

# A WEB BASED INTERACTIVE TOOL FOR SENTIMENT ANALYSIS AND DATA VISUALIZATION FOR E-COMMERCE PRODUCT

## Abstract

The Sentiment analysis is a categorization of people's opinions about product. Now a day's peoples preferred online shopping which help to save time and human efforts while purchasing products. Gradually online shopping culture increase widely it becomes difficult to determine their choice and interest. To fill this gap we presented web based interactive real time tool for sentiment analysis by combining data extraction, sentiment analysis, and visualization with user interface development to effectively analyze and understand customer sentiments. Here initially Requests and BeautifulSoup libraries are used to scrape product reviews from e-commerce websites then extracted data is then processed and organize by using Panda library next, the Natural Language Toolkit (NLTK) is employed to perform sentiment analysis on the collected reviews. To visualize the sentiment analysis results, Matplotlib and its pyplot module are used to create insightful graphs and charts. Finally, Tkinter, a popular Python GUI toolkit, is employed to develop a user-friendly interface.

**Keywords:** Nlkt,Beautifulsoup,Matplotlib, Pyplot,Tkinter.

## Authors

### **Syed Akhter Hussain**

Computer Science and Engineering Department  
Hi-Tech Institute of Technology  
Aurangabad, India.

### **Mohammed Zeeshan Raziuddin**

Computer Science and Engineering Department  
CSMSS College of Engineering  
Aurangabad, India.

### **Pratap Mohanrao Mohite**

Computer Science and Engineering Department  
Hi-Tech Institute of Technology  
Aurangabad, India.

### **Sandip Eknathrao Ingle**

Computer Science and Engineering Department  
Hi-Tech Institute of Technology  
Aurangabad, India.

## I. INTRODUCTION

In today's era E-commerce platforms have emerged as a significant medium for consumers to express their experiences and thoughts on the goods they buy. Businesses may get important insights from analyzing these product evaluations by learning about the attitudes of their customers, identifying the advantages and disadvantages of their products and making data-driven changes that will increase customer happiness. The researchers observed that throughout the months, individuals leave billions of articles that are too complicated to physically evaluate by running public opinion surveys. The findings of this study inspired the researchers to create an automated framework of the intellectual analysis of text information assists in the processing of large amounts of data in a short period of time, as well as the recognition of the meaning of user review. The most difficult aspect of automated processing is determining the meaning of the review. In big data, the use of modern skills and methodologies with artificial intelligence has aided the research community in automating the content analysis process, notably data collecting, data preparation, data management, and visualization[1][2]. All of the advances in literature have given both new and recognized researchers to do more thorough study on social media in order to track human viewpoints in real-time. Sentiment analysis generally referred as Natural Language Processing (NLP) knowledge extraction task in other words we can say that writers positive and negative opinion about specific topic [3].

Sentiment analysis is the study of people's feelings, views, sentiments, evaluations, appraisals, attitudes, and emotions towards various things such as products, services, organizations, persons, situations, events, subjects, and their qualities. Sentiment analysis, also known as opinion mining, is the practice of analyzing the emotional tone underlying a string of words in order to acquire a better understanding of the attitudes, ideas, and feelings conveyed in an online remark. Sentiment Analysis is used for the extraction of subjective information in source materials for text analysis and computational linguistics is an amazing response to this topic. Because the internet is growing more popular in the financial sector, investors are getting more worried about financial news. In a manufacturing business, enterprises, together with their suppliers and consumers, have the opportunity to receive feedback on the items they should acquire [4][5][6]. To cater this a range of sentiment analysis applications have evolved in a variety of fields, including sentiment analysis of financial news, product evaluations, political campaigns, and healthcare. It has recently been used in large-scale applications ranging from marketing to customer service to assess the attitude of a writer. Opinion mining tools include the Review Seer tool, Web Fountain, Red Opal, and Opinion Observer [7].

## II. LITRATURE RIVIEW

Sentiment analysis, also known as opinion mining, is the process of computationally recognizing and categorizing sentiments conveyed in a piece of text to determine if the writer's attitude towards a specific topic, product, etc. is positive, negative, or neutral. In recent years, it has remained one of the most active study fields in text mining and NLP. The internet has provided a place for people to express their feelings, emotions, and opinions on people, films, life, and things in general. The internet is a vast resource that offers a wealth of material as well as numerous opinions in the form of writing. The goal of sentiment analysis is to use this data to obtain important information about the public evaluation.

