internal-combustion engine vehicles to clean-fuel vehicles is going to happen but how fast it will happen and what are going to be implications for society, governmental policies, and industry. Based on the presented literature review, the paper also outlines the most significant open questions and challenges that are yet to be solved: (i) scarcity of trustworthy (open) data, and (ii) designing a generalized methodology for charging station deployme

* 1. **Statement of the Problem**

The statement of research study is as given herewith. To Electric Vehicle Price using electric vehicle dataset and provide novel predictive analytics for Vehicle Industry to understand their customers and provide Satisfied Price for vehicle. Vehicle Industries are facing lots of problems some are listed here due to digitization and acceptance of new economic policies in India.

1. Increased competition
2. Customer Satisfaction
3. Increased Petrol and Diseal Price
4. Charging Point
5. Attracting new customers and many more.

To improve Price experience and be in competition they need to provide more vehicle Charachteristics experience by knowing their Vehicles by analysis and prediction. Traditional methods are used for this but the problems associated with the traditional data analytics are listed herewith.

1. As the data comes at high speed it is unable to store due to storage capacity.
2. Due to unstructured format unable to transfer, curate and process using traditional data bases.

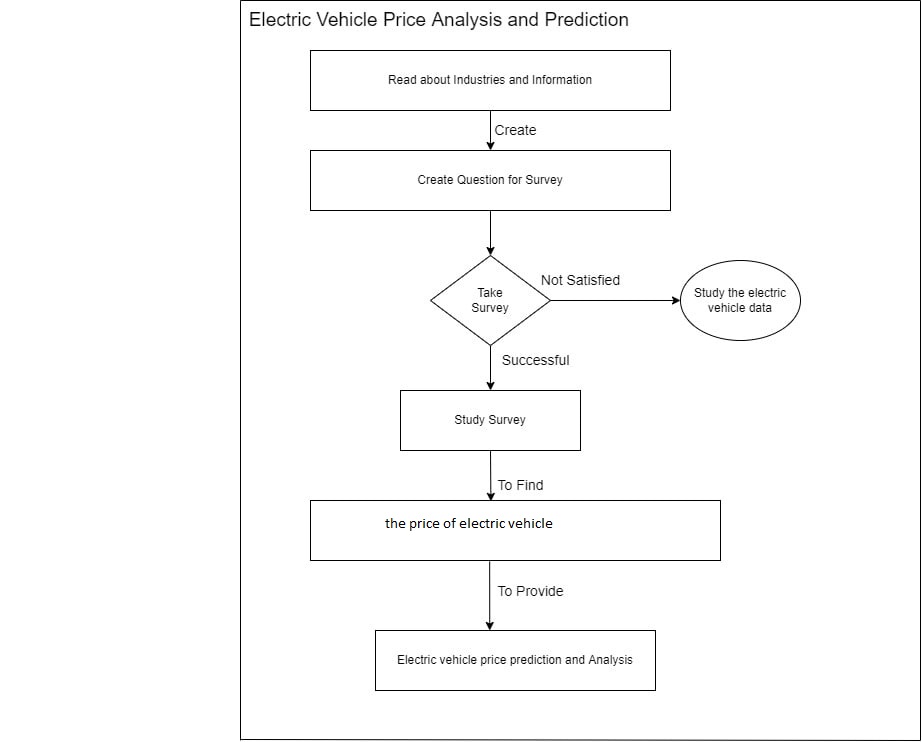
The proposed study helps banks to understand the need of problem and develop a predictive analytics accordingly. The study also gives the novel approach to know their customers and provide them the services as per they required to improve their experience. The study also helps customers to choose bank as per their requirement and convenience.

* 1. **Objectives of the Study**

The researcher has set the following main objectives of the study.

1. To study the need of Electric Vehicle Requirement.
2. To identify relevant Electric Vehicle Dataset and Machine Learning algorithms.
3. To effectively analyze the Electric Vehicle Price .
4. To design and implement Predictive Analytics for EV Price analysis and prediction.
5. To suggest an effective solution for EV Industries to create more Satisfied Market for vehicle.
   1. **Research Methodology**

The present study is an experimental study to identify user Electric Vehicle to provide more personalized experience and develop an inference predictive analytics to analyze and predict Electric Vehicle Price using different components with the help of Linear Regression Modelas described in the following figure 1.1. The study is further extended to check usefulness of the system.



**Figure 1.1: Problem Identification**

* + 1. **Sampling**

According to Gartner survey Electric vehicles are considered to be one of the most important ways for India to solve a series of energy problems such as urban smog, energy supply security and turning to green development. With the powerful support provided by the government, India has become the world's largest electric vehicle market and continues to maintain a high-speed growth. In the table 1.1 why people want to  purchase Electric Vehicle

* + 1. **Data Collection**

Selection of method of data collection is based on nature, scope, availability of money, time and precision factor etc. To carry out the research, researcher has collected secondary data. Researcher has collected massive amount of relevant multi-structured data from various social media sites, product review sites, newspapers, online forms and Vehicles’s websites using web crawler which downloads pages from web servers, indexer which builds a reverse index to the best pages and runtime which answers user’s queries by hitting keywords. Data collected in different formats and some sample data sets are Approximately 100000 comments are collected for each bank.

* + 1. **Research Work Stages**

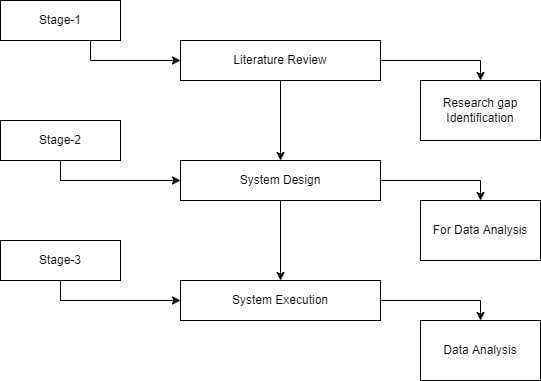
The research experiment is carried out in 4 different stages Literature Review, Pilot Study, System Design and System Execution as shown in Figure 1.2 and elaborated further. The output of previous step is taken as the input for the next step. The secondary data collected in earlier in the previous section is used to get the exact result at every stage.

**Stage-1: Literature Review**

Researcher has carried out literature review in stage 1 to identify the research gap between existing work and present study.

* To understand the theoretical concept of topic.
* To study the frameworks shared by researchers.
* To study the Methodology used by researchers.
* To identify the research gap between current research work and future research work

Researcher has identified, gathered and carried out detailed literature survey. Qualitative and quantitative information is obtained from various literature sources like national and international research journals, conference, thesis and books. Literature review is carried out in various sections like, Need for Personalization, Customer Behavior Analysis, Traditional Analytics and Big Data Analytics for prediction. Each section is concluded with some conclusion at the end. Research gap identified which is included in chapter Number 2.

****

**Figure 1.2: Different Stages of Research Work**

**Stage-2: System Design**

To design the system we have carried out following steps .

1. Customer Comments about home loan service are collected from the official web sites of banks and other social networking sites as mentioned in section 1.9.2.
2. Data is pre-processed by removing noise and converted in CSV format.
3. Small test with 100, 1000 and 10000 comments is performed to identify most relevant big data technologies suitable for our data. Researcher has selected pig and spark framework. For data visualization we have selected D3.js.
4. Researcher has conducted small test with 1000 comments to identify most reliable machine learning algorithms. We have selected k-mean clustering, Decision tree, Naive bays and neural network for building predictive analytics using spark framework.
5. Researcher has proposed pig behavior analytics and developed data analysis algorithm to analyze customer behavior with manual ratings on top of MapReduce.
6. Researcher has developed methodology to generate automatic ratings to analyze customer behavior using AIFFN and pig on top of MapReduce.
7. Hybrid predictive analytics is developed using Pig and Machine Learning algorithms to predict future behavior of customers.

**Stage-4: System Execution**

In this stage we have executed system with large data sets. Collected data is stored in HDFS and cleaned using Pig tool. Then we have selected feature attributes and tested on proposed system. Data set is divided in two sections training set and testing set. In every stage we have increased the size of data. Final results are compared to suggest suitable analytics.

* 1. **Limitations of the Study**

Following are the limitations of the study.

* The present study is limited to analyze and predict Electric Vehicle Price
* The study is confined to analyze only multi structure data not images, videos etc.
* Only classification machine learning algorithms are used to build predictive model.
  1. **Thesis Outline**

The thesis entitled “Electric Vehicle Price Prediction and Analysis” comprises of seven chapters described as follows.

**Chapter 1: Introduction**

In this chapter we have given the overall introduction of topic and research methodology. We have described the need for Electric Vehicle Price Prediction and Analysis. Further we discus about the topic in broad. This chapter briefs about Motivation of the research problem followed Justification of the Topic, Significance of the Topic, Scope of the Topic, Objectives, Hypothesis, Data Collection methods and Research methodology. Limitations of the study and Detailed Chapter Scheme are also mentioned here. The chapter is concluded with chapter summary.

**Chapter 2: Literature Review**

This chapter elaborates the research already carried out in the related area and the future studies are indicated with the review of literature. The chapter provides detail review of research topic with related fields such as need of Electric Vehicle survey, important of Electric Vehicle analysis and prediction, Traditional data analytics etc. The research gap between existing research and present research is identified in detail is mentioned in this chapter.

**Chapter 3: Therotical Framework**

For Electric Vehicle analysis using reviews and comments automatic rating prediction is essential. The large scale data analytics framework elaborates theoretical concepts of various b technologies for data analysis and prediction to discover of hidden behavioral patterns. The frame work focuses on the different steps required for developing predictive analytics and how Hadoop framework and spark are useful to achieve this. This chapter also describes the importance of machine learning algorithms in behavior prediction. .

