

SENTIMENT ANALYSIS OF TRAVELLING PASSENGERS USING MACHINE LEARNING

Amilanshu Gupta¹, Devanshi Dwivedi², Janhwi Singh³

¹ B.Tech 4th Year, Dept. of Computer Science & Engineering, ITM Gorakhpur, U.P, India

² B.Tech 4th Year, Dept. of Computer Science & Engineering, ITM Gorakhpur, U.P, India

³ B.Tech 4th Year, Dept. of Computer Science & Engineering, ITM Gorakhpur, U.P, 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", we are targeting to find the sentiment of roadways passengers using machine learning algorithm. we 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 we are 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.

1. INTRODUCTION

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 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

Wusheng Liu,¹ Qian Tan , ² and Wei Wu³ In this research paper author wants to tell the early warning of seat availability, time duration to reach the destination etc. the author had done the search very seamlessly and this research is currently implemented in various companies such as MakeMyTrip, RedBus etc.

Nur Khaleeda Othman , Masnida Hussin, Raja Azlina Raja Mahmood The significance of customer engagement in business operation becomes a new perspective in organization prospect.

The authors in (Windasari et al., 2017) had conducted the sentiment analysis on GoJekby using Twitter dataset and analyzed it using Support Vector Machine (SVM) method. In their research, it reveals that on average 86% in year 2000 most of comments is in Indonesian language. The author has to come up with the English language as well as it has been read by many people in the world.

S. Thirumurugan Halim, Zahid, presented a study on the existing approaches for the detection of unsafe driving patterns of a vehicle used to predict accidents. AI techniques are surveyed for the detection of unsafe driving style and crash prediction. A number of statistical methods which are used to predict the accidents by using different vehicle and driving features are also covered in this paper. The approaches studied in this paper are compared in terms of datasets and prediction performance.

3. MATERIAL AND METHODS

3.1 Data Pre-processing:-

In any Machine Learning measure, Data pre-processing 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.

3.3 Data Quality Assessment:-

Because information is frequently taken from various sources 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 Pre-processing Tasks:-

- **Casing:** - We will be converting all the letters in upper case or lower case.
- **Noise Removal:** - Here we will be eliminating unwanted characters such as HTML tags, Punctuation marks, special characters etc.
- **Tokenization:** - Here we will convert all the tweets into tokens. All the tokens will be in words that are separated in the text have.
- **Stop word Removal:** - Some of the words that are actually doesn't make sense or don't contribute much to the machine learning model.
- **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, we 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.

