Proceedings of the 8th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'22)
Prague, Czech Republic - July 28- 30, 2022
Paper No. MHCI 112
DOI: 10.11159/mhci22.112

Optimizing Business Sales and Improving User Experience by Using Intelligent User Interface

Sayli Arjun Pednekar[1], Swati Chandna[2]
SRH Hochschule Heidelberg
Heidelberg, Baden-Württemberg, Germany
saylipednekar19@gmail.com[1]; swati.chandna@srh.de[2]
Department of Applied Computer Science
Heidelberg, Baden-Württemberg, Germany

Abstract **–** This research explores the impact on the user experience when the users, that is, the people in business, are exposed to an improved version of an intelligent user interface of the review management software. Machine learning algorithms, such as Lexicon-based sentimental analysis and NRC Emotion recognition, are employed to assist the proposed review management software, Review Dock. To provide additional assistance, a Content-based Recommendation system is integrated. More than 17,000 Amazon reviews are used to generate the results. To improve the satisfaction level of the already created prototype, three iterations of usability testing were conducted on nine participants. The findings show that by following the Web Content Accessibility Guidelines (WCAG) standards, an average satisfaction score of 2.49 out of 5 on the first iteration is significantly improved to 4.9 on the last iteration.
Furthermore, the polarity categorization is similar across most evaluations, which are accomplished on previously unseen data sets. However, the results also reveal that the designs will only perform well for a small-medium industry. This research attempts to fill the limitations in the literature with respect to user experience. Regardless of the tools offered, the issue for businesses in utilizing an available solution that diminishes the engaging experience remains unchanged. As a result, a new solution should solve the limits, which will directly affect the company's sales. The research question states what steps the review management software may take to reduce the overly convoluted user interface? Therefore, proposing a solution called Review Dock will provide a plethora of responses and entirely focus on customer happiness by providing a comprehensive overview of how to enhance a product's sales.

Keywords**:** User Experience; Customer Satisfaction; Behavioural Detection; Lexicon-based Sentimental Analysis; Rule-based Recommendation System; NRC Emotion Recognition.

1. Introduction

The e-commerce industry has undergone a significant transformation in business [1]. All organizations have acknowledged that the web serves as a tool in moulding the productivity and efficiency of the business [2]. Researchers and scientists are attempting to find an efficient way to extract emotions from text or speech. Several methodologies dealing with words and phrases have been established, including natural language processing, machine learning, and so on [3]. Customer reviews are an important factor in business growth. A positive review can also showcase the incredible work of the manufacturers or businesses. Still, a negative review can have a severe influence on the business's reputation [4] and cannot be overlooked. To maintain a competitive advantage, the manufacturer can evaluate customer feedback better to understand their sentiments about the product [5].

According to Grand View study, the Business to Customer retailing category will have the largest revenue in 2020, accounting for over 99.4 percent of the total. It was also anticipated that by the end of the same year, the entire value would be 3.67 trillion US dollars. The study also underlines the fact that it will rise by 9.7 percent from 2021 to 2028 at a compound annual growth rate, or CAGR [6]. To elaborate on their behaviour, if the customer's transaction history is understood, it will be easier for the business to predict what their customers like the most [1]. To keep track of how well a business is selling its products online, review management software is used. If a product has a bad rating, the company must understand what emotion is influencing the total rating. Even if the feature has been implemented, it is not drawn visually appealing on the screen and brought forward. This research focus on developing an intelligent user interface mainly focuses on the customer satisfaction and the user experience while using the software. To obtain more accurate results, an attempt is made to segregate

the emotions more profoundly. To validate the user interface, three iterations of usability testings are performed till a consistent result is not obtained.

The objective is to improve the user experience of the proposed Review management software by finding a technique to categorize a sentiment such as positive, negative and neutral into visually appealing format by delving into technology by using the existing machine-learning algorithm. And to present it to the businessman in a very simple user interface format.