**Chapter 4: Analysis and Prediction model for Electrical Vehicle Price**

This chapter deals with understanding of Electric Vehicle using Drive type and Number of seats, Acceleration available on . To achieve this goal researcher has first carried out comparative performance analysis of big data technologies . She has proposed Pig Data Analytics Framework to No of Seats and then Drive type Acceleration , Killwatt hours . Due to null data of Price of india , uk this framework is unable to analyze data correctly. To overcome this problem researcher has developed data by fillna Analytics for prediction and data analysis. If the method is applied on a pandas series object, then the method returns a scalar value which is the mean value of all the observations in the dataframe. Machine Learning Algorithms in Chapter 5

**Chapter 5: Experimental Findings and Suggestions.**

Present research is about predicting Electric vehicle price using Machine Learning technologies to provide more personalized experience to them. This chapter presents the findings drawn by the researcher from the work done. To reach the conclusions the researcher has carried out various experiments at various stages starting from suitable technology selections to implementation of predictive analytics. The present research is helpful to the Electric vehicle organizations to know there vehicle and according to the make vehicle according to customer need.

* 1. **Chapter Summary**

This chapter gives the broad introduction about the research carried out. In this chapter we discussed motivation of the research, Origin of the problem, significance and methodology used.. The present research provides organizations with greater opportunities by exposing Electric Vehicle Price and helps to bridge gap between what customers want to do and what they actually do. This information can be useful to make business decisions and improve services for providing more personalized experiences to gain the customer’s satisfaction and maximize profit.

***Chapter 2***

**Literature Review**

**2.1 Introduction**

Whilst different traveller groups have different preferences, most policy incentives to make EVs more attractive are directed to all car drivers, without any particular group to target. Moreover, policy incentives will have implications that are likely to differ between different categories of car drivers. Considering behavioural change as a process, different car drivers are in different stages-of-change towards electric vehicle use. Moreover, they have different attitudes towards sustainable transport and electric vehicles in particular and different needs depending on where they live. All of these aspects may influence EV adoption in general, as well as the responsiveness to policy measures For broad overview few selected references are taken here. The review is broadly divided into three sections as follows,

1. Need for Electric Vehicle.
2. Traditional Analytical and Prediction Systems.
3. Big Data Analytics for Prediction and Behavior Analysis in Banking.

* 1. **Need for Electric Vehicle**

In this section we are focusing importance of Electric Vehicle. Due to the multiple Models, range drive Types helps to customer to choose vehicle. In this article, the effects of policy incentives on EV adoption are investigated, in order to evaluate the effectiveness and efficiency of different possible policy incentives. Socio-psychological constructs from the Transtheoretical Model of Change and the Protection Motivation Theory

Electric vehicles (EVs) can decrease the dependence of the transport sector on fossil fuels, which has environmental advantages. A decrease of local exhaust emissions can increase the air quality and decrease health problems that are related to air pollution (e.g. Vienneau et al., 2015), especially in urban areas. On a global level, EVs may decrease the CO2-emissions related to personal transport, especially if the electricity that is used can be generated using nuclear energy or renewable energy sources such as wind, water and solar power and if many EVs charge in off-peak electricity demand hours (Jochem et al., 2015). However, the uptake of EVs is still relatively low. One of the reasons for the slow uptake of EVs is that these vehicles have a comparatively high investment cost because of the high cost of batteries (Newbery and Strbac, 2015). Therefore, in many countries in Europe policy incentives are provided in order to stimulate the purchase of electric vehicles (e.g. Lieven, 2015).[16]

EVs offer substantial economic and environmental benefits, by substituting grid-based electricity for fossil fuels. Their operating cost is lower. They also reduce greenhouse gas (GHG) and other emissions, enhance energy security, and promote use of renewable energy (Egbue and Long, 2012). EVs have been identified as an important factor in the move toward sustainability in the automotive industry (Maitin and Lacy, 2011). Given these societal benefits, governments at all levels have good reasons to facilitate uptake of EVs. However, creating effective policy requires understanding consumer perceptions and attitudes toward these vehicles (Schuitema et al., 2013).[17]

After obtaining an estimation of EV range, it is started the calculation of the route optimization based on the current position. For route optimization this process may be iterative. This approach be complemented with can personalized one using a driving profile that acts as a training set for a Data Mining (DM) approach to estimate the EV range. The DM approach uses a regression model to find the best fitting estimation. This is done based on the current battery > State-of-Charge (SoC) level, the past driver behaviour (battery SoC level, weather information [18].

**Concluding Remark:** This section concludes that, due to high price of diseal and polutions people are demanding more and switching from fuel vehicle to Electric Vehicle Economic analysis suggests that an electric vehicle (BEV or PHEV) purchased today will have lifetime costs no higher than and likely lower than a conventional vehicle, due to significant fuel cost savings over the vehicle life. But the purchase price of an electric vehicle today is typically higher.

* + 1. **Traditional Analytical Systems in Electric Vehicle**

To provide with satisfied price experience understanding and predicting Electric vehicle is at the top priority. In a research community, a lot of work has been done to verdict on Electric vehicle The original data sets collected have nevertheless suffered from various data corruption problems, such as inconsistency, loss of data and segments, invalid ranges, abnormal patterns, etc. Data preprocessing is therefore crucial prior to any analysis or adoption to ensure that every data set adopted is healthy and consequential. The corrupted data were removed, and subsequently, the corresponding part was fitted by linear interpolation Predictive analytics is a stream of the advance analytics which make predictions about hidden future prize which is concerned with forecasting probabilities and trends. It involves Defining problem, Data collection, Data analysis, Statistics, Data Modeling and Deployment [19][20]. In this section we tried to list out analytical techniques and predictive models developed by other researchers.

* + 1. **Relational Data Base Management Systems**

In 1970 Database Management System (DBMS) was constructed using two approaches mainly hierarchical data model for storing enormous data generated by Apollo space program and then network data model to create a standard database and resolve some of the difficulties of hierarchical model such as inability to represent complex relationships. But both models had disadvantages.

1. For answering even simple query complex programs had to be written.
2. Minimal data independency.

In early 1980’s Relational Database Management System (RDBMS) was developed for commercial use but unable to handle increasingly complex data. So later two new data models had emerged, the Object Relational Database Management Systems (ORDBMS) and Object-Oriented Database Management System (OODBMS) to implement the relational and object data models respectively to represent the third generation of Database Management System [21]. The types of data generated by applications become richer than previous as the volume of data keeps growing. As a result, traditional relational databases are challenged to capture, store, search, transfer, analyse and visualize variety of bulky data. They focus on resolving the complexity of relationships among schema-enabled small amount of data only [22].

* + 1. **Data Mining Techniques and algorithms**

Data Mining is the process of discovering patterns, trend and behavior using Machine Learning Algorithms, Artificial Intelligence and Statistics etc. The data mining process aims to extract valuable information from large dataset for future use. Data Mining is an essential step in predictive analytics and various data mining techniques and algorithms performs data mining and statistical analysis in order to determine trends and patterns in data. Data Mining Process involves Data Cleaning, Data Integration, Data Selection and Transformation, Data Mining, Pattern Evaluation and knowledge representation [45]. Traditional database software based on SQL are not sufficient to discovering increased information Data mining technologies and techniques for recognizing and tracking pattern with in data helps business to know their customers[9]. Many researchers have analyzed data and developed predictive models using data mining techniques [78][79], some of them we have briefed here.

IBM report**” Predictive modeling techniques”** focuses on predictive analytics techniques, including Predictive Model, Neural networks (NNs), clustering, support vector machines (SVMs), and association rulesto convert big data available today into real value. All these techniques identify hidden patterns from large data set of past data and resulting into a predictive model. Once predictive model is validated it is applied to a current situation to predict future [56][59].

Based on their review of EV penetration rate studies, Al-Alawi and Bradley (2013) suggest that three market forecasting techniques are used in the literature: agent-based models (simulation), consumer choice models, and diffusion models (curve-fitting). The relevant agents in market penetration rate studies include: consumers, car manufacturers, policy makers and fuel suppliers. For instance, Driscoll et al. (2013) simulate consumer demand for EVs in Ireland using revealed preference data. Vehicle price, fuel cost, driving range, battery replacement cost, charging time and maintenance cost are among the important attributes used in consumer choice modelling.[21]

The main goal of the presented work, based on Information and Communication Technologies (ICT), is the interaction between the driver and the EV. This is done aiming planning the trip of the EV and their usage optimization, including. the energy use and the battery charging process. Through the developed solution, the EV driver can interact with the EV battery charging system using the information of the charging points. This information is related with the charging point position, the availability, the energy price, and the reservation with the associated parking place. The integration of this information associated with driver interaction devices and applications to be developed, will contribute for a better trip planning and energy usage. Consequently, it is reduce the range anxiety of the EV driver. In this research work we describe a personalized range prediction in a mobile application.[22]

Sweda and Klabjan [30] focus on using an ABM to solve the problem of charging station deployment. This is based on forecasting EV adoption and estimating where charging stations are required the most. Factors that determine an agent’s vehicle purchase decision include vehicle price, fuel cost, the agent’s greenness, social influence, an infrastructure penalty, and a distance penalty. However, the quantification of these parameters was not clearly stated in the paper. We present a detailed EV adoption and usage model that can be used not only EV charging station planners, but by utility operators and policymakers[23].

**Concluding Remark:** The literature review in this section reveals that, for successful Electric Vehicle, Industries should take efforts to know their customers by understanding what they are demanding and what we are giving. This can be possible by using different data analytics techniques. The most popular data analytics systems are RDBMS and Data mining in this category. The traditional data use proper statistical and machine learning methods to analyze huge structure data to discover knowledge. Many researchers have used various data mining process and techniques for data analysis. They have built efficient predictive models and implemented them on small data set of structured data only. The broadly used approaches and techniques are as follows.

