Grenze International Journal of Engineering and Technology, Jan Issue

A Novel Web Application of Product Recommender System based on Sentiment Analysis

Dr Lavanya Pamulaparty[1], Mohammed Adnan[2] and Abdul Moid Khan Mohammed[3]
[1-3]Methodist College of Engineering and Technology/CSE, Hyderabad, India
Email: lavanya.post@gmail.com, mohd.adnan2317@gmail.com, skhan852000@gmail.com

***Abstract*—In present scenario the data overload is a serious problem. for mere humans, its impossible to analyse data manually without error or bias. One prime example of this regarding the review of products on e-commerce websites is that when a customer buys a certain product, he first looks for its ratings and then decides whether to buy that item or not. But only ratings are insufficient to determine the quality of the product. Reviews give us a much clearer and more in-depth understanding of the product. There is a need to summarize the feedback in a meaningful manner. Here's where sentimental analysis comes into play. Sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition that isn't always right. This proposed model combines sentiment analysis and a recommendation system to provide users with a comprehensive recommendation of e-commerce products. The aim of this analysis is to determine the orientation of a review then the recommender module provides users with product recommendations by analyzing large amounts of user-supplied data that is ratings and the sentiments of users and other factors like price and the number of sales to provide the user with the best recommendation. This is developed in the Python language.**

***Index Terms*— Sentiment Analysis, Product Reviews, Vader Sentiment.**

I. INTRODUCTION

In a world driven by data, the limitations, and possibilities of what we can carry out with the data are unfathomable. Whether it is data from a football game or a student's database, there are times when we can give meaningful execution to the data at hand and use it to solve a variety of issues and make our lives easier.

Companies may have bought mountains of client feedback in today's data-overloaded atmosphere. However, it is still hard for normal people to manually analyse it without inaccuracy or bias. A perfect illustration of this relating to product reviews on e-commerce websites is the fact that when a buyer buys a product, he first checks its ratings before deciding whether to buy it. However, merely ratings are insufficient to adequate for deciding the product's quality. Reviews supply a far more precise and comprehensive picture of the product. Additionally, we may have our own prejudiced, predetermined notions on the issue at hand, which may affect how we understand the material we must analyse. We will also need to synthesise the input into a handful of actionable insights so that it is useful to the user. Users want insights to make an educated decision before buying a product. And they are inadequate. We must devise a more efficient strategy for gaining the most valuable ideas. Here, emotive analysis comes into play and aids the user in condensing evaluations and providing a concise insight. Sentiment analysis may be automated, and choices can be made based on a substantial quantity of data as

Grenze ID: 01.GIJET.9.1.757

opposed to simple intuition, which is not always accurate. In addition to using emotion scores, a recommender system also evaluates parameters such as pricing, reviews, and purchase. It finds the best product for the consumer and provides a mathematical reason for its selection.
They may be used to measure and record a vast array of internal and external company activity. While the data itself may not be particularly useful, it serves as the foundation for all reports and is therefore vital to company. Consumer data are measurements about customer engagement. It might be the number of jobs, the number of inquiries, the revenue earned, the costs incurred, and so forth. We need statistics to understand our interactions with the consumer.

II. Literature Survey

Sentiment analysis is a popular topic. It is a new technique to making effective use of linguistic knowledge because everyone has a unique perspective on things. The analysis of these opinions can provide useful information. The combination of web scraping, sentiment analysis, and the Recommender system model provides better recommendation outcomes. The sentiment analysis output helps in the generation of more intricate recommendation outcomes.
Alka Singhal et. al [5] have researched on how recommendation systems Work in an IT environment and in the real world. The two areas of text mining and sentiment analysis, which are often studied separately, are combined in this study. The textual tweet data we got from Twitter constitutes the sole subject matter of this article. Future additions to this data set could include photographs, audio, internet articles, multimedia, etc. Geospatial data is completely disregarded. All of these qualities may be taken into account in the future.
Satuluri Vanaja et. al. [6] Their work uses Amazon customer review data and focuses on finding aspect terms

