

How to Cite:

Kaladevi, P., Karthikram, V., & Kavinprabhu, N. (2022). Electric vehicles rating from you tube comments using machine learning. *International Journal of Health Sciences*, 6(S4), 4995–5005. <https://doi.org/10.53730/ijhs.v6nS4.9228>

Electric vehicles rating from you tube comments using machine learning

Kaladevi P

Assistant Professor, Department of Computer Science and Engineering, K. S. Rangasamy College of Technology, Namakkal, Tamil Nadu, India
Corresponding author email: kaladevi@ksrct.ac.in

Karthikram V

Department of Computer Science and Engineering, K. S. Rangasamy College of Technology, Namakkal, Tamil Nadu, India
Email: karthikramss2428@gmail.com

Kavinprabhu N

Department of Computer Science and Engineering, K. S. Rangasamy College of Technology, Namakkal, Tamil Nadu, India
Email: iamkavinprabhu007@gmail.com

Abstract---Electric motors play the notable position inside the automobile industry. A swiftly growing technology that ambitions to lessen using fossil fuels in the automobile quarter. The electric vehicles produced in enormous scale in now a day. Now many organizations are liberating their electric powered motors and in destiny it will become the most famous one. It reduces the air pollutants and need of fossil fuels to run a vehicle is constantly reduced. The Charging cost is comparatively lower than regular vehicles. The electric vehicles run on Lithium-ion batteries which will reduce the air pollutants in an effective manner. But in producing the Lithium-Ion batteries also can cause pollution. Electric powered motor's function which have fewer shifting parts and are therefore greater environmentally friendly. This challenge explains the people perspectives that the era of electric vehicles prompted. Perspectives and ideas that the humans think about electric powered motors via the remarks on YouTube may be explored. The positive and drawbacks of the electrical vehicles given by the people can be analysed. The peoples who work in industries can realize how electric powered cars creates an effect in surroundings. They percentage about their thoughts about electric powered motors through You tube comments. Those remarks will provide a result about the opinion of many humans if you want to be beneficial for the surroundings. This venture additionally analyses predominant features of the electrical

vehicles. The fundamental evaluation approximately electric powered cars is their mileage, fee, charging time and also the pollution level as a result of them. People decides to buy regular vehicles which gives them better mileage as per their uses . They do now not need to spend most of their time in Charging Stations. This will help humans to pick a higher choice in shopping for motors. Agencies also can get ideas from professionals how they want to layout their automobiles efficaciously to the human beings. Human beings also can understand about the merits and de-deserves of electrical automobile.

Keywords---machine learning, KNN algorithm, NLP.

Introduction

Electric vehicles which has created a great impact in Automobile industry. To reduce the air pollutants electric vehicles were introduced. The maintenance of electric cars is less when compared to cars runs on fossil fuels. The electric vehicles which runs on lithium-ion battery which requires lesser amount of fossil fuels to run the vehicle. Electric cars which draw the energy source to drive their wheels from battery parks which runs on electricity rather than mechanical engines. It is eco-friendly and it is less pollutant when compared to vehicles running on fuels. The adoption of electric cars reduces the amount of oil that a nation imports in order to provide energy for its nation. Electric vehicles which will reduce the high rate of pollution.

Electric fuelling stations still in the development stages. Lot of places does not have the charger facility to charge your electric vehicles at rural area. It may lead to some troubles for the family trips using electric cars. The electric vehicles have its own positives and negatives. The electric vehicles technology is on a boom and is still increasing. Because of this, people share their opinion about electric vehicles through commenting in various applications like You Tube. The electric vehicles are very much costly when compared to the fuel vehicles. It is not suitable for every person as it costs so high when it is compared with other non-electric vehicles. Many electric cars have a range 52-98 miles and it should be recharged again to drive further. The time taken to recharge an electric vehicle completely can takes 300 to 420 minutes. Most of the electric cars comes up with a smaller number of seats. It does be suitable to plan a trip with the family. The electric vehicles which run using the Lithium-ion batteries. The production of Lithium-ion Battery also causes air pollution. The pie-chart can help the people to better understand the positives and negatives of electric vehicles. The analysis dashboard can hold analysis of electric vehicles basic features in it. The project can help companies to improve the facilities of the electric vehicles.

