

Online Product Scoring based on Sentiment Based Review Analysis

Adith M Shetty¹, Dhruv R Makati², Monil Shah³, Bhargavi Dalal⁴

^{1,2,3}B.E Student, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India ⁴Assistant Professor, Dept. of IT Engineering, Shah and Anchor Kutchhi Engineering College, Chembur Mumbai India

Abstract - The Internet has accessed a whole new world of shopping for buyers via the web, where viewers can purchase products hassle free, without leaving their respective homes. This process gets easier as the number of customers increase on a daily basis that share their experience and their thoughts on the product. These experiences or reviews could be either positive or negative in regard to the product. These aspects give a clearer view to other potential buyers for making a decision of whether to buy the product or pass it. In this competitive market, this gives rise to agenda driven reviews that are marketed to either boost the sale of a particular product or dampen its sale. Sentimental analysis is a solution to such a pickle, by classifying authentic reviews and comments apart from spam, fake, and untrue replies. The Support Vector Machine (SVM) algorithm works in this case and helps the customers in not wasting hours and hours of their time trying to figure out which product is authentic. This paper proposes a system that provides the users an authentication of any product they wish, based on the online reviews and comments posted by the customers who have already had an experience of it. It aims to provide summarized positive and negative features about products and services by analyzing their online reviews

Key Words: online product scoring, sentiment analysis, support vector machine (SVM), sentiment classification, classification, online reviews.

1. INTRODUCTION

Sentimental analysis is the field of comprehending emotions and expression of humans which they express in the form of reviews, feedback, and suggestions. In this day and age, people have the freedom to express themselves on various platforms, that allow them to connect to a vast majority of the community. These can be shared via various methods, like discussion forums, YouTube channels, reviews blogs, and social media posts. This information can be shared by anyone, it does not necessarily have to be experts of the field. The analysis of this information in a condensed way can be classified into positive, negative or neutral in nature. And it may also be classified based on various variables.

Such freedom to express, may lead to fake information being shared around. This helps in generating false hype and endorsement of a personal product. Agenda driven reviews become a common problem when such a

platform is available. From manufacturers to retailers, anybody might try and perplex potential customers in buying their product. Neutral and veracious reviews help online users in making a right decision for themselves and obtaining the right product suited for them. A correct analysis helps every party involved in this particular instant. Reviews act as a trust-building mechanism and creates a good relationship between the involved entities.

2. LITERATURE SURVEY

In [1], the authors concentrated on classifying online reviews that have been extracted from Amazon by using different algorithms. These algorithms that include Naive Bayes Classifier, Logistic Regression and SentiWordNet algorithm help in determining the polarity of the online reviews. The authors of [2] focused on emotional sentiment analysis which not only will make it possible to classify the reviews as positive or negative, but also to classify their based on the emotion which can be of anger, fear, happiness, disappointment, etc.

[3], have used Naive Bayes Classifier for the polarity classification and analysis using R tool which is a data mining tool. The outcome of the process is presented in the form of a confusion matrix. Support Vector Machine (SVM) technique has been used by the authors of [4] for extracting the polarity of a review. Analysis of frequency, precision and recall has been done by the authors for the proposed algorithm. The Support Vector Machine (SVM) technique outperforms in every case.

When extracting online reviews, one of the major concerns is of handling slang words or misspelled words. Regarding this problem, [5] developed a feature extraction procedure which involved vectorization and preprocessing of the text. Product reviews can be used to extract whether specific features of the product are good or bad according to the customer. In [6], the authors carried out an opinion miner which was feature based. The main work of the miner is to find the significant features of the product by investigating the review and making the assessment profile of every product which can be utilized by the user.

The authors of [7] have exhibited three kinds of new features which incorporate review density, semantic and emotions. The authors have provided the model and

algorithm to build each feature. They have concluded that the proposed model, calculation and features are productive in fake review detection process. Regarding fake review detection, in [8], the authors justify that the utilization of ratings alone to determine whether the

3. PROPOSED WORK

The proposed work includes the Support Vector Machine (SVM) classifier. Support Vector Machine is a supervised machine learning algorithm that is used for classification or prediction based on the trained dataset. Classification is prediction of a label or a group for the entered content. It performs classification or categorization by mapping data to a dimensional space which can help to categorize the data points which normally are not separable linearly. A separating plane in the space is found, which separates between the categories. The data then is transformed in such a way that the found separator can be drawn or used as a hyperplane for categorization. Following are the steps for building a basic model.

- Extracting Data or Choosing the most appropriate dataset for testing and training.
- Pre-processing and Vectorizing the data chosen.
- Developing a Supervised Learning Model for performing the polarity classification.

The dataset chosen for our training and testing is an intensive collection of different online reviews of products from e-commerce giant, Amazon. The entire dataset would be split into training and testing dataset. The analysis is then done for the testing part of the dataset. Once the model is built, it would be used for predicting the real-time scraped reviews of the required product. After classifying the polarity of each review, an average score would be generated for the product which would be displayed to the user.

3.1 Advantages of the Approach

- The system helps in achieving max work with minimal efforts due to its simplicity.
- Real time data can be processed with minimal efforts.
- The performance is more accurate and efficient than other classifiers.
- If any changes are made in the data, it could be updated easily.

