Analysis of User's Behavioral Pattern for Various Brands Using Machine Learning

Farah Jamal[1], Ms. Kavita Agrawal[2]
Deptt of CSE[1] , Deptt of CSE[2]
Integral University[1], Integral University[2]
Email: farahjamal.369@gmail.com[1] , kavita@iul.ac.in[2]

Abstract-Sentiment Analysis in recent times has become the focus of many researchers due to availability of large amount of online data. The analysis of online data is very beneficial as well as it is in demand in different applications. Product based sentiment analysis is one of them. It helps the organization in getting up-to-date review of their products and helps in understanding user behavior towards their product. This paper introduces a novel approach to perform sentiment analysis of social media reviews of different products. The paper contribute to the field of sentiment analysis, which extracts opinions from the tweets. The main aim is to classify the text as positive, negative or neutral. Various sentiment classifier are used to analyze the data. But with increasing complexity of data, it's time to inquest at the same old sentiment extraction process and to re-figure them. We have proposed a hybrid approach. Here, Naïve Bayes classifier is used for classification, Long-Short term memory for predicting the outcome and decision tree for precision.

Keywords: - Sentiment Analysis, Naïve Bayes, Long-Short term memory, Decision tree, Hybrid

1. INTRODUCTION

Social media has become a platform for spreading out information across the world. People share their information, and other users share their views about that shared information.[2] People use to comment on every topics like movies, politics, products etc. Particularly, when we think about business domain, there are numerous feedbacks on the products. People share their reviews whether they liked the product or not. User's opinion present on social networking sites helps the organizations in analyzing the current market trends and can plan their future marketing strategy according to that.

"Sentiment analysis is a mechanized process of wrenching out of knowledge from the views expressed by the people on social media platforms"[1]. Different approaches have been introduced to perform sentiment analysis. Deep learning techniques have helped in getting more accurate and efficient result. In this paper, we proposed a new hybrid approach to perform sentiment analysis. We will be using Naïve Bayes, LSTM and Decision Tree.

Sentiment Analysis approaches can be categorized as: - Machine learning, lexicon-based and hybrid.

1.1 Machine learning is an Application of Artificial Intelligence with efficiency to be trained accordingly and helps software applications to become more exact with being programmed distinctly.[1] The main aim is

to allow system learn things naturally without human assistance. ML can be divided into three category: - Supervised, Semi-Supervised and Deep Learning. **Supervised Learning Algorithms** are used to predict the future events. It is performed on the basis of label data whereas unsupervised technique is used when the data is not labeled.

Naïve Bayes is one of the most used supervised learning algorithm. It is the powerful and scalable algorithm and gives great results when it is used in textual data. It is probabilistic classifier based on Bayes theorem. Without considering the position of a word, it calculates the posterior probability.

Decision Tree is the most simplest and useful tool used for classification and prediction. Decision tree perform classification without requiring much computation. It gives graphical characterization of possible solutions.

Deep learning is a machine learning approach that allows system to master things intuitively. We have applied Long-Short Term Memory. LSTM is an extension of RNN, this is because it remembers patter for long duration of time. The benefit of deep learning is that their progress improves as the size of data increases. [1]

1.2 Lexicon-based is a simple approach. It is performed by using a precompiled sentimental dictionary. Sentence is slit into tokens (words) and based on that data is

International Journal of Research in Advent Technology, Vol.7, No.5, May 2019
*E-ISSN: 2321-9637*
Available online at www.ijrat.org

classified as positive, negative or neutral. It is further divided into dictionary based or corpus based. The limitation of this approach is that with the growth of dictionary size its performance degrades. [2]

1.3 Hybrid technique is the combination of both machine learning and lexicon-based approach.
After analyzing all the approaches, we have used the combinations of supervised learning technique to perform more accurate sentiment analysis

2. SYSTEM MODEL
The data collected can be classified using supervised as well as unsupervised classifier. The unsupervised method includes lexicon based method. The supervised classifier is based on algorithms.
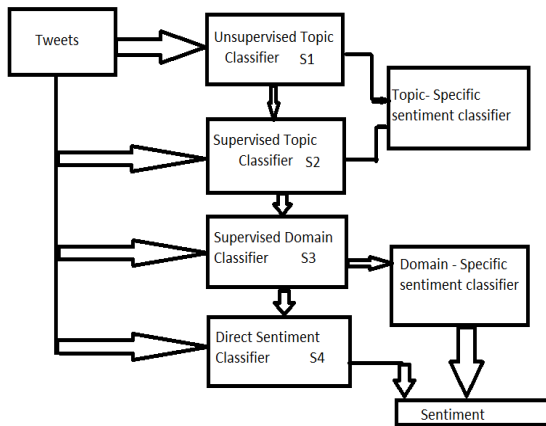
3. METHODOLOGY

3.1. Data collection

The very first step is the collection of data. The data are collected from the twitter with the help of Tweepy. We have collected almost 80,000 data sets.

3.2. Processing of data

Since the datasets consists of repeated tweets, so in this step we have fetched the unique tweets.

3.3. Training of data

With the help of pickle API we have converted the collected data sets into trained dataset. Pickling is used to convert python in character stream

3.4. Classification of data

We have used Supervised Learning Algorithm: -Naïve Bayes for the classification purpose. Naïve Bayes are useful when there are large datasets, as well as known for performing highly refined classification.

