All Sciences Proceedings
http://as-proceeding.com/

2nd International Conference on Engineering, Natural and Social Sciences

April 4-6, 2023 : Konya, Turkey

https://www.icensos.com/

Sentiment Analysis on Amazon Product Reviews using Text Analysis and Natural Language Processing Methods

Abdul Haseeb*, Rabia Taseen [2], Marryam Sani [3], and Qamas Gul Khan[4]

[1]*Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan*
[2] *Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan*
[3]*Department of Computer Sciences, COMSATS University Islamabad, Wah Cantt, Pakistan*
[4]*Department of Electrical Engineering, University of Engineering and Technology Taxila, Pakistan*

**Abdul.Haseeb@students.uettaxila.edu.pk*

***Abstract* –** In this study, customer reviews of goods are categorized and analyzed using a web-based tool, preserving the analyst the time and effort of reading through millions of reviews. Natural Language Processing (NLP) and text analysis (TA) techniques are used in the Sentiment Analysis on Product Reviews technology that has been developed. The completed gadget is made up of the following 5 parts. Customer Satisfaction Score (CSS), Lexicon-Based Sentiment (LBS), Text Analytics, Amazon Products, and Application Programming Interface (API) Scraping of New Reviews. The text analytics application cleans up the textual data and extracts sentiment. It is possible to determine the customer satisfaction number by averaging the emotional quantities. With the help of Python-based sentiment analysis, we can investigate and find assessments that analyzers want to examine.

Keywords – Text Analysis (TA), Natural Language Processing (NLP), Lexicon-Based Sentiment (LBS), Customer Satisfaction Score (CSS), and Application Programming Interface (API)

I. INTRODUCTION

Now a days, we can easily access products all over the world, and a huge number of reviews in various categories are available on various platforms daily, making it difficult for users to stay up to date on those reviews that are relevant to their product interests. In recent few years, the data set can be generated as much as the processing of humans. It will not be possible to read all of the data created in one day if everyone in the world stops working [1]. Customers' ratings and comments should be used to make the ultimate decision about a restaurant's or brand's product quality. As a result, managers must study the clients' hidden emotional reactions to meet their needs as well as provide expertise in terms of flavor, environment, cost, and customer care [2]. Our perception of purchasing any product and the user decision can be directly affected by the responses given by those who already experienced it [3]. As a result, we want a well-organized method for handling huge data. We are addressing this problem by utilizing sentiment analysis (SA) of product reviews. As we can see, users have multiple reviews against various products, but they are unable to study them all. To make it easier for customers to create a sentiment graph of reviews before purchasing a product, we use SA and feature selection algorithms. The article examines product review management software, which aids companies in compiling feedback and rankings. Additionally, it will improve how customers view goods like laptops made by HP, Dell, or Samsung, as well as men's and women's apparel, footwear, and services like shopping. This also allows buyers to access a website where they can read product reviews with a simple click. We create a review-based system by extracting internet product reviews and using a scraping technology called API to gather

data in real time. AMAZON data on reviews is extracted. Then we use sentiment algorithms to create reviews in the form of a chart, either a graph chart or a percentage chart. As a result, customers will benefit from reading all the reviews since they will be able to conveniently type in the name of the product and receive all of the reviews at once.

During SA, developers may take various text data and further identify and categorize the views. One issue is polarity classification; to address this, researchers might use review level and sentence level categorization to get successful results. The Support Vector Machine (SVM) and Naive-Bayes (NB) algorithms, which are commonly supervised classification approaches, are used to check the positive and negative evaluations of a product. The accuracy of favorable and negative reviews is 0.87 and 0.80, respectively, when using the NB. The precisions after applying the SVM technique are 0.89 and 0.87, respectively. These findings are compared to previous research [4][5]. With the aid of the POS algorithm, a data set of 5513 reviews is evaluated for adverb extraction. They apply the polarity to positive, negative, and neutral reviews by using WordNet. They got 0.52 precision for neutral, 0.37 precision for negative, and 0.22 precision for positive [6].

Text-based SA is an algorithm. Gathering data from several platforms and polarizing it is a difficult undertaking. The researchers provide data filtration, a strategy that calls all positive and negative evaluations into a single file. They delete all unneeded characters to increase the performance of a model since the file contains too many excess numbers or symbols [7]. Customers and business owners seek a platform where they can interact with one another face-to-face. They will be able to determine what the user's requirements are because of this. Now that the new technique has been implemented, Natural Language Processing (NLP) has entered a new phase known as the Lexicon approach. For movie reviews, the Text Blob and VADER tools are utilized [8]. There is a lot of material on the website that is in an unorganized style. In addition to the TF-IDF, the CNN-LSTM (Long Short-Term Memory) was given (term frequency-inverse document frequency). The CNN-LSTM has five layers: a weighted embedding layer, a convolution layer, a max-pooling layer, an LSTM layer, and dense layers. When compared to the literature, the suggested study yields better results [9]. In [10], SA with convolutional neural network (CNN) and Bidirectional gated recurrent unit is used in the hybrid model. The Neutro-VADER approach employs a set of 228 words. The model may be used to focus on individual attributes, and then score measures are used to rank the items [11].

