IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Reviewing Flipkart Product Comments using Methods Based on Sentiment Analysis

Ishan Raj[1], Manveer Singh Jhala[2], Pravin Vilas Kadam[3], Prof. Amruta Patil[4]
Students, Department of Computer Engineering[1,2,3]
Guide, Department of Computer Engineering[4]
Sinhgad Inst. Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *With the explosion of social networking sites, blogs and review sites a lot of information is available on the internet. This data contains sentiments and opinions about a wide variety of product features of such products. This form of opinion and feedback is crucial to the companies developing these products as well as the companies that want to develop these products better than the competitive products. Sentiment Analysis is the process of analyzing all this information, retrieving opinions about such products and services, and labeling them as good or bad. In other words, finding out whether the product is good or bad or positive or negative. The most crucial parts of any review of any particular product is the numeric rating and the description provided along with that product. In this project, we will take into consideration both these factors for product reviews to make a final decision on a classifier that is best suited for product reviews analysis. We have collected reviews based on the features that best describe the emotions and opinion for each review, we have built a feature set of 1000 features, and with this limited feature set, we will find out which classifier obtains the best result on review type data. To decide the best classifier, we perform evaluations on it, by running a wide variety of data set generators and calculating the resubstituting and generalization mistakes and errors for each classifier. We then use the mean of these findings to calculate the paired Student's t-test to relatively compare the output of the classifiers. Based on the findings of this evaluation, we can state which is the best suited classifier for our model.*

Keywords: Filpcart.

I. INTRODUCTION

Reviewing product using sentiment analysis is becoming popular for text mining. Research is also considering research in area of computational linguistics. Research work is focusing on correlation among Flipkart product reviews. Research is also considering rating of products provided by customers. Research has considered traditional machine learning algorithms along with SVM (Support Vector Machine). A huge quantity of both structured and unstructured multimedia data is being uploaded on the Internet due to rapidly growing ubiquitous web access over the world. However, analyzing those raw media resources to discover their hidden semantics is becoming a challenging task. As a result, it is difficult to retrieve the right type of media to satisfy multimedia content consumers. Sentiment Analysis is the most extensively researched Natural Language Processing (NLP) and Machine Learning implementations (ML). This field has exploded since the introduction of Web 2.0. Thanks to the Internet, people can now express their thoughts, ideas, and sentiments regarding products, people, and life in general. Because of this, the Internet has developed into a large repository of textual data with strong opinions. An aim of sentiment is to retrieve crucial information about public opinion that may be utilized to make informed corporate decisions, election strategy, and product consumption.

II. MOTIVATION

- **Businesses and organizations:** The motivation for performing sentiment analysis for various businesses is to get the sentiment of public regarding their product and then according to that improve their products and make huge profits and compete with their rivals accordingly.
- **Customers:** Another motivation is also to make things easier for customers. Customers getting the general sentiment of public can make better decisions for themselves whether to invest their money or time in a particular product or not.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

- **Advertisement:** Placing ads in the user-generated content is one of the key motivations behind performing sentiment analysis. Companies can select which place would be best to place advertisement. Like if their customers are generally kids then they use cartoon tv channels or magazines to place their ads. Place an ad when one praises a product-processing basically suggests that the transformations that are applied to our dataset before inserting it to the Machine Learning or Deep Learning algorithms. In data Preprocessing, the raw data is converted into clean data set, it is a very popular technique which basically converts the data which is gathered from different sources and it converts it into clean data which is feasible for analysis.

III. METHODS

3.1 Architecture

Expand Abbreviation: The chat language contains many abbreviations in short forms and expanded acronyms. The fast conditions are the shorter version of the word by replacing or removing few letters, such as nyt—night.

Data Cleaning: The data gathered from the source is cleaned step by step. The first step is to verify the conditions and frequencies: the comments could have the word n-gram or a single word. Every text in the statement should be converted to a single format lower to the higher case or higher to the lower point.