2. Literature Survey

Three methods were combined to detect the behavioural change in users after downloading the fitness or nutritional applications. The research was successful in finding out positive and negative reviews based on the mobile application, it also faced limitations such as the text mining tool was not able to perform appropriately in the presence of typographical errors or text inconsistencies [7]. A new sentimental analysis model SLCABG, which is a combination of the sentiment lexicon and Convolution Neural Network and attention-based Bidirectional Gated Recurrent Unit. Although model was demonstrated to be superior, but it could only create outcomes and categorize them as positive or negative, which is insufficient for most systems [8]. To extract the necessary sentiments from the text, a lexicon-based approach was used. The data was not pre-trained and was used to identify the polarity and subjectivity of the text given. The obtained results are in both positive and negative formats [9]. [10] proposed that the semantic information of the word is contained in the word vector, which has a low dimension. However, distributed word vectors do not contain sentiment information about words. The contribution of the word's sentiment information to text sentiment classification is embedded into the classical term frequency-inverse document frequency technique in this paper, and the weighted word vector is formed. This research proposes a sentiment analysis method for comments based on bidirectional long, short term memory (BiLSTM). However, the sentiment analysis method of comments based on BiLSTM consumes a long time in the training model.

3. Research Methodology

The overview of the system architecture consists of the front-end which communicates with the back end. The data fed to the system consists of 17,000 reviews which are collected from amazon.com using the free tool amazon web scrapper. Following the implementation of the sentimental analysis code, it was discovered that reviews more than 15,000 were providing findings that were close to accurate, otherwise, the results were ambiguous. Therefore, more than 17,000 reviews were scrapped in order to get accurate results in the sentimental analysis.

3.1. Competitor Analysis

A thorough competitor analysis was carried out, and limits were identified. Out of the various software, the apps and services in the top spots do have many designs and experience concerns. User Interface and User Experience related issues such as too much content on one page - complex for an eye to catch all the information in one go, repetition of data in different formats, asks to create an account before showing what they are offering, undefined Visual and Horizontal Hierarchy, incorrect use of white space, improper typography and not covering the entire age group were observed. Other technical limitations like Segregation of reviews only into positive, negative, and neutral sentiments, improper graph selection, no automatic reply to the preview function, it does not have a function to compare with similar products that were identified. A survey was carried out by having participants fill out Google forms.

3.2. Core Functionalities

Once all the requirements and necessary data were collected, the following features were shortlisted out of many.

1. To design a simple User interface, that is, avoiding a populated user interface and providing plenty of white space to help users distinguish between different components.
2. Limited use of graphs.
3. Letting the user know why a particular product is rated negatively in the form of pie chart and bar graphs. That is, recognition of emotion in particular sentiment.

4. Allowing the user to compare one product with another.
5. Showing a brief analytic of the selected product

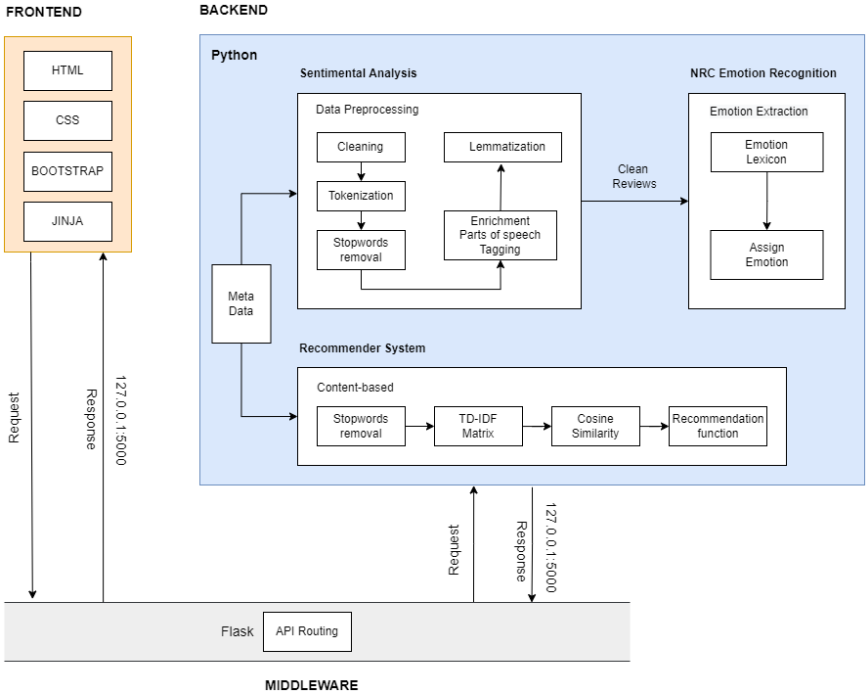

Fig. 1: System Architecture

4. Implementation

Python is chosen as the back end as it has already built-in libraries for Sentimental analysis and Recommender systems. The entire application will be built using Flask as the middleware, Python for the back end and HTML (HyperText Markup Language), CSS (Cascading Style Sheets), JINJA for the front end.

