“A STUDY ON SENTIMENT ANALYSIS OF THE TWO-WHEELER ELECTRIC VEHICLE USERS IN INDIA”

By

SAKSHI CHATURVEDI
107220437008

Under the supervision of
MRS. KALYANI GORTI

Submitted in partial fulfillment
For the award of the degree of

BACHELORS OF COMMERCE (HONOURS BUSINESS ANALYTICS)

Bharatiya Vidya Bhavan

Bhavan’s Vivekananda College
of Science, Humanities & Commerce
Autonomous College -Affiliated to Osmania University
Sainikpuri, Secunderabad-500094
2022 - 2023

DECLARATION

I hereby declare that this project entitled **"A STUDY ON SENTIMENT ANALYSIS OF THE TWO-WHEELER ELECTRIC VEHICLE USERS IN INDIA"** has been prepared by me in partial fulfillment of the requirement for the award of Degree, Bachelor of Commerce (Hons. Business Analytics).

I also hereby declare that this project report is the result of my own effort and that it has not been submitted to another university or institution for the award of any other degree or diploma.

Place : Hyderabad

Date :

Sakshi Chaturvedi

<u>CERTIFICATE</u>

This is to certify that the project report titled **“A STUDY ON SENTIMENT ANALYSIS OF THE TWO-WHEELER ELECTRIC VEHICLE USERS IN INDIA”** submitted in partial fulfillment for the award of B. Com (Hons. Business Analytics) course of Department of Commerce, Bhavan’s Vivekananda College of Science Humanities & Commerce, Sainikpuri, Secunderabad was carried out by **Sakshi Chaturvedi** under my guidance. This has not been submitted to another university or institution for the award of any other degree, diploma, or certificate.

Name of the Guide: Mrs. Kalyani Gorti Signature

Place of the Guide: Hyderabad

ACKNOWLEDGEMENT

I take this opportunity to express my deep sense of gratitude and indebtedness to **Dr G.S.V.R.K. Choudary,** Principal, Bhavan's Vivekananda College of Science, Humanities and Commerce and **Mrs. L.V. Kamala Devi,** Head, Department of Commerce for their co-operation extended to me in the preparation of this project work.

I sincerely thank the project coordinators **Dr. P.R. Venugopal and Mrs. Kalyani Gorti** for their support. I am profoundly thankful to my project guide **Mrs. Kalyani Gorti** for her valuable guidance and suggestions for the successful completion of my project.

I would also like to thank all other faculty members of Department of Commerce for giving valuable suggestions and guidance.

Sakshi Chaturvedi

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

A STUDY ON SENTIMENT ANALYSIS OF THE TWO-WHEELER ELECTRIC VEHICLE USERS IN INDIA

1. INTRODUCTION

In today's society, when new technology is being used everywhere on a daily basis, electric vehicles must be the transportation of the future. Electric vehicles are being promoted for a number of reasons, including pollution, rising fuel demand, global warming, and the promotion of environmentally benign modes of transportation. An electric vehicle is a type of electric-powered transportation (EV). As opposed to conventional automobiles and trucks that utilise a gasoline (petrol) or diesel-powered engine, electric cars and trucks employ an electric motor that is powered by electricity from batteries or a fuel cell.

Electric vehicles are crucial because they do not require non-renewable energy sources; instead, electricity may be produced using solar, hydro, biomass, thermal, wind, etc. Electric vehicles use electric engines and need a battery and electricity to operate. It is inevitable to turn on conventional sources of energy due to the rising rates of pollution and global warming. The world has been continuously exploited by our reliance on oil. The time has come to abandon our reliance on oil. Due to the hazardous gases, it releases into the environment, which are toxic to both flora and fauna, oil is a major source of pollution. We may reduce the release of dangerous gases into the environment by employing non-conventional energy sources like electricity in our vehicles. By 2030, India would require 1 crore electric vehicles, say experts. Infrastructure development is required for charging stations, service stations, etc. Electric vehicles are also economical, as is well known. To charge them, all they need is an electrical outlet and Electric Supply. Utilizing more electric vehicles can lessen our reliance on foreign nations for petroleum, consequently boosting our economy. An electric vehicle does not require engine oil changes, which contributes to a better future for our children and the sustainable growth of our nation. A strong push for greener mobility needs to be made.

India is one of many nations that have committed to removing petrol and diesel-powered internal combustion engine (ICE) vehicles from the road. Electric vehicles (EV), which are more efficient than ICE choices, emit no emissions, and require no maintenance, will replace these cars. There are currently numerous electric vehicles on the road, and more are considering going completely electric as fuel prices in India soar. However, the COVID-19 pandemic has had an impact on the Indian automobile business due to supply chain delays, the shutdown of manufacturing facilities as a result of ongoing lockdowns, travel bans across the nation, and consumer financial hardship.

The demand for electric vehicle (EV) markets, however, increased. Due to several government initiatives and policies, it is anticipated to expand quicker during the projection period. To encourage the EV industry, the government has also offered a number of incentives to both consumers and manufacturers. In addition to the general people, businesses and governments are becoming more interested in EVs. To lessen their carbon footprint, e-commerce companies (like Amazon) are starting to adopt e-Mobility for last-mile delivery. Electric intercity buses have been introduced in many major cities by the Government of India and several other state governments as part of its e-Mobility public transportation experimentation. India has the greatest untapped market in the world, particularly for electric two-wheelers. Since this industry is open to 100 percent foreign direct investment through the automatic route, the market is anticipated to grow during the projection period. The two-wheeler market offers purchasers a variety of options, however the four-wheeler electric category has yet to gain traction. You have the freedom to operate an electric bike without looking for a petrol station. And these days, that helps you save a lot of money. Consider brands that are meeting the needs of all riders in the EV market, such as Ola, Ather, Revolt, Joy, Tork, Cybor, etc.

Since the start of 2022, specific automakers have started to introduce hybrid and electric vehicle (EV) models in a variety of vehicle segments—2Ws, 3Ws, and 4Ws—and customers can now buy these models. Additionally, a handful of private businesses offer e-2Ws/3Ws to those who transport things. A tiny percentage of e-2Ws and e-3Ws can be spotted on city roads; e-4Ws are extremely uncommon. According to market trends, more automakers intend to enter the EV market. India is currently in the take-off phase in this regard. However, compared to ICE vehicles, the selection of EV models currently for sale is constrained. This lessens the likelihood that customers will consider EVs as their next vehicle.

There have been some instances where EVs have caught fire, which has generated concerns about its safety. An electric vehicle caught fire in a Mumbai suburb in June 2022. In another incident the same month, an electric scooter that was plugged into a charger caught fire in the city of Patan in north Gujarat.

After learning about these incidents, the national government launched an investigation, which turned up mistakes made by some manufacturing firms, including the creation of inadequate battery management systems, the absence of suitable heat release venting mechanisms, and the use of subpar battery cells. Another analysis suggests that lithium-ion batteries have issues. These have to do with toxicity, logistical difficulties, short lifespans, flammability, underperformance, and overheating.

The Ministry of Road Transport and Highways has created a new set of EV battery safety requirements that EV manufacturers must adhere to. These concern battery cells, battery management systems, on-board chargers, battery pack designs, and heat spread brought on by internal cell short circuits. The regulations will go into effect on December 1, 2022.

The distance that an electric vehicle (EV) can travel while its battery is completely charged depends on a variety of variables, including battery capacity, driving speed, the weight of the EV, and the outside temperature. Batteries with a higher capacity can store more energy and have a greater driving range. For instance, certain e-scooters with batteries that have a 3.97 kWh capacity may travel up to 181 kilometres on a single charge. While others have a 452 km range, certain e-cars having battery capacities of 39.2 kWh.

Due to the lack of public battery charging stations nationwide and the fact that it takes longer to charge an EV battery than it does to refuel an ICE vehicle with gasoline or diesel, many drivers, especially those who travel long distances, are uncomfortable with the idea of frequently charging EV batteries.

Electric vehicles (EVs) can be charged in two different ways: by plugging into a wall outlet, an AC wall box charger placed at home, or a DC charger put in a public space. The amount of power supplied by the charger determines how long it takes to charge using either of these techniques. Home chargers require more time (six to 19 hours) than public chargers (more than an hour) to fully charge a device.

The cost of charging time is increased by two factors. First, there is a difficulty with effective transmission and distribution, despite the fact that India produces enough power to meet demand. This leads to power interruptions, which are frequent in India and can be challenging for EV users. Second, the lack of public charging facilities in towns and along highways discourages consumers from purchasing EVs.

Toxic elements like manganese, nickel, lithium, and cobalt are found in lithium-ion batteries used in EVs. The right infrastructure and processes would be necessary for the safe disposal of end-of-life batteries with the anticipated increase in the number of EVs. Such battery disposal facilities would be valuable for various electronic products powered by lithium-ion batteries in addition to meeting the needs of the EV sector.

As India works to hasten the adoption of electric vehicles, there will be a rise in the need for electricity to charge EV batteries. India currently uses coal as its primary energy source for producing power. When coal is used to create power, hazardous pollutants are released that have an impact on the environment and people's health. Acid rain, fog, haze, respiratory and lung conditions, as well as neurological and developmental harm to both humans and animals, are some of the problems that result. Therefore, it will be important to draw on safer and cleaner sources in order to meet future electricity demand.

EV sales are still behind expectations despite rising public and private sector investments in the sector. The delayed uptake of retail lending to assist individuals and institutions in financing EVs is one of the issues seen in this regard. According to a 2022 study by Niti Aayog, RMI, and RMI India, financial institutions are reluctant to lend because of the risks involved, including the quality of the product and the unpredictability of its resale value.

The technical expertise of workers, R&D capabilities, ancillary auto components, backward linkages (with metal industries, capital equipment, trucking, warehousing and logistics), linkages with dealership, retail, credit and financing, repair and maintenance, and other issues affect the proper development and growth of India's EV industry.

The general public still has a relatively limited grasp of electric vehicles (EVs), including their advantages, risks, available subsidies, methods and prices for charging, battery life, maintenance expenses, and resale value. The majority of customers still favour ICE vehicles since the most popular media, including radio, television, newspapers, and magazines, do not adequately distribute information. The Niti Aayog's November 2021 introduction of the e-Amrit web site will be an important development in this area. The portal offers a wealth of knowledge on EVs and works to raise awareness among a range of stakeholders, including consumers, businesses, and service providers.

1.1 NEED FOR THE STUDY

The demand for transportation is rising proportionately to the dramatic daily population growth in India. Consequently, there is a rise in gasoline demand. Traditional automobiles emit too much smoke, which contributes to air pollution and the yearly deaths of many people. The sector for electric scooters is growing far too quickly right now. Due to their low maintenance requirements, electric scooters and bikes are the most popular new vehicle choices. As a result, businesses must be aware of customer expectations in order to assist customers in purchasing an EV.

The need for the study to understand customer expectations and preferences regarding the purchase of electric scooters and bikes in India. The demand for transportation is rising due to the increasing population, which has resulted in a corresponding rise in gasoline demand. However, traditional automobiles are a major source of air pollution and are responsible for many deaths each year. Therefore, there is a need for more environmentally friendly alternatives, such as electric scooters and bikes. The sector for electric scooters and bikes is growing quickly due to their low maintenance requirements, which makes them a popular choice among consumers. As a result, businesses need to understand customer expectations and preferences to assist customers in making informed decisions about purchasing an electric vehicle.

Overall, the need for the study is driven by the increasing demand for transportation in India and the need for more sustainable and environmentally friendly alternatives. The study aims to provide insights into customer attitudes and preferences towards electric scooters and bikes, which can help businesses develop more effective marketing strategies and improve customer satisfaction.

1.2 OBJECTIVES OF THE STUDY

1. To explore the user's sentiments about two-wheeler electric vehicles in India through Logistic Regression model.

2. To analyze the user's sentiments about two-wheeler electric vehicles in India through Decision Tree Classifier model.
3. To explore the user's sentiments about two-wheeler electric vehicles in India through Random Forest classification model.
4. To analyze the users' sentiments about two-wheeler electric vehicles in India through the Support Vector Machine classification model.
5. To explore the user's sentiments about two-wheeler electric vehicles in India through Naive Bayes classification model.
6. To analyze the user's sentiments about two-wheeler electric vehicles in India through Gradient Boosting Classifier model.
7. To identify the best model for predicting the user's sentiments about two-wheeler electric vehicles in India.

1.2 RESEARCH METHODOLOGY

Sources of data:

Data is scraped from various sites from Google (bikewale.com, bikedekho.com) using Web Scraper extension.

Methods of Analysis:

1. Natural Language Processing

Natural language processing (NLP) is a field situated at the convergence of data science and Artificial Intelligence (AI) that – when reduced to the basics – is all about teaching machines how to comprehend human dialects and extract significance from the text. This is additionally why Artificial Intelligence is regularly essential for NLP projects.

2. Sentiment Analysis

Sentiment analysis, often known as opinion mining, is a technique used in natural language processing (NLP) to determine the emotional undertone of a body behind a text. This is a common method used by organizations to identify and group ideas regarding a certain good, service, or concept. Text is mined for sentiment and subjective information using data mining, machine learning, and artificial intelligence (AI).