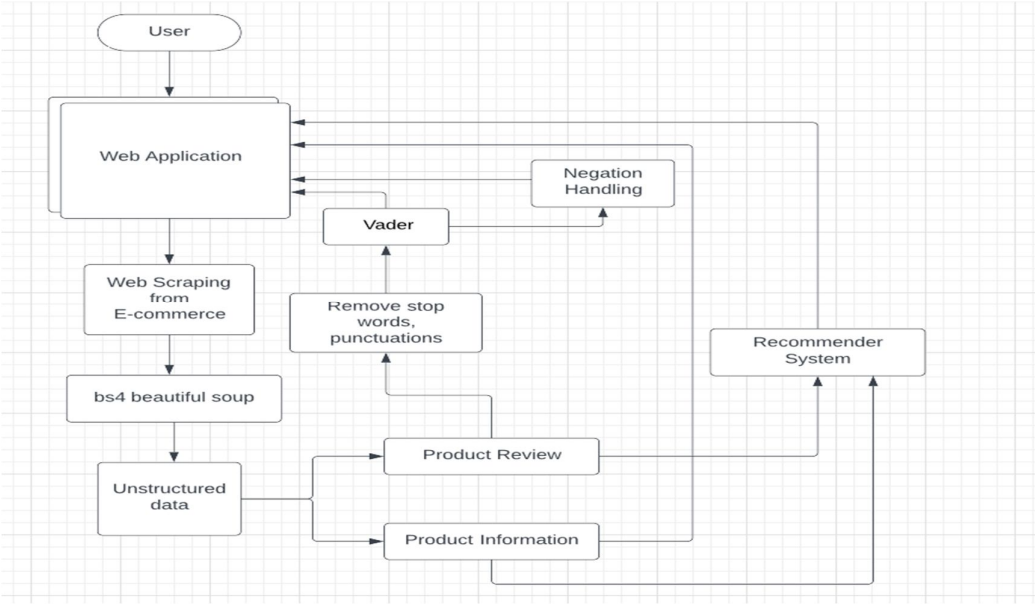

Figure-1: Proposed Model of Web App For Sentiment Analysis

from each review, identifying the Parts-of-Speech, applying classification algorithms to find the score of positivity, negativity, and neutrality of each review. In future the focus is on implementing advanced algorithm and compare with Naïve Bayes.
Shivaprasad T K et. al. [7] more research work is needed to improve the performance measures further. Sentiment analysis or opinion mining can be applied for any new applications which follow data mining rules.
The study of Prashast Kumar Singh et al. [8] entails the measurement of the sentiments of thousands of (good) reviews as well as the filtering of irrelevant and unhelpful remarks. Last but not least, giving the final user (business/manufacturer) a summary of the information regarding the expressed sentiments in the form of clear and understandable graphs, charts, and other visualisations. In the future, it will be possible to mine reviews from Twitter, which is rich in such slangs due to the 140-character limit on tweets, which forces users to use shortened forms, by accounting for Internet slang language and expressions like "LOL (laughing out loud)"[8].

Kim Schouten et. al. [9] research aims at finding the right implicit feature without this pre-knowledge. The proposed algorithm better in distinguishing between sentences that have an implicit feature and the ones that do not. In future work a shift to word-based methods like this one to concept-based methods will be intriguing.

III. Proposed System

The proposed system is intended to provide a user-friendly interface to the web application. Our project addresses the limitations encountered by data overload. The data cannot be analysed manually without error or personal prejudice. The sentiment analysis tool, in addition to saving time required for filtering through thousands of reviews, also overcomes the shortcomings of user biases and preconceived opinions about the product at hand, which can influence how users interpret the reviews they must analyse. Large amounts of relevant data will be scraped in real time from Flipkart. The data will be cleaned and pre-processed, the data is then analysed using VADER. The VADER sentiment score result is then processed and displayed to the user. The recommendation module reduces the user's time required to analyse product information and compare it with competitors by incorporating the relevant factors that is user sentiment, price, quantity sales, and ratings for a more enhanced recommendation. A comprehensive formula has been developed to include all relevant factors and recommend the best product.
Figure-1: Proposed Model of Web App For Sentiment Analysis
In the above Figure 1 explains the overview of the web app and how its modules work out. Firstly, the user engages the web application through his browser and sends a POST request to an e-commerce website, here in our case, Flipkart. To achieve this web scraping is done using the beautiful soap bs4 library to extract unstructured data from which product reviews and product info are retrieved.
Further division of this data is done, product info is used by the recommender system as well as is sent to the front-end. On the other hand, reviews are analyzed, and sentiment score is calculated