1. **Regression Algorithms:** Regression analysis is a statistical process for estimating the relationship among dependent and independent variables. [46]
2. **Association Algorithms:** Association rules are if/then statements that help to uncover hidden relationships between unrelated data in a relational database or other information repository.[47]
3. **Classification Algorithms:** Classification is a data mining function that assigns items in a collection to target categories or classes. [48]
4. **Clustering Algorithms:** Clustering  is a process  which  partitions a given data set  into  homogeneous  groups based on  given  features[49]
5. **Artificial Neural Network Algorithms**: Neural networks are computational structures consisting of an interconnected processing elements (PE) or nodes arranged on a multilayered hierarchical architecture. [50]

Though many traditional data analysis methods efficiently handle structured data, still they can be utilized for big data analysis in the combination with big data technologies [75].

* + - * 1. **Prediction and Analysis using Machine Learning**

V. Dhanalakshmi, D. Bino and A. M. Saravanan in **"Opinion mining from student feedback data using supervised learning algorithms,"** predict opinion mining to find out the positive and negative polarity of the student feedback based on pre-defined features of teaching and learning using supervised learning algorithms. They have developed an application of machine learning with the help of natural language processing techniques on student feedback data gathered from module evaluation survey results of Middle East College, Oman. They have analyzed data using Rapid Miner an open source data analytics tool and conducted comparative study of the algorithms like SVM, Naive Bayes, K Nearest Neighbor and Neural Network classifier to measure the performance. Depending on the results they found that, Neural Network gives the better performance followed by Naive Bayes [27].

Dr. Janardhana and M Manjunath in **”Sentiment Analysis and Opinion Mining using Machine Learning Techniques”**, elaborated that, a leading method for text message analysis is a sentiments or opinion analysis which gives the best results on opinions or sentiments. In this study, authors built the model and study the opinions associated with the movie review data in huge size. Movie reviews with labeled positive opinions are represented by 1 and negative opinions are represented by 0. The machine learning algorithms like Random Forest, Naive Bayes, and SVM are trained with 25,000 labeled movie review data to forecast the opinion associated with unlabeled test data more precisely. This model helpful to determine the customer opinion associated with the unstructured movie review data in digital format on web. The model is completely based on the NLP, Text Analysis, Machine Learning and Statistics [22].

In paper entitled **“Opinion Mining using Machine Learning”** authors R B Ravi Varma, Nagesh Y N, I M Umesh focused on Identifying the subjectivity and objectivity of information as the outcome of Opinion Mining and Sentiment Analysis which may be resulted in either positive or negative or a mix of both. According to authors, behavior identification is possible through sentiment analysis with machine learning. In recent days, opinion mining and sentiment analysis are being used to design innovative marketing strategies. Generated analog data is used to extract subjective contents and predict subjectivity such as positive or negative. This information is useful to build systems to understand customer’s demand and plan new business strategies accordingly. The chances of product failure can also be predicted using this. [36].

B. K. Bhavitha, A. P. Rodrigues and N. N. Chiplunkar in their paper entitled "**Comparative study of machine learning techniques in sentimental analysis"** state that Sentimental Analysis is a great way for determining the notion of people about distinct existence. Recently people are used to review the comments and posts on the product before buying which are known as opinion, emotion, feeling, attitude, thoughts or behavior of the user. According to them, Sentimental Analysis is a method for identifying the user behavior in positive and negative polarity which is expressed in texts. It attempts to divine the attitude or notion of a keynoter against assertive field or any product. [37].

Oscar Liombart in paper entitled **“Using Machine Learning Techniques for Sentiment Analysis”** [33]built test model and evaluate several machine learning methodslike Neural Network, Random Forest and Naive Baysfor the Sentiment Analysis task. The methodology he has used is shown in fig 2.11.



**Figure 2.1: Sentiment Analysis Model**

According to him, a text classification problem is the type of text and the words that we can see in the data because it has an important impact on the number of words that the machine learning methods will learn and, in consequence, the final number of features. Using machine learning they have proved the effect of transformations on the data which can improve the performance of the classification methods but the type of transformations depends on the dataset and the language of the data. For text classification they found Neural Network as a best algorithm.

**Concluding Remark** This data cannot be handling by traditional analytical tools due to their storage capacity and analytical ability but provides valuable insights. Big Data Analytics plays an important role in banking industry to handle this data and provide upcoming trends. Researchers have suggested various big data technologies to develop big data analytics strategy. Many machine learning algorithms are used by the researchers to develop accurate predictive models but still more advancement is required for large amount of multi structured data.

**2.3 RESEARCH GAP ANALYSIS**

A lot of research is still required on Electric Vehicle analysis and prediction using survey and comments available on social networking sites. The study concludes that,

* EV adoption is based on vehicle costs and utility, and EV usage – driving and charging – is dependent on an agent’s daily drive cycle. The results show that a high capacity battery would increase EV adoption slightly, but EVs are still too expensive for a significant increase in adoption .
* As use-based incentives cost less and are relatively effective compared to subsidies or registration tax rebates, these incentives might be more efficient. However, use-based incentives decrease the marginal cost of driving EVs and might cause rebound effects. At the same time, providing use-based incentives to car users gives a certain message favouring car use, be it electric car use, which governments might not intend to send.
* We use numerical case studies to demonstrate the benefits of our proposed pricing scheme, showing that it can find a nearsocially optimal equilibrium. While the price-only mechanism can as well, its performance is highly sensitive to the choice of the regularization penalty term, ξ. If ξ is not chosen properly, PEV charging costs in the price-only case can be considerably higher than with the price/quantity mechanism.
* Sources of EV-related data nowadays exist, but are still scarce. The most common way of acquiring the data is either through the cooperation with private companies or through proprietary devices Electronics 2019, 8, 1190 24 of 30 developed for the research purpose, what presents a major obstacle for producing high-quality data science research in the EV area even though the amount of data being generated from EVs and charging stations is growing every day.
* As a result, traditional analytics are challenged to capture, store, search, transfer, analyze and visualize variety of data.
* focus on resolving the complexity of relationships among schema-enabled small amount of data only.
* For large amount of multi structure data these traditional predictive analytics have limited scope.

From the study, it has been also observed that, many researchers have developed various methodologies for predicting and analyzing reviews and comments left on social networking sites. Many researchers have carried out sentiment analysis or opinion mining using comments and reviews using data mining techniques.

**Chapter Summary**

As local policy incentives decrease the marginal cost of EV-use rather than their fixed cost, those policy measures are of little help for the group of car buyers that can currently not afford to buy an EV. On the other hand, they stimulate EV-use by decreasing the absolute marginal cost as well as the marginal cost relative to other transport modes. Within the range limitations of EVs, EV trips will be more attractive than other transport modes.. Predictive Analytics help to find the gap between what predicted price want and actually the Industries give. To carry out the above study, we have gone through various related and admirable research papers, books, thesis, projects etc. In this chapter we tried to understand the different techniques, algorithms and tools used to predict customer behavior and concluded the research gap.