Sentiment analysis systems help organizations gather insights from unorganized and unstructured text that comes from online sources such as emails, blog posts, support tickets, web chats, social media channels, forums, and comments. Algorithms replace manual data processing by implementing rule-based, automatic or hybrid methods. Rule-based systems perform sentiment analysis based on predefined, lexicon-based rules while automatic systems learn from data with machine learning techniques. A hybrid sentiment analysis combines both approaches. In addition to identifying sentiment, opinion mining can extract the polarity (or the amount of positivity and negativity), subject and opinion holder within the text. Furthermore, sentiment analysis can be applied to varying scopes such as document, paragraph, sentence and sub-sentence levels.

3. Python

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented, and functional programming. Python programming is used for the analysis in the study.

4. Classification models

Classification models are a type of machine learning algorithm used for predicting the categorical class labels of new instances based on their input features. They are commonly used in a wide range of applications, such as image recognition, speech recognition, natural language processing, fraud detection, and customer segmentation.

The goal of classification models is to learn a mapping between the input features and the output class labels based on a set of labeled training data. The training data consists of input features and their corresponding class labels, and the model learns the relationship between the features and the labels using various techniques such as decision trees, neural networks, logistic regression, and support vector machines.

Once the model is trained, it can be used to predict the class labels of new instances based on their input features. The model takes the input features as input and outputs the predicted class label. The performance of a classification model is typically evaluated using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve.

Types of classification models to be used in the study:

a. Logistic Regression: Logistic regression is a statistical model used for binary classification problems, where the goal is to predict the probability of a binary outcome, such as success or failure, based on a set of input variables. It is a form of regression analysis that uses the logistic function to model the relationship between the input variables and the binary outcome.
b. Decision Tree Classifier: Decision Tree Classifier is a machine learning algorithm that can be used for classification problems. It works by creating a tree-like model of decisions and their possible consequences. The decision tree model consists of nodes that represent decision points and edges that represent the possible outcomes of each decision.
c. Random Forest: Random Forest is an ensemble learning model that uses multiple decision trees to improve the accuracy of classification. Random Forest works by training several decision trees on different subsets of the data and aggregating the results to make the final prediction.
d. Support Vector Machine: Support vector machine is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.
e. Naïve Bayes: Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering. This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods.
f. Gradient Boosting Classifier: Gradient Boosting Classifier is also a machine learning algorithm used for classification problems. It works by combining multiple weak classifiers to form a strong classifier that can make accurate predictions. Gradient boosting is an iterative approach, where each weak classifier is trained to correct the mistakes of the previous classifiers.

1.4 SCOPE OF THE STUDY

The study's main objective is to analyze customer attitudes toward electric vehicles in India to provide valuable insights into the factors that influence the adoption of EVs in the country. By using data from the popular websites from Google, the study aims to gather a diverse range of opinions and perspectives from a large sample of Indian drivers.

The study focusses specifically on two-wheeler users as two-wheelers are a popular mode of transportation in India and therefore have a significant impact on the overall adoption of EVs. The inclusion of several companies in the study such as Ola Electric, Benling, Ather, TVS iCube, Revolt, Bajaj Chetak, Hero Electric, Okinawa, Okaya, and Ampere also provides a comprehensive overview of the current players in the EV market in India.

1.5 LIMITATIONS OF THE STUDY

Accuracy issues in training models are often the source of sentiment analysis difficulties. Systems typically struggle with objectivity and are prone to misidentifying comments with a neutral sentiment. When systems are unable to comprehend the context or tone, sentiment can be difficult to detect as well. It's possible for someone to make contradicting claims. The majority of evaluations will contain both positive and negative feedback, which can be slightly managed by analyzing each sentence separately. But the more informal the medium, the more probable it is that humans will mix up diverse points of view in one statement, making it harder for a computer to comprehend. In some cases, the distribution of sentiment labels in the training data may be imbalanced, with one label being more prevalent than the others (neutral sentiments in this case). The sampling method may result in data biases and issues with data quality and accuracy because the data were gathered from Indian motorcycle websites.

CHAPTER II

REVIEW OF LITERATURE

2. REVIEW OF LITERATURE

A literature review is an essential component of any research study, as it provides an overview of previous research on the topic and identifies gaps, inconsistencies, and areas for further investigation. In the case of the current study, the literature review would need to cover previous research on people's attitudes towards motorcycles and the factors that influence their decision-making process when purchasing a motorcycle.

1. Felisia Handayani and Metty Mustikasari [2020] conducted a study titled ‘sentiment analysis of electric cars using recurrent neural network method in Indonesian tweets.’ This study used sentiment analysis of tweets on Twitter that contained the keyword "electric car." One of the pre-study steps and the data retrieval process using a Python library were Tweet Scraper. The method used to construct models is the Recurrent Neural Network, which has a Long Short Term Memory architecture. The final test outcomes demonstrated that datasets with 3000 data and a 70:30 ratio between training and testing data generated the best accuracy results. These datasets' precision, recall, and accuracy values—0.618, 0.507, and 0.722—indicate that adding more data improves results and that the distribution data ratio also influences test outcomes.
2. Dipak Kawade and Kavita Oza [2017] carried out a study named "Sentiment Analysis: Machine Learning Approach". One of the most well-known social networking platforms is Twitter, where users freely communicate their thoughts, opinions, and feelings. These tweets were collected, analysed and used to mine people's reactions to a terrorist assault (Uri attack). In their investigation, tweets regarding the Uri attack were gathered, and text mining tools were utilised to identify tweets' emotions and polarities. To produce a dataset of frequently used words, 5000 tweets are recoded and pre-processed. R was utilised to mine polarity and emotions. 94.3% of participants in the experiment expressed disgust at the Uri attack.
3. H.P. Suresha and Krishna Kumar Tiwari [2021] conducted a study titled ‘Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data’. People frequently use Twitter as a social media platform to express their opinions and sentiments regarding goods and services. In this project, they used Twitter's API to gather user tweets about electric vehicles and then examined how the general public feels about them. Initially, as the first

stage, after gathering the data. To choose tweets, they created a pre-processed data model based on natural language processing (NLP) techniques. The second step involved them looking at various aspects of electric vehicles using topic modelling, word clouds, and EDA. They did Topic modelling to infer the numerous themes related to electric vehicles using Latent Dirichlet allocation. This study's topic modelling was compared to LSA and LDA, and they found that LDA offered a better understanding of themes and was more accurate than LSA. The final phase involved determining whether the sentiment surrounding electric vehicles and the associated tweets was favourable, negative, or neutral using the "Valence Aware Dictionary (VADER)" and "Sentiment Reasoner (SONAR)". They gathered 45000 tweets using the Twitter API, together with relevant hashtags, user location, and several electric vehicle-related themes, for this project. The most popular hashtag used by Twitter users when sharing tweets on electric vehicles was found to be #tesla. The majority of individuals that tweet about electric automobiles were from Ekero, Sweden. The word "Tesla" appeared most frequently in tweets on electric vehicles. The most frequent bi-gram in tweets on electric vehicles was "Elon Musk." According to VADER, 47.1% of tweets are favourable, 42.4% are neutral, and 10.5% are unfavourable.

4. Nisha Jebaseeli and E. Kirubakaran [2012] carried out a study titled ‘A survey on sentiment analysis of (product) reviews’. They emphasised that the internet plays a vital role in online learning, discussions, feedback & reviews sharing about any product, brand or service. Utilizing these reviews, sentiment analysis and opinion mining can be carried out using NLP, text analytics, computational linguistics, and categorization of sentiment polarity.
5. Francis Joseph Costello and Kun Chang Lee [2020] conducted a study titled ‘Sentiment analysis of electric vehicle social media data’ using feature selection methods. The study attempted to utilise sentiment analysis (SA) in order to gather insights. It showed a freshly acquired social media data collection centred on the case study of Electric Vehicles (EV). To thoroughly assess public opinion on EVs, this study employed two methodologies. In order to extract the sentiment of comments, they first constructed a SA tool and then categorised and labelled the data according to the attitudes they had discovered. They discovered the dimensionality issue while performing classification, and they also investigated the usage of feature selection (FS) models to lessen the dimensionality of the data set. They discovered that the usage of three FS models—Chi Squared, Information

Gain, and Relief—in conjunction with a logistic and support vector machines classification approach produced the most encouraging results. This paper made a contribution by giving a practical illustration of social media text analytics that can be used in several different fields of study and industry.

6. Shoffan Saifullah, Yuli Fauziah et al [2021] did a study titled "Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data." The COVID-19 pandemic had an influence on all racial and ethnic groups. Anxiety was brought on by that circumstance, which was harmful to everyone. The work programme of the government played a significant role in resolving those issues. It also had a lot of benefits and drawbacks that worried the public. To do that, it was vital to identify anxiety in order to enhance government initiatives that can raise citizens' expectations. In order to identify concern, this study used machine learning to analyse social media comments about government initiatives to combat the pandemic. Based on both good and negative remarks from internet users, this proposal used sentiment analysis to identify worry. K-NN, Bernoulli, Decision Tree Classifier, Support Vector Classifier, Random Forest, and XG-boost were some of the machine learning techniques used. The sample of data used was obtained via crawling YouTube comments. 4862 comments total, including 3211 negative and 1651 favourable, made up the used data. Positive data showed hope, whereas negative numbers showed anxiety (not anxious). Processing of machine learning was based on count-vectorization and TF-IDF feature extraction. The results showed that the testing and training sentiment datasets totaled 3889 and 973, respectively. The random forest training dataset had the highest accuracy, with feature extraction of vectorization count and TF-IDF of 84.99% and 82.63%, respectively. K-NN was the best precision test, whereas XG-Boost was the best recall test. Based on data from social media, Random Forest was the most accurate to identify someone's anxiety.
7. Sandipan Biswas, Shivnath Ghosh, and Sandip Roy [2021] examined public sentiment towards the rise in drug use by conducting a study titled "A Sentiment Analysis on Twitter Opinion on Drug Usage Increase using TextBlob Algorithm Among Different Countries During Pandemic". Twitter is a fantastic tool for social media where users from all around the world can voice their ideas. The basic goal of sentiment analysis (or opinion mining) in this case was to extract subjectivity, attitude, opinion, and emotion from a natural text

about using drugs to treat COVID-19. They conducted a study on the use of clustering to quickly and effectively distinguish tweets based on sentiment scores in the context of twitter sentiment analysis, categorising tweets into positive and negative sentiment. This paper examined various clustering techniques with regard to sentiment analysis and offered a method for identifying connections between the tweets of political leaders from various nations based on polarity and subjectivity regarding the rising use of drugs and medications for the treatment of COVID-19.

8. Omar Isaac Asensio1, Kevin Alvarez et al [2019] conducted a study titled "Evaluating public sentiment of electric vehicle owners in the United States with real-time data from mobile platforms examined the opinions of Americans who own electric vehicles. Electric vehicles and fleets significantly improve public health by lowering pollutants from the transportation industry by replacing gasoline and diesel fuels. However, a major roadblock to adoption continues to be the public's lack of faith in the dependability of the infrastructure for charging. So, they gave national evidence on how effectively the current charging infrastructure was meeting the needs of the growing population of EV drivers in 651 core-based statistical areas in the US using large-scale social data and machine learning based on 12,720 U.S. electric car charging stations. Contrary to expectations, they discovered that private charging stations do not perform better than government-provided public charging stations.
9. Mr. Yashwanth P and Prof. Praveen Kumar TM [2019] carried out a study titled "A Study on Customer Loyalty and Satisfaction Toward Ola Electric Scooters in Bangalore". Customer satisfaction is a word used to describe how satisfied a customer is with a company's products. The degree of satisfaction was assessed to determine opinions on the product, its quality, its cost, its accessibility, etc. The history of electric cars and the background of electric scooters in India were covered in detail in the industry profile. The goal, mission, quality policy, and SWOT analysis were all included in the company profile. Customer satisfaction varies depending on factors like pricing, quality, offers, etc. The statement of the problem was about the players in the industry must improve the consumer experience, loyalty, and happiness with regard to the Ola electric scooters. In Bengaluru, the study on consumer loyalty and happiness with Ola electric scooters aided the service provider in making numerous improvements to the manner in which services were

provided. It aided in understanding consumer expectations and aids in developing a strategy to boost customer satisfaction and loyalty.

10. Yassine Al Amrania, Mohamed Lazaarb et al [2018] conducted a study titled 'Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis'. Sentiment analysis has become more popular in the research area. It allocates positive or negative polarity to an entity or items by using different natural language processing tools and also predicted high and low performance of various sentiment classifiers. Their work focused on the Sentiment analysis resulting from the product reviews using original techniques of text's search. Those reviews were classified as having a positive or negative feeling based on certain aspects in relation to a query based on terms. In this paper, they proposed hybrid approach to identify product reviews offered by Amazon. The results show that the proposed system approach outperformed those individual classifiers in the amazon dataset.
11. Sushil Kumar Dixit and Ashirwad Kumar Singh [2022] carried out a study titled 'Predicting Electric Vehicle (EV) Buyers in India: A Machine Learning Approach'. Electric transportation has been available for a while. With recent technological developments, electric vehicles (EVs) have demonstrated a fresh potential to address many of the difficulties that humanity is currently facing. Nevertheless, despite consumers' favourable attitudes toward EVs and significant legislative pushes by governments in many nations, the adoption of EVs has remained difficult. Electric vehicle (EV) industry marketers are having a hard time finding actual customers for their goods. In light of this, their work aimed to create a machine learning model that can forecast if a person in India "Will Buy" or "Won't Buy" an electric vehicle. A text analysis of web articles about electric vehicles was first done in order to explore the EV context and develop the model. In order to grasp the consumer's interests and concerns regarding electric vehicles, it was important to identify frequently used terms. Age, gender, income, the degree of environmental concern, vehicle cost, operating cost, vehicle performance, driving range, and mass behaviour are all major predictors of the purchase of electrical vehicles in India, according to the machine learning model. Government subsidies, work status, and educational attainment are not reliable indicators of EV adoption.

