SENTIMENT ANALYSIS ON PRODUCT REVIEW USING MACHINE LEARNING

R.Vignesh,
20275007@student.hindustanuniv.ac.in Department of Computer Application
School of Computer Science
Hindustan Institute of Technology &Science
Padur, Chennai - 603103.

Dr.D.Gandhimathi, Assistant prof.
dgandhim@hindustanuniv.ac.in
Department of Computer Application
School of Computer Science Hindustan Institute of Technology &Science Padur, Chennai - 603103.

Abstract --- The goal of this project is to perform the sentimental analysis on products reviews and implement the automation on rating by analyzing the review of that particular product, with a help of it, it will directly generate the rating, in the current world of the rating system is different from review and it can't be generated automatically and most of the people prefer rating rather reading reviews but with a help of this project we can implement the automation on the rating system and provide the rating based on the user's review and experience and, it would be a very easy task to know about defective on products and help the manufacture to figure out the problems on the product. In this way that will help the companies to produce better products according to the user's needs. The main important feature is it can detect the defective thing or some issues on that product on main features by ananalyzinghe review of that product, it uses machine learning "NLP" library that can detect the negative and positive words and gave values based on the amount of sentiment involved in that word accordingly with the help of the "tetetexblog library method in each review we will generate the rating depends on the reviews about the products are positive, negative or neutral.

Key Words: Python, Machine Learning, Text-Blob, Sentiment Analysis, Matplotlib, Seaborn, Visualization, Linear Regressio

1. INTRODUCTION

Sentimental analysis on products review is the method which can help product manufacturers and customers communicate and help them to understand their needs inefficient way that will help them to correct the mistakes and find the problems on the product faced by the buyers frequently by analyzing the sentiments and Repetitive keywords related to the features of the product with a help of its manufacturers can grow up their business the user's need and give a quick response to the buyers about defective parts of the product. The world scenario peoples often believe in rating before buying any product and rectify the rating is given by the rating is used on the user's review of the product and it also covers the main features of the product, if the features of the product had some defectives it will automatically import the manufacturers, this project performs the sentimental analysis with a help of a machine learning NLP library called TextBlop this machine learning library helps us to generate the result via implementing the word into it, it will givvalueses according to the emotions involved in that statements, TextBlop is a library which will give a value between -50 to +50, if the value is above 0 this will consider as a positive emotion if the value is below 0 that will consider as a negative emotion if the value is zero that will consider as a neutral value if one user put a review about the product we should read it for understand whether it is positive or negative and we cloud not do this for all reviews and it r time-consuming problem. To rectify the problem and save time to find the review is telling good or bad about the product. To execute is a project we have to analyze the string and in the modern world, text are having a lot of errors like spelling mistakes, grammars language differentiation, etc. Because textBlob major emotional conversion is words on English language and that is widely used by the product buyers so if some of the other language involved in the review sections we have to filtering it and translate it to make it feasible to perform the sentimental analysis the average of the product rating is calculated by applying logistic regression into the value we found by implanting the sentimental analysis via text blob in each comments(reviews), after implementing the logistic regression on those values we will get a average value which is the actual rating of the product, this process will perform by all reviews on products and that will give a average rating after find all those value we will visualize it through some visualization software , here in this project I used seaborn visualization tool , because seaborn has many features like changing columns x-axis and y-axis and this will get visualized in bar plot and that will give a visualize the average rating of every product with names.

2. LITERATURE SURVEY

Sentiment evaluation is continually a thrilling manner to investigate and apprehend people's reactions and opinions. in human nature, we can easily figure out someone's thoughts about the particular thing by looking at their reaction but in the growing modern world artificial intelligence does not much evolve to understand peoples thoughts and opinions and they are struggling to find the people whether they are happy or sad is likely all kinds of normal human behavior, but the sentimental analysis is helping artificial intelligence to find opinions and thoughts by monitoring their reactions. The traditional programming iteration process took a lot of time to analyze the result even though it's in linear programming format will become a drawback for emotional detection using the traditional programming method. After the arrival of the machine learning concept, people are trying to make a sentiment analysis by rendering a large amount of data even though researchers recognized several challenges in developing fixed sets of rules for expressions respectably. Much of the challenges in rule development stem from the nature of textual information. Six challenges have been recognized by several researchers.