4.1. Sentimental Analysis

Five steps to pre-process the data were followed to achieve a good result.

1. Cleaning the data:

The data was first imported using the Pandas. Pandas is a python library that helps to import and export or work with the data. The columns present in the .csv file are, ID, Rating, Reviews, Date, Tag, OrderID, OrderType. After importing, only the required columns such as Rating, Reviews and Tag were compressed, and a new array was created. Using the clean function, the data is cleaned. That is unnecessary punctuation marks were removed.

2. Tokenization:

With the use of the NLTK tokenize function, the entire sentence was tokenized into words, as followed.
Example - “A very good product I will order it again”
Tokenized review - [“A”, “very”, “good”, “product”, “I”, “will”, “order”, “it”, “again”].

3. Part of speech tagging:

The tokenized review is further tagged with a part of speech for lemmatization purpose. As the author is not training the data here, it is very necessary to understand the actual meaning a word is trying to represent and for this purpose identifying the part of speech is very important. This can be easily achieved by NLTK pos_tag() function in python.

4. Removal of the stop word:

As many tuples which already have been tagged consist of words like a, an, it, they, do not symbolize any meaning to derive the sentiments from the statement. Therefore, it will only be good enough to remove these articles and other unnecessary words which will only increase the computing time. These words are also known as the stop words.

5. Lemmatization / stem word:

Lemmatization is a process where a word is further cropped down to retrieve its stem value.

Example - office, offices, came, come, cook, cooking

Lemmatized word - office, come, cook.

After these five steps, sentimental analysis is performed using a python library TextBlob which also calculates the Polarity and Subjectivity of the sentence. This is achieved by using the method of NLTK Freq Dist, which displays the most occurred positive or negative words in all the reviews.

4.2. NRC Emotion Recognition

The metadata produced during the sentimental analysis is fed to the NRC Lexicons. The NRC Emotion recognition is used for word-level emotion segmentation. It has a list of English terms and their correlations with 8 major emotions [11].

4.3. Content Based Recommender System

In Content-based recommenders along with the basic information, metadata such as product description, product category, title, and so on is also considered to generate a similarity ratio. Rather focusing on what others like, a preference ratio is calculated based on the person's liked history [12]. In terms of data and its amount of disproportion, cosine similarity is used which produces a metric that assesses the level of similarity regardless of size. It measures a cos angle between vectors projected in three dimensions, as the name implies. Each word carries its semantic meaning. This is achieved by Term Frequency-Inverse Document Frequency. It provides with several occurrences of each word for each product. After having the matrix, it becomes very easy to fetch the cosine similarity.

5. Results

The algorithms such as Naive Bayes, Valence Aware Dictionary and Sentiment Reasoner and TextBlob were used to compare results in terms of accuracy. For the Naïve Bayes algorithm, the algorithm took longer to compile as it had to train the data first and then test it. Another reason for not choosing this algorithm is that the requirement to display behavioural aspects was not met. To avoid the time-consuming training phase, the next option was Valence Aware Dictionary and Sentiment Reasoner (VADER). Because VADER has few limits, such as being unable to recognize slang or errors, the output received ranged from severe to extreme. TextBlob, on the other hand, offers the capacity to spellcheck, which is superior to VADER. Although it has its own limits, after evaluating three algorithms, TextBlob produced the most similar findings with positive and negative results. Although the technical implementations were just to support the user interface, the focus was on achieving a greater user experience.

6. Usability Testing Evaluation

It was critical to recruit people who already had an idea about the firm, either directly or indirectly. As a result, the individuals were screened depending on whether they have direct or indirect contact. Following the screening of the applications, a formal invitation was extended by Outlook email. In order to evaluate the noted observations, a data logger was used to generate the Usability Metrics to measure satisfaction, effectiveness and efficiency. Finally, charts were generated based on the information provided. Charts such as the amount of time it takes to complete a task, the level of performance, and the level of satisfaction and usefulness were produced.

6.1 Usability Metrics

Usability is defined as the degree to which a product can be used to achieve stated goals with effectiveness, efficiency, and satisfaction by specified users according to the ISO 9241-11 standard. Usability is not a single, one-dimensional feature, but rather a collection of elements [14]. In the Usability testing carried on nine participants, a total of twelve tasks were

performed and three charts were generated to conclude the results. After filling the data logger, a SUS score of 96.9% was obtained. Amongst the 5 snippets, following are the 2 snippets of the user interface of the ReviewDock.

