International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 08 | Aug 2019 www.irjet.net p-ISSN: 2395-0072

Analysis of Brand Value Prediction based on Social Media Data

Roshani Yadav[1], Namrata Sharma[2]

[1]PG Student, Dept. of Computer Science and Engineering, Sushila Devi Bansal College Of Engineering, Indore.
[2]Asst. Professor, Dept. of Information Technology, Sushila Devi Bansal College Of Engineering, Indore.

Abstract - *The focus of this paper is to adopt different sentiment analyzers with machine-learning algorithms in order to develop the strategy with the highest precision for learning about brand values. Linguistic orientation is computed by words, phrases and phrases calculated in an extremely document in an extremely lexical sentiment analysis. Polarity is calculated in the Lexicon-based technology on the assumption of the dictionary which comprises a particular word's linguistic score. The approach of machine learning is, however, largely aimed at classifying the text through the use of algorithms such as Naïve Thomas Bayes and the choice box. The field of sentiment assessment from feeling lexicons or machine-learning methods has now been significantly expanded. However, this study focuses on comparing the lexicons of sentiment (SentiWordNet and TextBlob), which makes sentiment assessment the easiest possible. What we did not find in such a preceding assessment. In this way, we also tend to calculate the feelings of 2 analyzers named SentiWordNet and TextBlob. Although the results of their results were comparatively higher, we tested their results with 2 supervised machine learning classifiers, Naïve Thomas Bayes, and the choice treaties. We expanded the work to an application where we worked on Amazon Review Data, NLTK with Naïve Byes gave us accuracy of 82% which was good but we extended it with CNN due to large volume of data and achieved overall accuracy of 94.4%*

Key Words: Brand Value Prediction, Hybrid Methods, Sentiment Analysis, Naïve Bayes, Convolution Neural Network

1. INTRODUCTION

Sentiment analysis is circumstantial mining of data usually in text format which is being used by most of us to express our thoughts. It isolates and quotations the subjective information available in the material, It can have varied utilities for example it can be used in helping a business to comprehend the social sentiment of their, product, brand or service while observing online discussions. . Different techniques and software tools are being developed and are available to perform Sentiment Analysis. The aim of this work is to review and compare some freely accessible web services, examining their competencies to categorize and provide polarity score for different pieces of text with respect to the opinions enclosed within.

The applications of sentiment analysis are varied and covers broad domain. The capability of retrieve and analyze the perceptions of a person from social data is being widely embraced by organizations worldwide. Few common examples are presented below:

- **Stock Market Prediction:** Indian Stock Market Prediction Using Sensex and Nifty, Correlation of shift in stock exchange with social sentiment data is being used.
- **Public Sentiment on Government Policy:** The Obama administration used sentiment analysis to measure public opinion to policy announcements and campaign messages ahead of 2012 presidential election. [2]
- **Brand Value Prediction:** The understanding of analyzing the consumer attitudes and proactively utilizing the facts the way Expedia took benefit of when they observed that there was an upsurge in negative feedback to the music they used in one of their commercials. They analyzed the social media was full with the frustrating comments by its consumers. [3] Expedia was able to report the negative sentiment in a spirited and corrective way by airing a new version of the advert.
- **Sentiment classification on clinical narratives:** has been base to analyze patient's health status, medical condition and treatment. sentiment score of a sentences being calculated using language models with the equivalent coefficients for assessing the sentence's sentiment score.[5]

With social media networks such as Facebook and Twitter, it is becoming practicable to systematize and measure what public outlook is on a given issue, news story, product, or brand. Opinions that are extracted from above mentioned services can be valued. Datasets that are collected or repositories that are already available can be investigated and presented in such a way that it makes it easy to classify positive, negative or neutral. [6] These are few applications which explain the broad domains where sentiment analysis is already being used as practice and published as research, blog or article. Many other potential applications can be explored and developed to show practical usage of methodology.

1.1 Machine Learning Approach

Symbolic techniques uses web searches that can provide important statistical information about frequency of words and its correlation to derive the features that will be used as input to classifiers. Machine learning is branch of Artificial Intelligence which enables machine to learn and make decisions. It is used on selected feature and then classifier is used to categorize the data.