Despite the explosive increase of Chinese E-commerce structures in current years, studies specialize in the sentiment category of Chinese files in assessment to their western counterparts. [1] This paper appears into the nascent place of Natural Language Processing (NLP) within side the Sentiment Analysis of Chinese Text. The proposed Deep Learning technique is using a sentence-primarily based totally technique within side the sentiment evaluation of on line critiques to benefit extra granularity and growth category accuracy via way of means of the use of a linear category algorithm. [1]

Recognizing emotions in conversations is a difficult task that has recently gained popularity due to potential applications [3]. Until now, however, there has been no large-scale multimodal emotional conversation database containing more than two speakers per dialogue. To fill this gap, we propose the Multimodal Emotion Lines (MELD) dataset, an extension, and improvement of Emotion Lines. MELD contains approximately 13,000 phrases from 1,433 dialogues from the TV series Friends. Each expression is annotated with emotion and sentiment labels and includes audio, visual, and text modes. [2]

In this digital world, expertise sharing, information exploration, tutored posting and different connected social exploitations are common to each individual furthermore as social media/network comparable to Face Book, Twitter, etc plays a significant role in such forms of activities. In general, several social network-based sentimental feature extraction details and logic are out there as well as many analysisers work on it domain for the previous couple of years. however, all those research specifications are narrowed within the sense of building some way for estimating the opinions and sentiments regarding the tweets and posts the user raised on the social network or the other connected net interfacing medium. Several social network schemes offer the flexibility to the users to push the voice tweets associate degreed voice messages, so the voice messages might contain some harmful furthermore as traditional and necessary content. during this paper, a replacement methodology is meant known as Intensive Deep Learning-based Voice Estimation Principle (IDLVEP), during which it's accustomed establish the voice message content and extract the options supported the language process (NLP) logic. [3]

Detecting hate speech on Twitter is critical for applications such as extracting controversial events, building AI chatter bots, content recommendations, and sentiment analysis. We define this task as being able to classify a tweet as racist, sexist, or both. The complexity of natural language constructs makes this task very challenging. We have performed extensive experiments with several deep learning architectures to learn semantic word embeddings to deal with this complexity. [4]

3. EXISTING SYSTEM

Sentiment analysis is always an interesting way to analyze and understand people's reactions and opinions. in human nature, we can easily figure out someone's thoughts about a particular thing by looking at their reaction but in the growing modern world artificial intelligence does not much evolve to understand peoples thoughts and opinions and they are struggling to find the people whether they are happy or sad likely all kind normal human behavior, but the sentimental analysis is helping artificial intelligence to find the opinion and their thoughts by monitoring their reactions. In the existing system, we can only analyze the negative and positive but in this project, we can find the neutral views and visualize is in a various amount of reactions about the product At the beginning of sentiment analysis, people used to analyze the people's emotions by using particular keywords. In the early stage of sentimental analysis, only those words are used to analyze the emotion of people, those concepts are only

based on the iteration process with the traditional programming method. By using keywords in sentimental analysis the accuracy will be too lesser compared to applying machine learning algorithms that will be a sentiment analysis with keyword and iteration process.

1.Admiration, 2.Empathic Pain, 3.Adoration, 4.Nostalgia, 5.Appreciation Of Beauty, 6.Satisfaction, 7.Interest, 8.Fear, 9.Amusement, 10.Excitement, 11.Anger, 12.Horror, 13.Anxiety, 14.Awe, 15.Awkwardness, 16.Boredom,17.Calmness, 18.Confusion, 19.Craving, 20.Disgust, 21.Scissors, 22.Relief, 23.Entrancement, 24.Sadness, 25.Surprise, 26.Joy

They use more than 1500 word to analyze the reaction for that they need to achieve several iterations and that need's lot of work to figure out the sentiment.

DISADVANTAGES OF EXISTING SYSTEM

- Programmer needs to put a lot of keywords to find the matching's and determine the reactions.
- It's not capable of solving the spelling mistake, so the result might get affected.
- It does not mind the grammatical errors and it's mostly available in the English language so if the text from other languages is placed in sentimental analysis it can't be get detected or else translated to English to get proper analysis.
- They use char-n gram to detect hate speech. But it only delivers the positive or negative as values and neutral emotions are out of consideration.

