WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Review Analyzer, a Tool Which Analyzes Sentiments of Reviews

Paresh Shah[1], Shruti Agar[2], Tanishq Goyal[3], Varsha Valwani[4]

Computer Science and Engineering, Acropolis Institute of technology and research, Indore.India

ABSTRACT: -

Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentimental Analysis has gain much attention in recent years. The growing popularity of social media sites has generated a massive amount of data that attracted researchers, decision makers and companies to investigate people's opinions and thoughts in various fields. This research paper aims to obtain dataset of tweets and apply different Machine Learning algorithms to analyze and classify texts. This research paper explored text classification accuracy while using different classifiers for classifying balanced and unbalanced datasets.

Key-Words:- Sentiment Analysis, Natural Language Processing, Classification, Balanced Dataset, Unbalanced Dataset.

Introduction

The recent widening expansion of social media has changed communication, sharing, and obtaining information. In addition, to this, many companies use social media to evaluate their business performance by analysing the conversations' contents.
This includes collecting customers' opinions about services, facilities, and products. Exploring this data plays a vital role in consumer retention by improving the quality of services.It also allowsuser to analyze his/her reviews in terms of positive, negative and neutral score.Social media sites such as Twitter offer valuable data that can be used by business owners not only to track and analyse customers' opinions about their businesses but also that of their competitors. Moreover, these valuable data attracted decision-makers who seek to improve the services provided.
In this research paper, several research papers that studied Twitter's data classification and analysis for different purposes were surveyed to investigate the methodologies and approaches utilized for text classification. The authors of this research paper aim to obtain open-source datasets then conduct text classification experiments using machine learning approaches by applyingdifferent classification algorithms, i.e., classifiers. The authors utilized several classifiers to classify texts of two versions of datasets. The first version is unbalanced datasets, and the second is balanced datasets. The authors then compared the classification accuracy for each used classifier on classifying texts of both datasets.

Problem Formulation

Around quintillions of data is produced every day. And also there is a huge amount of data given by users/customers as a feedback towards a particular commodity or service. Here is a need to analyze what views or thoughts they have towards the same. So sentiment analysis helps us to do the polarity detector for the data feuded by the users. Sentiment analysis plays a very important role in business strategies. Is helps them to do the betterment in services what they are providing currently.
If we talk about product analysis, it will find out what the public is saying about a new product or analyze negative feedback you may have never seen
In terms of social media monitoring around 6,000 tweets are sent every second. Whichever industry you work in –tech, finance, health, government – you probably receive a lot of feedback on social media. And, you're looking at hours, maybe even days, to process all that data manually.
So this models aims to saves manpower to read each and every review and come to the point.

Thus the proposed model has following objectives:
Obj1: To detect the polarity of Sentence.
Obj2: To analyze user reviews about the product.

Obj3: To extract expressions of opinion describing a target feature and classify it as positive, negative orneutral.
Obj4: To extract polarity of speech by media influencers.

Solution Proposed

The Proposal is to deploy a system that takes user reviews or comments either from a product-based company or from social media influencers and apply NLP tools to obtain the polarity of typed sentence. Also, it will extract the most common words, extract total words entered, extract total characters entered,total number of tokens, unique number of tokens and performing a basic as well as semantic analysis.