Machine learning algorithms widely used for sentiment analysis, two types are supervised and unsupervised methods. A supervised algorithm performs better and works on labeled data. Most popular one in supervised category is Naïve Bayes classifier, Maximum Entropy; Memory based learners KNN, Decision Tree and Support Vector Machine. Other supervised learning methods include Decision Rule classifiers, (Artificial) Neural Networks, Deep Learning, Regression and Random Forests.

2. Dataset: Amazon review data (Actual Reviews)

We have applied this model on any data available we have chosen review from Amazon, which is mostly from the user of their products. This dataset consists of a few million Amazon customer reviews (input text) and star ratings (output labels) for learning how to train fastText for sentiment analysis. The idea here is a dataset is more than a to real business data on a reasonable scale - but can be trained in minutes on a modest laptop.
Almost, 400K text reviews and ratings for testing in fastText format.

3. System Architecture

The validity of the model can be observed using error or accuracy of the model along with "false positive" and "false negatives"

$$Error = \frac{Number\,of\,Errors}{Total\,Number\,of\,Data\,Points} \quad \text{eq. (1)}$$

$$Accuracy = \frac{Number\,of\,Correct\,Prediction}{Total\,Number\,of\,Data\,Points} \quad \text{eq. (2)}$$

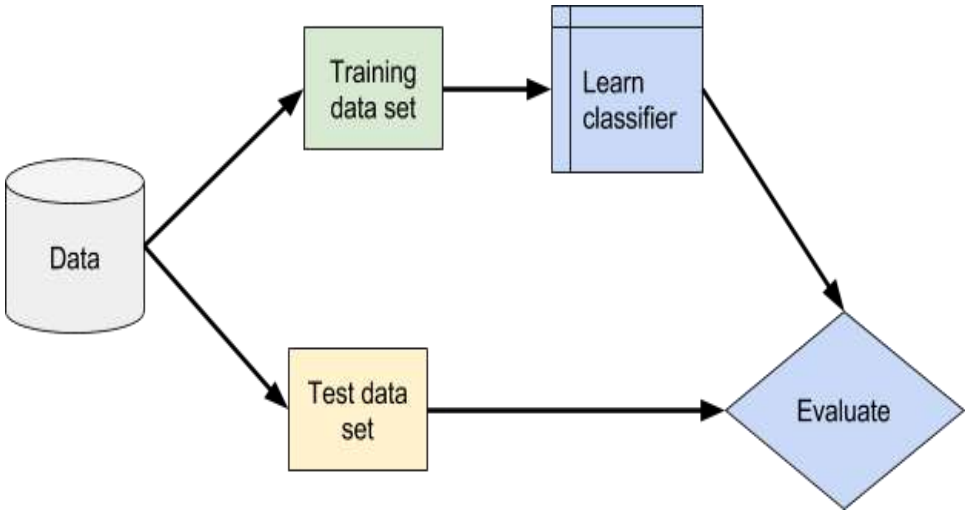


Fig -1: System Architecture

4. Implementation and Result Analysis

4.1 CNN based Experiments

1. We have used tensorflow backend, and time taken to process the training and testing data is presented below with time stack

2. Loading the data from the train and test split files

3. The following set parameters

max_features = 8192
maxlen = 128
embed_size = 64

4. Applying Tokenizer on max features

5. Padding 'Post and setting maxlen parameter

6. Setting CNN parameters and generate a model

The overall time taken to train and test the CNN model was 15.5 minutes; CNN shows its capabilities of working on large dataset of 4000000 reviews in just 15.5 minutes. The system accuracy of our own model presented above with NLTK- Naïve Bayes by more than 10%. The previous model achieved accuracy of 82%, while CNN model achieved accuracy of 94.14%

