

Product reviews and sentiment analysis using machine learning

Prof. D. S. Shingate¹, Harshada Godse², Snehal Shinde³, Ankita Kshirsagar⁴, Vidya Kale⁵

¹ Professor, Department of Information Technology, Matoshri college of engineering and research centre, Nashik, Maharashtra, India

^{2,3,4,5} Students, Department of Information Technology, Matoshri college of engineering and research centre, Nashik, Maharashtra, India

ABSTRACT

Abstract - Natural language processing, opinion mining, text summarization, and other related research fields are hot right now. Sentiment analysis or opinion mining utilizes many NLP techniques and tools to analyze unstructured data in order to obtain people's emotions toward specific entities. Essentially, these strategies allow a system to understand what humans are saying. Sentiment analysis employs many ways to detect the sentiment of a text or sentence [1]. People communicate their subjective opinions and experiences with one another on various online platforms. Obtaining relevant information on a product can be a time-consuming process. Companies may be unaware of all of their customers' needs. Product reviews can be used to determine how people feel about a particular topic. However, because these are lengthy, a summary of positive and negative reviews is required. The main focus of this paper is a study of the methods and techniques used to extract a feature-wise summary of the product and analyze it to generate an accurate review. More product review websites will be included in the future, as well as higher-level natural language processing jobs. Using the best and most innovative strategies or tools. for more accurate results, the system extracts only those keywords in the dataset and eliminates the rest of the terms.

Keywords: - Sentiment analysis, social media, Twitter, tweets.

1. INTRODUCTION

The Internet is changing the way people express their thoughts and feelings. It is now done primarily through blogging, internet forums, product online surveys, social media, and so on. Nowadays, thousands of individuals use social networking sites such as Twitter, Facebook, and Google Plus are places where people can express their emotions, and thoughts, and share information about their daily lives. [2]. We obtain interactive material from internet communities, where customers inform and influence others through forums. Twitter, status updates, blog posts, comments, reviews, and other forms of social media generate a significant volume of sentiment-rich data. Furthermore, social media gives a platform for businesses to communicate with their customers for advertising purposes. People rely heavily on user-generated content on the internet in order to make judgments [3][4]. For example, before purchasing a product or using a service, people will study its reviews online and discuss it on social media. The volume of content provided by users is just too large for a typical user to analyse. As a result, numerous sentiment analysis approaches are extensively utilised to automate this.

2. LITERATURE SURVEY

This paper [5] is mainly concerned with analysing customer feedback from the e-commerce industry. According to a survey of leading e-commerce websites, the product rating supplied by a consumer is not always compatible with the product review published by him/her. Text reviews, in fact, provide an accurate representation of the product. To address this issue, the mentioned system will provide a boolean result, indicating whether the product is good or bad, and the user will not be required to read all of the reviews in order to examine the product. They used supervised machine learning algorithms such as Support Vector Machine (SVM) and Nave Bayes, as well as binary model (BM) and TF-IDF, to investigate the effect of various term weighting functions on sentiment analysis accuracy.

Deep learning has recently recognized as an efficient method for handling sentiment detection issues. Without human intervention, a neural network learns a meaningful representation on its own. However, the availability of large-scale training data is critical to the success of deep learning. The author introduces a new deep learning architecture for product review sentiment categorization in this research [6], which employs widely available ratings as weak supervision signals. The framework is divided into two steps: (1) developing a high-level representation (an embedding space) that captures the overall sentiment distribution of sentences using rating information; and (2) adding a classification layer on top of the embedding layer and using labelled sentences for supervised fine-tuning. We study two types of low-level network structure for simulating review phrases: convolutional feature extractors and long short-term memory. To test the proposed approach, we created a dataset from Amazon that included 1.1 million weakly labelled review sentences and 11,754 tagged review sentences. The experimental results demonstrate the efficacy and superiority of the proposed framework over baselines using deep learning-based tools and techniques.