* 1. **REFERENCES**

**Research Papers:**

1. Chris Rygielski, Jyun-Cheng Wang , David C. Yen(2002), Data mining techniques for customer relationship management, Technology in Society 24 (2002) 483–502, www.elsevier.com/locate/techsoc
2. Arne Mauser, IljaBezrukov, Thomas Deselaers, Daniel KeysersLehrstuhlf¨urInformatik (2004), Predicting Customer Behavior using Naive Bayes and Maximum Entropy – Winning the Data-Mining-Cup 2004” VI, Computer Science Department RWTH Aachen University, D-52056 Aachen, Germany
3. J. Dean, S. Ghemawat(2004), MapReduce: Simplified Data Processing on Large Clusters, In Proc. of the 6th Symposium on Operating Systems Design and Implementation, San Francisco CA, 137 -149
4. Arun Mittal and Priya Mittal (2009), Consumer Behavopr Aspects and Promotion of Banking Services.
5. AbhijeetRaorane, Dr. R.V. Kulkarni(2011), Data Mining Techniques: A Source For Consumer Behavior Analysis, Cornell university Library,arXiv:1109.1202
6. P.Isakki alias Devi, S.P.Rajagopalan (2012), Analysis of Customer Behavior using Clustering and Association Rules Customer Behavior Prediction using Artificial Neural Network, International Journal of Computer Application(0975-8887), vol 43
7. Jinbao Zhu, Principal Software Engineer, and Allen Wang, Manager, Software Engineering, CA Technologies(2012), Data Modeling for Big Data
8. Stephen Kaisler, Frank Armour and Alberto Espinosa(2013), Big Data: Issues and Challenges Moving Forward , 46thHawaiInternatinal Conference on System Sciences, DOI 10.1109/HICSS.2013.645
9. Wei Fan Huawei, Albert Bifet (2013), “Mining Big Data: Current Status, and Forecast to the Future”, SIGKDD Explorations, volume 14, Issue 2 ,2013, Pp 1 -5
10. JosephO. Chan (2013), An Architecture for Big Data Analytics,communications of the IIMA, Volume 13 Issue 2, <https://scholarworks.lib.csusb.edu/ciima/vol13/iss2/1>
11. Dr. A. C. Pandey, Mithilesh Kumar Pandey(2013), Impact of Lifestyle on Brand Preference of Buyer Behaviour A Case Study of Uttarakhand International Monthly Refereed Journal of Research In Management & Technology 151 Volume II, ISSN – 2320-0073
12. Abdul Brosekhan, Dr. C. MuthuVelayutham (2013), Consumer Buying Behaviour – A Literature Review IOSR Journal of Business and Management (IOSR-JBM) e-ISSN : 2278-487X, p-ISSN : 2319-7668, PP 08-16 [www.iosrjournals.org](http://www.iosrjournals.org)
13. Nada Elgendy and Ahmed Elragal ICDM (2014), Big Data Analytics: A Literature Review Paper, LNAI 8557, pp. 214–227, 2014. © Springer International Publishing Switzerland
14. Pinki Rani (2014), Factors Influencing Consumer Behavior, Excellent Publishersa, Int. J. Curr. Res.Aca. Rev. 2014;2(9);52-61). 2014
15. Dr.PolonaTomic, Ms. RichaPandit (2014), Study of the Impact of Service Quality on Consumer Behavior in Internet Banking Services, PARIPEX- Indian Journal of Research, Vol. 3, Issue 1
16. Dr. Siddaraju1, Sowmya C L, Rashmi K, Rahul M (2014), Efficient Analysis of Big Data Using Map Reduce Framework, IJRDET, Vol 2 , Issue-6
17. Mr. V. Vivekanandan Ms. N. Karpagavalli(2014), Efficient Data Analysis Scheme For Increasing Performance In Big DataIJAICT, Vol 1, Issue 2
18. X. Jin, Benjamin W. Wah, Xueqi Cheng and Yunzhuo Wang (2015), Significance and Challenges of Big Data Research, Big Data Research, ScienceDirect, Elsevier, 59-64
19. UtkarshSrivastavaa and Santosh Gopalkrishnan(2015), “Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks”,Procedia Computer Science 50, 643-652, DOI 10.1016/j.procs.2015.04.098
20. A. Chandan, M. Mehta, B. Neeraja and Om Prakash (2015), Banking On Big Data: A Case Study ,ARPN Journal of Engineering and Applied Sciences, Vol 10
21. Aftab Yaseen1 , P. M. Khan2(2015), Big Data Analytics – An Overview Of Research Opportunities And Challenges, International Journal of Advance Research inScience and Engineering, Vol 4
22. D R, JanardhanaMulimani, Manjunath (2015). Sentiment Analysis and Opinion Mining using Machine Learning Techniques, International Journal of Innovative Research in Computer and Communication Engineering. 3297. 9321/9329
23. J. Ramsingh, Dr. V. Bhuvaneswari (2015), An Insight on Big Data Analytics Using Pig Script, IJETTCS, Volume 4, Issue 6, ISSN 2278-6856
24. ArushiJaina, Vishal Bhatnagara(2015), “Crime Data Analysis Using Pig with Hadoop”, International Conference on Information Security & Privacy (ICISP2015), 11-12, Nagpur, INDIA, Science Direct,Elsevier
25. Krati Bansal, Priyanka Chawla (2016), “A Study of Big Data Analysis Using Apache Pig” , I J C T A*,* 9(17) 2016, pp. 8665-8672 © International Science Press
26. A. Shrivastva, Chandan Kumar, NehaMangla (2016), Analysis of Diabetic Dataset and Developing Prediction Model by using Hive and R, Indian Journal of science and Technology, Vol 9(47), ISSN 0974-6846
27. V. Dhanalakshmi, D. Bino and A. M. Saravanan(2016), "Opinion mining from student feedback data using supervised learning algorithms," *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, pp. 1-5.  
    doi:10.1109/ICBDSC.2016.7460390
28. AninditaKhade (2016), Performing Customer Behavior Analysis using Big Data Analytics , 7th International Conference on Communication, Computing and Virtualization, ScienceDirect, procedia Computer Science 79(2016)986-992
29. Ms.SarikaRathi(2017) ,”A brief Study of Big Data Analytics using Apache Pig and Hadoop Distributed File System” ,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)Volume 6, Issue 1
30. Pooja Jain, Prof. Jay Prakash Maurya (2017), “Comparative Analysis Using Hive and Pig on Consumers Data”, International Journal of Computer Science and Information Technologies, Vol. 8 (2) , 285-291
31. Anand Gupta, H.Thakur,(2017) “A Big Data Framework Using Apache Spark and Deep Learning”, arXIV:1711.09279v1.
32. A.Baltas, A. Karnavos, A. Tsakalidis (2017), “An Apache Spark Implementation for Sentiment Analysis on Twitter Data”, Springer International Publication, pp. 15-25, DOI: 10.1007/978-3-319-57045-7-2
33. Oscar Liombart (2017), “Using Machine Learning Techniques for Sentiment Analysis”, School of Engineering
34. WararatSongpan (2017), “The Analysis and Prediction of Customer Review Rating Using Opinion Mining” IEEE, Computer Society, 978-1-5090-5756-6/17,
35. Poonam Sawant, Dr. AbhijeetKaiwade (2017), “Pig: Novel Approach to Analyze Customer Behavior in Banking”, IPASJ International Journal of Computer Science (IIJCS), Volume 6, Issue 2,, Page No. 43-49-111
36. R B Ravi Varma , Nagesh Y N , I M Umesh , Pradeep N C (2017) , “Opinion Mining using Machine Learning” , International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) Volume VI, Issue XII, ISSN 2278-2540
37. B. K. Bhavitha, A. P. Rodrigues and N. N. Chiplunkar (2017), "Comparative study of machine learning techniques in sentimental analysis," *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, ,pp.216-221. doi: 10.1109/ICICCT.2017.7975191
38. A Hasan, S. Moin, A. Karim, S. Shamshirband(2018), “Machine Learning Based Sentiment Analysis for Twitter Accounts”, Mathematical and Computer Application

**Links:**

1. Banking in Western India getting customer centric -<http://economictimes.indiatimes.com/industry/>banking/finance/banking/banking-in-western-india-getting-customer-centric/articleshow/2498342.cms
2. It’s Time For Personalization in Financial Services By Matthew Lifshotz, Director of Global Business Developmenthttps://thefinancialbrand.com/37391/bank-personalization-product-development/
3. Personalization in Banking: From Novelty to Necessity By [Jim Marous,](http://jimmarous.com/) Co-Publisher of The Financial Brand and Publisher of the [Digital Banking Report](http://www.onlinebankingreport.com/)https://thefinancialbrand.com/58063/personalization-in-banking-digital-research-study/
4. “Personalization: The secret ingredient to customer engagement for banks” May 6, 2016http://www.the-future-of-commerce.com/2016/05/06/customer-engagement-banks/
5. Dept. of Financial Services Ministry of Finance, Government of India
6. <http://www.predictiveanalyticstoday.com/what-is-predictive-analytics/>
7. <https://en.wikipedia.org/wiki/Predictive_analytics>
8. <http://www.comp.dit.ie/btierney/oracle11gdoc/datamine.111/b28129/regress.htm>
9. <http://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining>
10. <https://en.wikipedia.org/wiki/Classification>
11. <https://en.wikipedia.org/wiki/Cluster_analysis>
12. <https://en.wikipedia.org/wiki/Artificial_neural_network>
13. www.zeroprotraining.com
14. <http://www.edureka.co/blog/the-hype-behind-big-data/>
15. [www.sakal.com](http://www.sakal.com), page No.7, 4/02/2015
16. <http://www.cloudnetworkinnovations.com/hadoop-analytics-for-financial-services.html> Hadoop analytics For Financial Services Clients
17. <https://www.mapr.com/solutions/industry/big-data-and-apache-hadoop-financial-services> Big Data and Apache Hadoop for Financial Services
18. <http://www.predictiveanalyticstoday.com/big-data-analytics-and-predictive-analytics/>
19. http://hadoop.apache.org/docs/r1.2.1/hdfs\_design.html, Retrieved 2015.
20. http://hortonworks.com/hadoop/yarn/, Retrieved 2015.

**Articles:**

1. IBM report” Predictive modeling techniques” Capgemini Consulting (2014), Big Data Alchemy: How can Banks Maximize the Value of their Customer Data?
2. Victoria Zagorsky (2014), Unlocking the Potential of Big Data in the Banking Sector
3. JostHoppermann and Martha Bennett ( 2015), Big Data in Banking: It’s time to act
4. TomGroenfeldt (2013), More Personalized Banking through Big Data and Analytics, slideshare.neti
5. Big Data Analytics Delivers Customer Centricity for Banks,Bloomberg Business week (2015)
6. Challenges and Opportunities with Big Data A community white paper developed by leading researchers across the United States Executive Summary
7. Big Data Planning Guide for Financial Services Elizabeth Dias
8. Predictive Modeling With Big Data(2012) IBM
9. V. Agniswaram Big Data Analytics Beyond Hadoop, dezon.com
10. Changing the game with big data using big data and analytics to build today’s financial business.
11. Big Data at the bank Jon Ginovsky ABA Banking journal
12. Big data has big implications for knowledge management

**Books:**

1. Consumer Behavior in Indian Perspective by Suja R. Nair.
2. Hadoop the definitive guide by Tom White, O’REILLY
3. Digital Banking Consumer Demand Hyper Personalization 2017,Digital Banking Report thefinacialbrand.com
4. RajendraAkerkar, Sajja(2016), Intelligent Techniques for Data Science, Springer, ISBN 978-3-319-29205-2
5. A. Bagha,V. Madisetti, cloud computing: a handson approach (University Press India Private Limited).

**Thesis:**

1. Large Scale Data Analytics of User Behavior for Improving Content Delivery, AthulaBalachandran, CMU-CS-14-142, December 2014
2. Mining Frequent Patterns In Data Streams, K.S.Oza, 2010

***Chapter 3***

**Theoretical Framework**

* 1. **Introduction**

In recent years we have seen the dramatic increase in the growth of information due to collection of data from various applications and services. Very popular example is internet data which is growing at lightning speed. The information stored on the net is huge in amount. All over the world approximately millions of bytes are added in every second. The data and knowledge on the net is growing at an amazing rate by proving estimates wrong in every second. This massive amount of data is called as a Big Data which is a collection of large, complex or required data which becomes difficult or impossible to process, analyze and store using existing tools, standard database management and analytical solutions. This huge amount of data is very valuable to improve quality of life and make our world at a better place by extracting meaningful associations, trends and patterns which can be possible using big data analytics. In this chapter we have given detailed introduction about big data technologies we used for Prediction and Analysis

* 1. **Python**

Python is a very popular general-purpose interpreted, interactive, objectoriented, and high-level programming language. Python is dynamically-typed and garbage-collected programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL).