IV. Implementation

A. Web scraping

We used a Self-built Web Scraper that extracts underlying HTML code, and, with it, data stored in a database the scraper then copies the entire website's content, which is then processed.
Although many huge websites, like as Google, Twitter, Facebook, Stack Overflow, etc., supply APIs that
allow you to access their data in a structured fashion which is the ideal choice, Flipkart does not allow customers to obtain big volumes of data in an organized format, or they are not technologically advanced enough, therefore we scraped their website for data using web scraping.
The Web scraper was built in Python using the Beautiful Soup package, which is ideally suited for Web Scraping. It generates a parse tree that may be used to extract HTML data from a website. Additionally, Beautiful Soup includes many navigations, search, and modification tools for these parse trees. In the web app, web scraping is used to scrape the following data
Product title
Product image
Product price
Product stars
No of people who bought the product
No of reviews on the products
Link of the page having the reviews

B. Data processing

In the proposed research we have used Natural Language Toolkit, or more commonly called NLTK, is a Python-based toolkit and set of programs for symbolic and statistical natural language processing of the English language. It includes libraries for tokenization, parsing, classification, stemming, tagging, and semantic reasoning

C. Negation handling for word jargons

The part of the sentence that negation effects are called the vicinity or scope of negation. We performed explicit Negation handling. It is proposed for negation and inverting the polarities of words, These words are opinionated.

Figure 4.1: A Negative review with positive words

```
[[], [('Not great', 1), ('Not fit.', 1), ('Not best', 1)], [], []]
```

Figure 4.2: Concatenating" Not" to positive adjectives

D. Sentiment Analysis

Sentiment Analysis, as its name implies, is the process of extracting sentiment and opinions from a collection of reviews or the emotions behind a situation. It further condenses the data into the behaviour that the data senses. The purpose of this study is to automatically determine the direction of a review. We have employed VADER to perform sentimental analysis.

E. Vader Sentiment

VADER makes use of a variety of a list of lexical properties that have been classified as having a positive or negative semantic orientation. It displays the strength of the emotion, or how positive or negative a sentiment is, in addition to the positivity and negativity scores. The NLTK package contains VADER, which can be used to instantly process unlabeled text data.

F. Recommendation System

A recommender system is a way to tell the user based on certain calculations and algorithms the product which the system recommends. Here we recommend the user a product based on the following formula.

$$\alpha + \left[\frac{\text{No. of orders} \times 2}{\text{Average of all orders}}\right] + \left[\text{Ratings} \times \beta\right] + \left[\text{Sentiment Score} \times \Omega\right]$$

Figure 4.3 : Recommender System Formula

α = price coefficient
β = ratings constant (8.4)
Ω = sentiment constant (0.4)
The formula in Fugure 4.3 is applied to all the products and a score is calculated; the highest score is the recommended product.

Price Coefficient

Price	Ratings	Sentiment Score	Price Coefficient
↓	↑	↑	8
↓	↑	↓	7
↓	↓	↑	6
↓	↓	↓	5
↑	↑	↑	4
↑	↑	↓	3
↑	↓	↑	2
↑	↓	↓	1

Figure 4.4 : Price Co-efficient table

Each price, rating and sentiment score is compared against their averages and the price coefficient is calculate. For ex, if the price is higher, ratings are higher and the sentiment score is lower, then the price coefficient is 3.