Literature Survey

Twitter sentimental analysis is used for identifying the similar text patterns present in the input text. The system has been tested with the live data set as well as offline dataset, and the result showed. With the help of geolocation as parameter for filtering the tweets, we can identify whether the tweets from India

or other country. Converting it to an excel file and removing special characters and misspelled words and removing hashtags, numbers and link to make data effective. The term feeling defined as the personal explanation of emotion, private in experiencing them specifically which exists for very small duration. The term mood is defined as sentimental state of mind which is related to emotion exists for long duration but less concentrated. The existing classification techniques giving less importance in pre-processing to filter noise which plays a major role in increasing the accuracy of the algorithm. Further, the existing works are limited in the field of crowd-sensing which requires attention. In this work, real time Twitter data is utilized by the framework to address the above problems. Lemmatizing and Stemming- stemming reduce the words to their stems, lemmatize done the same work as stemmer but keeping the linguistics of the word context. The analysis is classified with emotional, positive and negative tweets. Finally, text is collected as input and converted into reactions using lexicon-based classification. And then finally the positive and negative responses are classified.

The amount of people increased significantly taking their foods or meals in restaurants due to rise their financial capability, miscellaneous food habits and improved lifestyle. Any restaurant with plenty of positive reviews attain symbol of faith, ensure quality of foods and services among the customers. They can give the food reviews in their natural language. It is challenging to create a corpus that comprises a big number of Bangla text reviews on restaurants due to a lack of resources in Bengali and researchers who do not publish their datasets. To find and classify the reviews of the Bengali text and to help the user or the restaurant cooks to classify the positives and negatives of the food in this. Any restaurant with large number of positive reviews can attain a symbol of faith, ensure quality of foods and services among the customers. Through this first we want to import some positive and negative of the Bengali words in the training set preparation. All the reviews are pre-processed and unwanted which does not denote about the reviews are removed from the reviews. Extracted features are used to train the proposed model that could classifies reviews into positive and negative sentiment. Multinomial naive bayes performs really well when multiple occurrences of the words have significant impact in the classification problem.

For every subject, multiple chats are taken and translated and using a neural network, each sentence emoji is scored in a dimensional form. The emotions classified into Categorical or Dimensional. The composition of the emotions expressed by the subject (out of Happy, Sad, Bored, Fear, Anger and Excitement) are also defined. The scores are added up for each subject. All the subjects are asked to share their conversations without media, since the model requires only textual data. WhatsApp allows different skin colours for emojis. Regex library in combination with the Unicode Library can be used to extract these emojis the analysis results that, the behavioural traits are extracted. The words can be used to train a model to predict the emotion category. Emotion analysis detects type of feelings and identifies the temperament of the user as a combination of emotions, happy, sad, angry, fear, excitement or boredom. It is determined that, if the subject likes to use emojis and if they use it as a replacement for words or as an add-on to express their emotions better. To process sentences, first translate each sentence using Google Trans as sentences are in a combination of English and Hindi. After this each of the translated sentences is passed into a Neutrality

Classifier. Sum of all emojis Scores and Sentence score are added together, to receive the final score. These values are plotted to examine and compare a person's emotional scale. Parallel Dots which is an Artificial Intelligence Platform created Using advanced Deep Learning and Machine Learning techniques.

The majority of current research on content data preparing, centres in the genuine area as opposed to assessment space. Break down feelings, attributes of clients about any items, subjects, or issue. The popular feeling, web is turning into a spreading and exceptionally wide stage where online gatherings, social locales, websites and different destinations contains sentiment and audit of individuals in type of posted messages. Content mining assumes a fundamental job in online gathering feeling mining Opinion mining from online discussion is significantly more troublesome than unadulterated content procedure because of their semi organized qualities. Information mining is registering procedure of finding designs in huge informational collections including strategies at the convergence of AI, measurements, and database frameworks. Opinion examination is generally utilized book mining application that can follow client assumption about an organization. Feeling investigation is one of the most worried parts about recognizing and characterizing suppositions or feelings that are communicated inside a content. We can actualize conclusion mining investigation utilizing Support Vector Machine (SVM) calculation. It is a directed AI calculation utilized essentially for characterization and relapse issues. Adversary classes having twofold highlights SVM draws a line between the classes and for classes having different highlights hyperplanes are drawn. Content conclusion examination, likewise as passionate extremity calculation, has become a thriving outskirts in the content mining network. The expectation of client information over the top patterns encourages the client to know about the assessment of other clients' human conduct.