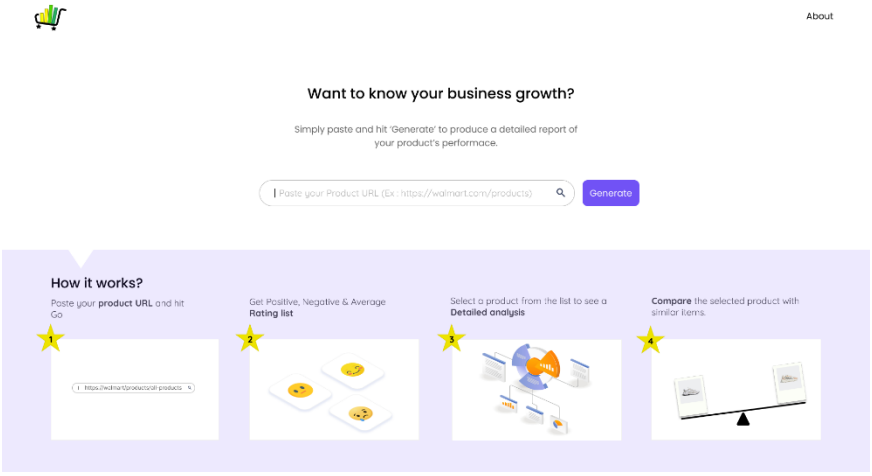

Fig. 2: Landing Page

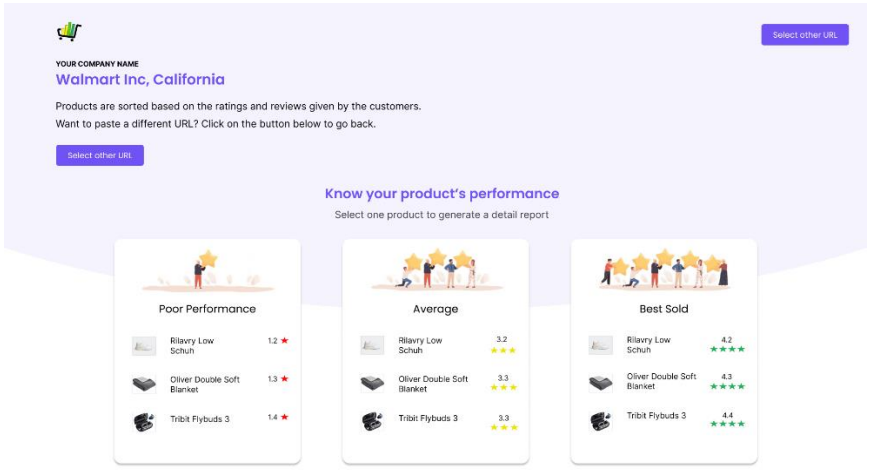

Fig. 3: Home Page – With Sentimental Analysis

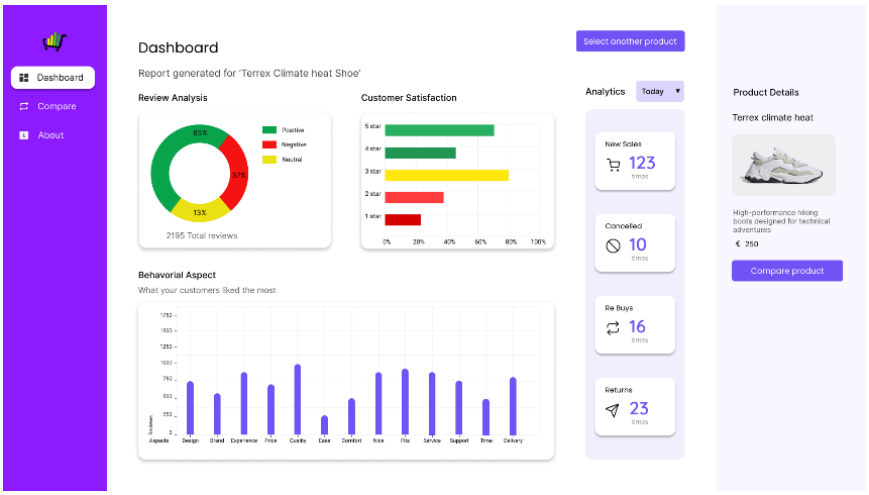

Fig. 4: Dashboard depicting the Review analysis, Behavioural aspects and Analytics