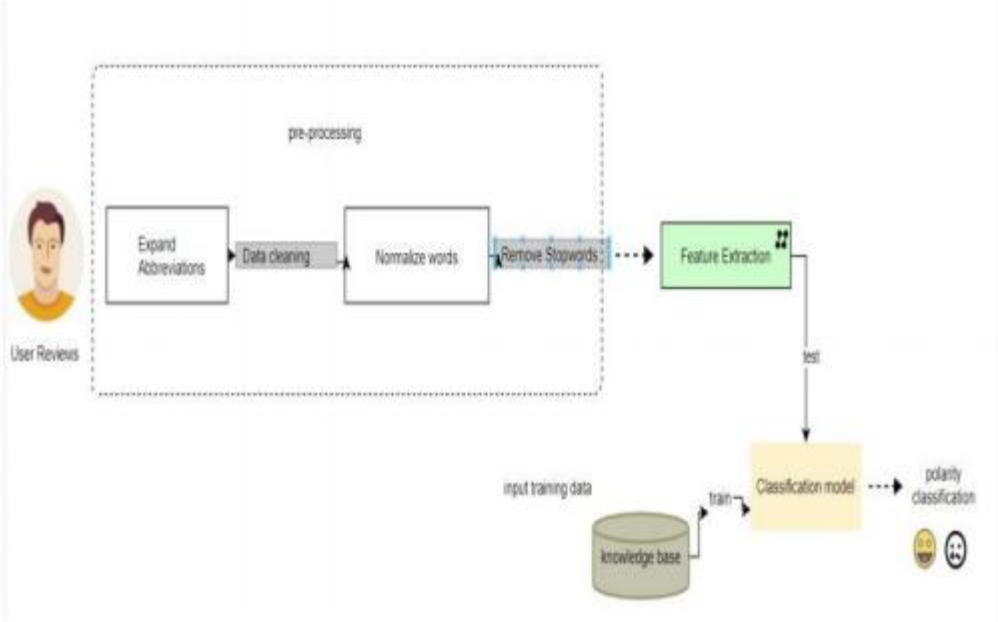

Figure 1 Block Diagram

Figure 2 below shows what actually sentiment analysis and polarity detector do towards the data user haveentered.

The implementation is as follows

1.The user will enter his/her data in English language for analyzing. User can also use mix case sensitive words. Figure 3 below shows use case diagram for the same..

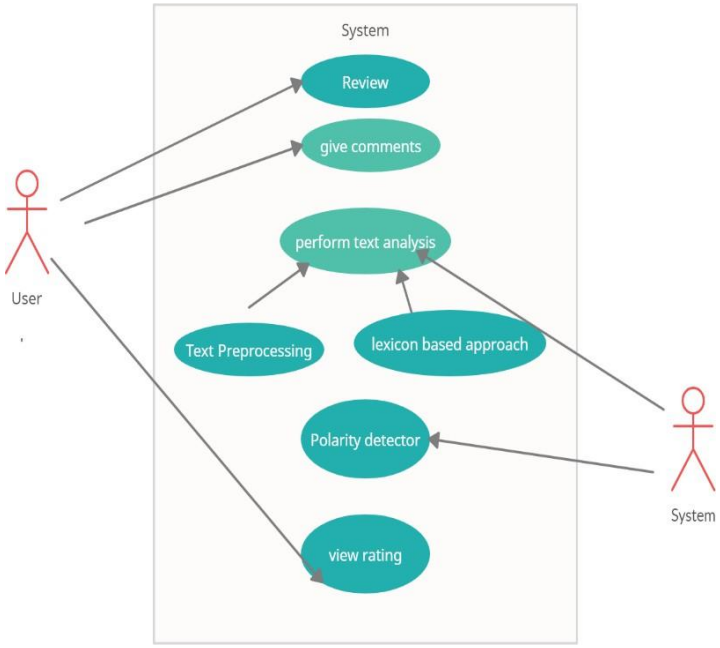

Figure 1. use case diagram

1.Pre-processing of data will be carried out like removal of stop words and punctuation marks which might will add noise to the data.
2.Stemming is carried out.
3.Lemmatization is done.
4.Whole text will be converted into lower case letters.
5.After, pre-processing analyzing will be done via NLP (Natural Language Processing). During the analysis polarity and sentiment will be identified of the entered data.
6.Basic as well as semantic analysis will performed to the entered data.
7.A report on the results will be show to the user via help of front-end module created with (HTML, CSS)
Below attached figure 4 shows the basic analysis and figure 5 shows semantic analysis

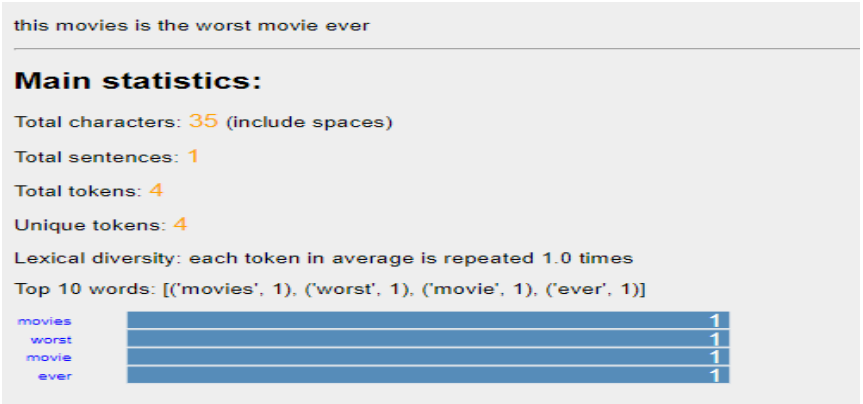

Figure 4.Basic Analysis

Figure 5. Polarity Detector

Literature Review

One fundamental problem in sentiment analysis is categorization of sentiment polarity.Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral). Based on the scope of the text, there are three levels of sentiment polarity categorization, namely the document level, the sentence level, and the entity and aspect level. The document level concerns whether a document, as a whole, expresses negative or positive sentiment, while the sentence level deals with each sentence's sentiment categorization; The entity and aspect level then targets on what exactly people like or dislike from their opinions.