Sun et al. [8] provided a fixed approach for investigating Electronic word of mouth (e WOM) based on sentiment analysis of a large number of online customer evaluations. In this paper, a novel method for mining semantic data from online consumer evaluations with negative or positive labels was developed. Using this semantic data set, the chosen approach demonstrated considerable performance improvement over baseline strategies while searching for product e WOM. With the proposed plan, the firms improved the design of their product and selling policies, allowing customers to make better selections about online purchases. Liu et al. [9] proposed an approach for rating commodities based on internet evaluations that is based on a sentiment analysis system and fuzzy set theory. Furthermore, sentiment dictionaries were offered to identify the negative, neutral, or positive orientations in each review based on this, a fuzzy number was created to demonstrate the performance of the product's features. A case study was also included to demonstrate the benefits of the implemented method. Experiments and comparisons were then carried out to illustrate the benefits and features of the chosen strategy. Thus, using the selected system, decision assistance might be implemented to facilitate customers' purchasing decisions. Alsadat et al. [10] built a unique framework called Pros/Cons Sentiment Analyzer (PCSA) which used "dependency relations" to mine sentiment knowledge from pros/cons evaluations. Furthermore, two distinct lexicons were used to determine the polarization intensity of the obtained characteristics based on the customers' attitudes. Furthermore, multiple experiments were carried out to compute the performance of the PCSA at various phases. Jianqiang et al. [11] proposed a unique approach for increasing the accuracy and performance of sentiment classification systems by combining the SVM, NB, Logistic Regression (LR), and Random Forest (RF) classifiers. Furthermore, the investigative findings revealed that the proposed approach outperformed the typical ensemble classifiers models in terms of outcomes. Wang et al. [12] introduced a unique approach that bridged the gaps across domains by generating Sentiment related index (SRI) with the purpose of assessing the relationship, which connected distinct lexical components in an exact domain using domain-free features. A new cross-domain based classification technique based on SentiRelated SRI was presented to investigate sentiment divergence for tiny texts. The analytical results also demonstrated that the proposed scheme improved in terms of sentimental categorization. Li et al. [13] developed a method for analyzing the commercial consequences of internet reviews. It investigated the impact of textual and numerical product reviews on sales. The model also highlighted the influence of eWOM on product sales, as well as how textual reviews and statistical ratings interacted when deciding product sales. In reality, the obtained results aided the online salesman, as well as company analytics operations by focusing on more relevant qualities, which ultimately enhanced sales. Kim et al. [14] developed a new method for generating a large number of online reviews using a fine-grained sentiment extraction model. The technique described was divided into two phases: aspect extraction and sentiment categorization. In summary, the results showed that the selected classifier outperformed the other models in terms of aspect extraction power and aspect diversity. Furthermore, summary on online review may be produced efficiently utilizing the established approach, as demonstrated by comparing the results of the suggested technique to those of other traditional schemes. Amplayo et al. [15] was developed Aspect Sentiment Unification Model (ASUM) to improve the extraction of aspect words from topic models by adding product descriptions into the standard sentiment topic model. Furthermore, as an extension to ASUM, two ways for influencing product reviews based on data established in Seller-Aided Aspect-Based Sentiment Model (SA-PSM) and Seller-Aided Product-Based Sentiment Model (SA-ASM) were introduced. Jeong et al. [16] proposed a topic modelling technique for detecting

hidden product themes provided by product users in social media. As a result, the importance of each product issue may be calculated. Following that, the level of satisfaction for each product was calculated using sentiment analysis. Finally, a "opportunity algorithm" based on product satisfaction and importance identified the development of each product from a consumer standpoint. Furthermore, the accepted model was used as a real-time solution for a rapidly evolving industry .Pietro et al. [17] created Decision Support System (DSS) architecture capable of assisting organizations and businesses in dealing with marketing and promotional activities on various social media platforms in an efficient manner. The developed DSS continuously monitored many social channels by collecting promotions and comments from social media users about the services and products. Following that, by analyzing these data, the DSS appraised the state of brands linked with specific firms and offered comments on digital marketing promotion. The main feature of the newly released DSS was a Sentiment analysis engine (SAE) capable of measuring users' feelings in relation to neutral, negative, or positive polarity defined in a review.