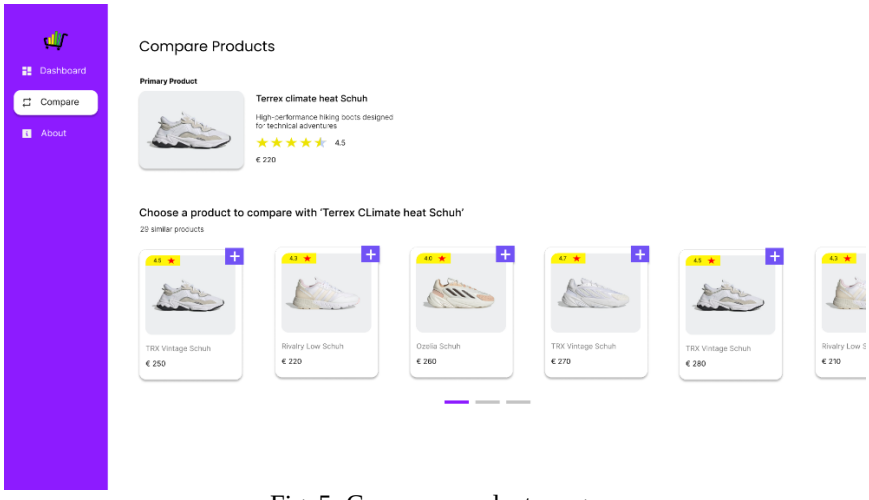

Fig. 5: Compare products page

7. Discussions

Due to the lack of user sentiments identification in present solutions, businesses are having problems identifying the actual emotion of their clients. As a result, they want tools to better monitor and improve their business growth by focusing on user pain points.

7.1 Findings

The results might suggest that it is okay to just display positive, negative and neutral results to the user for more customer satisfaction. However, based on the findings from the survey and results, a more plausible explanation is if the sentiments are displayed in a more segregated form, then businessperson will have a broader view of why a specific product is lacking than others.

In terms of customer satisfaction, it can only be met if the company strives to preserve its relationship with its customers; after all, it is customer centric. Following the approach of the proposed solution, all nine participants were satisfied based on

the minimum content provided on the screen. Almost all were pleased with the website's minimalist design, which displayed only what was necessary. To have a brief overview, below are the listed findings,

- It was discovered that even without training, satisfactory outcomes might be obtained.
- A smart system that includes a recommendation system amazed the audience.
- Less data (clean user interface) on the screen was favoured by the participants.
- The participants were confused by the use of slang. This is because they were expecting corporate terminology rather than a slang.
- Many of the participants only skimmed the website, not reading the context.
- They liked the colour design and said it was simple to distinguish between different products on the compare screen.
- Few participants noticed the lack of a back button.
- Because the majority of users skimmed the web page, the links to the preceding page that were present were mostly ignored. This was improved by using a button to emphasize the text.
- The most important factor was ease of usage.
- Participants were pleased because they were able to focus on the vital information rather than becoming lost in the user interface.
- Behavioural aspects, which displays the influence on product sales astonished the participants.

7.2 Limitations

Despite receiving a System Usability scale score of roughly 96%, there are a few noticeable drawbacks.

- The proposed system will be ideal for small businesses with few products. As a result, it can't be employed on a bigger scale.
- Because the targeted group was involved in business operations, the number of participants were limited to nine. As a result, more insights on revenue management and sales were not obtained, resulting in the development of a solution for smaller-scale industries.
- Apart from their correctness, the chosen algorithms take a longer time to compile, hence, the results printed on the implemented version are pre-compiled already in separate projects to avoid time constraints.
- Even though the interface offers various buttons for returning to the previous page, it lacks the back button. It perplexed the majority of users.

8. Conclusion and Future Scope

Even though the proposed solution is still in the implementation phase, it cannot be said that it is the best answer for the business; of course, many other aspects will have an impact on scalability on a wider scale. However, when the findings and limitations are considered, it is evident that proposed solution, when properly implemented, may be a good solution for improving and expanding business sales.