V. EXPERIMENTATION

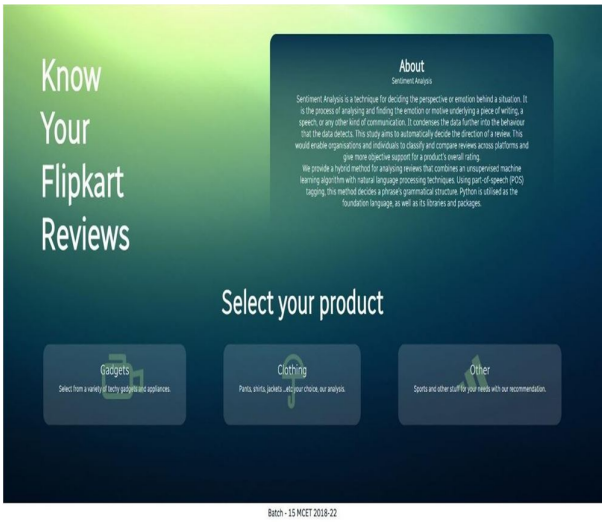

Figure 5.1 : Landing page

Figure 5.2: Search of a bottle

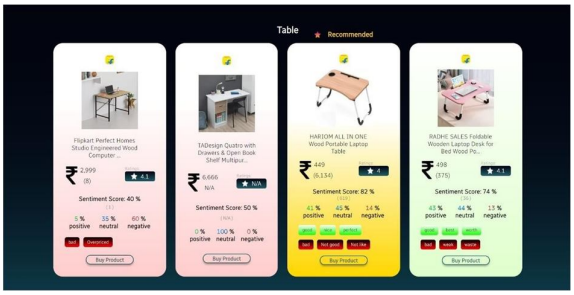

Figure 5.3: Search of a table

Figure 5.4 : A positive product

VI. CONCLUSION AND FUTURE SCOPE

A recommender system that considers not just the product ratings and price but also the overall sentiment score of the reviews. An intricate formula with just calculations that puts proper weightage on its parameters to get a perfect recommender score. The product with the largest recommender score is the recommended product. The whole idea behind this computation is to save the users time which he/she takes to analyse the reviews and find the right product.

Content-based filtering can be integrated into this model which works with the data that we take from the user, either explicitly or implicitly. With the data, we create a user profile, which is then used to suggest actions to the user. As the user provides more input or takes more action on the recommendation, the engine becomes more accurate. Furthermore, a more comprehensive formula can be formulated to enhance the accuracy.

REFERENCES

[1] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 298-301, doi: 10.1109/ICICCT.2017.7975207.

[2] P. K. Singh, A. Sachdeva, D. Mahajan, N. Pande and A. Sharma, "An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites," 2014 5th International Conference - Confluence the Next Generation Information Technology Summit (Confluence), 2014, pp. 329-335, doi: 10.1109/CONFLUENCE.2014.6949312.

[3] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1275-1279, doi: 10.1109/ICIRCA.2018.8597286.

[4] https://www.globalscientificjournal.com/researchpaper/Research_Paper_On_Recommendation_System. pdf
[5] Sharma, Ankita and Udayan Ghose. “Sentimental Analysis of Twitter Data with respect to General Elections in India.” Procedia Computer Science 173 (2020): 325-334.
[6] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1275-1279, doi: 10.1109/ICIRCA.2018.8597286.
[7] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), 2017, pp. 298-301, doi: 10.1109/ICICCT.2017.7975207.
[8] P. K. Singh, A. Sachdeva, D. Mahajan, N. Pande and A. Sharma, "An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites," 2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 329-335, doi: 10.1109/CONFLUENCE.2014.6949312.
[9] Schouten, K., Frasincar, F. (2014). Finding Implicit Features in Consumer Reviews for Sentiment Analysis. In: Casteleyn, S., Rossi, G., Winckler, M. (eds) Web Engineering. ICWE 2014. Lecture Notes in Computer Science, vol 8541. Springer, Cham. https://doi.org/10.1007/978-3-319-08245-5_8