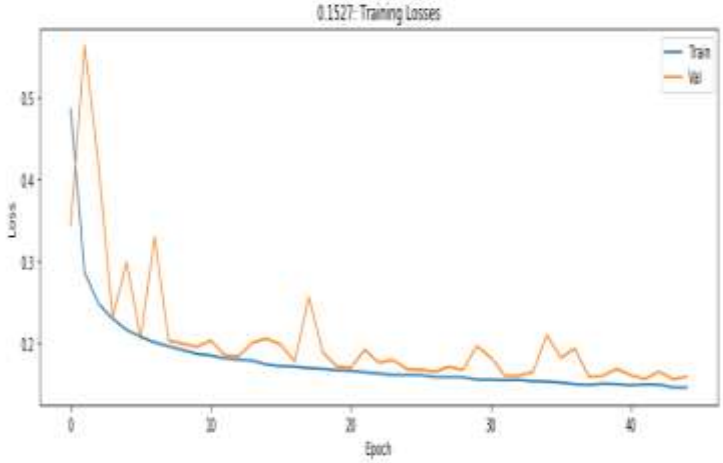


Fig -2: Training Accuracy

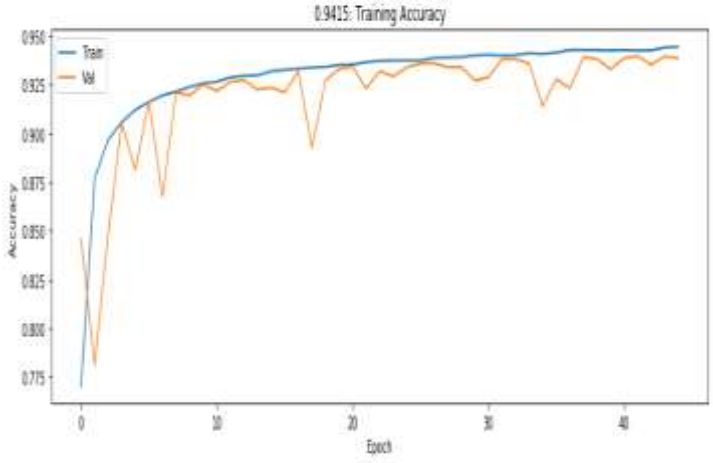
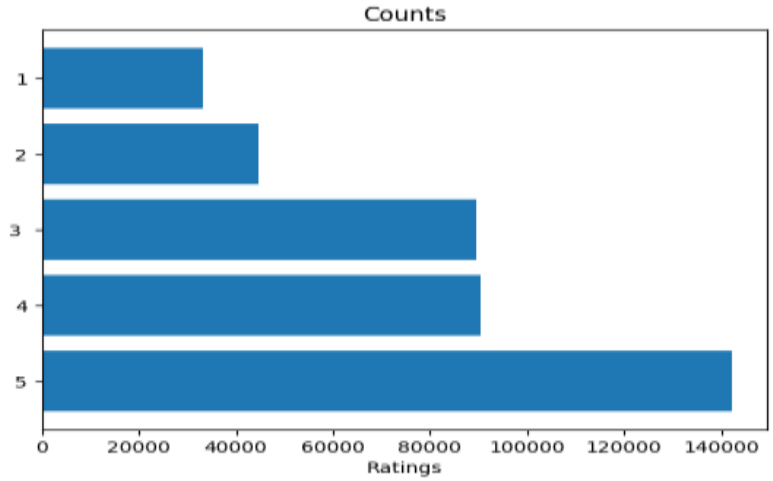


Fig -3: Training Loss

4.2 Observations

The graph shows that there are very few reviews with rating 1 or 2. The number of reviews in rating category 3,4 and 5 are way more than that of low rating. This shows the liking of people towards amazon products.



4.3 Using NLTK Naive Bayes

To further investigate the performance of method we have opted NLTK library and Nave Bayes Classifier which is based on conditional probability.

Split the data in 90:10 train and test parts (3600000 Training Sample and 400000 Testing Samples)

- Separate Positive and Negative tweets
- Remove the data with missing values i.e. 'na'
- Cleaning the data and Feature Extraction
- Extracting word features

5. Conclusion and Future Work

5.1 Conclusions

There has now been a significant expansion of the field of evaluation of feeling by lexicons or machine learning techniques. However, this research aims to compare feeling lexicons (SentiWordNet and TextBlob), making it as easy as possible to evaluate the feelings. What we have not found in such a previous evaluation. Thus, the emotions of the two analyzers called SentiWordNet and TextBlob are also calculated. Although the results were comparatively higher, the results were tested with 2 supervised machine classification systems, Naïve Thomas Bayes and the agreements of choice.

5.2 Future Work

Sentiment assessment can be implemented in many fields but it can be difficult to declare the declaration category positive or negative. [12] People's feelings are, moreover, complete. Machine learning systems can comprehend the human language multifaceted. For instance, feelings that are unordered and have the feeling of irony or sarcasm make it harder to comprehend.