4. PROPOSED SYSTEM

In this project, I'm going to perform the sentimental analysis for Amazon product review and implement the automation in a rating system that will automatically figure out the sentiments involved in each public review about the product which they used. In the modern world lot of people share their opinion on social media and all other platforms. Let's take one example, Amazon is the most popular website which millions of people used to buy the thing and it receives ratings and reviews at each movement, those ratings and reviews help the users to find the better product quickly. Sometimes putting rating is not always accurate enough to determine the product but some people are not given an accurate rating as well and now a day's lot of bots used to give a fake rating to promote the fake product at the market, but with a help of this project, we can implement the automation on the rating system, by performing sentimental analysis on a product that will facilitate scan every comment s and generate the rating supported whether or not the review regarding the merchandise is positive or negative.

To analyze the sentiment of a product review I used TextBlob machine learning NLP library that can find each word involved In that review and provide the numerical values to the words accordingly if the word comes under negative reaction that will give a value between 0.00 to 0.49 and if the word comes under positive reaction that will get a value between 0.51 to 1.00 if the word carries the neutral emotion that will provide that value as 0.50 from that we can figure out the emotions involved in that product reviews and after applying logistic regression to the values we will get an average rating the product by analyzing each review.

ADVANTAGES OF PROPOSED SYSTEM

- We will classify the approaches in various ways and describe the techniques used intuitively.
- We will also provide various aspects and evaluate any kind of techniques to figure out the problems in the product by analyzing the keywords reviews that belong to the product's significant features, which didn't achieved by the old rating system.
- And it only follows NLP library TextBlop's states so this will reduce the iterations for finding keywords.

5. METHODOLOGY

This project is mostly based on text process and needs to perform a lot of filtrations to get a proper result. here I used a machine learning NLP library called TextBlob that can detect emotions by processing the strings, textBlob contains a lot of words and their emotions that will indicate in number format, to perform the sentimental analysis through textBlob we need to follow the tokenization, tokenization is the method which separates the words and terminates the grammatical errors and plurals, By implementing the sentence into the text blob that will give a mean value of the whole review as a rating.

Linear classification is an algorithm that I used to figure out the positive and negative rating based on the rating if the rating is below average that will consider as a negative rating and above-average will consider as a positive rating if the values are at the exact average rating that will be the neutral rating.

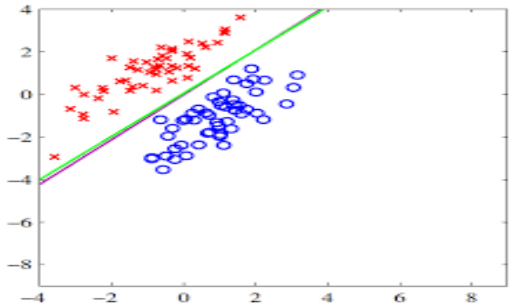
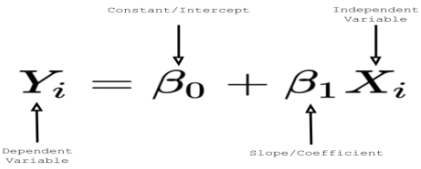


FIG 1: Linear Classification

Fig 1 suggests two types of dots the red dots (plus marks) indicates the above-average values and the blue dots (circle marks) indicate the negative values and the middle line is an average value. From that linear classification, [5] we can find the difference between positive and negative comments if the dots placed in the above-average line is high then that will consider as a positive review and the product got a high impression from the users, and if the dot is below the average line got more count that will consider as a negative review which most got the negative impression from the users



With the help of the linear regression algorithm, we can find the mean for each sentimental rating average; this will calculate the mean for the value we found via text blob.