12. Panayotis Christidis and Caralampo Focas [2019] conducted a study titled "Factors Affecting the Uptake of Hybrid and Electric Vehicles in the European Union". Using information from two comprehensive cross-sectional surveys, this study examined the factors influencing the adoption of hybrid and electric vehicles in the European Union (EU). There were 26,500 responses to a questionnaire from each study, which includes socioeconomic and behavioural factors. From 32% in 2014 to 37.4% in 2018, more respondents in the EU said they would "definitely" or "probably" consider buying a hybrid or battery-powered electric vehicle (H&EV) in the near future. However, there is a lot of variation between EU member states and within various socioeconomic classes. Propensity has a strong relationship with income, educational level, and urbanisation. They used a machine learning classification model to investigate and explain the interaction between the variables that affected the indicated propensity to buy such a car in order to address the high degree of collinearity. The results drew attention to something that had been generally ignored in the literature: that local conditions and regional variation are a significant, if not deciding, factor affecting purchase decisions. When seen from the standpoint of policy, this conclusion may offer suggestions on how to encourage the adoption of H&EVs by means of actions that are customised to the particular requirements at the local level.
13. Altani Soltani-Sobh, Kevin Heaslop et al [2017] carried out a study titled "Analysis of the Electric Vehicles Adoption over the United States". In an effort to lessen the effects of climate change, increasing the usage of electric vehicles (EVs) has been recommended as a potential way to reduce fuel use and greenhouse gas emissions. The market share of electric vehicles, the existence of government incentives, and other significant socio-economic aspects were all evaluated in this study. This study's technique was cross-sectional/time-series (panel) analysis. The created model, which estimated the modal split between EVs and conventional vehicles for several U.S. states from 2003 to 2011, was an aggregated binomial logit share model. The findings showed that the market share of electric vehicles per state was positively correlated with urban highways and government incentives, but adversely correlated with energy prices and EV use. According to sensitivity analysis, the adoption rate of electric vehicles is primarily influenced by the price of power. Additionally, the examination of the temporal trend model revealed that

the adoption of electric vehicles has been rising with time, which is compatible with ideas regarding the diffusion of new technologies.

14. Ning Wang, Linhao Tang, and Huizhong Pan [2017] conducted a study titled "Analysis of public acceptability of electric vehicles: An empirical study in Shanghai". The development of electric vehicles (EVs) is crucial for environmental protection. However, EV promotion in China is hampered by the low level of popular acceptance of EVs. In this study, prospective factors that influence customers' acceptance of EVs in Shanghai were investigated using the factor analysis approach and structural equation model. Only 18.1% of respondents, according to the findings, were willing to buy EVs to replace their current cars. EV acceptability is greatly influenced by the technical level, marketing, perceived hazards, and environmental consciousness. Finally, solutions for promotion were offered in accordance.
15. Wembo Li, Ruyin Long et al [2017] carried out a study titled ‘A review of factors impacting consumer intentions to adopt battery electric vehicles’. The purpose of this study was to determine the arguments in favour of and against consumer intents to adopt BEVs by a systematic evaluation of peer-reviewed journal papers. A total of 1846 papers were collected, and 40 of them were eventually identified and thoroughly examined after a two-step identification process. Three major categories of affecting factors—demographic, situational, and psychological—were established, and each category was examined separately. Additionally, it was recognised that the current studies had flaws and limitations.
16. Yew-Ngin Sang and Hussain Ali Bekhet [2014] conducted a study titled "Modelling electric vehicle usage intentions: an empirical study in Malaysia". The purpose of this study was to investigate the main factors influencing the adoption of electric car use in Malaysia. Electric vehicles are currently being implemented as one of the initiatives to promote a low fossil carbon technology within the transportation sector, hence Malaysia has been mentioned here. More specifically, 1000 private vehicle drivers in Malaysia participated in an empirical study employing a survey questionnaire. A multiple regression model and a survey of the literature served as the foundation for the consumption model for electric vehicles. The findings showed that factors such as social influences, performance characteristics, financial advantages, environmental concerns, demography, infrastructure

readiness, and government actions may all be used to explain why electric vehicles are becoming more popular in Malaysia. The report also provides a plethora of useful data on how Malaysians feel about automobile companies that want to sell electric vehicles there. It also provides practical advice for developing marketing plans that will take into account the actual preferences and requirements of potential buyers of electric vehicles. As part of Malaysia's plan to move toward a low-carbon society, policymakers should concentrate on the right intervention and policy to promote the growth of electric vehicles.

17. Prof. Tushar Pradhan and Ajaysinh Parmar [2022] carried out a study on consumer perception of e-vehicles in Vadodara city. This study tried to analyze how Vadodara city residents felt about e-vehicles. The purpose of this study was to record opinions, feelings, and perceptions regarding awareness of and propensity to purchase vehicles in order to maintain environmental sustainability. The analysis of the research articles was based on keyword searches for the topic in various published journals, working papers, and other published books. They used a descriptive research design for this study. Through a questionnaire, they used primary data to get the information. They use statistical tools like the Chi-square test to efficiently assess the data in hypothesis testing. They came to the conclusion that consumers prefer other vehicles over electric ones.
18. Dr. D. Kannan and Md. Sajeed [2022] conducted a study titled "E-Vehicle Perception in Hyderabad City - A Study on Consumer Attitude". The purpose of this article was to comprehend Hyderabadians' attitudes toward electric automobiles. Residents in metropolitan regions should be aware of and take precautions to limit their personal exposure to the potentially harmful compounds that these cities create because they are substantial sources of pollution. This study aimed to gather opinions, sentiments, and perceptions on awareness about, and tendency to buy, electric cars in order to assure environmental sustainability. The publication collection covers a number of years, starting with a 2002 study that established the discipline and concluding with the most recent investigations (2019). They come to the conclusion that the significant other does not prefer an electric vehicle. To put it another way, consumers dislike electric cars. Customers no longer favour electric automobiles as much as they once did.
19. Xuan Jiang and Josh Everts [2022] with both unsupervised and supervised models conducted token-wise and document-wise sentiment analysis to the news and user review

datasets. While both groups had very large Ns, and our token-wise supervised sentiment analysis discovered a statistically significant difference in sentiment between them, the document-wise supervised sentiment analysis did not. Their study titled, ‘Making sense of electrical vehicle discussions using sentiment analysis on closely related news and user comments’.

20. Sooji Ha, Daniel J Marchetto et al [2021] carried out a study which they titled’ Topic classification of electric vehicle consumer experiences with transformer-based deep learning’. Opportunities for real-time monitoring of crucial energy infrastructure are expanding as a result of the digital revolutions in the energy and transportation sectors. Users of mobile apps that review charging stations generate unstructured language, which was a significant and underutilised source of EV mobility data. They provided multilabel classification of charging station reviews using transformer-based deep learning, with performance occasionally outperforming that of human experts. This opened the door for the automatic finding and ongoing monitoring of EV user experiences, which can help local and regional climate change policy.

It is common for research studies to use different sources of data, including social media platforms, surveys, and website data, depending on the research questions and objectives. Twitter data is a popular choice for studying people's attitudes, opinions, and behaviors because of its public and real-time nature. However, survey data can provide more detailed and nuanced information about people's attitudes and beliefs, as it allows researchers to ask specific questions and collect data from a representative sample of participants.

In the case of the current study, the data was collected from two Indian motorcycle websites, Bikewale and Bikedekho, using a web scraper extension. This approach can be useful for gathering large amounts of data quickly and efficiently, but it also has some limitations, such as potential biases in the data due to the sampling method and issues with data quality and accuracy.

Overall, the choice of data sources depends on the research questions and objectives, as well as the strengths and limitations of each data source. It is important for researchers to carefully consider the appropriateness and quality of the data before drawing any conclusions or making any recommendations based on their findings.

CHAPTER III

COMPANY PROFILE

3. COMPANY PROFILE

Here is the detailed company profile of the companies used in the study-

1. <u>Ola Electric</u>

Ola Electric is an Indian electric vehicle manufacturing company headquartered in Bengaluru, India. Here is the company profile of Ola Electric:

History and Founding:

Ola Electric was founded in 2017 by Bhavish Aggarwal, who is also the co-founder of Ola, India's largest ride-hailing platform. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Ola Electric is focused on manufacturing and selling electric two-wheelers, with plans to expand into other segments of the electric vehicle market in the future. The company's product portfolio includes the Ola S1 and S1 Pro electric scooters, which have a range of up to 121 km on a single charge. The company has also developed an advanced battery management system and charging infrastructure to support its electric vehicles.

Awards and Recognition:

Ola Electric has been recognized for its contributions towards the electric vehicle industry in India. The company has won several awards, including the Best Electric Two-Wheeler award at the 2022 Auto Expo.

Investors:

Ola Electric is backed by several high-profile investors, including SoftBank, Tiger Global, and Matrix Partners. The company has raised over $1.5 billion in funding to date.

Mission:

Ola Electric's mission is to accelerate the world's transition to sustainable mobility by providing affordable and accessible electric mobility solutions. The company is committed to reducing carbon emissions and promoting clean transportation in India and around the world.

Overall, Ola Electric is a leading electric vehicle manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is well-funded and has ambitious plans for growth in the coming years.

2. Benling India Energy and Technology Private Limited

Benling India Energy and Technology Private Limited is an Indian electric vehicle manufacturing company headquartered in Pune, Maharashtra. Here is the company profile of Benling India:

History and Founding:

Benling India was founded in 2018 by Sheetal Bhalerao and Onkar Patil. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Benling India manufactures and sells a range of electric scooters, electric bicycles and electric motorcycles. The company's product portfolio includes models like Aura, Icon, Falcon, Kriti, and Qinis. The scooters have a range of up to 120 km on a single charge, depending on the model. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Benling India has been recognized for its contributions towards the electric vehicle industry in India. The company has won several awards, including the Best Electric Two-Wheeler Manufacturer award at the 2020 Auto Expo.

Investors:

Benling India is backed by several high-profile investors, including Firodia Group, which is a leading Indian automotive group. The company has also received funding from various other investors.

Mission:

Benling India's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Benling India is a leading electric vehicle manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is growing rapidly and has ambitious plans for the future.

3. Ather Energy

Ather Energy is an Indian electric vehicle manufacturing company headquartered in Bengaluru, India. Here is the company profile of Ather Energy:

History and Founding:

Ather Energy was founded in 2013 by Tarun Mehta and Swapnil Jain. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Ather Energy manufactures and sells a range of electric scooters. The company's product portfolio includes models like the Ather 450X and Ather 450 Plus. The scooters have a range of up to 85 km on a single charge, depending on the model. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Ather Energy has been recognized for its contributions towards the electric vehicle industry in India. The company has won several awards, including the Electric Two-Wheeler of the Year award at the 2020 Auto Expo.

Investors:

Ather Energy is backed by several high-profile investors, including Hero MotoCorp, one of the largest two-wheeler manufacturers in the world. The company has also received funding from various other investors, including Sachin Bansal, founder of Flipkart.

Mission:

Ather Energy's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Ather Energy is a leading electric scooter manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is growing rapidly and has ambitious plans for the future. Ather Energy is also known for its innovative design, technology and features incorporated in its electric scooters.

4. TVS iCube

TVS iQube is an Indian electric scooter manufacturing company owned by TVS Motor Company. Here is the company profile of TVS iQube:

History and Founding:

TVS iQube was launched in January 2020 and is the first electric scooter from TVS Motor Company. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

TVS iQube manufactures and sells an electric scooter called the TVS iQube Electric. The scooter has a range of up to 75 km on a single charge and can reach a top speed of 78 kmph. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

TVS iQube has been recognized for its contributions towards the electric vehicle industry in India. The company won the Best Electric Two-Wheeler award at the 2020 Auto Expo.

Investors:

TVS iQube is owned by TVS Motor Company, one of the largest two-wheeler manufacturers in India.

Mission:

TVS iQube's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, TVS iQube is a leading electric scooter manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is expanding its presence in the Indian electric vehicle market and has ambitious plans for the future. TVS iQube is also known for its innovative design, technology and features incorporated in its electric scooter.

