SENTIMENT ANALYSIS OF TRAVELLING PASSENGERS USING MACHINE LEARNING

Amilanshu Gupta[1], Devanshi Dwivedi[2], Janhwi Singh[3]

[1]*BTECH(pursuing), Department of Computer Science Institute of Technology and Management(ITM), Gida, Gorakhpur, U.P, 273001, India*

[2]   *BTECH(pursuing), Department of Computer Science Institute of Technology and Management(ITM), Gida, Gorakhpur, U.P, 273001, India*

[3]   *BTECH(pursuing), Department of Computer Science Institute of Technology and Management(ITM), Gida, Gorakhpur, U.P, 273001, India*

---***---

Abstract –

Sentiment analysis is a vast and customer centric technology used with the help of machine learning algorithm. Nowadays Sentiment analysis are used in Social media monitoring, customer support, customer feedback, Brand monitoring and reputation management, Voice of Customer(VoC), Voice of employee, Product management, Market research and competitive research and in various fields. In this project of sentiment analysis which is "Sentiment Analysis of Travelling Passengers using Machine Learning".I am targeting to find the sentiment of travelling passengers using machine learning algorithm. I took a review of apps such as Red Bus, Make My Trip, Yatra.com from both the google play store and app store and based on the sentiment of those reviews I am able to find sentiments such as: App based rating, App based sentiments, Country based sentiment, Author based sentiment, Subject based sentiment, Language based sentiment by this sentiment analysis an organization can easily predict the behaviour of the customer based on his previous Reviews , by doing the sentiment on the wide range of customer one can easily suggests ads, products and other things based on his analysis.

KEYWORDS: sentiment analysis, machine learning, data preprocessing, user.

1.INTRODUCTION

As Internet is developing greater and its viewpoints are becoming more extensive. Online media and miniature contributing to a blog stages alongside the video and App based stages, for example, Facebook, Twitter, Instagram, Tumblr and other web-based media stages overwhelm in spreading in capsulated news and moving point across the globe at an exceptionally fast speed with simply a one ticks these days anything can be viral across the globe and inside a subsequent it used to reach among many people groups. Estimation investigation is the forecast of feelings in a word, sentence or corpus of archives. It is proposed to fill in as an application to comprehend the perspectives, sentiments and feelings communicated inside an online notice. The expectation is to

acquire an outline of the more extensive popular assessment behind specific points. Correctly, it is a worldview of arranging discussions into positive, negative or impartial marks. Many individuals utilize web-based media locales for systems administration with others and to keep awake to-date with news and recent developments. These destinations (Twitter, Facebook, Instagram, google+) offer a stage to individuals to voice their conclusions. Estimation investigation is the forecast of feelings in a word, sentence or corpus of archives. It is proposed to fill in as an application to comprehend the perspectives, sentiments and feelings communicated inside an online notice.

The expectation is to acquire an outline of the more extensive popular assessment behind specific points For instance, individuals rapidly post their surveys online when they watch a film and afterward start a progression of remarks to examine about the acting abilities portrayed in the film. This sort of data frames a reason for individuals to assess, rate about the presentation of any film as well as about different items and to think about if it will be a triumph. This sort of huge data on these locales can utilized for promoting and social investigations . Along these lines, estimation examination has wide applications and incorporate feeling mining, extremity, grouping and impact investigation. Estimation Analysis is frequently completed at two levels

 1) coarse level and 2) fine level.

In coarse level, the examination of whole archives is done while in fine level, the investigation of traits is done . The feelings present in the content are of two kinds: Direct and Comparative. In similar assessments, the correlation of items in a similar sentence is included while in direct estimations, objects are autonomous of each other in a similar sentence. Nonetheless, doing the investigation of tweets communicated in not a simple work. A ton of difficulties are engaged with terms of resonance, extremity, vocabulary and punctuation of the tweets. They will in general be profoundly unstructured and non-linguistic. It gets hard to decipher their significance. In addition, broad use of slang words, abbreviations and out of jargon words are very normal while tweeting on the web. The classification of such words per extremity gets intense for regular processors included. This venture utilizes Apache Spark's quick handling capacities to investigate opinion from such high speed ongoing tweets.

2. LITERATURE REVIEW

The authors proposed a effective method of predicting polarity using sentiment analysis – **Akhil and Pavithra, year(2018) .**
The authors proposed an approach to the prediction of French election based – **Lei Wang, year(2017) .**
The authors proposed various techniques for automatically classifying the sentiment of twitter message – **Sarang, year(2016).**
The author proposed an approach for a survey on sentiment analysis of twitter data – **Akhil and Pavithra, year(2017).**
The authors presented the warning of seat availability – **Wusheng Leu, year(2016).**
The authors proposed a rating concept that allows the evaluation of public transport – **Nur Khaleeda, year(2019)**

3. SENTIMENT ANALYSIS PROCESS

3.1 Data Preprocessing:-

 In any Machine Learning measure, Data preprocessing is that progression where information gets changed or encoded. To carry it to such an express that now the machine can undoubtedly parse it. At the end of the day, the highlights of information would now be able to be effectively deciphered by the calculation .