6. MODULE

i. Data collection

ii. Data filtering

iii. Data classification

iv. Data visualization

i. DATA COLLECTION

Data is collected from Amazon's product reviews repository and those data contains many columns like rating, sub review, product name, Meta value, time, date, customer name, customer ID, product ID, etc so we need to filter those unwanted columns to process the sentimental analysis.

ii. DATA FILTERING

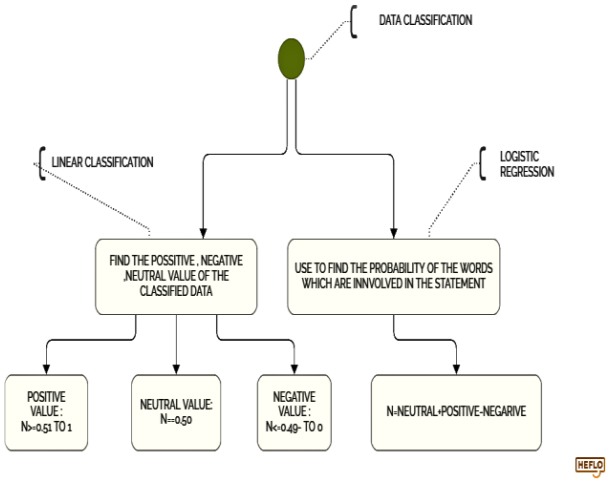


FIG 2: DATA FILTERING

[6] This project's major concentration is optimizing the strings and analyzing the sentiments of the product reviews, performing the filtrations on text, TextBlob is used, and optimizing the string to make it feasible to do sentimental analysis. Noun Phrase Extraction is used to separate the objective words, tokenization is the method used to spate the sentence and makes it meaningful full to do sentimental analysis. As we (Figure 2.0) mentioned, Words Inflection and Lemmatization are used to accurate grammatical mistakes and spelling mistakes.

iii. DATA CLASSIFICATION

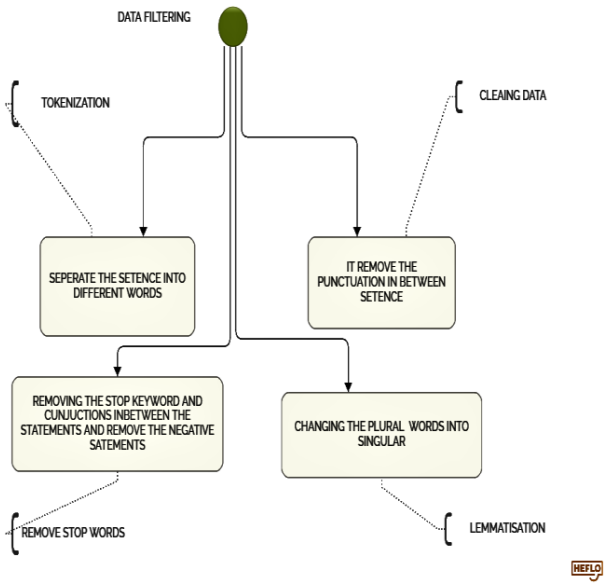


FIG 3: DATA CLASSIFICATION

Text blob may be a python library that won't categorize the word whether or not that's negative-positive or neutral. The above (Figure 2.1) as it's suggests [7] If the word comes under positive reaction that will assign the value from 0.51 to 1. If the word come under a negative reaction that will assign the value between 0.49 to 0, if the word is not negative nor positive that will become neutral and assign the

value 0.50. After applying all filtering steps we will get a set of data that only carry the meaning of full words, from that words we have to clarify which word belongs to which side whether it is “positive or negative and neutral”.

iv. DATA VISUALIZATION

After applying all methods mentioned above we can get two types of positive and negative data set that involves all values, from that data set we can find the difference ratio by visualizing it using “mathplotLib” and “seaborn” python visualization libraries. This will get visualized in the bar plot x-axis carry the rating values and the y-axis carry the name of the product.