3. SYSTEM ARCHITECTURE

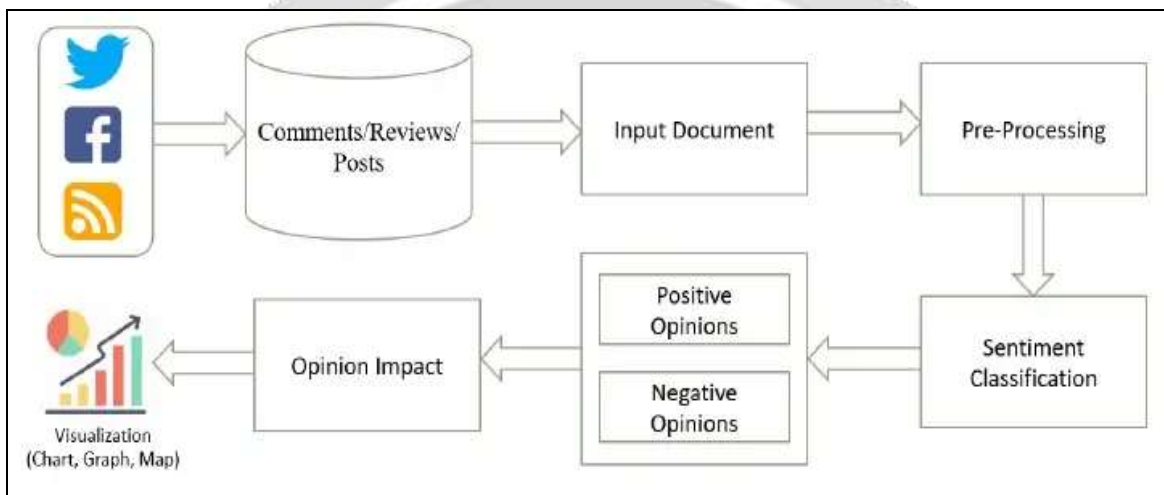


Figure 1 - System Architecture

We mined tweets in this system using Twitter's Search API and then processed them for further analysis, which included Natural Language Processing (NLP) methods like tokenization, stop word removing, stemming/lemmatization etc. Following that, we used the Text Blob library to predict the sentiment of each tweet which is pre-trained using the naïve bays algorithm with millions of worlds [9]. After predicting the sentiment of each tweet, we grouped the reviews into the three categories i.e. positive, negative, and neutral. Then, using the bar graph and the word cloud we are visualizing the result. In our project, we used Python programming to extract tweets from Twitter. Python is a popular programming language for statistical computing and machine learning algorithms. Twitter API was used to create an application and obtain authorization in order to extract tweets from Twitter. To complete this task, we use NumPy, Pandas, Matplotlib, Text blob, and request libraries.

The complete framework designed for sentiment analysis is as follows:

1. API Setup and Fetching data using API – In this step, we are accessing Twitter API using Python's TWEETPY Library. To access the Twitter API, you will need 4 things from your Twitter account.
 - consumer key
 - consumer secret key
 - access token key
 - access token secret key

2. Pre-processing or Data cleaning

- Tokenization - Tokenizing sentences to break the text down into sentences, words, or other units by detecting whitespace in words.
- Removing stop words - Removing stop words like “if,” “but,” “or,” and so on.
- Lemmatization - Normalizing words by condensing all forms of a word into a single form.

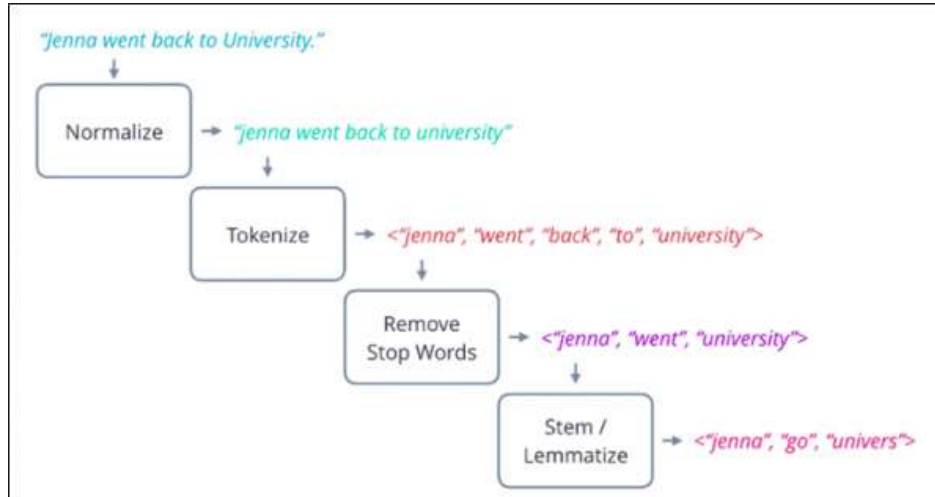


Figure 2 - Pre-processing

3. Perform Sentiment Analysis – In this step, we are checking the sentiment of a given tweet using the TEXT BLOB library which is widely used in NLP.
4. Visualization – After successfully detecting sentiment for the given word we are showing the outcome in the form of a bar graph using matplotlib library and word cloud using the seaborn library.