3.5. Polarity detection

The next step is to determine the polarity of the tweet, whether they are positive, negative or neutral. For this we have used deep learning algorithm i.e. LSTM. LSTM is an extension of RNN. LSTM enables RNN to remember input for long period of time. This is the best algorithm for prediction.

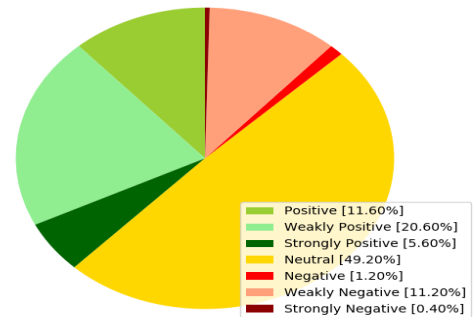### 3.6. Graphical representation
Finally, we have applied decision tree for graphical representation of data. It gives us the clear analysis about the data.

4. RESULT ANALYSIS
Datasets of different products were used in the experiment. Data sets on shoes which include puma, Nike, reebok and adidas brand.

4.1. Result Analysis of Nike Shoes

How people are reacting on nike shoes by analyzing 500 Tweets.

Positive [11.60%]
Weakly Positive [20.60%]
Strongly Positive [5.60%]
Neutral [49.20%]
Negative [1.20%]
Weakly Negative [11.20%]
Strongly Negative [0.40%]

GENERAL REPORT: - Weakly Positive

4.2. Result Analysis of Reebok Shoes

No. of Tweets Fetched	500
Positive	54
Weakly Positive	85
Strongly Positive	6
Negative	5
Strongly Negative	3
Weakly negative	54

International Journal of Research in Advent Technology, Vol.7, No.5, May 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Neutral	312

How people are reacting on reebok shoes by analyzing 500 Tweets.

- Positive [10.80%]
- Weakly Positive [17.00%]
- Strongly Positive [1.20%]
- Neutral [62.40%]
- Negative [1.00%]
- Weakly Negative [7.00%]
- Strongly Negative [0.60%]

GENERAL REPORT: - Weakly Positive

No. of Tweets Fetched	500
Positive	58
Weakly Positive	103
Strongly Positive	28
Negative	6
Strongly Negative	2
Weakly negative	58
Neutral	252
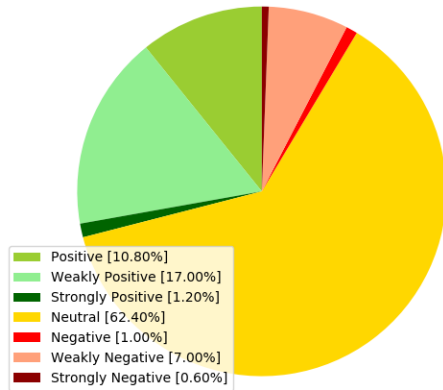
5. CONCLUSION

Sentiment analysis is emerging as a challenging field with diverse application. Different approaches have been introduced till now. Social networking sites, twitter and data set are the best medium for collecting user's data. This methodology will help the organization to know about the current situation of the market for same products and plan their marketing strategy.The analysis is done to find out whether the product is liked by the consumer or not. Once the mindset of the user is known it becomes easy for the organization to adjust and improve their marketing strategy . We have surveyed different approaches of categorizing and analyzing sentiments of users on social media. The use of LSTM with other techniques has helped in getting better performance yet with different cons. So here we have proposed our hybrid approach for sentiment analysis of various products. The combination of Naïve Bayes, LSTM and decision tree has helped in getting better accuracy by predicting the polarity of the data and getting the general review about the products. The analysis is done to find out whether the product is liked by user or not. They could easily track the opinions of customer over time. This will help the organization to get the actual issues related to product, so that they could improve their services.

6. FUTURE WORK

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Similarly, this work will help in analyzing users review or opinion related to different brands or product. Thus it will help the organization in planning their marketing strategy.

REFERENCES

[1] Farah Jamal, Ms.Kavita Agarwal. (2018), "Analysis of user behavioral pattern using sentiment analysis: A Survey" in International Journal of Science and Research. Vol 8, Issue 4, April 2019, pp 843-847.

[2] Achira Jeewaka Shamal, Rankothge Gishan Hiranya Pemathilake, Sachith Paramie Karunathilake and Gamage Upeksha Ganegoda. (2018), "Sentiment Analysis using Token2Vec and LSTMs" in International Conference on Advances in ICT for Emerging Regions.

[3] Lin Yue, Weitong Chen, Xue and Wanli Zuo. (2018), "A survey of sentiment analysis in social media" in Springer-Verlag London Ltd. ,pp 1-47

[4] Nurulhuda Zainuddin, Ali Selamat, Roliana Ibrahim. (2017), "Hybrid sentiment classification on twitter aspect-based sentiment analysis" in Springer Science+Business Media,LLC,part of Springer Nature.

[5] Avinash Surnar, Sunil Sonawane. (2017), "Review for Twitter Sentiment Analysis Using Various Methods " in International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol- 06.