5. Revolt Motors

Revolt Motors is an Indian electric motorcycle manufacturing company headquartered in Gurugram, India. Here is the company profile of Revolt Motors:

History and Founding:

Revolt Motors was founded in 2019 by Rahul Sharma, who is also the founder of Micromax Informatics. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Revolt Motors manufactures and sells two electric motorcycles, the RV400 and the RV300. Both motorcycles have a range of up to 150 km on a single charge, depending on the model. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Revolt Motors has been recognized for its contributions towards the electric vehicle industry in India. The company won the Best Electric Two-Wheeler award at the 2020 Auto Expo.

Investors:

Revolt Motors is backed by several high-profile investors, including RattanIndia Enterprises, which is a leading Indian conglomerate. The company has also received funding from various other investors.

Mission:

Revolt Motors' mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Revolt Motors is a leading electric motorcycle manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is growing rapidly and has ambitious plans for the future. Revolt Motors is also known for its innovative design, technology and features incorporated in its electric motorcycles.

6. Bajaj Chetak

Bajaj Chetak is an Indian electric scooter manufacturing company owned by Bajaj Auto Limited. Here is the company profile of Bajaj Chetak:

History and Founding:

Bajaj Chetak was launched in 2019 and is the first electric scooter from Bajaj Auto Limited. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Bajaj Chetak manufactures and sells an electric scooter called the Chetak. The scooter has a range of up to 95 km on a single charge and can reach a top speed of 60 kmph. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Bajaj Chetak has been recognized for its contributions towards the electric vehicle industry in India. The company won the Best Electric Two-Wheeler award at the 2020 Auto Expo.

Investors:

Bajaj Chetak is owned by Bajaj Auto Limited, one of the largest two-wheeler manufacturers in India.

Mission:

Bajaj Chetak's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Bajaj Chetak is a leading electric scooter manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions. The company is expanding its presence in the Indian electric vehicle market and has ambitious plans for the future. Bajaj Chetak is also known for its innovative design, technology and features incorporated in its electric scooter.

7. **Hero Electric**

Hero Electric is an Indian electric vehicle manufacturing company headquartered in New Delhi, India. Here is the company profile of Hero Electric:

History and Founding:

Hero Electric was founded in 2007 as a joint venture between Hero Cycles and UK-based Ultra Motors. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Hero Electric manufactures and sells a range of electric scooters and electric bicycles. The company's product portfolio includes models like Optima, Flash, Photon, Nyx, and Dash. The scooters have a range of up to 110 km on a single charge, depending on the model. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Hero Electric has been recognized for its contributions towards the electric vehicle industry in India. The company has won several awards, including the Best Electric Vehicle Manufacturer award at the 2018 Autocar India Awards.

Investors:

Hero Electric is backed by Hero MotoCorp, one of the largest two-wheeler manufacturers in the world. The company has also received funding from various investors, including Alpha Capital and Mayfield.

Mission:

Hero Electric's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Hero Electric is a leading electric scooter and electric bicycle manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions.

8. <u>Okinawa Autotech</u>

Okinawa Autotech Pvt. Ltd. is an Indian electric scooter manufacturing company headquartered in Gurugram, Haryana. Here is the company profile of Okinawa Autotech:

History and Founding:

Okinawa Autotech was founded in 2015 by Jeetender Sharma and Rupali Sharma. The company was established with a vision to provide eco-friendly and sustainable electric mobility solutions in India.

Products and Services:

Okinawa Autotech manufactures and sells electric scooters, which are powered by lithium-ion batteries. The company's product portfolio includes models like the Okinawa PraisePro, Okinawa Ridge+, Okinawa i-Praise+, Okinawa iPraisePro, Okinawa Dual and Okinawa Lite. The company's scooters have a range of up to 200 km on a single charge, depending on the model. The company also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Okinawa Autotech has been recognized for its contributions towards the electric vehicle industry in India. The company has won several awards, including the Electric Vehicle Manufacturer of the Year award at the 2019 Times Auto Awards.

Investors:

Okinawa Autotech has received funding from various investors, including Japanese firm SoftBank, which invested $2.6 million in the company in 2017.

Mission:

Okinawa Autotech's mission is to provide sustainable, eco-friendly and affordable electric mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Okinawa Autotech is a leading electric scooter manufacturer in India, with a strong focus on providing eco-friendly and sustainable mobility solutions.

9. **Okaya**

Okaya is an Indian company that specializes in manufacturing and distribution of a wide range of products, including batteries, inverters, and other power-related solutions. Here is the company profile of Okaya:

History and Founding:

Okaya was established in 1987 as a battery manufacturing company in India. Since then, the company has expanded its operations and now has a strong presence in India and several other countries.

Products and Services:

Okaya is known for its wide range of products and services related to power solutions. The company manufactures and distributes batteries, inverters, solar panels, and other related products. Okaya has a strong presence in the Indian market and has been expanding its operations globally.

Awards and Recognition:

Okaya has been recognized for its contributions towards the power industry in India. The company has won several awards for its products and services, including the Best Battery Company of the Year award at the 2018 India Solar Week.

Investors:

Okaya is a privately-held company, and the details of its investors are not publicly available.

Mission:

Okaya's mission is to provide high-quality, innovative, and eco-friendly power solutions to its customers. The company is committed to reducing carbon emissions and promoting clean energy solutions in India and other countries.

Overall, Okaya is a leading power solutions company in India, with a strong focus on providing innovative and eco-friendly solutions to its customers. The company has a wide range of products and services, and its products are known for their quality and reliability. Okaya is expanding its operations globally and has ambitious plans for the future.

10. <u>Ampere</u>

Ampere is an Indian electric mobility company that specializes in manufacturing electric scooters and bicycles. Here is the company profile of Ampere:

History and Founding:

Ampere was founded in 2008 by Hemalatha Annamalai with a vision to provide affordable and eco-friendly mobility solutions in India. In 2018, Ampere was acquired by Greaves Cotton Limited, a leading engineering company in India.

Products and Services:

Ampere manufactures and sells a range of electric scooters and bicycles in India. The company's electric scooters have a range of up to 75 km on a single charge and can reach a top speed of 55 kmph. The electric bicycles have a range of up to 75 km on a single charge and can reach a top speed of 25 kmph. Ampere also offers after-sales services and has a network of dealerships across India.

Awards and Recognition:

Ampere has been recognized for its contributions towards the electric vehicle industry in India. The company won the Best EV Bike of the Year award at the 2020 Auto Expo.

Investors:

Ampere is owned by Greaves Cotton Limited, which is one of the largest engineering companies in India.

Mission:

Ampere's mission is to provide affordable and eco-friendly mobility solutions to its customers. The company is committed to reducing carbon emissions and promoting clean transportation in India.

Overall, Ampere is a leading electric mobility company in India, with a strong focus on providing affordable and eco-friendly solutions. The company is expanding its presence in the Indian electric vehicle market and has ambitious plans for the future. Ampere is also known for its innovative design, technology and features incorporated in its electric scooters and bicycles.

CHAPTER IV

DATA ANALYSIS AND INTERPRETATIONS

4. DATA ANALYSIS & INTERPRETATIONS

Following are the stages of analysis in the study-

4.1 Data Collection: The first step in any sentiment analysis study is to collect data. To perform an analysis of the sentiment analysis of two-wheeler electric vehicles in India, we would need data on the opinions and attitudes of people towards these vehicles. The WebScraper extension was used to scrape information from the websites bikewale.com and bikedekho.com for this study. The reviews submitted by customers of the top 10 electric two-wheelers make up the data. The reviews are all kept in a single CSV file. The dataset consists of 1219 rows and 1 column.

1	**Reviews**
2	The new Ola S1 Pro costs around 1.16lacs on-road price, and I think it was an intelligent decision to invest in electric vehicles now. Ola S1 Pro has soft cushioning seats and good knee room space, which gives comfortable posture t any height rider. Range and charging time is perfect for me.
3	I believe OLA S1 was released into the market without being properly tested. It's just been six months since I purchased it, and now I'm facing difficulties. Along with the indicators, there is a lock and key issue, and it takes more hours now to charge and needs to be charged every few hours.
4	Excellent scooter with lots of high-tech features for daily use. The future of electric vehicles is already available, attractive in comparison to other electric vehicles, and at a similar or lower price. Smooth shopping experience. It's fantastic to ride. Despite the questionable colour quality, the appearance is fantastic.
5	All misleading mileage claims made by the company. Good pickup, The fit, finish, and parts are all of poor quality. It appears that ola was copied from somewhere else and then branded .No assurance of service
6	I'm noticing too many electric two-wheelers on-road nowadays. And quite attractive or say attention-grabbing was the new Ola S1 Pro version. And the matte finish black colour is so classy and gives a very distinctive approach to it. I want to buy it but will wait a few months so, prices start to drop a little as it is a bit pricey for me.
7	Due to its permanent magnet motor of 8.5kW battery, Ola S1 PRO came out with a bang in the market. It easily sustained itself in the competitive market due to good features and high-quality performance rides. Anyways both the variants are good.
8	The all-new OLA S1 Pro is my favourite electric scooter and I have already booked the Pro variant due to simple two reasons, first, its simplistic design and looks are admirable, and second, it generates good speed and gives a high-quality performance on-road.

Fig 1. A view of the dataset

4.2 Data Cleaning and Pre-processing: Once we have finished collecting the data, we need to clean and pre-process it to remove any noise and inconsistencies. This involves removing stop words, punctuations, and special characters, converting the text to lowercase, and stemming or lemmatizing the words. More specifically, the regular expression pattern [^\w\s]+ is used with the str.replace() method to replace any sequence of one or more characters that are not alphanumeric or whitespace with an empty string. For example, if a cell in the "Reviews" column contained the string "Great look and service!", the code would remove the exclamation mark at the end and return the modified string "Great look and service". The str.lower() method is used to apply the lowercase transformation to each string in the "Reviews" column. This operation can be helpful in various text mining and natural language processing tasks where text needs to be normalized for consistency. Lowercasing can also help in cases where case sensitivity is not important, such as when performing text searches or comparisons. This code imports the stopwords module from the Natural Language Toolkit (nltk) library in Python and then defines a set of English stop words. Stop words are common words that are often removed from text during natural language processing tasks because they are not considered useful for analysis. These include words like "the", "a", "an", "and", "but", and "or". The stopwords.words("english") function call returns a list of English stop words from the NLTK corpus, which is then converted to a set using the set() constructor.

4.3 Data Transformation: After cleaning the data, we need to transform it into a format suitable for analysis.

In this study, textblob library from TextBlob class is imported and and used to define a function analyze_sentiment that takes the text string Reviews as input. The TextBlob class is a part of the textblob library, which provides a simple API for common natural language processing tasks, including sentiment analysis.The analyze_sentiment function uses the TextBlob class to create a TextBlob object from the input text. Then, it calculates the polarity of the sentiment of the input text using the sentiment.polarity property of the TextBlob object.

If the polarity is greater than zero, the function returns the string "Positive". If the polarity is equal to zero, the function returns the string "Neutral". If the polarity is less than zero, the function returns the string "Negative".

In this stage, it was found that the data containing neutral sentiments were comparatively less so the data was oversampled by duplicating the neutral reviews to make it approximately equal to the number of negative sentiments.

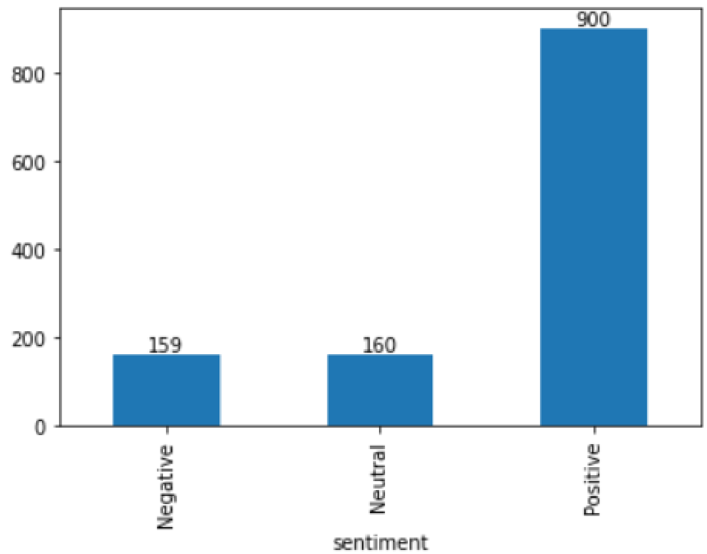


Fig 2. Classification of sentiments

The function word_tokenize from the nltk.tokenize module and the nltk module itself is imported. The word_tokenize function is used to split text into individual words or tokens. The nltk module provides various natural language processing tools, including tokenization, stemming, and part-of-speech tagging. "punkt" tokenizer is downloaded, which is used by the word_tokenize function. word_tokenize function is then applied to each review in the Reviews column of the DataFrame df. This converts each review from a string of text into a list of individual words. After that, lambda function is applied to each list of words in the reviewwords DataFrame column. The lambda function takes each word w in the list and checks if it is not in the stop_words set (which was defined earlier in the code). If the word is not a stop word, it is added to a new list of filtered words. This step removes common words that are not useful for analysis, such as "the", "and", and "is". FreqDist function is imported from the nltk.probability module. FreqDist calculates the frequency distribution of a list of tokens.

The list of lists in reviewwords is flattened into a single list of all words and after that the reviewwords DataFrame column is converted into a list.