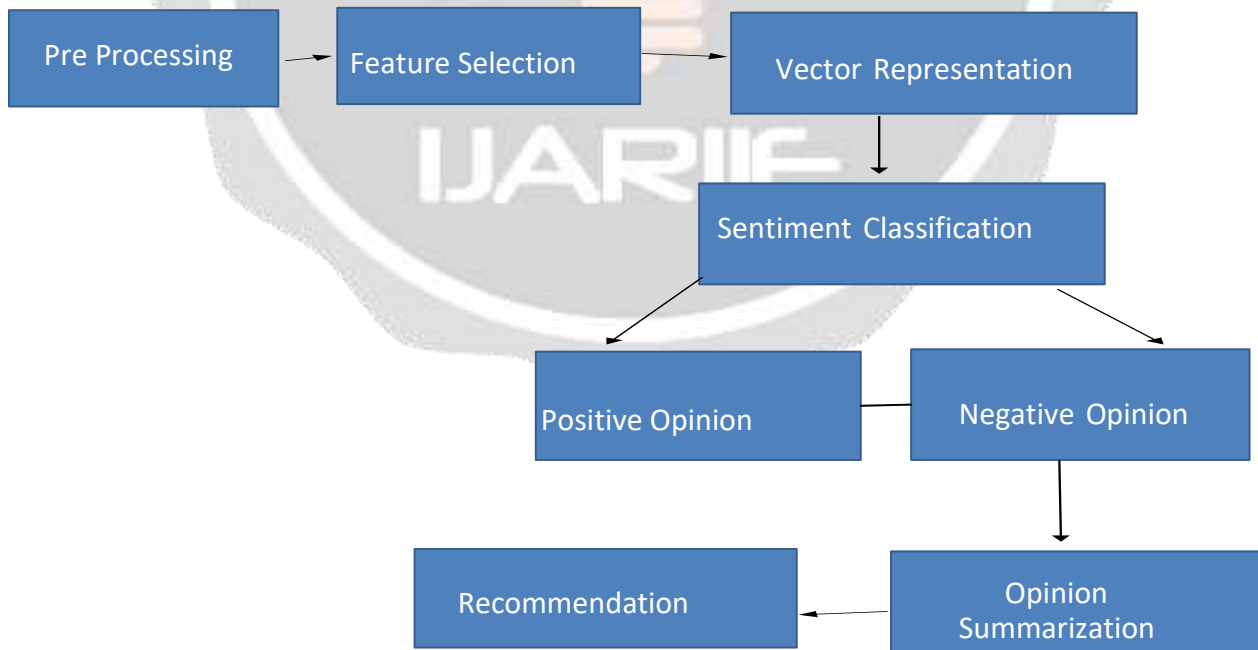


Fig 1: Flowchart of Proposed Model

The implementation part of the project are as follows: -

1. Load in and visualize the data: - Firstly, I have downloaded a data from App Bot which is a raw data. I took three Apps reviews for the analysis of roadways passengers They are as follows Red Bus, Make My Trip and Yatra.com .
2. Data Processing: - In data processing the raw data is processed on the jupyter notebook.
3. Data Imputation and Manipulation
4. Conversion of Categorical Data to mode values
5. Categorical values converted to Numerical Values

5..EXPERIMENTAL RESULTS

I have successfully done the sentiment analysis on the roadways passengers based in the App reviews here are the desired results based on the sentiments.

Rating Based Sentiment: -

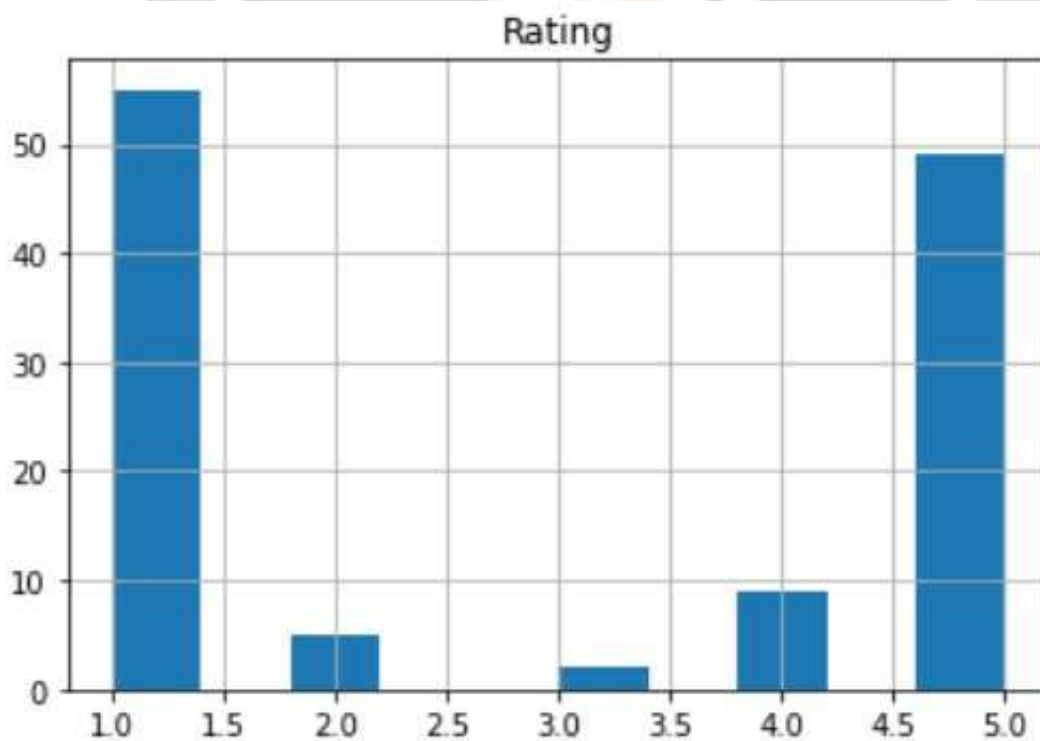


Fig 2: Rating based sentiment

6. 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. Presently, I am performing sentiments by artificially creating it to check the sentiments if the roadways passengers but this sentiment analysis can be very helpful for the companies to make their service everlasting. It can improve this system by building a Memory-Based .

7. 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* , vol. 11, no. 11, 2016.
- [5] P. M. Di Gangi, M. M. Wasko, and R. E. Hooker, "Getting Customers' ideas To Work For You: Learning From Dell How To Succeed With Online User Innovation Communities.," *MIS Q. Exec.*, vol. 9, no. 4, 2010
- [6] P. C. Verhoef, W. J. Reinartz, and M. Krafft, "Customer engagement as 29 a new perspective in customer management," *J. Serv. Res.*, vol. 13, no. 3, pp. 247–252, 2010.
- [7] S. Kabadayi and K. Price, "Consumer--brand engagement on Facebook: liking and commenting
- [8] R. J. Brodie, L. D. Hollebeck, B. Juriü, and A. Iliü, "Customer engagement: conceptual domain, fundamental propositions, and implications for research," *J. Serv. Res.*, vol. 14, no. 3, pp. 252–271, 2011.
- [9] L. R. Men and W.-H. S. Tsai, "Infusing social media with humanity: Corporate character, public engagement, and relational outcomes," *Public Relat. Rev.*, vol. 41, no. 3, pp. 395–403, 201