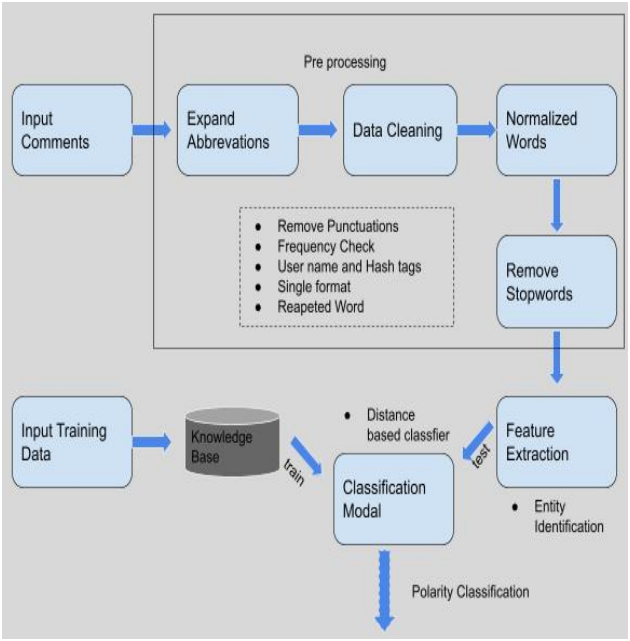

Figure 1: System architecture model

A. Normalize Words and Eliminate Stop Words

Normalizing the word is nothing but removing the repeated words. It is recurrently found in social media text that some words are frequently repeated more than one time. Eliminating the stop words was applied as the fourth pre-processing step. After the word "not," if any entity occurs, then the entity is concatenated with "not" to form one single word.

B. Feature Extraction

The cleaned data may contain many distinctive properties like target objects, adjectives, nouns, kinds of sentences. These properties are all termed entities. These identified entities are then fed into the classifier stage as a feature. For determining the entities, we made a Sent WordNet database. It is built in two phases.

1. WordNet terms: relationships like antonyms, and synonyms.
2. WordNet: relates the similar adjective pictures of noun and verb groups.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

IV. Results and Analysis

We have trained the Flipkart reviews dataset of more than 10000 sentences to give the sentiment of any text. In general, we can say that it gives a good accuracy score of 89% and gives a good prediction whether the sentiment of text is positive or negative.

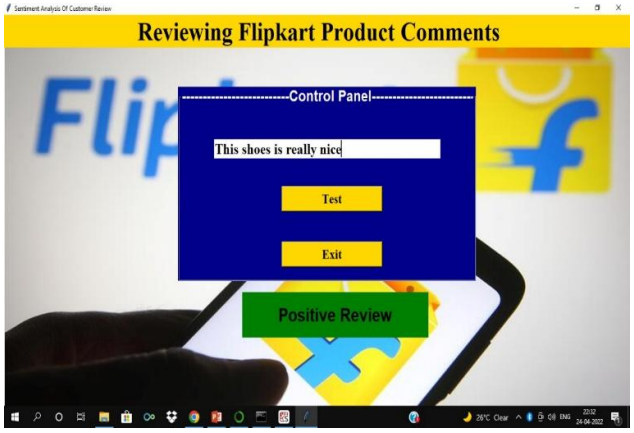

Figure 2: Sentiment of a review

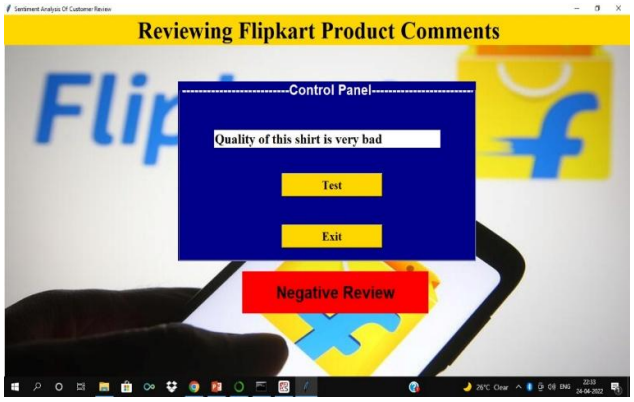

Figure 2: Sentiment of a review

The results of our model can be seen in above images (Fig. 2 and 3). When we give an input in the form of a text than our model predicts whether the emotions or opinion of the text is positive or negative. Here in the above example our model accurately predicts the sentiment of the given input.

V. Application

In future, a computer will be able to sense human emotion by sensory means and then create an environment that will help in the complete advancement of human life. Apart from that sentiment analysis has many applications in present world. It can be used by businesses to compete with their rivals and improve their product. It can be used in social media monitoring. Emotions and sentiments of people can be tracked by doing sentiment analysis of their social media posts and activities. Nowadays, sentiment analysis is used for customer support as well. Quick and urgent requests of customers can get a automated response using sentiment analysis.