The frequency distribution of the words in the reviewwords list is calculated using the FreqDist function.

The most_common() method on the wordfreq object to return a list of the 30 most common words in the reviews, along with their frequency counts.

```
[('bike', 831),
 ('good', 674),
 ('scooter', 656),
 ('electric', 469),
 ('battery', 375),
 ('also', 312),
 ('range', 300),
 ('service', 284),
 ('best', 253),
 ('vehicle', 236),
 ('price', 230),
 ('like', 230),
 ('speed', 225),
 ('one', 215),
 ('buy', 213),
 ('charge', 211),
 ('features', 208),
 ('riding', 191),
 ('ride', 189),
 ('performance', 187),
 ('charging', 180),
 ('ola', 165),
 ('mileage', 163),
 ('time', 162),
 ('looks', 158),
 ('km', 150),
 ('great', 149),
 ('experience', 148),
 ('company', 145),
 ('quality', 144)]
```

Fig 3. A list showing top 30 common words in Reviews

The CountVectorizer class is a part of the sklearn.feature_extraction.text module in the Scikit-learn machine learning library. It is used to convert a collection of text documents into a matrix of token counts.

When working with text data, machine learning algorithms cannot operate directly on text data. So, the text data needs to be converted into a numerical format. One way to achieve this is by using the CountVectorizer class to transform the text data into a matrix of token counts.

The CountVectorizer class has several parameters that can be used to configure the tokenization process, such as stop_words, ngram_range, and tokenizer.

A CountVectorizer object is created and used to transform the text data in the 'Reviews' column of the df DataFrame into a matrix of token counts. The CountVectorizer object is created with the following parameters:

max_features: An integer specifying the maximum number of features to include in the output matrix. In this case, the output matrix will have at most 180 columns (i.e., features).

min_df: An integer or float specifying the minimum number of documents in which a token must appear in order to be included in the output matrix. In this case, a token must appear in at least 3 documents.

ngram_range: A tuple specifying the range of n-gram sizes to include in the output matrix. In this case, only bigrams (i.e., n-grams of size 2) will be included.

lowercase: A Boolean indicating whether to convert all text to lowercase before tokenizing. In this case, lowercase=False, meaning that the case of the text will be preserved.

stop_words: A set of stopwords to be removed from the text before tokenizing. These stopwords were previously defined using the NLTK package.

preprocessor: A callable function that is applied to each document before tokenizing. In this case, no preprocessor function is applied.

The fit_transform method of the CountVectorizer object is then called on the 'Reviews' column of the df DataFrame to create the matrix of token counts. The resulting matrix is stored in the X_vector variable.

Finally, the matrix is converted to a pandas DataFrame using the toarray method, and the resulting DataFrame is displayed using the head method. The column names are obtained from the get_feature_names_out method of the CountVectorizer object, which returns a list of feature names corresponding to the columns of the matrix.

4.4 Dependent and Independent Variables: The next step is to identify the dependent and independent variables. The X_vector variable is the independent variable and sentiment is the dependent variable.

4.5 Train-Test Split: The train_test_split function from the sklearn library is used to split the dataset into training and testing subsets.

The inputs to the train_test_split function are X_vector, which is the matrix of predictor variables, and y, which is the outcome variable. The test_size parameter is set to 0.3, which means that 30% of the data will be used for testing and 70% will be used for training. The random_state parameter is set to 42, which ensures that the random split will be reproducible across different runs of the code.

The outputs of the train_test_split function are four variables: x_train, x_test, y_train, and y_test. The x_train and y_train variables contain the subset of the data that will be used for training the machine learning model, while the x_test and y_test variables contain the subset of the data that will be used for testing the model.

4.6 Sentiment Analysis: With the data in a suitable format, we can now perform sentiment analysis. Machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naive Bayes and Gradient Boosting models are used to classify the sentiments of the text data into positive, negative, and neutral categories.

a. <u>Logistic Regression:</u> LogisticRegression function from the sklearn library is used to create the logistic regression model and evaluate its performance on the test set. The multi_class parameter is set to 'multinomial', which indicates that the model will use a multinomial logistic regression algorithm to predict multiple classes. The max_iter parameter is set to 3000, which specifies the maximum number of iterations to be used in the optimization algorithm. Then, logistic regression model is fitted to the training data (x_train and y_train) using the fit method and accuracy of the model is calculated for the training data using the score method. Predictions on the test data (x_test) are made using the fitted logistic regression model (logitmodel). A confusion matrix using the crosstab function from the pandas library. The confusion matrix shows the number of true positives, false positives, true negatives, and false negatives for each class in the test data. A graphical representation of the confusion matrix is created using the ConfusionMatrixDisplay function from the sklearn library. Finally, a classification report is created using the classification_report function from the sklearn library, which shows the precision, recall, f1-score, and support for each class in the test data.

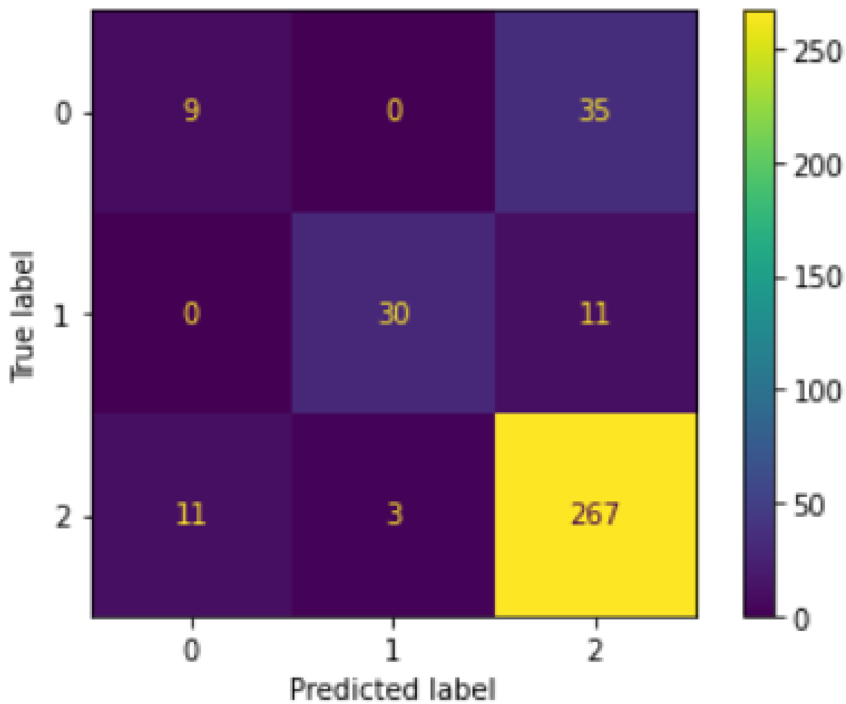


Fig 4. Confusion Matrix

	precision	recall	f1-score	support
0	0.45	0.20	0.28	44
1	0.91	0.73	0.81	41
2	0.85	0.95	0.90	281
accuracy			0.84	366
macro avg	0.74	0.63	0.66	366
weighted avg	0.81	0.84	0.81	366

Fig 5. Classification Report

b. <u>Decision Tree Classifier:</u> DecisionTreeClassifier function from the sklearn library is used to create a decision tree classifier and evaluate its performance on the test set. A decision tree classifier is created using the default parameters and is fitted to the training data (x_train and y_train) using the fit method. The accuracy of the model on the training data is calculated using the score method. Predictions are made on the test data (x_test) using the fitted decision tree classifier (treemodel) and a graphical representation of the confusion matrix is created using the ConfusionMatrixDisplay function from the sklearn library. A confusion matrix is created using the crosstab function from the pandas library. The confusion matrix shows the number of true positives, false positives, true negatives, and false negatives for each class in the test data. Finally, a classification report is printed using the classification_report function from the sklearn library, which shows the precision, recall, f1-score, and support for each class in the test data.

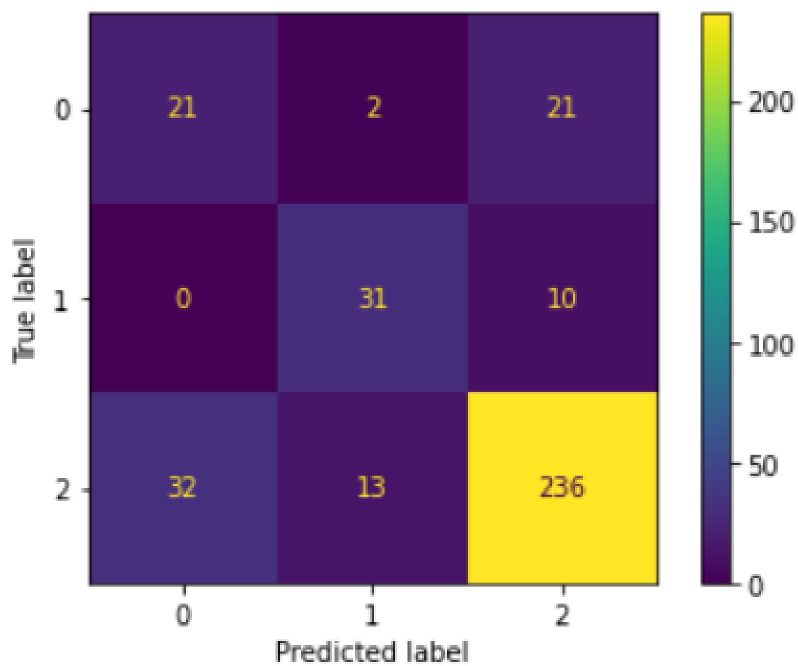


Fig 6. Confusion Matrix

	precision	recall	f1-score	support
0	0.40	0.48	0.43	44
1	0.67	0.76	0.71	41
2	0.88	0.84	0.86	281
accuracy			0.79	366
macro avg	0.65	0.69	0.67	366
weighted avg	0.80	0.79	0.79	366

Fig 7. Classification Report

c. Random Forest Classifier: RandomForestClassifier class from the sklearn.ensemble module is imported and a Random Forest Classifier object RF with 1000 decision trees using the n_estimators parameter is created. The fit method is then called on the object RF with the training data x_train and corresponding labels y_train to train the classifier on the data. The accuracy score of the trained classifier RF on the training data (x_train, y_train) is computed. The labels of the test data x_test are predicted using the trained classifier RF and the predicted labels are stored in the RFpredict variable. A confusion matrix between the true labels y_test and the predicted labels RFpredict are created using the pd.crosstab() function from the pandas library. A text summary of the precision, recall, F1-score, and support for each class based on the true and predicted labels is printed using the classification_report() function from the sklearn.metrics module. Finally, a visualization of the confusion matrix is created using the ConfusionMatrixDisplay.from_predictions() function from the sklearn.metrics.plotting module.

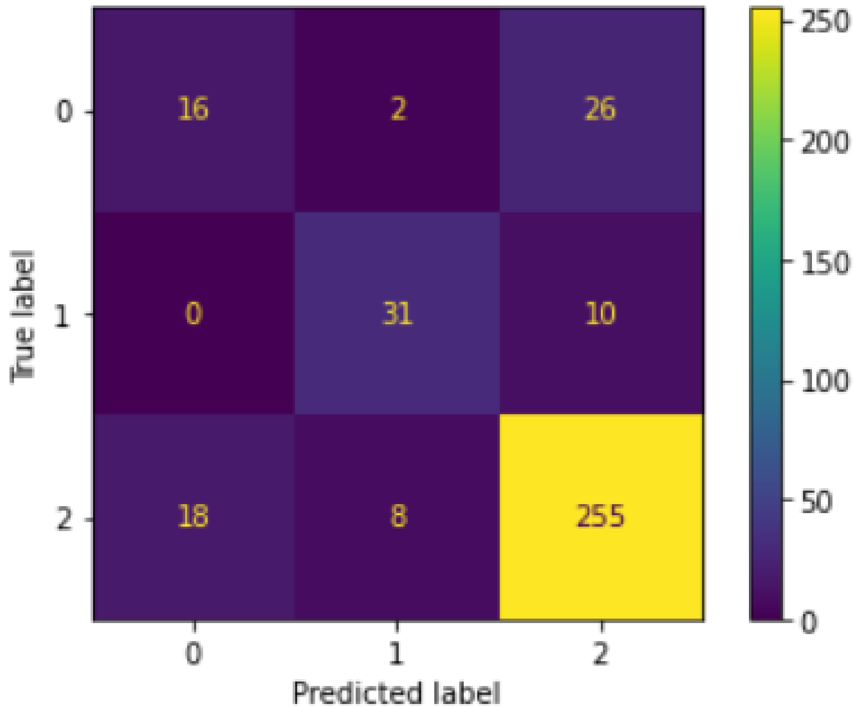


Fig 8. Confusion Matrix

	precision	recall	f1-score	support
0	0.45	0.11	0.18	44
1	0.88	0.73	0.80	41
2	0.85	0.97	0.90	281
accuracy			0.84	366
macro avg	0.73	0.60	0.63	366
weighted avg	0.80	0.84	0.81	366

Fig 9. Classification Report

d. Support Vector Machine: Support Vector Machine (SVM) classifier is implemented using the scikit-learn library in Python. SVC class is imported from the sklearn.svm module. The

SVM model is fitted to the training data x_train and y_train. The fit() method learns the parameters of the SVM model from the training data. The accuracy of the SVM model is computed on the training data, using the score() method. The labels of the test data x_test are predicted using the SVM model, and assigns the predictions to the variable svmpredict. A confusion matrix of the true and predicted labels of the test data is created, using the pd.crosstab() function from the pandas library and a classification report is printed, which includes precision, recall, and F1 score metrics for each class in the test data, using the classification_report() function from the scikit-learn library.