* 1. **Pandas**

Pandas is an open-source library that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library. Pandas is fast and it has high performance & productivity for users.

* 1. **Numpy**

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It is open-source software.

* 1. **Matplotlib**

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.

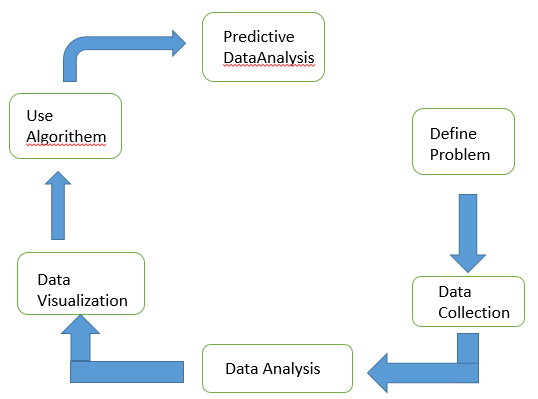
One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

* 1. **Seaborn**

Seaborn is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on the top of [matplotlib](https://www.geeksforgeeks.org/python-introduction-matplotlib/) library and also closely integrated to the data structures from [pandas](https://www.geeksforgeeks.org/introduction-to-pandas-in-python/).  
 Seaborn aims to make visualization the central part of exploring and understanding data. It provides dataset-oriented APIs, so that we can switch between different visual representations for same variables for better understanding of dataset.

* 1. **Predictive Analytics**

To analyze customer behavior and provide more personalized experience predictive analytics is a greater way. Predictive analytics is a stream of the advance analytics which make predictions about hidden future events from past occurrences. Predictive analytics is concerned with forecasting probabilities and trends. Predictive analytics involved Defining problem, Data collection, Data analysis, Statistics, Data Modeling and Deployment as shown in the figure 3.1. Predictive analytics plays an important role in prominence alongside the emergence of big data where enterprises store larger and broader pools of data in Hadoop clusters and other big data platforms. Therefore, predictive analysis can be done using big data technologies and Machine Learning Algorithms [1][2].

****

**Figure 3.1: Predictive Analytics Process**

* 1. **Machine learning algorithms**

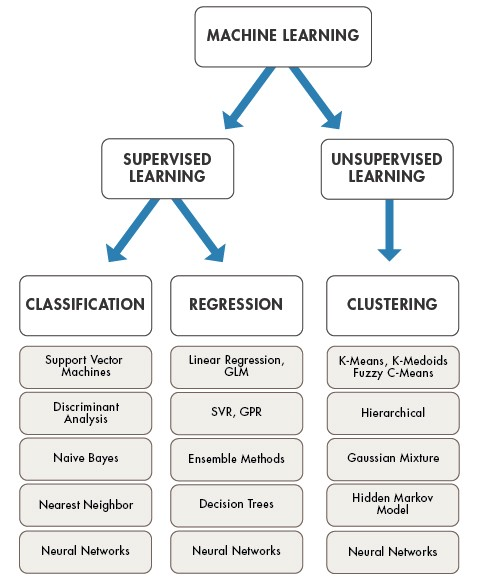
Machine learning provides systems to automatically learn and improve the experience without programmed. It is an application of artificial intelligence (AI) that focuses on the development of computer programs which can access data for learning. It is categorized as supervised learning and unsupervised learning. Classification of Machine Learning Algorithms is shown in following diagram[10][11].

* + 1. **Supervised Learning**

Machine learning uses supervised learning most frequently than unsupervised learning. In supervised learning we have input variables (x) and an output variable (Y) and an algorithm to learn the mapping function from the input to the output in such a way that if you have new input data(x) you can predict the output variables (Y).

Y = f(X)

It is called as supervised learning because the procedure of algorithm learning from the training dataset is similar to a learning process supervised by a teacher. Supervised learning problems can be further categorized into regression and classification problems. Some popular examples of supervised machine learning algorithms are; Naïve Bays, Neural Network, Linear Regression, Random Forest, Decision Tree and Support Vector machines [10].



**Figure 3.14: ML Algorithms**

(Source https://www.mathworks.com/discovery/machine-learning.html)

* + 1. **Unsupervised Learning**

Unsupervised learning does have only input data (X) and no corresponding output variables. The main aim of the unsupervised learning is to construct the model for the underlying structure or distribution in the data in order to know more about the data. In unsupervised learning there is no correct answers and supervision like teacher. Algorithms handles the things their own to discover the knowledge and present the interesting structure in the data. Unsupervised learning problems can be further divided into clustering and association problems. Mainly used algorithms are k-mean clustering and Apriori algorithm [10].

* 1. **.Linear Regression Model:**

For types of prediction analysis, linear regression is commonly used. The goal of regression is to look at two different impacts. Can a group of predictor factors perfectly predict the outcome of an outgrowth value?Which factors, in particular, are significant predictors of the outgrowth variable, and how do they influence the outgrowth variable – as indicated by the size and sign of the beta estimates? The link between one dependent variable and one or more independent variables is explained using these retrogression estimations. The simplest version of the retrogression equation with one dependent and one independent variable is

1. ***y = c b \* x,***

where y is the estimated dependent variable score, c denotes the constant, b denotes the regression measure, and x denotes the score on the retrogression measure.

* 1. **Chapter Summary**

For Electric Vehicle analysis using reviews and comments automatic rating prediction is essential. The large scale data analytics framework elaborates theoretical concepts of various b technologies for data analysis and prediction to discover of hidden behavioral patterns. The frame work focuses on the different steps required for developing predictive analytics and how Hadoop framework and spark are useful to achieve this. This chapter also describes the importance of machine learning algorithms in behavior prediction.

***Chapter 4***

**Analysis and Prediction model for Electrical Vehicle Price**

* 1. **Introduction**

In today’s world, there is a huge amount of climate change we are observing. Different committees are established on a world level for getting the solution. The electric vehicle is one of the solutions for tackling Climate change. Eclectic vehicles are the next-generation cars that everyone will utilise in their daily lives. Which is good for our health and environment. As EV will be in use more so mass production of these vehicles are in progress. Many automobile companies are launching their EV in the market. Many different features are given by them which decide the cost of the vehicle. Battery life, Speed, Acceleration, performance, safety features, and many factors are there to calculate the cost of the EV. We decided to refer to different research papers focusing on specific features of EV and on the basis of this basis of this we can predict the cost of the EV using Machine learning.

The science of teaching computers to perform tasks without being told what to do is known as machine learning. Machine Learning is a subset of Artificial Intelligence, which is the most often used word in the late 20th century. We provide computers with artificial intelligence so that they can do tasks independently. They work with extraordinary precision and speed. We develop a model and then train it via machine learning. Machine learning is divided into three forms that are widely used across the world.

* 1. **Survey in Electric Vehicle**

With global electric vehicles (EVs) sales expected to boom in the next 12 months, nearly 90 per cent of consumers in India are willing for buying an EV, according to a survey . Consumers are willing to pay extra for an added value of being environmentally responsible.

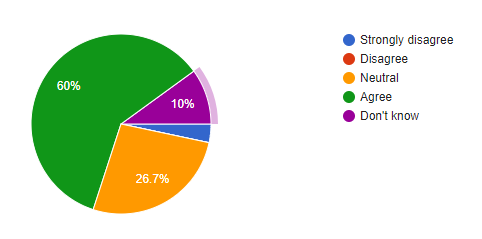


Figure 4.1 Environmental concern

According to the survey, the top reason for buying an EV is environmental concern, with 86.7 per cent also stating that the COVID-19 pandemic has heightened awareness and concerns about environmental issues.

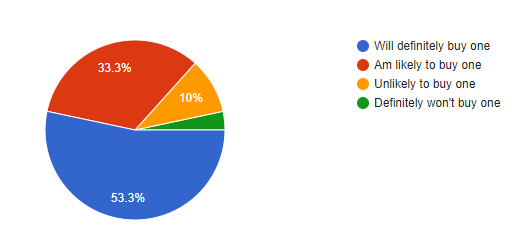


Figure 4.2 Buying one vehicle

About 86.9 per cent of those looking to buy an EV feel it is their responsibility to reduce their personal environmental impact, and 53.3 per cent feel buying an EV is one way to achieve this goal.

* 1. **Analytics Framework for Behavior Analysis with Manual Ratings**

**4.3.1Data pre-processing**

The data set collected from the internet sites was consists of noise and incomplete data, using pre-processing techniques noisy and inconsistent data is reduced and converted into CSV format. In the consideration of nature of data, in the initial stage, researcher has used multiple data sets individually to draw conclusions.