VI. Conclusion And Future Works

Research is supposed to provide more flexible and accurate solution. Naive Bayes provided solution for classification. The data set would be trained and pre-processed to provide more accuracy in solution. This research is supposed to solve the problem of previous research that was faced during sentiment analysis. To eradicate misclassifications in the design of prediction models used to interpret feedback and reviews, more extensive methodologies are required. The results of a extensive variety of hybrid sentiment mining approaches are empirically examined on datasets of different sizes in this

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

study. According to different accuracy criteria, the hybrid ensemble technique (HEM1) is the most resilient of the methodologies employed for balanced data models I, II, and III. According to the findings, a compound blend of unigram, bigram, and trigram performs effectively in practically all prediction algorithms. Even though SVMs can cope with any amount of data imbalance, the findings imply that data imbalance can have an influence on the employment of SVMs in real-time applications for class prediction. The recommended revised bagging process has just been shown to be efficient and superior to many alternative approaches employing different data sampling methods in extensive experiments with benchmark and actual device datasets. PCA is a strong dimension reduction strategy for both balanced and unbalanced datasets when utilizing mixed techniques, as per the findings. Other feature reduction methodologies, including such latent Dirichlet distribution, might be researched throughout the upcoming. Mostly in future, more study may be done to evaluate the influence of various domain and region-specific characteristics. Extending sentiment mining to new areas might result in some surprising results. In the future, further n-gram variants and attribute weighting might be offered to achieve a better level of accuracy. The work of this study is mainly focused on dividing sentiment into two categories: positive and negative (binary classification). A multiclass emotion categorization system including positive, negative, neutral, and other categories might be implemented in the future. The key purpose of our research is to find traits that appear as nouns or noun sentences in feedbacks. Future testing will discover inferred habits. Because ensemble learning procedures take a very long time to decide, parallel computing technologies can be investigated as a solution. Because the knowledge gained by ensembles is not available to humans, the inability to examine the findings of ensemble learning approaches is a key disadvantage. As a result, improvement in the interpretability of ensembles is an important research area. Future opinion-mining systems will require a much more comprehensive and diversified knowledge and understanding in general knowledge. This will give more accurate understanding of natural language perspectives as well as a strong link between multi-model and machine-processable data. More microbe methodologies towards the configuration of intelligent opinion-mining systems capable of handling semantic information, making analogies, discovering new affective knowledge, and detecting, perceiving, and "feeling" emotions will result from combining theoretical theories of emotions with practical engineering goals of analyzing emotions in natural language text.

REFERENCES

[1]. B. Liu, “Sentiment analysis and opinion mining,” Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, May 2012.

[2]. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, ser. EMNLP’02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[3]. V. Ng, S. Das gupta, and S. M. N. Arifin, “Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews,” in Proceedings of the COLING/ACL on Main Conference Poster Sessions, ser. COLING-ACL ’06. Stroudsburg,PA, USA: Association for Computational Linguistics, 2006,pp. 611–618

[4]. J. Zhao, K. Liu, and G. Wang, “Adding redundant features for crfs-based sentence sentiment classification,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing, ser.EMNLP ’08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008, pp. 117–126.

[5]. P. Melville, W. Gryc, and R. D. Lawrence, “Sentiment analysis ofblogs by combining lexical knowledge with text classification,” in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD’09. New York,NY, USA: ACM, 2009, pp. 1275–1284.

[6]. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts,“Learning word vectors for sentiment analysis,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, ser. HLT’11.Stroudsburg, PA, USA: Association for Computational Linguistics,2011, pp. 142–150.

[7]. Alexander Pak and Patrick Paroubek, “Twitter as a corpus for Sentiment Analysis and Opinion Mining”, Proceedings of the Seventh conference on International Language Resources and Evaluation, pp. 1320-1326, 2020

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

[8]. Das, Sourav, and Anup Kumar Kolya (2019)"Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm." Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), IEEE, 239-244.

[9]. Garg, Neha, and Rinkle Rani (2020) "Analysis and visualization of Twitter data using k-means clustering." International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 670-675.

[10]. Mishra, Prerna, Ranjana Rajnish, and Pankaj Kumar. (2019) "Sentiment analysis of Twitter data: Case study on digital India." International Conference on Information Technology (InCITe)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds. IEEE, 148 – 153.

[11]. Devika, R., and S. Revathy. (2020) "Survey on clustering techniques in Twitter data." Second International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 1073-1077.

[12]. McCallum, Andrew, and Kamal Nigam. "A comparison of event models for Naive Bayes text classification." AAAI-98 Workshop on learning for text categorisation. Vol. 752. 1998.

[13]. Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 324.

[14]. Zhu, Maoran, et al. "Identification of Opinion Leaders in Social Networks Based on Sentiment Analysis: Evidence from an Automotive Forum." (2016).

[15]. Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[16]. Pang, B., L. Lee, and S. Vaithyanathan, Thumbs up?: "sentiment classification using machine learning techniques", Acl-02 Conference on Empirical Methods in Natural Language Processing, 2002, p. 79--86..