4. RESULTS

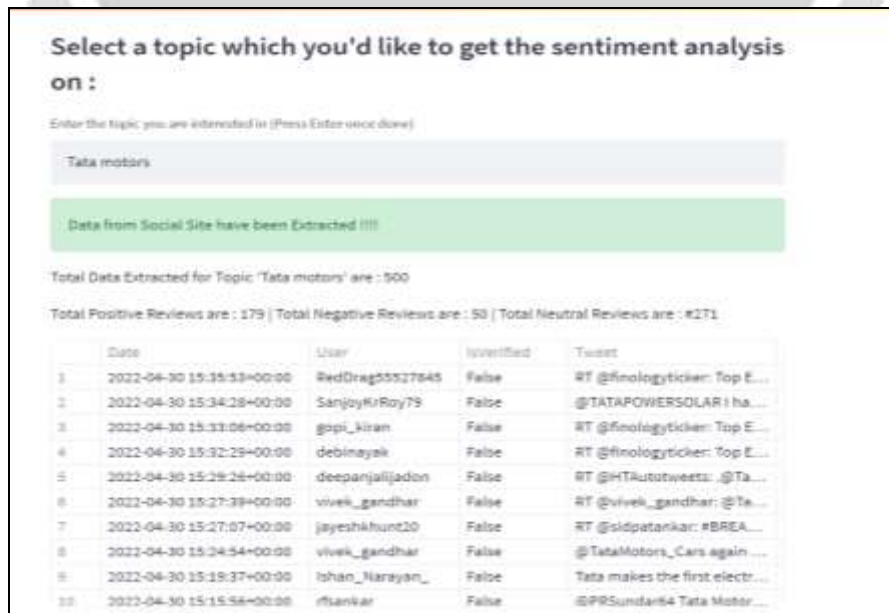


Figure 3 – Result -1

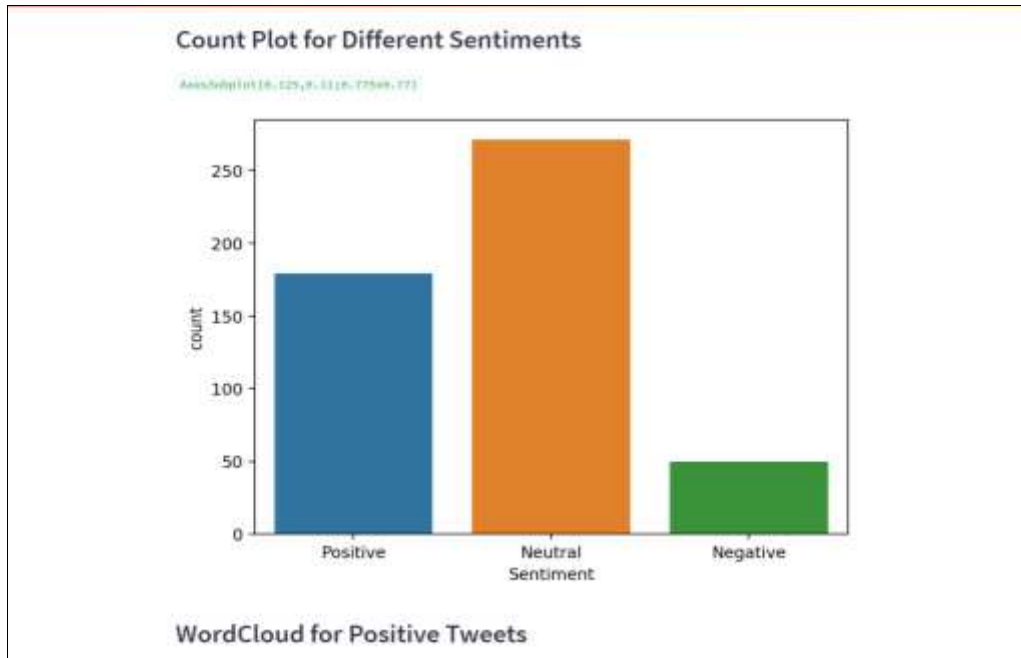


Figure 4 – Visualization

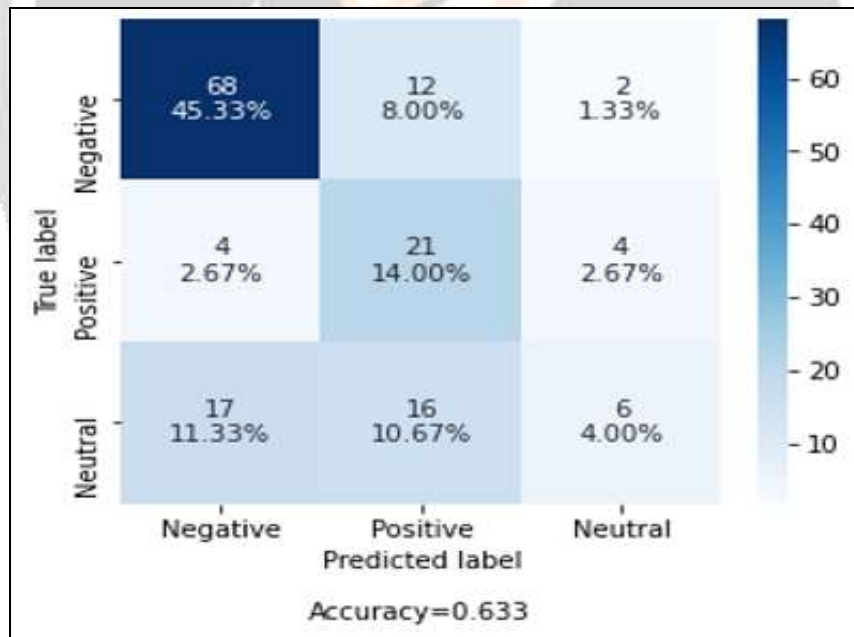


Figure 5 - Analysis

For this exercise, I chose 20 goods and gave them 150 random reviews to categorize them as Positive, Negative, or Neutral. I calculated a final sentiment rating based on the average of 20 product reviews. That was the benchmark. We can now make a comparison. Based on this experiment, I constructed confusion matrices to determine the accuracy of an algorithm in comparison to human-analyzed texts.

5. CONCLUSIONS

In this work, we developed Twitter sentiment analysis web application to evaluate customers' perceptions of the essentials to market success. The application utilizes a natural language-based approach that is more accurate for sentiment analysis; machine learning techniques are also employed. As a consequence, the computer categorizes reviews as positive, negative, or neutral, which is represented in a bar graph and word cloud. The application is developed as a web application utilizing Python, streamlit, and the Twitter API.

6. REFERENCES

- [1] Linguamatics, What is Text Mining, Text Analytics and Natural Language Processing? [Online- Blog]
- [2] Available at: <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing>.
- [3] Bakopoulos, Vassilis & Baronello, John & Briggs, Rex. (2017). How Brands Can Make Smarter Decisions in Mobile Marketing: Strategies for Improved Media-Mix Effectiveness And Questions for Future Research. *Journal of Advertising Research*. 57. 447-461. 10.2501/JAR-2017-052.
- [4] Ballestar et al., 2019, M.T. Ballestar, P. Grau-Carles, J. Sainz, Predicting customer quality in e-commerce social networks: A machine learning approach, *Review of Managerial Science*, 13 (3) (2019), pp. 589-603.
- [5] Balaji et al., 2016, M.S. Balaji, K.W. Khong, A.Y.L. Chong, Determinants of negative word-of-mouth communication using social networking sites *Information & Management*, 53 (4) (2016), pp. 528-540.