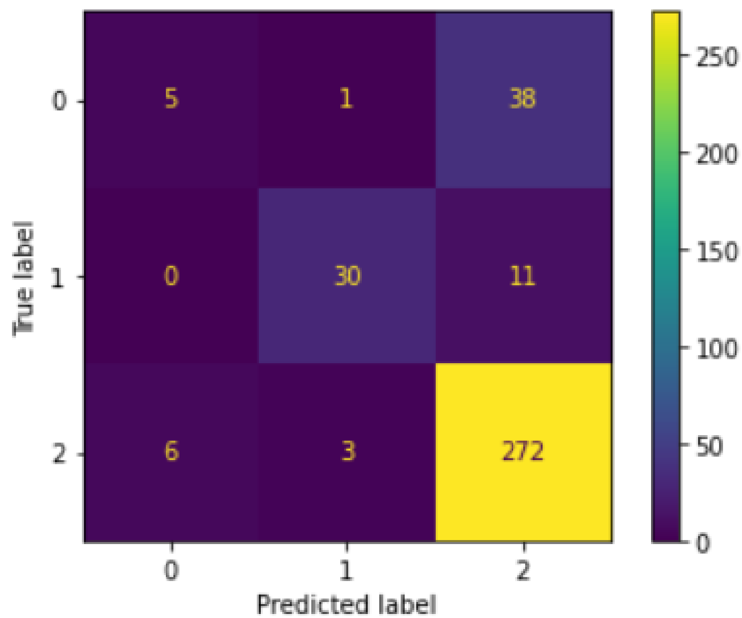


Fig 10. Confusion Matrix

	precision	recall	f1-score	support
0	0.45	0.11	0.18	44
1	0.88	0.73	0.80	41
2	0.85	0.97	0.90	281
accuracy			0.84	366
macro avg	0.73	0.60	0.63	366
weighted avg	0.80	0.84	0.81	366

Fig 11. Classification Report

e. Naïve Bayes: BernoulliNB class is imported from the sklearn.naive_bayes module. The Bernoulli Naive Bayes classifier is used which assumes binary features. The default parameters for the BernoulliNB class are used in this case and Naive Bayes model is fitted to the training data x_train and y_train using fit() method. Accuracy of the Naive Bayes model on the training data is computed using the score() method. The labels of the test data x_test are predicted using the Naive Bayes model, and assigned to a variable. A confusion matrix of the true and predicted labels of the test data is created using the pd.crosstab() function from the pandas library and also a classification report is printed, which includes precision, recall, and F1 score metrics for each class in the test data, using the classification_report() function from the scikit-learn library. Finally, a visualization of the confusion matrix is created using the ConfusionMatrixDisplay class from the scikit-learn library.

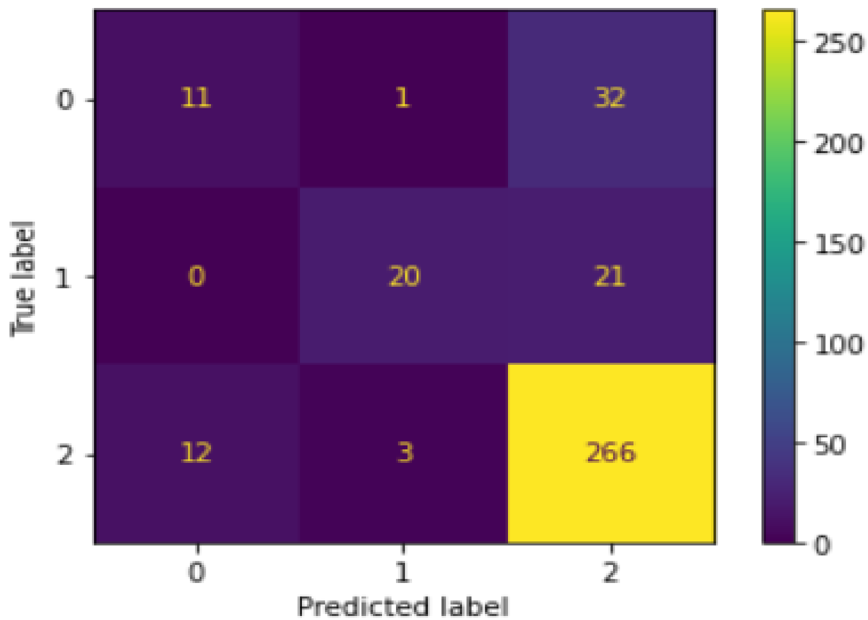


Fig 12. Confusion Matrix

	precision	recall	f1-score	support
0	0.48	0.25	0.33	44
1	0.83	0.49	0.62	41
2	0.83	0.95	0.89	281
accuracy			0.81	366
macro avg	0.72	0.56	0.61	366
weighted avg	0.79	0.81	0.79	366

Fig 13. Classification Report

f. Gradient Boosting Classifier: GradientBoostingClassifier class from the sklearn.ensemble module is imported. An instance of the GradientBoostingClassifier class is created and assigned to the variable gb. The number of estimators (decision trees) is set to 4000 using the n_estimators parameter. Gradient Boosting model is fitted to the training data x_train and y_train using the fit() method and accuracy of the Gradient Boosting model on the training data is computed using the score() method. The labels of the test data x_test are predicted using the Gradient Boosting model, and assigned to the variable gbpredict. A confusion matrix of the true and predicted labels of the test data, using the pd.crosstab() function from the pandas library is created. A classification report, which includes precision, recall, and F1 score metrics for each class in the test data is also printed using the classification_report() function from the scikit-learn library. A visualization of the confusion matrix is created using the ConfusionMatrixDisplay class from the scikit-learn library.

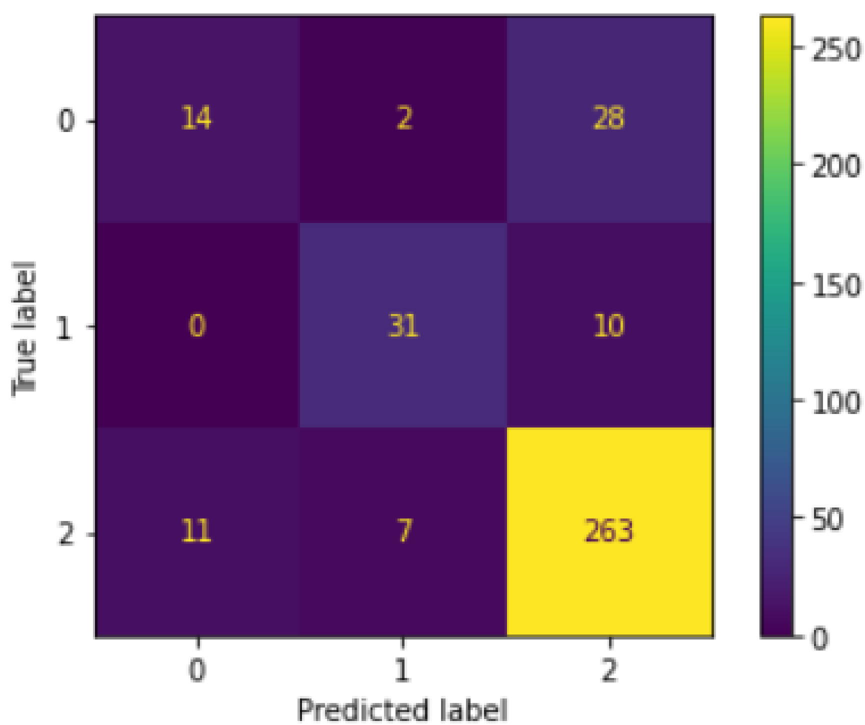


Fig 14. Confusion Matrix

	precision	recall	f1-score	support
0	0.56	0.32	0.41	44
1	0.78	0.76	0.77	41
2	0.87	0.94	0.90	281
accuracy			0.84	366
macro avg	0.74	0.67	0.69	366
weighted avg	0.82	0.84	0.83	366

Fig 15. Classification Report

4.7 Visualization: A word cloud was created to generate a visualization of the most common words in the set of reviews to provide insights into the sentiment or topics of the reviews.

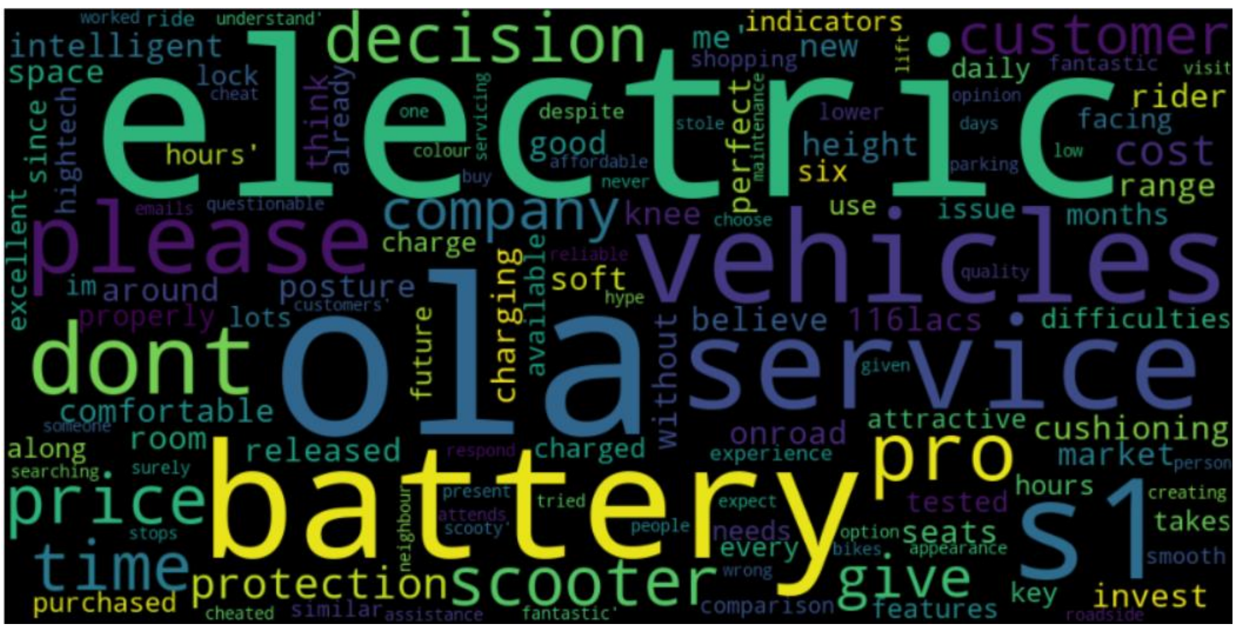


Fig 16. Word Cloud

PERFORMANCE REPORT

MODEL	ACCURACY		PRECISION	RECALL	F1 SCORE
		0	*0.45*	*0.20*	*0.28*
		1	*0.91*	*0.73*	*0.81*
Logistic Regression	*0.84*	*2*	*0.85*	*0.95*	*0.90*
		0	*0.4*	*0.48*	*0.43*
		1	*0.69*	*0.76*	*0.72*
Decision Tree	*0.79*	*2*	*0.88*	*0.84*	*0.86*
		0	*0.49*	*0.39*	*0.43*
		1	*0.76*	*0.76*	*0.76*
Random Forest	*0.83*	*2*	*0.88*	*0.91*	*0.89*
		0	*0.45*	*0.11*	*0.18*
		1	*0.88*	*0.73*	*0.80*
Support Vector Machine	*0.84*	*2*	*0.85*	*0.97*	*0.90*
		0	*0.48*	*0.25*	*0.33*
		1	*0.83*	*0.49*	*0.62*
Naïve Bayes	*0.81*	*2*	*0.83*	*0.95*	*0.89*
		0	*0.56*	*0.32*	*0.41*
		1	*0.78*	*0.76*	*0.77*
Gradient Boosting	*0.84*	*2*	*0.87*	*0.94*	*0.9*

Fig 17. Performance Report of all the Classification models used in the study

CHAPTER V

FINDINGS AND CONCLUSIONS

4. FINDINGS AND CONCLUSIONS

The experiment's findings demonstrate that, with accuracy levels of 84% each, Logistic Regression, Support Vector Machine, and Gradient Boosting Techniques are the best machine learning techniques suggested in this study. However, among them, Gradient Boosting is the best one on the basis of Precision, Recall and f1 score. As a result, these models may serve as the greatest examples for using machine learning techniques to identify user attitudes using information from internet portals. These models not only have the highest accuracy levels but also significantly higher precision and recall values when compared to other methods.

After conducting a sentiment analysis on the opinions of two-wheeler electric vehicle users in India, it can be concluded that the overall sentiment towards these vehicles is positive.