3.2 Operational Components

- Web Scraping - Web Scraping is basically a process in which the data is being harvested / gathered from the various places on the web and then the gathered data is being stored in the local storages of your computers for further processing. The library used for the same is BeautifulSoup (bs4).

review is phony or fake is insufficient, as the data that can be extracted is restricted.

- Data Pre-processing - As the data is been extracted from various sources the format of the incoming data can be different if the source is different. Thus, pre-processing is done to give the data a single unified format due to which performing various operations becomes easy.
- Vectorizing - Vectorizing is a very essential and complex step to be performed on the data. The basic work that happens is that it maps the words or phrases with the actual numbers which makes it easy for the system to understand the content and perform complex analysis. TfidfVectorizer is used for vectorization. The above two steps together come under Feature Extraction.
- Classification - The final step is to predict the polarity of the processed and transformed data. This is done by our SVM classifier.

3.3 System Workflow

Step by step breakdown of the project workflow.

1. The user is supposed to copy the URL of the product he/she wants to get the rating of from the original website such as flipkart, amazon, etc.
2. User is required to enter the copied URL and click on the button to get the rating of that particular product.
3. The reviews of the entered product would be scraped and stored.
4. The scraped reviews would then be featured extracted to our model' suitable format.
5. Each scraped review would be classified to a polarity.
6. Based on the classification of the online reviews, an overall rating would be computed by us and provided to the user. Based on the rating, the user can determine whether she/he wants to buy the product or not.

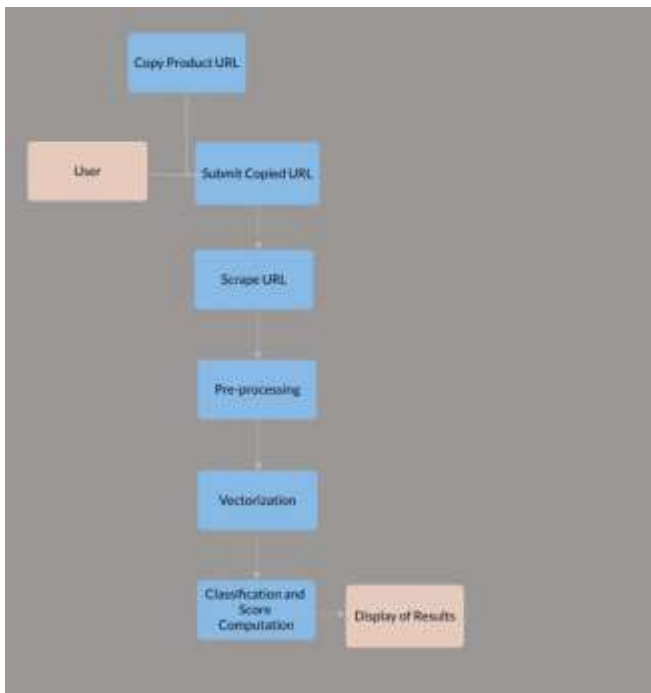


Fig -1: System Workflow

4. CONCLUSION

The system pertaining to this paper is a state-of-the-art review analysis system that is much better than the evaluated review systems of some E-commerce giants. Overall, this paper pursues a faster process, SVM algorithm is the best choice for this problem without sacrificing too much accuracy. On the basis of different features being selected, we will be able to determine the accurate performance and customer satisfaction of the product. This would help to establish a proper review system for genuine customers, who are worried about the product performance. Also, this would be making it possible for different E-commerce brands to establish them and the products they sell. This system would enable users to make a faster, more accurate, and by far a more desirable choice. It would also eliminate agenda driven campaigns, and would move the online shopping platform towards an honest and more trust-based direction.

5. REFERENCES

[1] K. L. S. Kumar, J. Desai and J. Majumdar, "Opinion mining and sentiment analysis on online customer review," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCCIR), Chennai, 2016, pp. 1-4.

[2] Z. Singla, S. Randhawa and S. Jain, "Statistical and sentiment analysis of consumer product reviews," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6.

[3] P. P. Surva and B. Subbulakshmi. "Sentimental Analysis using Naive Bayes Classifier," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-5.

[4] Abinash Tripathya, Ankit Agrawal, Santanu Kumar Rath " Classification of Sentimental Reviews Using Machine Learning Techniques ", 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[5] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, 2013, pp. 1-5.

[6] A. Angelpreethi and S. B. R. Kumar, "An Enhanced Architecture for Feature Based Opinion Mining from Product Reviews," 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, 2017, pp. 89-92.

[7] Y. Li, X. Feng and S. Zhang, "Detecting Fake Reviews Utilizing Semantic and Emotion Model," 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), Beijing, 2016, pp. 317-320.

[8] Wahyuni, Eka & Djunaidy, Arif. (2016). Fake Review Detection from a Product Review Using Modified Method of Iterative Computation Framework. MATEC Web of Conferences. 58. 03003. 10.1051/mateconf/20165803003.

[9] A. Agarwal, V. Sharma, G. Sikka and R. Dhir, "Opinion mining of news headlines using SentiWordNet," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, 2016, pp. 1-5.

[10] P. Kherwa, A. Sachdeva, D. Mahajan, N. Pande and P. K. Singh, "An approach towards comprehensive sentimental data analysis and opinion mining," 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 606-612.