- [6] S. N. Singh and T. Sarraf, "Sentiment Analysis of a Product based on User Reviews using Random Forests Algorithm," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 112-116, doi: 10.1109/Confluence47617.2020.9058128.
- [7] W. Zhao et al., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 185-197, 1 Jan. 2018, doi: 10.1109/TKDE.2017.2756658.
- [8] R. M. Duwairi and I.Qarqaz, "A framework for Arabic sentiment analysis using supervised classification", *Int. J. Data Mining, Modelling and Management*, Vol. 8, No. 4, pp.369-381, 2016.
- [9] Shoukry, Amira & Rafea, Ahmed. (2012). Preprocessing Egyptian Dialect Tweets for Sentiment Mining.
- [10] [Online Blog] Sentiment Analysis using NLTK – A Practical Approach, Sriniketh J — June 25, 2021, <https://www.analyticsvidhya.com/blog/2021/06/sentiment-analysis-using-nltk-a-practical-approach/>.
- [11] L. Wasser and C. Farmer, "Sentiment Analysis of Colorado Flood Tweets in R", Earth Lab, 2018. [Online]. Available: <https://earthdatascience.org/courses/earth-analytics/get-data-usingapis/sentiment-analysis-of-twitter-data-r/>. [Accessed: 01- Mar- 2018].
- [12] Sentiment Analysis using TextBlob, Parthvi Shah Jun 27, 2020, [Online]. Available: <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>.
- [13] How to build a real-time live dashboard with Streamlit, By AbdulMajedRaja RS, Posted in Community, April 21 2022 [Online] Available at: <https://blog.streamlit.io/how-to-build-a-real-time-live-dashboard-with-streamlit/>.
- [14] A Beginner's Guide to matplotlib for Data Visualization and Exploration in Python, Aniruddha Bhandari — February 28, 2020 [Online] Available at: <https://www.analyticsvidhya.com / blog / 2020 / 02/ beginner - guide -matplotlib -data - visualization -exploration - python/>.
- [15] Product Quality Assessment using Opinion Mining in Persian Online Shopping Fatemeh HosseinzadehBendarkheili; Rezvan MohammadiBaghmolaei; Ali Ahmadi 2019 27th Iranian Conference on Electrical Engineering (ICEE) Year: 2019 | Conference Paper | Publisher: IEEE