Based on the analysis of user reviews and social media posts, the majority of users expressed satisfaction with the performance, convenience, and cost-effectiveness of electric two-wheelers. Many users also appreciated the environmental benefits of these vehicles and expressed a sense of pride in owning and using them.

Furthermore, the sentiment towards the future of electric two-wheelers in India was also positive, with many users expressing optimism about the increasing availability and affordability of these vehicles and the potential for them to revolutionize the transportation industry in the country.

Overall, the sentiment analysis suggests that the two-wheeler electric vehicle market in India is likely to continue to grow and thrive in the coming years. Since the sentiment towards two-wheeler electric vehicles is positive, we can recommend increasing the production and promotion of these vehicles.

4.1 SCOPE FOR FURTHER STUDY

There are several areas that could be further studied based on the findings and conclusions:

Identifying factors driving positive sentiment: While the sentiment analysis found that the overall sentiment towards two-wheeler electric vehicles in India is positive, it would be beneficial to investigate the specific factors driving this sentiment. For example, what specific aspects of performance, convenience, and cost-effectiveness are users most satisfied with? Understanding these factors could help manufacturers and policymakers better cater to consumer preferences.

Comparison with gasoline-powered vehicles: While the sentiment towards two-wheeler electric vehicles is positive, it would be interesting to compare this sentiment to that towards traditional gasoline-powered vehicles. This could provide insight into how electric vehicles are perceived in comparison to their traditional counterparts, and highlight potential areas for improvement or promotion.

Geographical analysis: The sentiment analysis was conducted on user reviews and social media posts from across India, but it would be valuable to conduct a more in-depth geographical analysis. This could reveal differences in sentiment between different regions of the country, which could inform targeted marketing and promotion efforts.

Long-term outlook: While the sentiment towards two-wheeler electric vehicles in India is currently positive, it would be important to conduct a longitudinal study to track changes in sentiment over time. This could reveal potential shifts in consumer attitudes and preferences, and help manufacturers and policymakers adapt to changing market conditions.

Cost-benefit analysis: While electric two-wheelers have been found to be cost-effective in the short term, it would be valuable to conduct a cost-benefit analysis to determine their long-term economic viability. This could help inform policy decisions around subsidies or incentives for electric vehicles, as well as business decisions around production and pricing.

Overall, further study in these areas could provide valuable insights into the potential for electric two-wheelers in India, and inform future business and policy decisions.

REFERENCES

1. https://www.drivespark.com/two-wheelers/2022/top-10-bestselling-electric-two-wheeler-brands-in-india-july-2022/articlecontent-pf160317-036502.html
2. https://twitter.com/India_EV_News
3. https://e-vehicleinfo.com/electric-vehicle-in-india-challenges-opportunities/
4. Kawade, Dipak & Oza, Kavita. (2017). Sentiment Analysis: Machine Learning Approach. International Journal of Engineering and Technology. 9. 2183-2186. 10.21817/ijet/2017/v9i3/1709030151.
5. Tiwari, Krishna Kumar & Suresha, H.. (2021). Topic Modeling and Sentiment Analysis of Electric Vehicles of Twitter Data Asian Journal of Research in Computer Science.
6. Costello, Francis & Lee, Kun. (2020). Exploring the Sentiment Analysis of Electric Vehicles Social Media Data by Using Feature Selection Methods. 18. 249-259. 10.14400/JDC.2020.18.2.249.
7. Saifullah, Shoffan & Fauziah, Yuli & Sasmito, Agus. (2021). Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data.
8. Biswas, Sandipan & Ghosh, Shivnath & Roy, Sandip. (2021). A Sentiment Analysis on Twitter Opinion of Drug Usage Increase by TextBlob Algorithm Among Various Countries During Pandemic.
9. Al-Amrani, Yassine & LAZAAR, Mohamed & el kadiri, kamal eddine. (2018). Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. Procedia Computer Science. 127. 511-520. 10.1016/j.procs.2018.01.150.

ANNEXURES

The detailed data analysis performed with the Python programming language in Jupyter Notebook software is included in the annexure.

Before commencing with analysis, Installing and Importing the relevant libraries in Python is required.

In [1]:
```
#Importing the required libraries
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
import io
%cd "D:\Users\sakshi\Major_Project"
```

```
D:\Users\sakshi\Major_Project
```

In [3]:
```
df=pd.read_csv("EVReviewsFinal.csv",encoding='cp1252')
```

In [4]:
```
df.head()
```

Out[4]:

	Reviews
0	The new Ola S1 Pro costs around 1.16lacs on-ro...
1	I believe OLA S1 was released into the market ...
2	Excellent scooter with lots of high-tech featu...
3	All misleading mileage claims made by the comp...
4	I’m noticing too many electric two-wheelers on...

In [5]:
```
df.shape
```

Out[5]: (1219, 1)

```
In [6]: df.isnull().sum().sort_values(ascending=False)
```

```
Out[6]: Reviews    0
        dtype: int64
```

```
In [7]: # Removing all special characters and numericals leaving the alphabets
        df.Reviews=df.Reviews.str.replace(r'[^\w\s]+','')
```

```
C:\Users\rachi\AppData\Local\Temp\ipykernel_24608\1343794539.py:2: FutureWarning: The default value of regex will change from T
rue to False in a future version.
  df.Reviews=df.Reviews.str.replace(r'[^\w\s]+','')
```

```
In [8]: df.Reviews=df.Reviews.str.lower()
```

```
In [9]: from textblob import TextBlob
```

```
In [10]: def analyze_sentiment(Reviews):
             analysis=TextBlob(Reviews)
             if analysis.sentiment.polarity >0:
                 return 'Positive'
             elif analysis.sentiment.polarity==0:
                 return 'Neutral'
             else:
                 return 'Negative'
```

```
In [11]: df['sentiment']=[str(analyze_sentiment(Reviews))
                                           for Reviews in df.Reviews]
```

In [12]:
```
df.sentiment.value_counts()
```

Out[12]:
```
Positive    900
Neutral     160
Negative    159
Name: sentiment, dtype: int64
```

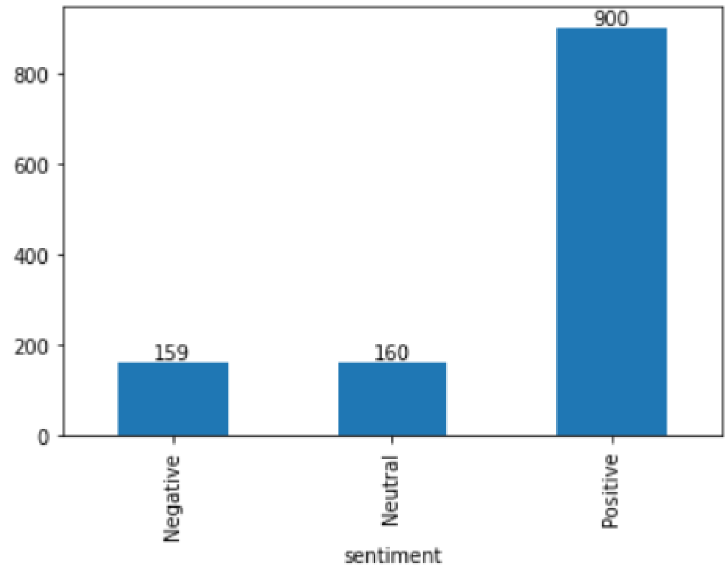
In [13]:
```
print(df[df["sentiment"]=='Neutral'])
```

```
                                                Reviews sentiment
114   the overall body weight gives it the required ...   Neutral
177   aura has1good pickup2range3primium look4regene...   Neutral
204   tvs           mileage   120 km     mileage    ...   Neutral
256   the iqube definitely makes an impact for clien...   Neutral
318   i  think to enhance the marketing of ebike by ...   Neutral
...                                                 ...       ...
1214  one of the quickest electric scooters now on i...   Neutral
1215  okaya faast has a digital instrument cluster a...   Neutral
1216  they cheated customers by creating hype it sto...   Neutral
1217  please give protection to the battery because ...   Neutral
1218  servicing and maintenance is very low cost in ...   Neutral

[160 rows x 2 columns]
```

In [14]:
```
print(df)
```

```
                                                Reviews sentiment
0     the new ola s1 pro costs around 116lacs onroad...  Positive
1     i believe ola s1 was released into the market ...  Positive
2     excellent scooter with lots of hightech featur...  Positive
3     all misleading mileage claims made by the comp...  Positive
4     im noticing too many electric twowheelers onro...  Positive
...                                                 ...       ...
1214  one of the quickest electric scooters now on i...   Neutral
1215  okaya faast has a digital instrument cluster a...   Neutral
1216  they cheated customers by creating hype it sto...   Neutral
1217  please give protection to the battery because ...   Neutral
1218  servicing and maintenance is very low cost in ...   Neutral

[1219 rows x 2 columns]
```

In [15]:
```
ax=df['sentiment'].groupby(df['sentiment']).count().plot(kind='bar')
ax.bar_label(ax.containers[0])
plt.show()
```



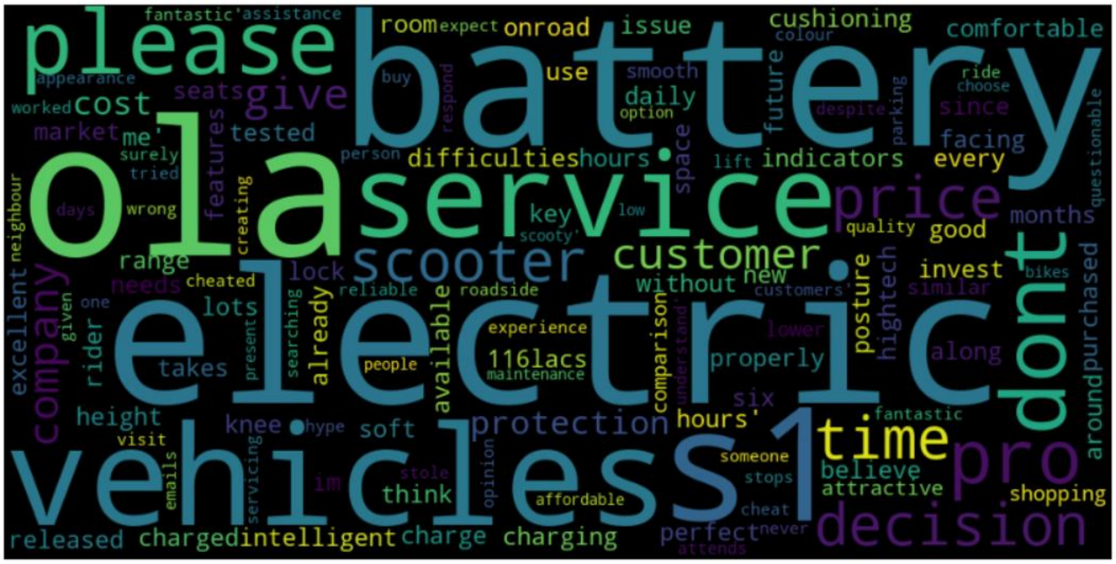
In [16]:
```
from nltk.corpus import stopwords
```

In [17]:
```
stop_words=set(stopwords.words("english"))
```

In [18]:
```
from wordcloud import WordCloud
```

In [19]:
```
wordcloud=WordCloud(width=1000,height=500,stopwords=stop_words,
                    max_words=3000,repeat=False).generate(
    str(df.Reviews.values))
```

In [20]:
```
plt.figure(figsize=(15,8))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



In [21]:
```
import nltk
nltk.download('punkt')
```

In [22]:
```
from nltk.tokenize import word_tokenize
```

In [24]:
```
reviewwords=df.Reviews.apply(word_tokenize)
```

In [25]:
```
reviewwords=reviewwords.apply(lambda
                              x:[w for w in x if w not in stop_words])
```

In [26]:
```
from nltk.probability import FreqDist
from nltk import flatten
```

In [27]:
```
reviewwords=reviewwords.to_list()
```

In [28]:
```
reviewwords=flatten(reviewwords)
```

In [29]:
```
wordfreq=FreqDist(reviewwords)
```

```
In [34]: X_vector.shape
Out[34]: (1219, 180)
```

```
In [30]: wordfreq.most_comm
Out[30]: [('bike', 831),
 ('good', 674),
 ('scooter', 656),
 ('electric', 469),
 ('battery', 375),
 ('also', 312),
 ('range', 300),
 ('service', 284),
 ('best', 253),
 ('vehicle', 236),
 ('price', 230),
 ('like', 230),
 ('speed', 225),
 ('one', 215),
 ('buy', 213),
 ('charge', 211),
 ('features', 208),
 ('riding', 191),
 ('ride', 189),
 ('performance', 187),
 ('charging', 180),
 ('ola', 165),
 ('mileage', 163),
 ('time', 162),
 ('looks', 158),
 ('km', 150),
 ('great', 149),
 ('experience', 148),
 ('company', 145),
 ('quality', 144)]
```

```
In [31]: from sklearn.feature_extraction.text import CountVectorizer
```

```
In [32]: vectorizer=CountVectorizer(max_features=180,min_df=3,ngram_range=(2,2),lowercase=False,stop_words=stop_words,preprocessor=None)
```