**4.3.2Data Analysis Algorithm**

**Input:** Input file in CSV Format

**Output:** Electric Vehicle Price

**Begin**

**Step-1:** Import the Libraries

**Step-2:** Checking the Column list of table

**Step-3:** Replacing Currency sign and changing Datatypes

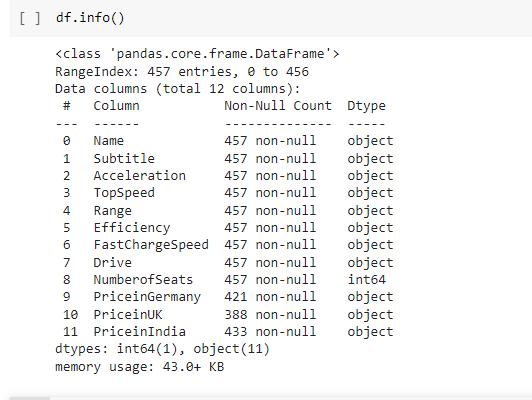
**Step-4:** Fill the Null Value

**Step\_5:** Visualize the Data

**End**

**4.3.3Experimental Results**

1. **List of Evaluation Parameters and concluding Remarks**

****

**Figure 4.3: Evaluation Parameters**

|  |  |
| --- | --- |
| **Sr. No.** | **Column Name** |
| 1 | Subtitle |
| 2 | Acceleration |
| 3 | Top Speed |
| 4 | Range |
| 5 | Efficiency |
| 6 | Fast Charge Speed |
| 7 | Drive |
| 8 | Number Of Seats |
| 9 | Price in Germany |
| 10 | Price in UK |
| 11 | Price in India |

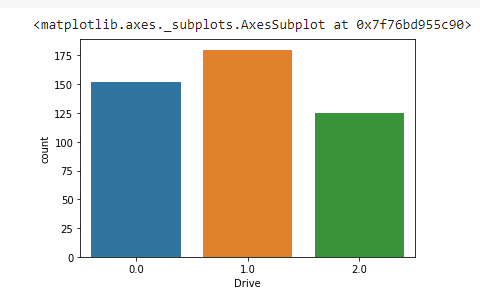
**Table 4.2: Evaluation Parameters and Concluding Remarks**

**Interpretation:** Figure 4.3 and table 4.2 give an overview of various evaluation parameters which can be used for analysis of behavior; it is observed that, the different evaluation parameters have same concluding remarks. On the basis of this parameters Electronic Vehicle Price going to determine

1. **Vehicle Drive count**

|  |  |  |
| --- | --- | --- |
| **Sr.no** | **Drive Type** | **Count** |
| 1 | Front Wheel Drive | 150 |
| 2 | All Wheel Drive | 175 |
| 3 | Third Wheel Drive | 125 |

**4.3 Types of Drive**

****

**4.4 Types Of Drive Graphical**

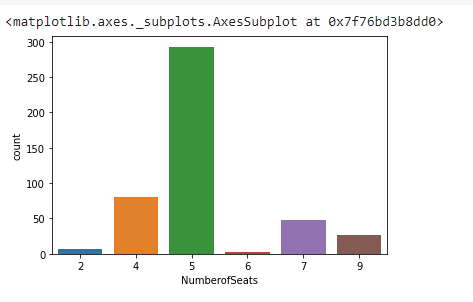
**Interpretation:**

Figure 4.3 and table 4.4 shows the Drive type according to number of it. Figure 4.4 gives the graphical interpretation of Drive concludes that, the highest drive type according to graphical data All Wheel Drive.

**3 No of Seats In Vehicle**

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Number Of Seats** | **Count** |
| **1.** | **2** | **7** |
| **2** | **4** | **80** |
| **3** | **5** | **293** |
| **4** | **6** | **3** |
| **5** | **7** | **48** |
| **6** | **9** | **26** |

**4.5 Number Of Seats In vehicle**



4.6 Number of seats

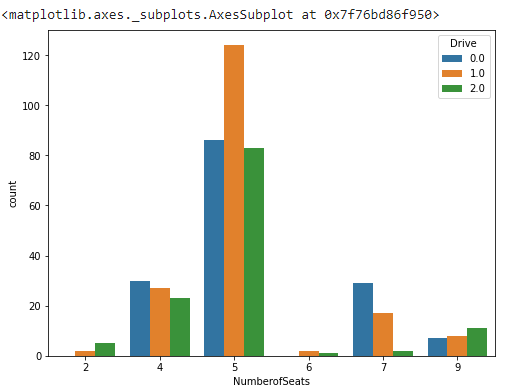
**Interpretation:**

Table 4.5 and figure 4.6 give a count of Number Of Seats in Electric Vehicle . It Observed that 5 seats have high number 243 . and low Number of seats is 6 it only have 3 count of Electric Vehicle

**4 Numbers of Seat and Drive**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **No Of Seats** | **Drive** | | |
| **Front Wheel Drive** | **All Wheel Drive** | **Third Wheel Drive** |
| **1.** | **2** | **0** | **2** | **5** |
| **2** | **4** | **27** | **25** | **22** |
| **3** | **5** | **85** | **130** | **80** |
| **4** | **6** | **0** | **3** | **2** |
| **5** | **7** | **30** | **17** | **3** |
| **6** | **9** | **5** | **6** | **8** |

**4.7 Drive and No of seats**

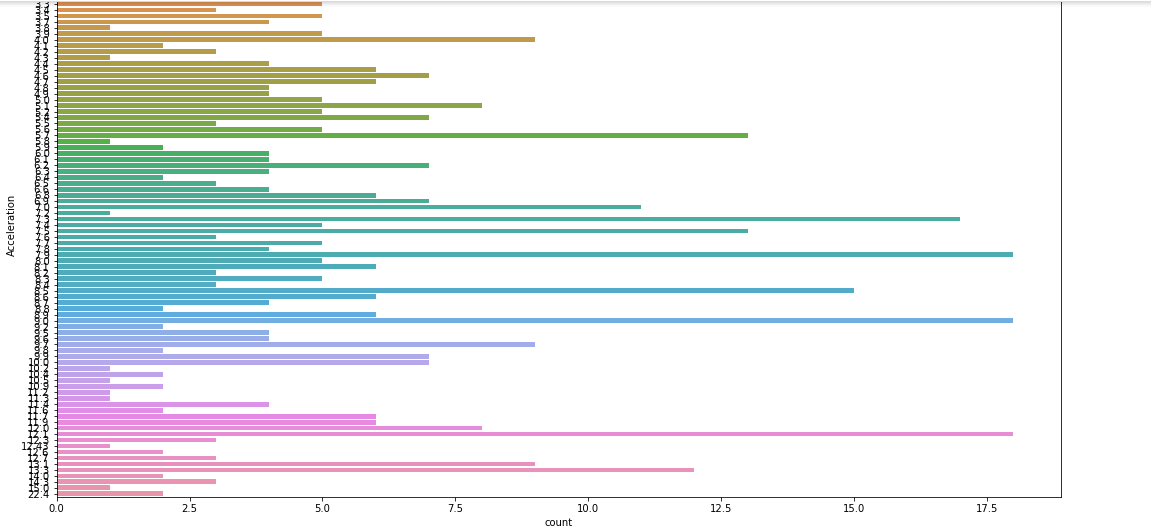
****

**4.8 Number of seats and drive**

**Interpretation:**

Figure 4.7 and table 4.8 shows the number of seats and drive . 5 No of Seats most requirement or manufactured in Market with all Drive type.

**5.Acceleration Count**

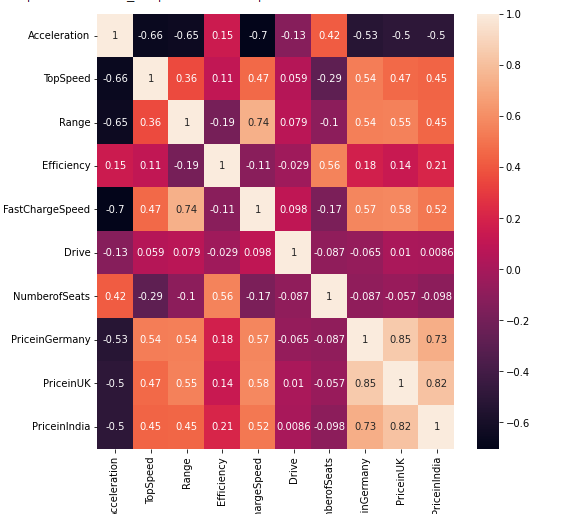
****

**Fig 4.9 Acceleration**

**Interpretation:**

In fig 4.9 As you can there arr no of acceleration in a dataframe . high level acceleration is start with 12 . middle level Acceleration is the7.5 to 12 acceleration. Low level Acceleration is the 1.0 to 7.5

**6. Coreleation between the Data with each other**

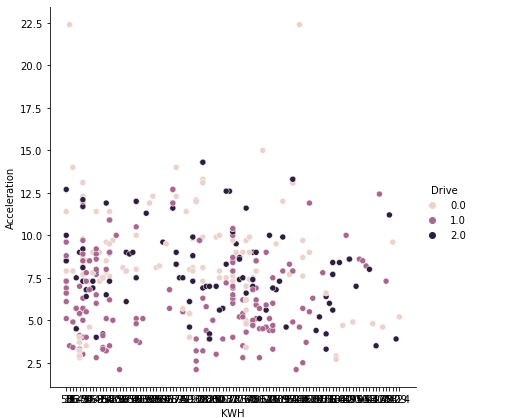
****

**Fig 4.11 heatmap of DataFrame Column**

**Interpretation:**

Table 4.10 and Figure 4.11 is used to show the correlation between each dataframe .It is important to discover and quantify the degree to which variables in your dataset are dependent upon each other. Heatmap is defined as a graphical representation of data using colors to visualize the value of the matrix.