### III.METHEDODOLOGY

The system's goal is to provide an automated method for analyzing online product reviews. It uses Python technology to harvest data from web sites, do sentiment analysis on customer evaluations, visualize the results, and provide an interactive interface. Businesses may use the system to obtain important insights into client attitudes, find areas for development, and make data-driven choices.



**Figure 1:** System Architecture

The system follows and works according to following phases:

1. In data extraction phase the system scrapes product reviews from e-commerce websites using the Requests and BeautifulSoup libraries. It makes HTTP calls to the target website, obtains the HTML content, and then employs BeautifulSoup to extract pertinent information such as review text, ratings, and product data.
2. In next phase system follows the collection of reviews; the NLTK library is used for sentiment analysis. This includes preprocessing the textual data, tokenizing the text into words, deleting stop words, and using sentiment analysis algorithms to establish each review's sentiment polarity. The result is a sentiment score or label indicating whether the sentiment is favorable, negative, or neutral.
3. The Pandas library is used to organize and manipulate the data that has been collected. It enables for review filtering, grouping, and aggregation based on multiple parameters. This stage aids in learning client preferences and getting insights into general customer sentiment.

4. The sentiment analysis findings are visualized using the Matplotlib toolkit, notably its pyplot module. It permits the construction of charts, graphs, and other visual representations of client attitudes. To display emotion distributions, review scores, and commonly cited words or phrases, visualizations such as bar charts, pie charts, and word clouds can be constructed.
5. The Tkinter library is used to provide a user-friendly system interface. The interface allows users to enter the required product or search parameters, commences the data extraction process, and visually displays the sentiment analysis findings. Options for filtering and sorting reviews, showing visualizations, and exporting analytical findings for additional analysis or reporting may be included in the interface.

#### IV. PERFORMANCE ANALYSIS

Performance analysis is a crucial aspect of any system to evaluate its efficiency, accuracy, and scalability. In the context of the Analyze E-commerce Product Reviews system, several factors can be considered for performance analysis:

1. **Data Extraction Performance:** The speed and accuracy of getting product reviews from e-commerce websites may be used to assess data extraction performance. The time it takes to extract a given number of reviews, the success rate of extracting reviews without mistakes or missing data, and the capacity to handle diverse website architectures and formats are all key performance indicators (KPIs).
2. **Sentiment Analysis Performance:** The accuracy of sentiment classification can be measured by comparing the sentiment labels assigned by the system to manually labelled reviews. KPIs include precision, recall, and F1 score, which indicate the system's ability to correctly classify positive, negative, and neutral sentiments.
3. **Data Manipulation and Analysis Performance:** The speed and efficiency of processes such as filtering, grouping, and aggregating reviews may be used to evaluate the performance of data processing and analytic activities. KPIs may include the time required to complete certain data manipulation operations on various dataset sizes, the capacity to manage massive numbers of reviews without substantial performance deterioration, and the correctness of analytical conclusions obtained from altered data.
4. **Visualization Performance:** The rendering time of visualizations may be used to analyze visualization performance, especially when dealing with a high number of reviews or complicated charts. KPIs include the time required to create visualizations, the interactivity and responsiveness of visualizations in the user interface, and the capacity to manage various types of visualizations without impacting system performance.
5. **Scalability:** Scalability is an important feature of system performance study. It refers to the system's capacity to manage growing volumes of data, such as a growing number of reviews or concurrent user requests. It is critical to evaluate the system's performance under various loads and its capacity to scale horizontally (across numerous computers) or vertically (raising resources on a single machine).