Since reviews of much work on sentiment analysis have already been included in, in this section, we will only review some previous work, upon which our research is essentially based. Hu and Liu summarized a list of positive words and a list of negative words, respectively, based on customer reviews. The positive list contains 2006 words and the negative list has 4783 words. Both lists also include some misspelled words that are frequently present in social media content. Sentiment categorization is essentially a classification problem, where features that contain opinions or sentiment information should be identified before the classification. For feature selection, Pang and Lee suggested to remove objective sentences by extracting subjective ones. They proposed a text-categorization technique that is able to identify subjective content using minimum cut. Gann et al. selected 6,799 tokens based on Twitter data, where each token is assigned a sentiment score, namely TSI (Total Sentiment Index), featuring itself as a positive token or a negative token. Specifically, a TSI for a certain token is computed as:

$$TS\,I = p - tp / tn \times n / p + tp / tn * n \qquad ((1))$$

where p is the number of times a token appears in positive tweets and n is the number of times a token appears in negativetweet. tptntptn is the ratio of total number of positive tweets over total number of negative tweets.

Methodology

Your most unhappy customers are your greatest source of learning." — Bill Gates

This line of Bill Gates has a very important role in business tactics and also this is one of the only reason why Microsoft is among the top organization in these computational era. This paper presents a method to analyze users reviews and sentiment specifically to designed to work with movie reviews, hospital reviews, product reviews, college reviews, hotel/rooms reviews. The model is cop up with self-intelligent to preprocess the data user may enter. We have use NLP based approach to preprocess the data like:

1) Removal of stop words- it aims to remove words like a, an, as, at which has not very much important rolein analysis rather it introduces noise in the data.
2) Removal of punctuations- it will remove the punctuation marks from the data.
3) Tokenization- turning of sensitive data to no sensitive data.
4) Stemming- it is used to obtain the root word. Ex: running-run(root), walking-walk(root) etc.
5) Lemmatization- grouping together different infected forms of word.

Now the noise and irrelevant data is washed off from the sentence and hence pre-processing is done. Nowit is ready to fed the data to model to analyze the polarity of sentence towards negative, positive and neutralscore.

IV. Result & Discussions

The results or outcomes are as follows:

1. We are able to detect the polarity of sentence as well as basic analysis and report of entered data is generated successfully.
2. We are able to pre-process the data which contains lots of noise which will further used for betterment of accuracy of our model and prevents model from overfitting.

Conclusion

After the development of project our team came towards following conclusion:

1. It will help consumer and produce sector in terms of betterment of their product.
2. Brand monitoring is one of the most popular applications of sentiment analysis in business. Bad reviews can snowball online, and the longer you leave them the worse the situation will be. With Sentiment analysis tools, you will be notified about negative brand mentions immediately.
3. User is able to get 360-degree analysis of entered sentence
4. It will help in getting a better idea about the service and quality of entities like hotels/resorts,hospitality, college etc.

Acknowledgment

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Prof. Shivshankar Rajput & Project coordinators for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I am very much thankful to other faculty and staff members of CSE Dept., AITR Indore for providing me all support, help and advice during the project. We would be failing in our duty if do not acknowledge the support and guidance received from Dr S.C Sharma, Director, AITR, Indore whenever needed.

I would like to express my gratitude towards my parents & my group member Paresh Shah, Varsha Valwani, Tanishq Goyal for their kind co-operation and encouragement which help me in completion of this project. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

References

Reference Format for Journal Paper

[1] S.K. Gupta, L. Sharma, "*An Approach for Real Time Machine Learning Environment*", International Journal ofComputer Science and Engineering, Vol.1**6**, Issue.**22**, pp.4-10, **2020**.

Reference Format for Book/Book Chapter

[2] G. Sharma, "A Proposed Novel Approach for Internet of Things Protocol Execution Environment", IJCSERT Publisher,India, pp. **142-145**, **2020**.

Reference Format for Conference Paper

[3] P.B. Verma, "A Proposed Approach for Real Time Cloud Environment using Machine Learning", In the Proceedings ofthe 2020 International Conference on Data Science and Engineering, India, pp.**542-545**, **2020**.