**7. Relational Graph of Acceleration and Kilowatt Hours, Drive**

** Fig 4.12 Acceleration and kilowatt and drive**

**Interpretation:**

In fig 4.12 there is relation between acceleration , kilowatt hours and Drive type with each other easy and understandable way. To easy understanding and easy to use

**8 Fast Charge Speed and Efficiency**

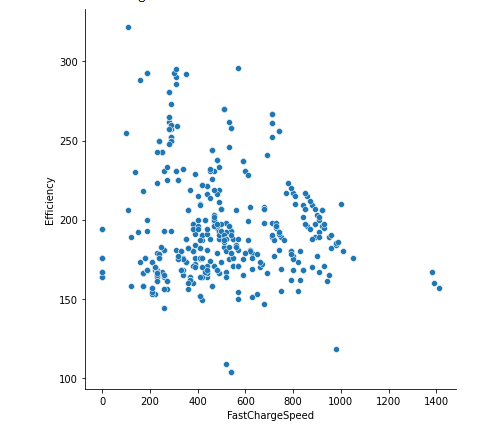


Fig 4.13 Fast Charge Speed and Efficiency

**Interpretation:**

Fig 4.13 It is showing the relation between the Fast Charge and Efficiency. In the forms of dot in same color

* 1. **Chapter Summary**

This chapter deals with understanding of Electric Vehicle using Drive type and Number of seats, Acceleration available on . To achieve this goal researcher has first carried out comparative performance analysis of big data technologies . She has proposed Pig Data Analytics Framework to No of Seats and then Drive type Acceleration , Killwatt hours . Due to null data of Price of india , uk this framework is unable to analyze data correctly. To overcome this problem researcher has developed data by fillna Analytics for prediction and data analysis. If the method is applied on a pandas series object, then the method returns a scalar value which is the mean value of all the observations in the dataframe. Machine Learning Algorithms in Chapter 5.

***Chapter 5***

**Experimental Findings and Suggestions**

* 1. **Introduction**

The problem is identified through the study of literature reviewed in chapter 2 and based on the secondary data collected predictive analytics is designed and implemented in chapter 5. This chapter presents experimental findings of the research and suggestions given. The present study has been carried out in various stages to analyze and predict Electric Vehicle Price from the Price in india and Drive Type and no of Seats an effective solution to the Electric Vehicle Price using machine learning algorithms.

* 1. **Experimental Findings**

To serve the purpose we have carried out some experiments. Experimental findings are listed in below categories.

* + 1. **Selection of Technologies**

To select the suitable technologies researcher has carried out literature review and done performance analysis. Findings are as follows.

* We found Python as a suitable tool along with Data Science Programming model to serve as to clean and process unstructured data in the form of SQL Though . Python is a high level language and requires minimum lines of code .
* We found Spark framework is providing fast execution environment to build a predictive model with Machine Learning algorithms . Its 10 time faster than Hadoop framework.
* We found Decision Tree, Neural Network, and Naive Bayes algorithms suitable for our research for correct predictions and further analysis.
* Data Analytics is not complete without visualization. So we found D3.js for visualize data of Electric Vehicle
  + 1. **Price Analysis with Machin Learning**

After selection of technologies Python Electric vehicle price Analytics is proposed and implemented in chapter 4 to analyze Electric Vehicle price using no of components using like No of seats, Acceleration , drive type , Price in india

* In Using data frame we got to know there are different type of Parameters on the basis of that Electric vehicle price going to Predict for the future use
* We have analyzed that parameter like acceleration , kilowatt hours , price ,Fast charge speed , no of seats , drive types ,Efficiency , ratings and find out that
  + All wheel drive type have high data compare to the front wheel drive and the third wheel drive
  + In 65% Electric vehicle have the 5 no of seats in the vehicle campartive vehicle . second is the 9% have the 4 no of seats in Electric vehicle.
  + People are more prefer to buying 5 -4 no of seats electric vehicle.
* Comparing two components to each other by using graphical method we got to know what is relation or affecting each other
* All wheel Drive with no of seats is 5 have the high Prefer in the Electric vehicle market . In some no of seats there is not front wheel drive is available in the wlwctric vehicle
* 7.0 , 9.0 and 12.0 have the high Acceleration count upto the 18
* Front wheel drive type have high type of acceleration compare to all wheel Drive and rear wheel drive with no of seats is the 5
* High range is the 900 and it is in All wheel drive .No of vehicles range between the 200 to 500 range.
  + 1. **Electric Vehicle Price Prediction and Analysis with Predictive Analytics**

Here we have build hybrid Model to analyze and predict customer behavior. Automatic data is generated by identifying frequently used price which resulted in decreasing missing error rate between prediction which concludes as follows.

* A scatter plots are a form of mathematical or graphic diagram that displays values for two variables for a collection of data using Coordinates. One and many variables can be presented if the points are coded
* Fundamental purpose of a scatter plot are to observe and display relationships of two numeric variables. When looking at the data as a whole, the dots in a scatter plot reflect not just the values of individual data points, but also patterns
* the correlation between all the columns with other columns, dark color in the heatmap shows the lowest values and lighter shows highest values.
* There we can see many correlations tending +1 which means they are perfect positive correlation in which both variables move in the same direction together.
* Also there are some points which are tending to -1 or perfect negative correlation, meaning that as one variable goes up, the other goes down.
* Int the scatter point graph we can see the points are in one line and many points are collaging on each other which means out Actual price of the EV Vehicle and predicted values are almost same which is a good result.
* We got 1.0 R Square value for linear and lesso model and for Random Forest mode we got 0.99 which means our dataset is a better fit for the model.
* Mean Absolute Error(MAE), Mean Squared Error(MSE), Root Mean Squared Error(RMSE) is in the range 1 to 4 which are lower values indicate better fit.
  1. **Suggestions**

Based on the conclusions drawn in the previous sections the suggestions are listed below in two sections case based and analytics based.

* + 1. **Case Based Suggestions**

From the conclusion drawn in the section 6.2, the researcher has suggested following case based suggestions as follows,

* 1. No doubt, that no of components affect to the price prediction of vehicle because of that data should be clean and analyzed properly for the prediction of electric vehicle price
  2. The prediction model is good enough to predict the price of Electric Vehicle
  3. It is ready for deployment. However, there are still space of improvement in the future. For example Body seats, Model, Rapid Charge, Power Train, etc.
  4. Brand is an important features generated from one of the column in the dataset (Brand).
  5. Another Variables important is Acceleration, Fast Charge, Top Speed and Range.
     1. **Analytics Based Suggestions**

From the conclusion drawn in the section 6.2, the researcher has suggested following case based suggestions as follows,

1. The study reveals that, to predict data linear regression is the one of the best solution.
2. The study suggests that, pig on the top of python is the best choice for extract, transform and load.
3. There are the other ways to predict the data using different algorithms
4. The present study suggests that, data science is the powerful technique to visualize data in web application format one can use.
   1. **Contribution Made to The Body of Knowledge and Outcome of Research.**

The present study is confined to predict Electric vehicle price by analyzing given data of Acceleration , efficiency , range . Till date various methods and techniques are used to predict customer behavior. In the present research we tried to develop our won hybrid methodology to predict customer behavior. Our research work is able to

* Help Elctric vehicle company to bridge gap between what customers want and what they actually get in market.
* Help customer to buy new car by comparing the electric vehicle price.
  1. **Scope for Further Study**

Research is an unbroken and never-ending process. Improvement and enhancement can be done continuously. It was a very good experience to work on this research topic. Always there is a research exposure, as the new dimensions of research are introduces. By considering such dimensions future work can be carried out to find out hidden exposure.

1. In the present research, the emphasis is given on building predictive model for predicting Electric vehicle price using different algorithm with better performance.
2. The present study is carried out on stand-alone machine. The same research can be carried out on multiple machines.
3. The present study analyzes and predicts the behavior about home Electric vehicle charging station. further it can be extended to study other services.

iv The study can be also extended to build model using unsupervised and Lexicon based approach.

* 1. **Chapter Summary**

Present research is about predicting Electric vehicle price using Machine Learning technologies to provide more personalized experience to them. This chapter presents the findings drawn by the researcher from the work done. To reach the conclusions the researcher has carried out various experiments at various stages starting from suitable technology selections to implementation of predictive analytics. The present research is helpful to the Electric vehicle organizations to know there vehicle and according to the make vehicle according to customer need.

**Bibliography**

**Thesis**

1. Large Scale Data Analytics of User Behavior for Improving Content Delivery**,** Athula Balachandran, CMU-CS-14-142, December 2014
2. Mining Frequent Patterns In Data Streams, K.S.Oza, 2010
3. Machine Learning For Data Mining In Medicine, M. Solomon Pushparaj, 2008

**Books**

1. Consumer Behavior in Indian Perspective by Suja R. Nair.
2. Hadoop the definitive guide by Tom White, O’REILLY
3. Digital Banking Consumer Demand Hyper Personalization 2017,Digital Banking Report thefinacialbrand.com
4. Rajendra Akerkar, Sajja(2016), Intelligent Techniques for Data Science, Springer, ISBN 978-3-319-29205-2
5. Big Data Analytics with R and Hadoop

**Research Papers**

[1] Ümit Agbulut, Ali Etem Gürel, Yunus Biçen “Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison” 2021 Renewable and Sustainable Energy Reviews 135 (2021) 110114

[2] Raphael Couronné, Philipp Probst & Anne-Laure Boulesteix “ Random forest versus logistic regression: a large-scale benchmark experiment ” 2018 SpringerLink.

[3] Gulden Kaya Yanik, Nese Guler , “A Study on Multiple Linear Regression Analysis” 2013 The Authors. Published by Elsevier Ltd

[4] Rahim Alhamzawi, Haithem Taha Mohammad Ali, “The Bayesian adaptive lasso regression ” June 2018 ScienceDirect Mathematical Biosciences Volume 303.

[5] Kanwal Noor et.al. “Vehicle Price Prediction System using Machine Learning Techniques” International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 201.

[6] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric “ Car Price Prediction using Machine Learning Techniques” TEM Journal. Pages 113-118, February 2019

[7] Sameerchand Pudaruth “Predicting the Price of Used Cars using Machine Learning Techniques”, 2014 International Journal

**Links**

1.An empirical analysis of electric vehicle cost trends: A case study in Germany

https://www.sciencedirect.com/science/article/abs/pii/S2210539522000463#!