The reviews based on the features of the smart phone like battery back-up, camera, storage, cost, etc., in our project we are going to use reviews as input and mainly concentrating on positive and negative reviews. The random forest technique decision tree is used as the base learners. Base learner will take the random data as input from the data set by using row sampling with replacement. In boosting each learner will take the data from the data set randomly and finds the incorrectly classified classifiers or weak classifiers and these correctly classified classifiers are combined and trained with the base learners in order to make them as strong classifiers which will increase the accuracy of the model. The stump or root node is having only two leaf nodes, one is correctly classified to another one is incorrectly classified. Initially for all sample classifier weights are same. Learner takes the data from the dataset randomly and finds the incorrectly classified classifiers and those classifiers will be trained again with the decision tree in order to make them into strong classifiers. To understand the reviews of the customer and the business company can improve the product by their reviews. The hybrid model incorporates rules-based, lexicon, based, and machine learning approaches into a unified scheme that the powerful fields of individual classifiers while concurrently attempting to prevent their weaknesses.

Modules and Algorithms

Data Collection and Preprocessing

You tube, one of the leading Social Media platforms. The data is collected from the You tube. The dataset for the project is taken out from You tube by extracting the comments. The video which holds numerous comments about the opinion of the electric vehicle is found and the comments are taken out for the process of analysing the opinion of the electric vehicles. Analyzing the opinion of electric vehicles is an informative thing for the consumers and the automobile companies. The comments about the electric vehicles are taken out using the You tube API. The piece of code which will verify the credentials of the Google API. After successful verification It will allow the user to collect the comments from You tube. The Comments are extracted using python code. The data collected are uncleaned and has many irregular words which can affect the accuracy when it was implemented in Machine Learning algorithms. The data is pre-processed and cleaned for better accuracy. The collected comments written in a CSV file.

Steps involved in Pre-processing of data

- The dataset is to be pre-processed for the implementation of the algorithm.
- The strip () function in the string, which is used to remove the trailing and leading whitespaces from the string.
- The lower () function which is used to convert all the uppercase characters into lowercase. If it already in lowercase it will return the original string without a change.
- These inbuilt functions in python which will help to clean a dataset in a suitable manner.

Here the comments from the You tube can be converted into cleaned dataset by implementing the functions. The data frame is cleaned and it will be moved to the next step for analyzing the opinion.

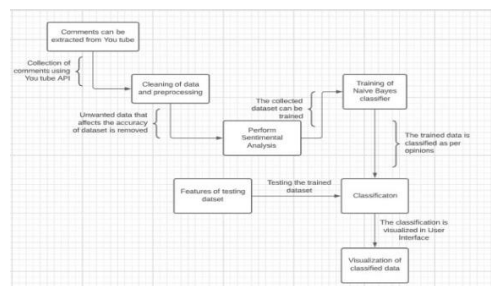


Fig 3.1 Project Architecture

Analyzing the opinion

Sentiment analysis, which is a machine learning tool that analyses texts for polarity, from positive to negative. By training machine learning tools with some examples of emotions in text, machines automatically learn how to detect

sentiment without human input. Machine learning is a method of analyzing the data that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from identify and learn patterns and make decisions with minimal human intervention. Resurging interest in machine learning is due to some of the same factors that have made data mining and Bayesian analysis more popular than ever. Things like growing volumes and varieties of available data, computational processing that is cheaper and more powerful, and has an affordable data storage. Sentiment analysis techniques used to analyze opinionated text, which contains opinions of people toward entities such as products, organizations, individuals, and events. Businesses increasingly capturing more data about their customers' sentiments that has led to the proliferation of sentiment analysis.