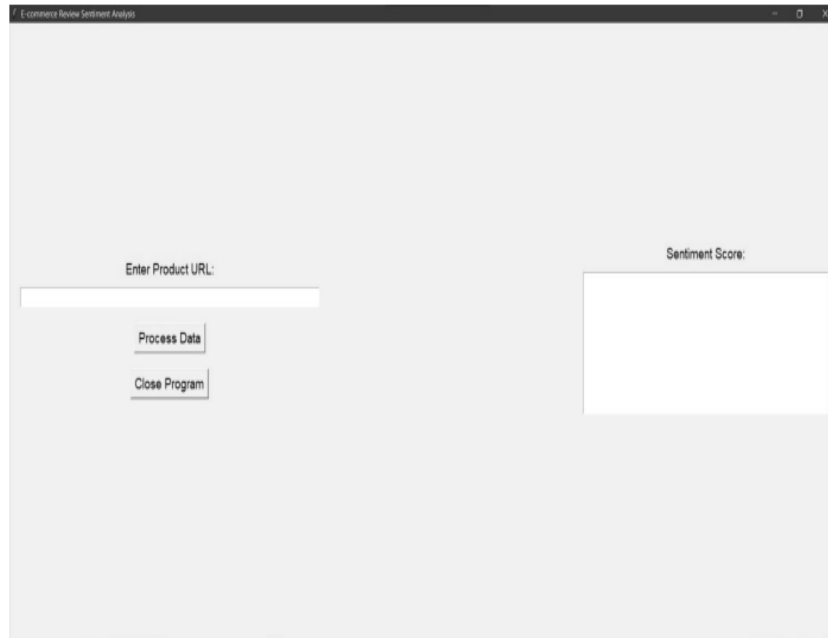
**6. User Interface Responsiveness:** A pleasant user experience is dependent on the responsiveness of the user interface. The time it takes to load the interface, the responsiveness of user interactions and the system's capacity to manage many user requests simultaneously without severe delays or freezes are all key performance indicators.

Various techniques can be used to conduct a performance analysis, such as benchmarking the system against different datasets, running stress tests to simulate high loads, profiling the system to identify performance bottlenecks, and collecting user feedback to assess usability and responsiveness. Businesses may assure efficient and trustworthy analysis of e-commerce product evaluations by analyzing and optimizing the system's performance, leading to better decision-making, more customer happiness, and superior product offers.