2.Electric Vehicles in India: Market Analysiswith Consumer Perspective,Policies and Issueshttps://www.researchgate.net/publication/325801124\_Electric\_Vehicles\_in\_India\_Market\_Analysiswith\_Consumer\_PerspectivePolicies\_and\_Issues

3.https://www.researchgate.net/publication/327103592\_Comparison\_of\_Electric\_and\_Conventional\_Vehicles\_in\_Indian\_Market\_Total\_Cost\_of\_Ownership\_Consumer\_Preference\_and\_Best\_Segment\_for\_Electric\_Vehicle

4.Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach https://link.springer.com/article/10.1007/s12626-022-00109-9

5. Electric Vehicles Gain Ground but still Face Price, Range, Charging Constraints

Michael D. Plante and Sean Howard https://www.dallasfed.org/research/economics/2022/0222

6. https://theicct.org/sites/default/files/publications/EV\_cost\_2020\_2030\_20190401.pdf

**ANNEXURE**

**Annexure A**

**Bibliography**

**Thesis**

1. Large Scale Data Analytics of User Behavior for Improving Content Delivery**,** Athula Balachandran, CMU-CS-14-142, December 2014
2. Mining Frequent Patterns In Data Streams, K.S.Oza, 2010
3. Machine Learning For Data Mining In Medicine, M. Solomon Pushparaj, 2008

**Books**

1. Consumer Behavior in Indian Perspective by Suja R. Nair.
2. Hadoop the definitive guide by Tom White, O’REILLY
3. Digital Banking Consumer Demand Hyper Personalization 2017,Digital Banking Report thefinacialbrand.com
4. Rajendra Akerkar, Sajja(2016), Intelligent Techniques for Data Science, Springer, ISBN 978-3-319-29205-2
5. Big Data Analytics with R and Hadoop

**Research Papers**

[1] Ümit Agbulut, Ali Etem Gürel, Yunus Biçen “Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison” 2021 Renewable and Sustainable Energy Reviews 135 (2021) 110114

[2] Raphael Couronné, Philipp Probst & Anne-Laure Boulesteix “ Random forest versus logistic regression: a large-scale benchmark experiment ” 2018 SpringerLink.

[3] Gulden Kaya Yanik, Nese Guler , “A Study on Multiple Linear Regression Analysis” 2013 The Authors. Published by Elsevier Ltd

[4] Rahim Alhamzawi, Haithem Taha Mohammad Ali, “The Bayesian adaptive lasso regression ” June 2018 ScienceDirect Mathematical Biosciences Volume 303.

[5] Kanwal Noor et.al. “Vehicle Price Prediction System using Machine Learning Techniques” International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 201.

[6] Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric “ Car Price Prediction using Machine Learning Techniques” TEM Journal. Pages 113-118, February 2019

[7] Sameerchand Pudaruth “Predicting the Price of Used Cars using Machine Learning Techniques”, 2014 International Journal

**Links**

1.An empirical analysis of electric vehicle cost trends: A case study in Germany

https://www.sciencedirect.com/science/article/abs/pii/S2210539522000463#!

2.Electric Vehicles in India: Market Analysiswith Consumer Perspective,Policies and Issueshttps://www.researchgate.net/publication/325801124\_Electric\_Vehicles\_in\_India\_Market\_Analysiswith\_Consumer\_PerspectivePolicies\_and\_Issues

3.https://www.researchgate.net/publication/327103592\_Comparison\_of\_Electric\_and\_Conventional\_Vehicles\_in\_Indian\_Market\_Total\_Cost\_of\_Ownership\_Consumer\_Preference\_and\_Best\_Segment\_for\_Electric\_Vehicle

4.Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach https://link.springer.com/article/10.1007/s12626-022-00109-9

5. Electric Vehicles Gain Ground but still Face Price, Range, Charging Constraints

Michael D. Plante and Sean Howard https://www.dallasfed.org/research/economics/2022/0222

6. https://theicct.org/sites/default/files/publications/EV\_cost\_2020\_2030\_20190401.pdf

**Annexure B**

**Sample Data Set**

**Program Code**

**EvProject.py**

**#import Libraries**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.linear\_model import Lasso

from sklearn import metrics

**#import CSV file**

df = pd.read\_csv('EVDatabase.csv',encoding='cp1252')

df

**#Read the data**

df.info()

df.dtypes

**# Replacing Currency sign and changing Datatypes**

df['Efficiency']=df['Efficiency'].astype(str)

Efficiency=[]

for item in df['Efficiency']:

     Efficiency+=[int(item.replace(' Wh/km',''))]

df['Efficiency']=Efficiency

df['Range']=df['Range'].astype(str)

Range=[]

for item in df['Range']:

    Range+=[int(item.replace(' km',''))]

df['Range']=Range

df['FastChargeSpeed']=df['FastChargeSpeed'].astype(str)

FastChargeSpeed=[]

for item in df['FastChargeSpeed']:

    FastChargeSpeed+=[int(item.replace(' km/h','').replace('-','0'))]

df['FastChargeSpeed']=FastChargeSpeed

df['TopSpeed']=df['TopSpeed'].astype(str)

TopSpeed=[]

for item in df['TopSpeed']:

    TopSpeed+=[int(item.replace(' km/h',''))]

df['TopSpeed']=TopSpeed

df['Acceleration']=df['Acceleration'].astype(str)

Acceleration=[]

for item in df['Acceleration']:

    Acceleration+=[float(item.replace(' sec',''))]

df['Acceleration']=Acceleration

df['PriceinGermany']=df['PriceinGermany'].astype(str)

PriceinGermany=[]

for item in df['PriceinGermany']:

    PriceinGermany+=[item.replace('€','').replace(',','')]

df['PriceinGermany']= list(map(float,PriceinGermany))

df['Subtitle']=df['Subtitle'].astype(str)

Subtitle=[]

for item in df['Subtitle']:

    Subtitle+=[item.replace('Battery Electric Vehicle | ','').replace(' kWh','').replace('      ','')]

df['Subtitle']=Subtitle

df['PriceinIndia']=df['PriceinIndia'].astype(str)

PriceinIndia=[]

for item in df['PriceinIndia']:

    PriceinIndia+=[item.replace('?','').replace(',','')]

df['PriceinIndia']=list(map(float,PriceinIndia))

df['PriceinUK']=df['PriceinUK'].astype(str)

PriceinUK=[]

for item in df['PriceinUK']:

    PriceinUK+=[item.replace('£','').replace(',','')]

df['PriceinUK']=list(map(float,PriceinUK))

**#Checking the Null Values**

car\_dataset=df.copy()

car\_dataset.isnull().sum()

**#Checking fill Value**

car\_dataset['PriceinIndia']=car\_dataset['PriceinIndia'].fillna(car\_dataset['PriceinIndia'].mean())

car\_dataset['PriceinUK']=car\_dataset['PriceinUK'].fillna(car\_dataset['PriceinUK'].mean())

car\_dataset['PriceinGermany']=car\_dataset['PriceinGermany'].fillna(car\_dataset['PriceinGermany'].mean())

car\_dataset.isnull().sum()

**#Encoading Drive Type**

car\_dataset.replace({'Drive':{'All Wheel Drive':1,'Front Wheel Drive':0,'Rear Wheel Drive':2}},inplace=True)

car\_dataset['Drive']=list(map(float,car\_dataset['Drive']))

df=car\_dataset.copy()

car\_dataset.head(170)

#**Visulize the Data**

print(df.Drive.value\_counts())

sns.countplot(x = 'Drive', data = df)

print(df.NumberofSeats.value\_counts())

sns.countplot(x = 'NumberofSeats', data = df)

plt.figure(figsize=(8,6))

sns.countplot(x = 'NumberofSeats', hue='Drive', data=df)

plt.figure(figsize=(18,10))

sns.countplot(y = 'Acceleration', data = df)

plt.figure(figsize=(8,8))

sns.heatmap(car\_dataset.corr(), annot=True)

sns.relplot(x="KWH", y="Acceleration", height=6,hue="Drive",data=df)

sns.relplot(x="KWH", y="Acceleration", size="NumberofSeats", height=6,sizes=(15, 100),data=df)

sns.relplot(x="TopSpeed", y="Range",height=6, hue="Drive",data=df)

sns.relplot(x="FastChargeSpeed", y="Efficiency", height=6,data=df)

X = df.drop(['Name','PriceinGermany','PriceinUK'],axis=1)

Y = df['PriceinIndia']

#car\_dataset.to\_csv("newev.csv")

 X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size = 0.1, random\_state=2)

**#Linear Regression Model**

lin\_reg\_model = LinearRegression()

lin\_reg\_model.fit(X\_train,Y\_train)

pred = lin\_reg\_model.predict(X\_test)

pred

lin\_reg\_model.score(X\_test, Y\_test)

training\_data\_prediction = lin\_reg\_model.predict(X\_train)

training\_data\_prediction

error\_score = metrics.r2\_score(Y\_train, training\_data\_prediction)

print("R squared Error : ", error\_score)

plt.scatter(Y\_train, training\_data\_prediction)

plt.xlabel("Actual Price")

plt.ylabel("Predicted Price")

plt.title(" Actual Prices vs Predicted Prices")

plt.show()

test\_data\_prediction = lin\_reg\_model.predict(X\_test)

error\_score = metrics.r2\_score(Y\_test, test\_data\_prediction)

print("R squared Error : ", error\_score)

plt.scatter(Y\_test, test\_data\_prediction)

plt.xlabel("Actual Price")

plt.ylabel("Predicted Price")

plt.title(" Actual Prices vs Predicted Prices")

plt.show()