This research aims to improve customer satisfaction by upgrading the user interface, as well as to examine the impact of sentiments and emotions on businesses sales. A review scraper was used to collect data from Amazon.com's product reviews. The sentiments of the reviews were extracted using three algorithms, Naive Bayes, VADER (Valence Aware Dictionary and sentiment Reasoner) and TextBlob. These algorithms were evaluated using NRC emotion recognition to gain a better understanding of the emotion lurking beneath the hood. The performance of integrating Rule-based lexicon analysis with TextBlob and NRC was somewhat better than the other two. Although it was superior to the other two, not all of the reviews were properly classified. It was also limited by the fact that manual intervention was required. For a better user experience, a few people proposed using the conventional top navbar user interface, although some preferred the old-school dashboard. Author believes that the experience could be improved if more users participated and there was a bigger audience base than just nine. However, the results of the sentimental analysis may still be inaccurate, but such improvements would necessitate different pre-processing of the data, such as maintaining appropriate stop words or using various representation methods.

Usability research revealed that users prefer a user-friendly solution to operate rather than one that is more complex and full of jargon. The final testing yielded a score of 4.9, compared to a score of 2.49 in the first round of iteration. It may be concluded that customer satisfaction and sentiment segregation are essential elements to consider when designing and targeting e-commerce customers**.**

Future research could be performed to better understand the importance of these findings. More research is needed to determine the causes, effects and relationship between customers and the products or brands.

Considering the minimalist approach, some participants may like a minimalist design, while others may require more insights; hence, with user experience as the highest priority, and ability to personalize the options shown to users will be provided in the future.

- The use of artificial intelligence to develop smart chatbots that directly respond to unanswered reviews can help increase customer satisfaction significantly.
- Understanding the customer's purchasing history and recommending relevant products.
- Customer purchase history and the use of collaborative filtering can assist in predicting the price of a few products, which will boost the rate of interest.
- Back button implementation from a User Experience standpoint.

References

[1] N. Kahraman, M. A. Tunga, S. Ayvaz, and Y. B.Salman, "Understanding the purchase behaviour of turkish consumers in b2c e-commerce," International Journal of Intelligent Systems and Applications in Engineering, vol. 7, no. 1, pp. 52–59, 2019.

[2] U. Sivarajah, Z. Irani, S. Gupta, and K. Mahroof, "Role of big data and social media analytics for business to business sustainability: A participatory web context," Industrial Marketing Management, vol. 86, pp. 163–179, 2020.

[3] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhajj, "Emotion detection from text and speech: a survey," Social Network Analysis and Mining, vol. 8, no. 1, pp. 1–26, 2018.

[4] S. Dixit, A. J. Badgaiyan, and A. Khare, "An integrated model for predicting consumer's intention to write online reviews," Journal of Retailing and Consumer Services, vol. 46, pp. 112–120, 2019.

[5] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, "Mining product innovation ideas from online reviews," Information Processing & Management, vol. 58, no. 1, p. 102389, 2021.

[6] G. V. Research, "B2c e-commerce market size, share & trends analysis report by type (b2c retailers, classifieds), by application (clothing & footwear, media & entertainment), by region, and segment forecasts, 2021 - 2028." https://www.grandviewresearch.com/industry-analysis/b2c-e-commerce-market, 2021. Accessed on 09.01.2022.

[7] F. Pimenta, L. Lopes, F. Gonçalves, and P. Campos, "Designing positive behavior change experiences: a systematic review and sentiment analysis based on online user reviews of fitness and nutrition mobile applications," in 19th International Conference on Mobile and Ubiquitous Multimedia, pp. 152–161, 2020.

[8] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in Chinese based on sentiment lexicon and deep learning," IEEE Access, vol. 8, pp. 23522–23530, 2020.

[9] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[10] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on bilstm," Ieee Access, vol. 7, pp. 51522–51532, 2019.

[11] S. M. Mohammad and P. D. Turney, "Nrc emotion lexicon," National Research Council, Canada, vol. 2, 2013.

[12] A. Sharma, "Beginner tutorial:Recommender systems in python." https://www.datacamp.com/community/tutorials/recommender-systems-python, 2020. Accessed on 16.01.2022.

[13] S. Prabhakaran, "Cosine similarity – understanding the math and how it works (with python codes)." https://www.machinelearningplus.com/nlp/cosine-similarity/, 2018. Accessed on 19.01.2022.

[14] J. Mifsud, "Usability metrics – a guide to quantify the usability of any system." https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability, 2022. Accessed on 14.02.2022.