A more sophisticated method or algorithm for an in depth assessment of why a specific phrase belongs to a positive, negative or natural category. For example, we look at descriptive statistics for feeling score, re-tweet, preferred and limited data from tweets in this section of the study. All feelings are negative because they are only adverse tweeting feelings from the dataset sieved down. There is a median feeling of all adverse tweets, which is the real sense of process individuals.

REFERENCES

[1] Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, Enrique Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Information Sciences, Volume 311, 2015, Pages 18-38,

[2] Kristian Bannister Understanding Sentiment Analysis: What It Is& Why It's Used: https://www.brandwatch.com/blog/understanding-sentiment-analysis/ Jan 2015

[3] SUSAN KRASHINSKY MARKETING REPORTER, No strings, please: Expedia Canada ad falls flat, THE GLOBE AND MAIL, May 2018

[4] https://www.growthaccelerationpartners.com/blog/sentiment-analysis

[5] Tran-Thai Dang and Tu-Bao Ho, Mixture of Language Models Utilization in Score-Based Sentiment Classification on Clinical Narratives, Trends in Applied Knowledge-Based Systems and Data Science: 29th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2016, Morioka, Japan, August 2-4, 2016, Proceedings (pp.255-268)

[6] Sentiment Analysis: Overview, Applications and Benefits https: //www.growthaccelerationpartners.com/blog /sentiment-analysis

[7] AnaisCollomb, A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation, University of Lyon, INSA-Lyon, F-69621 Villeurbanne, France

[8] Aurangzeb Khan, BaharumBaharudin, and Khairullah Khan. Sentiment classification from online customer reviews using lexical contextual sentence structure. In Software Engineering and Computer Systems, pages 317–331 Springer, 2011.

[9] Devika M D, Sunitha C and AmalGanesha, Sentiment Analysis: A Comparative Study On Different Approaches, 1877-0509 © 2016 Published by Elsevier Procedia Computer Science 87. Pp 44-49. Fourth International Conference on Recent Trends in Computer Science & Engineering. Chennai, Tamil Nadu, India

[10] http://dx.doi.org/10.1016/j.asej.2014.04.011

[11] Asghar, Dr. Muhammad &MasudKundi, Fazal& Khan, Aurangzeb & Ahmad, Shakeel. (2014). Lexicon-Based Sentiment Analysis in the Social Web.journal of basic and applied scietific research. 4. 238-248.

[12] Duwairi, Rehab & A. Ahmed, Nizar& Y. Al-Rifa, Saleh. (2015). Detecting Sentiment Embedded in Arabic Social Media – A Lexicon-based Approach.Journal of Intelligent and Fuzzy Systems. 29. 10.3233/IFS-151574.

[13] Jurek, A., Mulvenna, M.D. & Bi, Y. Secur Inform (2015) 4: 9. https://doi.org/10.1186/s13388-015-0024-x

[14] R. Rezapour, L. Wang, O. Abdar and J. Diesner, "Identifying the Overlap between Election Result and Candidates' Ranking Based on Hashtag-Enhanced, Lexicon-Based Sentiment Analysis," 2017 IEEE 11th International Conference on Semantic Computing (ICSC), San Diego, CA, 2017, pp. 93-96.

[15] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," Expert Systems with Applications, vol. 57, pp. 117–126, Sep. 2016.

[16]Alexandra Balahur, José M. Perea-Ortega, Sentiment analysis system adaptation for multilingual processing, Information Processing and Management: an International Journal, Volume 51 Issue 4, July 2015 , Pages 547-556

[17] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A Review of Relational Machine Learning for Knowledge Graphs," Proceedings of the IEEE, vol. 104, no. 1, pp. 11–33, Jan. 2016.

[18] Kai Zhao and Yaohong Jin, "A hybrid method for sentiment classification in Chinese Movie Reviews based on sentiment labels," 2015 International Conference on Asian Language Processing (IALP), Suzhou, 2015, pp. 86-89.

[19] Twitter sentiment analysis using hybrid cuckoo search method, Information Processing & Management, ISSN: 0306-4573, Vol: 53, Issue: 4, Page: 764-779, 2017

[20] G. Liu, X. Xu, B. Deng, S. Chen and L. Li, "A hybrid method for bilingual text sentiment classification based on deep learning," 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai, 2016, pp. 93-98.