Naive Bayes classifier works on the principles of conditional probability. Conditional probability, calculated by multiplying the probability of the preceding event by the updated probability of the succeeding, or conditional, event. Naïve is independent of the occurrence of other features. Bayes' theorem, also called as Bayes' Rule or Bayes' law. The data collected after data pre-processing can be handled separated using Bayes classifications. Naïve Bayes Algorithm works on the principle on conditional probability. Multinomial Naïve Bayes used for classification.

Formula of Bayes Algorithm

$$P(A|B) = P(B|A) P(A)/P(B)$$

$P(A|B)$ is the Probability of hypothesis A on observed event B.

$P(B|A)$ is the Probability of the evidence is true.

$P(A)$ is the Probability of hypothesis before observing the given evidence.

$P(B)$ is the Probability of the Evidence.

The Multinomial Naive Bayes algorithm is a Bayesian learning approach which is popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as newspaper or email story, using the Bayes theorem. The Multinomial Naïve Bayes classifier used for multinomial distributed data. It is primarily used to classify document problems, it means a that the document belongs to which category such as Marketing, Sports, Politics. It is analysed as one of the best algorithms in text classification. Naïve Bayes algorithm classifies the comments of the people into positive and negative. Text classification is best in the Naïve Bayes Algorithm. The data accuracy is better than the KNN algorithm. The Multinomial Naïve Bayes algorithm which is better in performing the text classification. The Multinomial Naïve Bayes approach which is the very popular in the Natural Language Processing Techniques. It analyses the given sample and produce the high accuracy output in the text classification methods.

```
[ ] from sklearn.naive_bayes import MultinomialNB

model = MultinomialNB()
model.fit(xtrain, ytrain)

MultinomialNB()
```

3.2 Implementation of Naïve Bayes Algorithm

Visualization

Visualization is a technique for creating images, diagrams, or animations to communicate a message. The use of visualization is to be present the data or information of the particular product in an effective manner. Django, a collection of Python libraries allowing you to quickly and efficiently create a quality Web application, and is suitable for both frontend and backend. Django, a high-level Python web framework that enables rapid development of secure and maintainable websites. Charts, which is an essential part of working with data, as they are a special way to condense large amounts of data into an easy-to-understand format. The pie chart will be very effective in displaying the results effectively for the user. The charts will bring the understanding visualization for the project. It will help the people and companies to understand the how much it is positive and how much it is negative.

Data Visualization is the presentation the data in graphical format. It helps both educated and uneducated to people understand the significance of data by summarizing and presenting a huge amount of data in a simple and easy-to-understand format and it helps communicate information clearly and effectively. Developers can also use plugins to extend the web application. The information collected about the product should be visualized for further. The visualization of information can give a better view for the consumers and the automobile companies. Django which is good for search engine optimization. Django is highly scalable. Django is highly securable. The data collected from the machine learning algorithms can be visualized using Django for better understanding for the companies about the opinion of the product. Charts, which are used in critical situations where a simple table won't adequately demonstrate important relationships or patterns between data points. The charts can hold all the analysis details of the Electric Vehicles such as Mileage, Price, Capacity, Charging time in it.

Procedure

The You tube video uploaded by an expert on views of electric vehicles chosen out for analysing the comments. Then the numerous numbers of comments of that video took out using the You tube API provided by Google. Notebook documents contains the inputs and outputs of an interactive session as well as additional text that accompanies the code but is not meant for execution. This API allows the developers to take out the comments of every video in google by entering the You tube video id. This API can be generated by creating it in google developer's website. After the collection of comments, it has enormous emojis and unnecessary words in the comments. It can affect the accuracy of the ML algorithms. The emojis in the dataset can be removed by a program written in python. This python program removes all the emojis from the dataset. Then the data is cleaned and unwanted words are removed by programs written in python. The dataset is cleaned and it is effective for implementing ML algorithms. There are many ML algorithms that can be used for text classification. The Naïve Bayes Algorithm is used for classification of comments into positive and negative comments. The Naïve Bayes Algorithm took less time for training the data and gives out a better accuracy for the dataset. After the classification of comments, it

electric vehicles. The pie chart can hold the people's opinion in form of graphical representation which improves understanding the people's opinion. The pie chart representation of comments. The Dashboard can also hold the other analysis details of Electric vehicles such as Mileage, Capacity, Charging time and pollution range. It will help the common people to understand it by graphs. It will help people to their perfect vehicles for them. And it also shows companies that what they want to improve in their Electric Vehicles.