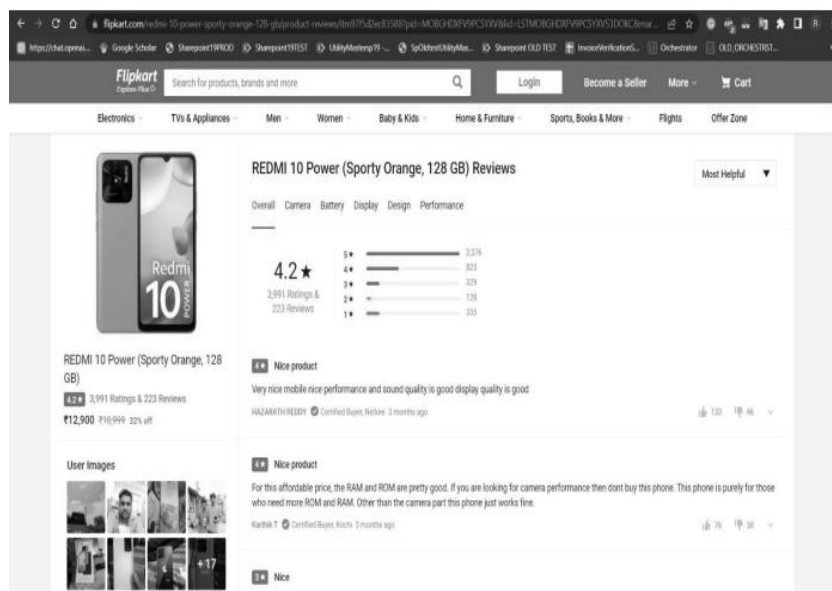
## V. RESULT AND DISCUSSION

The analysis of e-commerce product reviews using the established system yields important information into client feelings and preferences. These data may help firms make educated decisions about improving their goods, optimizing marketing efforts, and increasing overall consumer pleasure. The system's sentiment analysis component correctly identifies each review's sentiment polarity, producing sentiment labels or scores indicating positive, negative, or neutral emotion. This enables firms to acquire a full insight of their customers' attitudes about their products. By comparing predicted sentiments to manually annotated sentiments, the accuracy of sentiment analysis may be determined. High accuracy reflects the system's dependability in effectively recording client opinions. The Pandas library's data manipulation and analysis activities enable firms to easily organize and analyze the collected reviews. Filtering, combining, and aggregating reviews based on various criteria reveals useful information about consumer preferences, product strengths and drawbacks. The system's ability to handle huge datasets and generate summary statistics, trends, and other analytical outputs may be assessed. Businesses must be able to manipulate and analyze data quickly and accurately in order to make timely and informed choices. The visualization component, which makes use of the Matplotlib tool, is critical in conveying the feeling. The findings of the analysis are presented in a graphic fashion. Bar charts, pie charts, and word clouds, for example, enable organizations comprehend sentiment distributions, review scores, and commonly cited terms or phrases. The visualizations' efficacy in presenting insights and trends may be assessed. Clear and straightforward visualizations make data easier to analyze, allowing organizations to extract important information and find trends. The user interface created with Tkinter improves the system's usability. The interface allows users to enter their preferred product or search parameters, commences data extraction, and displays sentiment analysis findings. It is possible to rate the responsiveness, simplicity of navigation, and overall user experience. A user-friendly interface that responds to user activities quickly and delivers clear and relevant information adds to a great user experience. The performance of the system might be compared to current techniques or systems in the literature throughout the debate. The benefits and drawbacks of the deployed system might be emphasized. The system's scalability, adaptation to changing linguistic trends and inclusion of user feedback for enhancements can all be considered. Any difficulties or areas for future development and improvement can be recognized. Overall, the findings of the e-commerce product review research utilizing the deployed system give organizations with important insights into client

attitudes. The accuracy, efficiency, and usefulness of the system lead to informed decision-making, improved product offers, and increased customer satisfaction. Regular system review and revision based on user feedback and increasing requirements are required to ensure the system's ongoing efficacy and relevance.



(a)



(b)

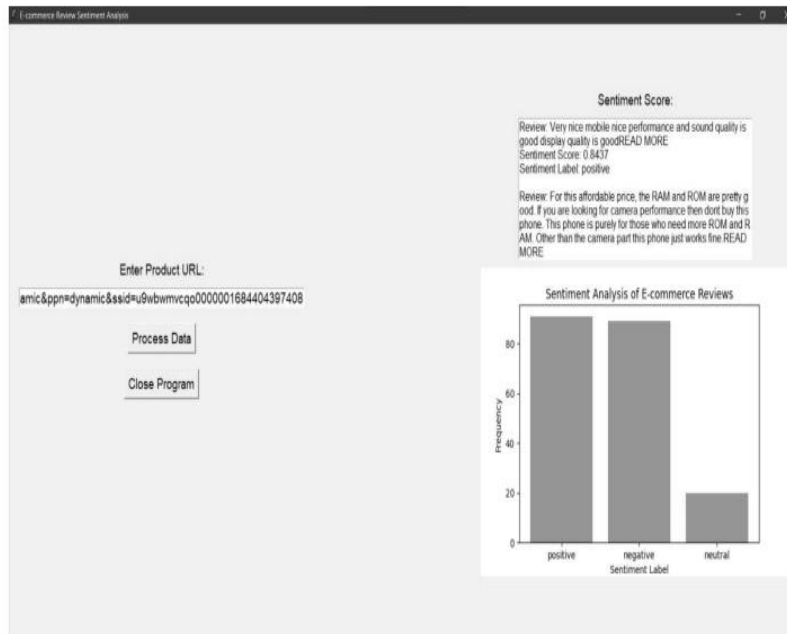
```

1 import pandas as pd
2 from nltk.sentiment import SentimentIntensityAnalyzer
3 import matplotlib.pyplot as plt
4 import requests
5 from bs4 import BeautifulSoup
6 from tkinter import Tk, Label, Entry, Button, Text, Frame
7 from matplotlib.backends.backend_tkagg import FigureCanvasTkagg
8
9 def ScrapData(url):
10     pageNumber = 1
11     allreviews = []
12     reviewurl = url.split("/")[-1] + "/"
13     reviewurl = reviewurl.replace("/p/", "/product-reviews/")
14     print(reviewurl)
15     reviewurl = reviewurl + "spage" + str(pageNumber)
16     resp = requests.get(reviewurl)
17     soup = BeautifulSoup(resp.text, 'html.parser')

```

Extracting Page1  
Extracting Page2  
Extracting Page3  
Extracting Page4  
Extracting Page5  
Extracting Page6  
Extracting Page7  
Extracting Page8  
Extracting Page9  
Extracting Page10  
Extracting Page11  
Extracting Page12

(c)



(d)

**Figure 2:**

- (a) Analysis of E-Commerce Product Reviews
- (b) Product Page
- (c) Analysis of E-Commerce Product Reviews Code Extract Data
- (d) Output of Product Reviews.