```
In [33]: X_vector=vectorizer.fit_transform(df.Reviews)
```


```
In [34]: X_vector.shape
Out[34]: (1219, 180)
```

In [35]:
```
pd.DataFrame(X_vector.toarray(),columns=vectorizer.get_feature_names_out()).head()
```

Out[35]:

	10 days	100 kms	45 kmph	95 km	also good	also great	amazing bike	ampere magnus	ather 450x	attends service	...	top speed	tvs iqube	two people	using tvs	value money	vehicle good	want buy	waste money	worst experience	w recomm
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	0	0	

5 rows × 180 columns

In [36]:
```
y=df.sentiment
```

In [37]:
```
y.value_counts()
```

Out[37]:
```
Positive    900
Neutral     160
Negative    159
Name: sentiment, dtype: int64
```

In [38]:
```
from sklearn.preprocessing import LabelEncoder
```

In [39]:
```
y=LabelEncoder().fit_transform(y)
```

In [40]:
```
from sklearn.model_selection import train_test_split
```

In [41]:
```
# Splitting the dataset, 80% for training and 20% for testing
x_train, x_test, y_train, y_test= train_test_split(X_vector, y, test_size=0.3, random_state=42)
```

In [42]:
```
#Logistic Regression
```

In [43]:
```
from sklearn.linear_model import LogisticRegression
```

In [44]:
```
logit=LogisticRegression(multi_class='multinomial',max_iter=3000)
```

In [45]:
```
logitmodel=logit.fit(x_train,y_train)
```

In [46]:
```
logitmodel.score(x_train,y_train)
```

Out[46]: 0.8909730363423212

In [47]:
```
logitpredict=logitmodel.predict(x_test)
```

In [48]:
```
pd.crosstab(y_test,logitpredict)
```

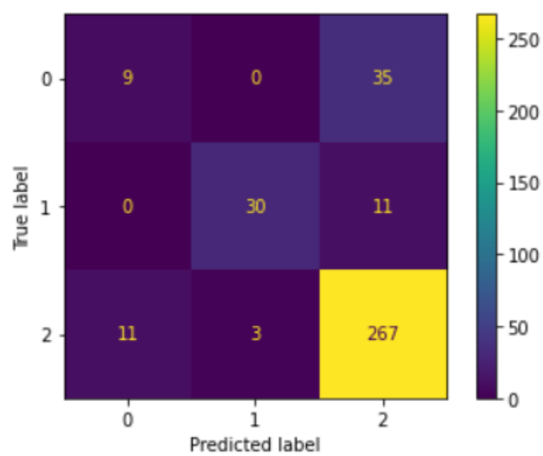
Out[48]:

col_0	0	1	2
row_0			
0	9	0	35
1	0	30	11
2	11	3	267

In [49]:
```
from sklearn.metrics import ConfusionMatrixDisplay
```

In [50]:
```
ConfusionMatrixDisplay.from_predictions(y_test,logitpredict)
```

Out[50]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b7d09f250>



In [51]:
```
from sklearn.metrics import classification_report
```

In [52]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test,logitpredict))
```

```
              precision    recall  f1-score   support

           0       0.45      0.20      0.28        44
           1       0.91      0.73      0.81        41
           2       0.85      0.95      0.90       281

    accuracy                           0.84       366
   macro avg       0.74      0.63      0.66       366
weighted avg       0.81      0.84      0.81       366
```

In [53]:
```
from sklearn import metrics
print('Model Accuracy: ',metrics.accuracy_score(y_test,logitpredict))
```

```
Model Accuracy:  0.8360655737704918
```

In [54]:
```
#Decision Tree
```

In [55]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [56]:
```
tree=DecisionTreeClassifier()
```

In [57]:
```
treemodel=tree.fit(x_train,y_train)
```

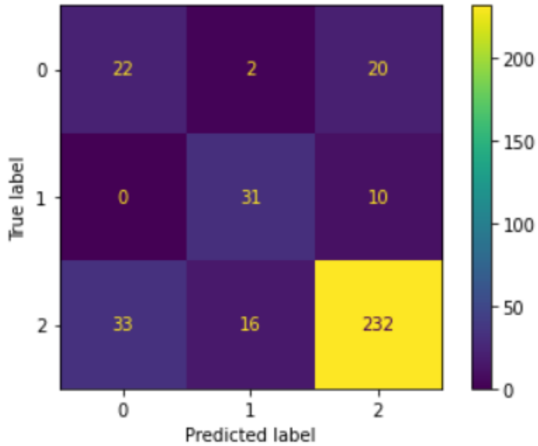
In [58]:
```
treemodel.score(x_train,y_train)
```

Out[58]:
```
0.9249706916764361
```

In [59]:
```
treepredict=treemodel.predict(x_test)
```

In [60]:
```
ConfusionMatrixDisplay.from_predictions(y_test,treepredict)
```

Out[60]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b7f4fd6a0>



In [62]:
```
print(classification_report(y_test,treepredict))
```

```
              precision    recall  f1-score   support

           0       0.40      0.50      0.44        44
           1       0.63      0.76      0.69        41
           2       0.89      0.83      0.85       281

    accuracy                           0.78       366
   macro avg       0.64      0.69      0.66       366
weighted avg       0.80      0.78      0.79       366
```

In [63]:
```
from sklearn import metrics
print('Model Accuracy: ',metrics.accuracy_score(y_test,treepredict))
```

```
Model Accuracy:  0.7786885245901639
```

In [64]:
```
#Random Forest
```

In [65]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [66]:
```
RF=RandomForestClassifier(n_estimators=1000).fit(x_train,y_train)
```

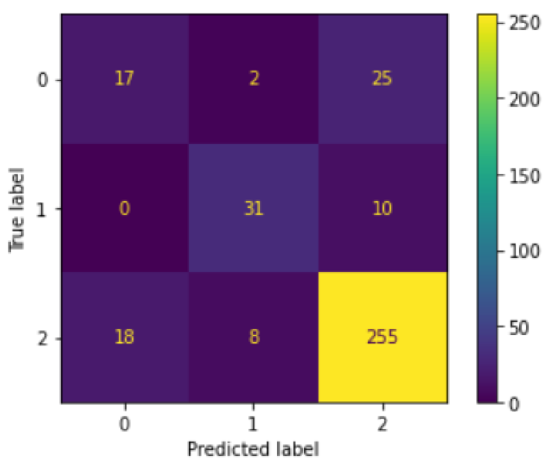
In [67]:
```
RF.score(x_train,y_train)
```

Out[67]:
```
0.9249706916764361
```

In [68]:
```
RFpredict=RF.predict(x_test)
```

In [71]:
```
ConfusionMatrixDisplay.from_predictions(y_test,RFpredict)
```

Out[71]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b7f8cbbe0>
```



In [70]:
```
print(classification_report(y_test,RFpredict))
```

```
              precision    recall  f1-score   support

           0       0.49      0.39      0.43        44
           1       0.76      0.76      0.76        41
           2       0.88      0.91      0.89       281

    accuracy                           0.83       366
   macro avg       0.71      0.68      0.69       366
weighted avg       0.82      0.83      0.82       366
```

In [72]:
```
from sklearn import metrics
print('Model Accuracy: ',metrics.accuracy_score(y_test,RFpredict))
```

```
Model Accuracy:  0.8278688524590164
```

In [73]:
```
#Support Vector Machine
```

In [74]:
```
from sklearn.svm import SVC
```

In [75]:
```
svm = SVC()
```

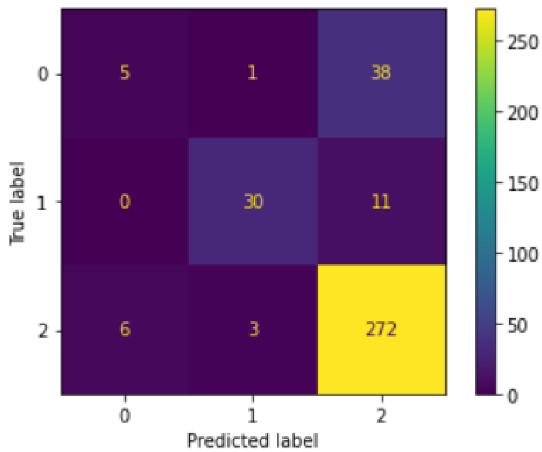
In [76]:
```
svmmodel = svm.fit(x_train,y_train)
```

In [77]:
```
svmmodel.score(x_train,y_train)
```

Out[77]:
```
0.8944900351699883
```

In [78]:
```
svmpredict = svmmodel.predict(x_test)
```

```
In [81]: ConfusionMatrixDisplay.from_predictions(y_test,svmpredict)

Out[81]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b023e6370>
```




```
In [80]: print(classification_report(y_test,svmpredict))

              precision    recall  f1-score   support

           0       0.45      0.11      0.18        44
           1       0.88      0.73      0.80        41
           2       0.85      0.97      0.90       281

    accuracy                           0.84       366
   macro avg       0.73      0.60      0.63       366
weighted avg       0.80      0.84      0.81       366
```

```
In [82]: from sklearn import metrics
         print('Model Accuracy: ',metrics.accuracy_score(y_test,svmpredict))

Model Accuracy:  0.8387978142076503
```

In [83]:
```python
#Naive Baye's
```

In [84]:
```python
from sklearn.naive_bayes import BernoulliNB
```

In [85]:
```python
nb = BernoulliNB()
```

In [86]:
```python
nbmodel = nb.fit(x_train,y_train)
```

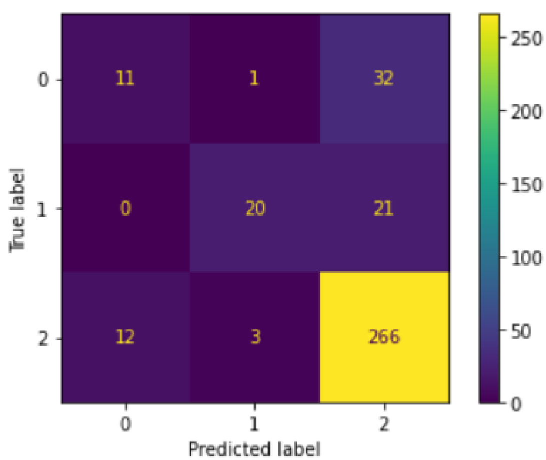
In [87]:
```python
nbmodel.score(x_train,y_train)
```

Out[87]: 0.8382180539273154

In [88]:
```python
nbpredict = nbmodel.predict(x_test)
```

In [91]:
```python
ConfusionMatrixDisplay.from_predictions(y_test,nbpredict)
```

Out[91]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b7f54c760>



In [90]:
```
print(classification_report(y_test,nbpredict))
```

```
              precision    recall  f1-score   support

           0       0.48      0.25      0.33        44
           1       0.83      0.49      0.62        41
           2       0.83      0.95      0.89       281

    accuracy                           0.81       366
   macro avg       0.72      0.56      0.61       366
weighted avg       0.79      0.81      0.79       366
```

In [92]:
```
from sklearn import metrics
print('Model Accuracy: ',metrics.accuracy_score(y_test,nbpredict))
```

```
Model Accuracy:  0.8114754098360656
```

In [93]:
```
#Gradient Boosting
```

In [94]:
```
from sklearn.ensemble import GradientBoostingClassifier
```

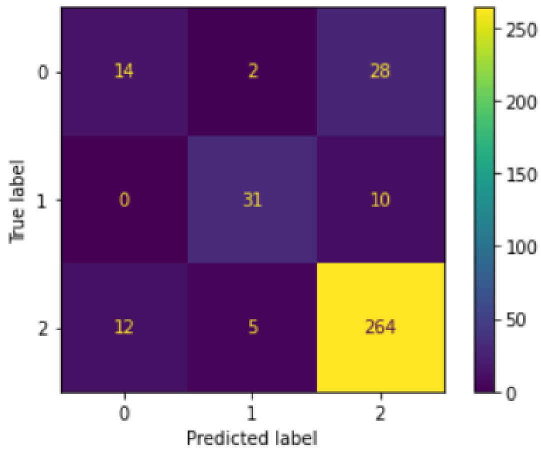
In [95]:
```
gb = GradientBoostingClassifier(n_estimators=4000)
```

In [96]:
```
gbmodel = gb.fit(x_train,y_train)
```

In [97]:
```
gbmodel.score(x_train,y_train)
```

Out[97]:
```
0.9249706916764361
```

In [98]:
```
gbpredict = gbmodel.predict(x_test)
```

In [101]:
```
ConfusionMatrixDisplay.from_predictions(y_test,gbpredict)
```
Out[101]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x25b0353ffa0>
```


In [100]:
```
print(classification_report(y_test,gbpredict))
```
```
              precision    recall  f1-score   support

           0       0.54      0.32      0.40        44
           1       0.82      0.76      0.78        41
           2       0.87      0.94      0.91       281

    accuracy                           0.84       366
   macro avg       0.74      0.67      0.70       366
weighted avg       0.83      0.84      0.83       366
```

In [102]:
```
from sklearn import metrics
print('Model Accuracy: ',metrics.accuracy_score(y_test,gbpredict))
```
```
Model Accuracy:  0.8442622950819673
```