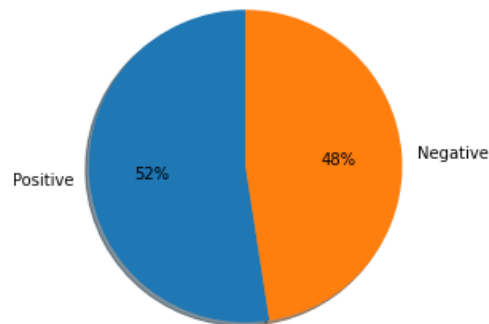


Fig 5.3 Pie Chart Analysis of You tube Comments

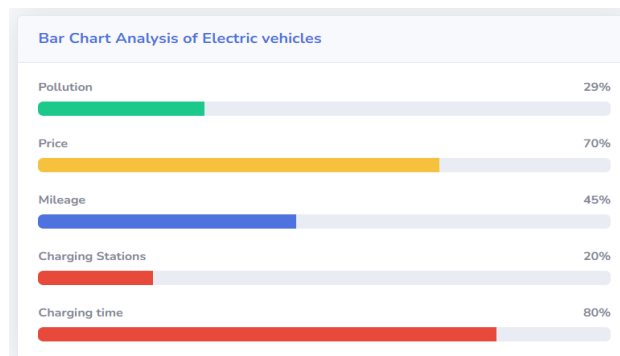
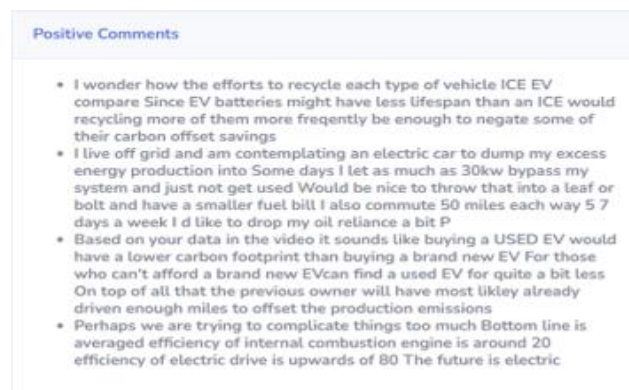
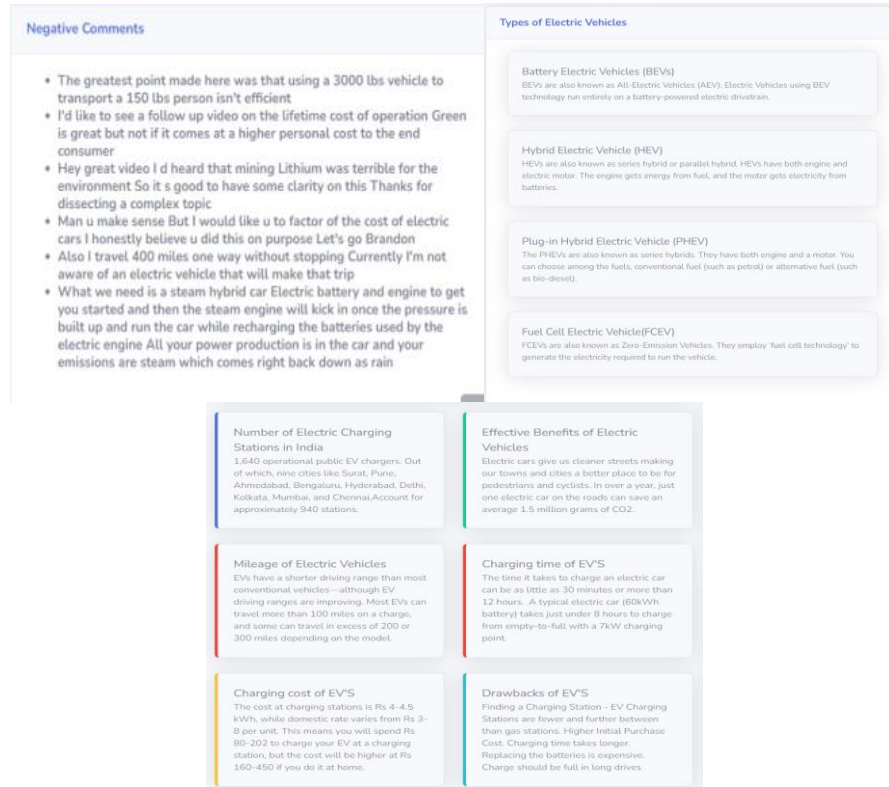