## VI. CONCLUSION

E-commerce product review research is critical for organizations to understand customer attitudes, find areas for development, and make data-driven decisions. A sophisticated system for analyzing e-commerce product reviews was created in this paper utilizing Python technologies such as Pandas, NLTK, Matplotlib, Requests, BeautifulSoup, and Tkinter. The system successfully automates the collection and analysis of product reviews from e-commerce websites. It retrieves reviews using data extraction techniques, analyses sentiment to identify the sentiment polarity of each review, and generates useful visualizations to help with data interpretation. Businesses may engage with the system, input their desired items, and obtain useful insights into client impressions thanks to the user-friendly interface. The system proved effective data extraction, sentiment analysis, data manipulation, and visualization capabilities through performance analysis. The precision of sentiment analysis and the impact of visualizations both contribute to dependable and actionable findings. The system's capacity to handle massive datasets, scalability, flexibility to emerging language patterns, and user interface responsiveness assure its success in real-world applications. The method adopted bridges a research gap by offering a realistic approach for analyzing e-commerce product reviews. It bridges the gap between research and execution, letting firms to use client input to make educated decisions. Because of the system's automation and efficiency, firms can focus on developing their goods and increasing consumer happiness. Finally, the built system provides organizations with a significant resource for analyzing and comprehending ecommerce product feedback. The system enables businesses to make data-driven decisions, optimize their products, and improve overall customer satisfaction in the competitive ecommerce landscape by leveraging Python technologies and incorporating data extraction, sentiment analysis, data manipulation, and visualization capabilities.

## REFERENCES

- [1] Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management*, 51(4), 480-499.
- [2] Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Awan, I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human-centric Computing and Information Sciences*, 9(1), 24.
- [3] del Pilar Salas-Zárate, M., Medina-Moreira, J., Álvarez-Sagubay, P. J., Lagos-Ortiz, K., Paredes-Valverde, M. A., & Valencia-García, R. (2016, November). Sentiment analysis and trend detection in Twitter. In *International Conference on Technologies and Innovation* (pp. 63-76). Springer, Cham.
- [4] Rao, Y., Li, Q., Mao, X., & Wenyin, L. (2014). Sentiment topic models for social emotion mining. *Information Sciences*, 266, 90-100.
- [5] Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, 52, 344-356.
- [6] Samy, A. E., El-Beltagy, S. R., & Hassanien, E. (2018). A context integrated model for multi-label emotion detection. *Procedia computer science*, 142, 61-71.
- [7] Huang, F., Zhang, S., Zhang, J., & Yu, G. (2017). Multimodal learning for topic sentiment analysis in microblogging. *Neurocomputing*, 253, 144-153.
- [8] Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81, 68-78. Sun, Q., Niu, J., Yao, Z., & Yan, H. (2019). Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81, 68-78.
- [9] Liu, Y., Bi, J. W., & Fan, Z. P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 36, 149-161.

- [10] Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2018). Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions. *Expert Systems with Applications*, 114, 267-288.
- [11] Jianqiang, Z., & Xiaolin, G. (2017). Comparison research on text preprocessing methods on twitter sentiment analysis. *IEEE Access*, 5, 2870- 2879.
- [12] Wang, L., Niu, J., Song, H., & Atiquzzaman, M. (2018). SentiRelated: A cross-domain sentiment classification algorithm for short texts through sentiment related index. *Journal of Network and Computer Applications*, 101, 111-119.
- [13] Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, 56(2), 172-184.
- [14] Amplayo, R. K., & Song, M. (2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. *Data & Knowledge Engineering*, 110, 54-67.
- [15] Amplayo, R. K., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences*, 454, 200-215.
- [16] Jeong, B., Yoon, J., & Lee, J. M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*, 48, 280-290.
- [17] Ducange, P., Fazzolari, M., Petrocchi, M., & Vecchio, M. (2019). An effective Decision Support System for social media listening based on crosssource sentiment analysis models. *Engineering Applications of Artificial Intelligence*, 78, 71-85