7. BASIC ARCHITECTURE DIAGRAM

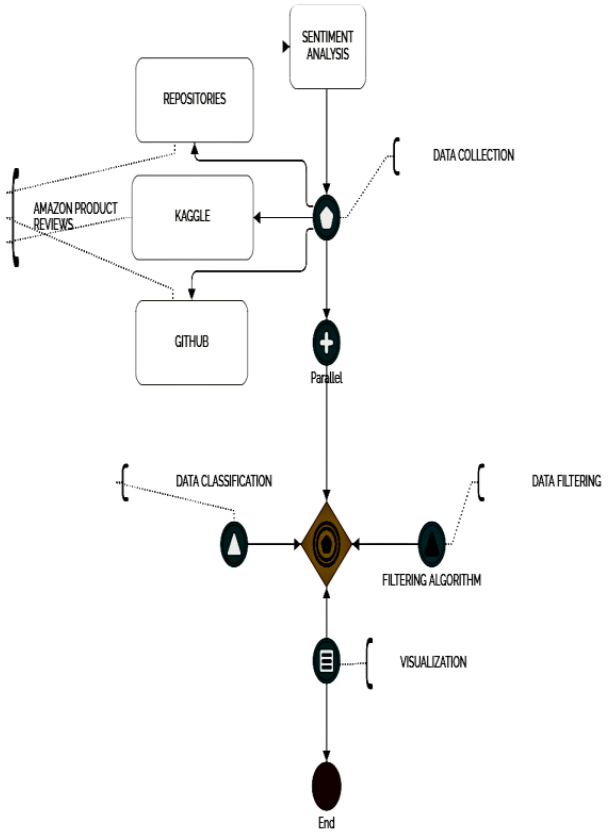


FIG 4: ARCHITECTURE DIAGRAM

8. SYSTEM SPECIFICATION

Hardware Requirements

- Windows 7,8,9,10 64 bit
- Ram 4 GB(min)
- Speed 2.5 GHz
- Processor I5

Software Requirements

- Operating system : 64-bit operating system, x64-based processor
- Coding Language : python
- Tool : Anaconda Navigator, Jupyter notebook.

9. CONCLUSION

To implement the automation on rating can achieve by following these implementations, with a help of that peoples often find the rating and their specific requirements with a help of this project, this will help the manufacturers to understand the problems via a focus on main features defective parts and rectify it much faster than before,

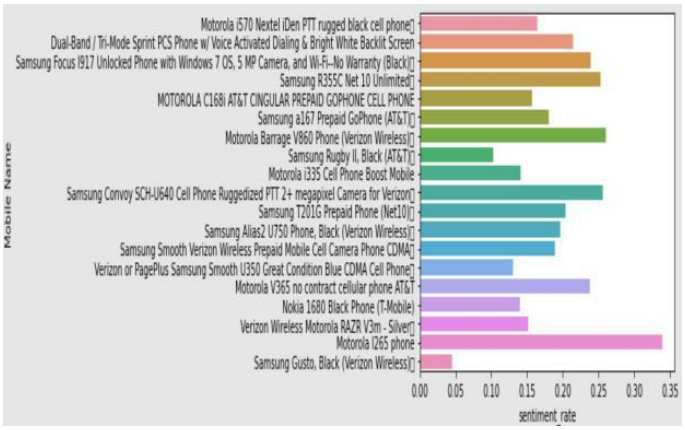


FIG 5: Product Rating After Analysis

the business more progressive and have a high reaction rate while receiving the bad impressions,(figure 3.0) this will accelerate the business and provide the better products in markets, it will help buyers to buy a product with an experienced result as a rating and it will avoid more spamming on rating and help the users to buy better products.

10. REFERENCES

[1] Lee, Jun & Zuba, Denis & Pang, Yan. (2019). Sentiment Analysis of Chinese Product Reviews using Gated Recurrent.

[2] Amba, Manika, Madhusudhan, Margam (2018). "Application of sentiment analysis in libraries to provide temporal information service: a case study on various facets of productivity". “Social Network Analysis and Mining”.

[3] MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations(2020) Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, Rada Mihalcea

[4] Deep Learning for Hate Speech Detection in Tweets(2019) - Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma IIIT-H, Hyderabad, India2Microsoft, India.

[5] Recent Advances of Large-scale Linear Classification Guo-Xun Yuan, Chia-Hua Ho, and Chih-Jen Lin.

[6] Text Classification Using Machine Learning Techniques M. IKONOMAKIS Department of Mathematics University of Patras, GREECE ikonomakis@mailbox.gr S. KOTSIANTIS Department of Mathematics University of Patras, GREECE sotos@math.upatras.gr V. TAMPAKAS Technological Educational Institute of Patras, GREECE tampakas@teipat.gr

[7] Zhang Y, Xiang X, Yin C, Shang L (2013) Parallel sentiment polarity classification method with substring feature reduction In: Trends and Applications in Knowledge Discovery and Data Mining, volume 7867 of Lecture Notes in Computer Science, 121–132.. Springer Berlin Heidelberg, Heidelberg, Germany.