Fig 5.4 General Analysis on Features of Electric Vehicles



It also hold the details of the Electric Vehicles such as its types and basic details should be described in separate rows. So, dashboard can hold all the basic details of its Electric vehicles. Time taken for an electric vehicle to be charged. The dashboard implementation can provide a better understanding for the public.



Conclusion

Sentimental analysis is a powerful way of finding the opinion of every product. It may be a review of a product or a rating of the product. It will help people to identify the positive, negative, neutral things of the product. Electric vehicles are one of the best innovations in this scientific world. Identifying the opinion of the electric vehicles can help companies how to improve the features of their electric vehicles. It is proposed that multiple comments about electric vehicles from YouTube could be taken out and analysed to find out the opinion of the common people. The comments taken out can be pre-processed and moved to the process of implementation of Multinomial Naïve Bayes algorithm. In these comments can be separated and it is proceeded for visualization. In this visualization part, the graph is plotted with the help of the separated comments. Then the dashboard can also hold the important features of the electric vehicles such as Mileage, Time taken for an electric vehicle to be charged. The dashboard implementation can provide a better understanding about the opinion of the public to the corporate companies.

References

- Lopez, M. M. L., Herrera, J. C. E., Figueroa, Y. G. M., & Sanchez, P. K. M. (2019). Neuroscience role in education. *International Journal of Health & Medical Sciences*, 3(1), 21-28. <https://doi.org/10.31295/ijhms.v3n1.109>
- Omar Sharif, Mohammed Moshikul Hoque, Eftekhari Hossain., Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naive Bayes, 2019, International Journal on Advances in Science, Engineering and Robotics technology.
- Poornima A, K. Sathiya Priya., A Comparative Sentiment Analysis of Sentence Embedding using Machine Learning Techniques, 2020, International Journal on Advanced Computing & Communication Systems.
- R. Jebersonretna Retna Raj, Prasanjeet Das, Prabat Sahu. , Emotional Classification on Twitter Data Using Word Embedding and Lexicon Based Approach., 2020, International Journal on Communication Systems and Network Technologies.
- Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
- Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
- Sathish Kumar, Mohamed Nabeem P, Manoj C K, Jeyachandran K Sentimental Analysis (Opinion Mining) in Social Network by using SVM Algorithm, International Journal on Computing Methodologies and Communication.
- Siva Kumar Pathuri, Dr.N.Anbazhagan , Dr.G.Balaji Prakash Feature Based Sentimental Analysis for Prediction of Mobile Reviews using Hybrid Bag-Boost Algorithm, 2020, International Journal on Smart Structures and Systems.
- Sonika Dahiya, Astha Mohta, Atishay Jain., Text Classification based Behavioral Analysis of WhatsApp Chats, 2020, International Journal on Communication Systems and Network Technologies.
- Yin Fu Huang, Yi-Hao Li., Sentiment Translation Model for Expressing Positive Sentimental Statements, 2019, International Journal on Machine Learning and Data Engineering.
- Yongxixi Jing Heping Gou, Chuanyi Fu, Wei Sun., Sentiment Classification of Online Reviews Based on LDA and Semantic Analysis of Sentimental work, 2019, International Journal on Computational Intelligence and Design.