3.2 Features of Machine Learning:-

A dataset can be seen as an assortment of information objects, which are regularly additionally called as a record, focuses, vectors, designs, occasions, cases, test, perceptions or substances. Information objects are depicted by various highlights that catches the essential qualities of an item, like the mass of actual articles or at the time at which the occasion happened. Highlights are regularly called as factors, qualities, fields, traits or measurements.

3.3 Data Quality Assesment:-

Because information is frequently taken from various ources which are typically not very solid and that too in various arrangements, the greater part within recent memory devoured in managing information quality issues when chipping away at an AI issue. It is basically unreasonable to expect that the information will be awesome. There might be issues because of human mistake, limits of estimating gadgets, or defects in the information assortment measure. We should go over a couple of them and techniques to manage them.

3.4 Preprocessing Tasks:-

3.4.1 Casing: - We will be converting all the letters in upper case or lower case.

3.4.2 Noise Removal: - Here we will be eliminating unwanted characters such as HTML tags, Punctuation marks, special characters etc.

3.4.3 Tokenization: - Here we will convert all the tweets into tokens. All the tokens will be in words that are separated in the text have.

3.4.4 Stop word Removal: - Some of the words that are actually doesn't make sense or don't contribute much to the machine learning model.

3.4.5 Text Normalization (Stemming & Lemmatization)

4. METHODOLOGY

We have used the App reviews which is taken from Google play store and App Store. Data of Make My Trip, Red Bus, Yatra.com, Firstly downloaded data from App Bot, after downloading the data we have preprocessed the data using NumPy, Pandas, Matplotlib, Seaborn. After preprocessing the data, I have changed categorical values to the numerical one's and applied statistics mode on the preprocessed dataset. After doing the preprocessing task, Feature selection has been done, Feature selection is the process of reducing the number of input variables when developing a predictive model. After doing the feature selection vector representation takes place where vectors are used to represent numeric characteristics, called features, of an object in a mathematical and easily analyzable way. After doing the vector representation sentiment classification takes place where labelling of the dataset has been done in which it is classified as Positive, Negative and Neutral. In the sentiment classification we got a positive and the negative opinion based on the sentiment we have with the particular dataset. After sentiment classification, Sentiment summarization takes place in which we have summarized all the positive, negative and neutral at one place in which later on recommended the sentiments of the roadways passengers which is rating based sentiment, country-based sentiment, subject based sentiment, author-based sentiment, device-based sentiment, version-based sentiment etc.
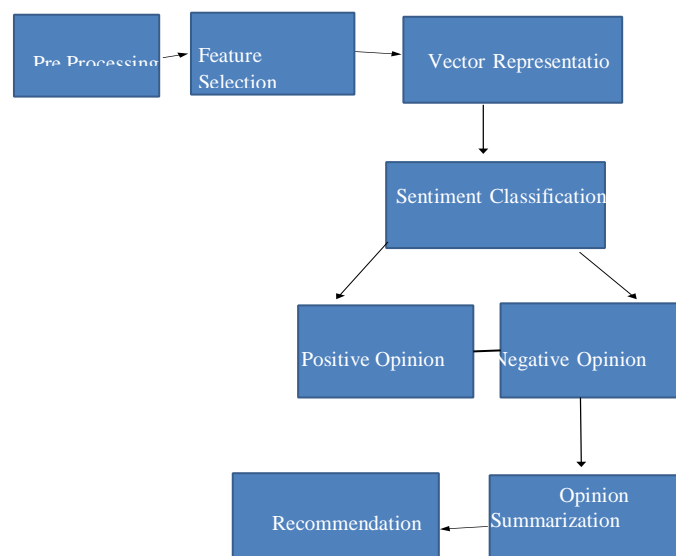
Flowchart of Proposed Model

FIGURE-1

5. MERITS AND DEMERITS OF SENTIMENT ANALYSIS

Merits –

1) Agent monitoring
2) Identifying key emotional triggers
3) Handling multiple customers
4) Adaptive customer service
5) Live insights
6) Quick escalations

Demerits –

1) Incorrectly targeted sentiment
2) Review language is dissimilar

4. CONCLUSIONS

Sentiment Analysis of Travelling Passengers Using Machine Learning has been created with the goal to check the emotions of travelling passengers while they travel in a roadways.
This sentiment can be helpful for the companies to know there customers much better so that they can further make their system customer oriented based on their reviews. It can improve this system by building a Memory-Based.

ACKNOWLEDGEMENT

REFERENCES

[1] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," Bus. Horiz., vol. 53, no. 1, pp. 59–68, 2010.

[2] P. H. Keckley and M. Hoffmann, "Social networks in health care: Communication, collaboration and insights," New York Deloitte Cent. Heal. Solut., 2010.

[3] S. Valenzuela, N. Park, and K. F. Kee, "Is there social capital in a social network site?: Facebook use and college students 'life satisfaction, trust, and participation1," J. Comput. Commun., vol. 14, no. 4, pp. 875–901, 2009.

[4] M. H. Saragih, "The Impact of Social Relationship One-Women Intention in Social Networking Sites (SNS) in Indonesia," Int. Rev. Computer Software ,A Review Paper on Machine Learning Based Recommendation System-International Journal of Engineering Development and Research in 2014.