Multiple techniques are used in [3] for vectorization of data, such as BOW (bag-of-Words, Glove), and then different ML techniques for classification (e.g., Logistic Regression (LR), NB, and the experiments are taken by the multi-class classification algorithm and finalized on the binary classification. By installing dictionaries, the recurrent neural network (RNN)-LSTM model is utilized to train the model for extracting characteristics such as emotions. They can then use continuous BOW to create the vector representation [12][13]. The SA may also be used to analyze restaurant reviews. In [1] the three methodologies of lemmatization, word correction, and keyboard are applied in the research. Their study focuses on data processing overfitting and performance concerns. In [14], to check the restaurant's reviews, a hybrid technique is applied. Because reviews allow business owners to meet their customers' expectations. Following the SA, Particle Swarm Optimization (PSO) assisted in the weight optimization, and the SVM approach was employed for classification. In [15], SA implements the convolutional attention-LSTM (CA-LSTM) approach. First, using fuzzy mathematics, map the link between positive reviews and sentiment characteristics. The CA-LSTM was then used in the second step. 83.3% of cases, the prediction is correct.

II. Behavioural Requirements

It provides a summary of the system's capabilities, outlines the behavioral requirements, hardware, software, and user interfaces, including GUI interfaces for application software, as well as the functional and non-functional requirements.

In addition to the system design, we have an API that will crawl websites for products and extract customer reviews using an API written in PHP. We then use sentiment analysis, which combines machine learning techniques and feature selection, to store the results in a database (MYSQL) and create web applications Fig. 1.

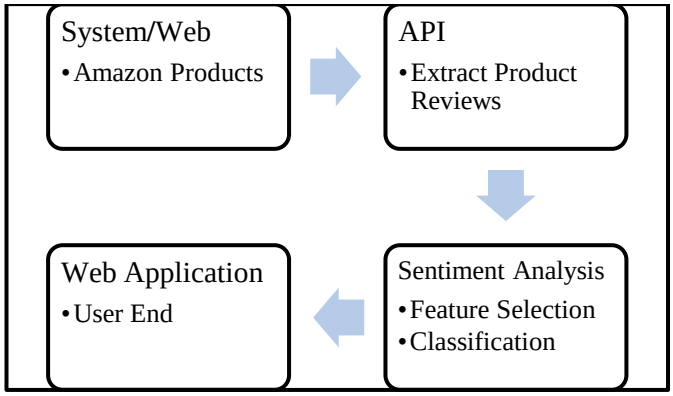

Fig. 1 System Design Diagram

Using NLP methods and approaches, this system analyses textual input from customers. With the aid of NLP and text mining methods, the product will analyse user comments and reviews. This will enable us to better comprehend how a consumer feels or perceives a particular merchandise, right? Powerful visualization tools will be used to show the analytics from real-time data, which will enable us to forecast both the product success score and the customer happiness score. It is a brand-new creation that is not a component of any existing structure. User interfaces, system diagrams, use cases, and other information are provided in the literature that follows.

A. *Different Cases Parameters and their Descriptions*

- *Registration or Sign Up*

Table 1. Registration Parameters

Actors	User
Description	User can register with the system with to perform certain tasks.
Pre-Condition	Users have an email.
Steps	Enter Email Id, Name, or other related details. (When he enters all details then user will be registered).
Post-Condition	Successfully register with the system.

- *Log In*

Table 2. Log in Parameters

Actors	User.
Description	User can login with the system with to perform certain tasks.
Pre-Condition	User have registered via email.
Steps	Enter email Id and password.
Post-Condition	Successfully login into the system.

- *Search Products*

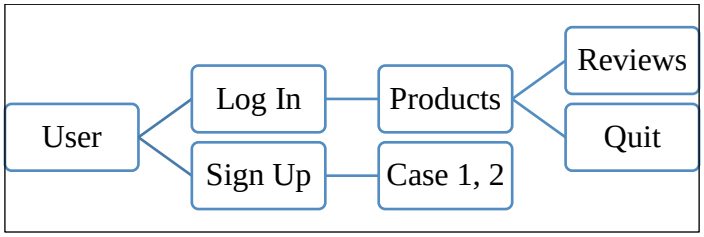

Fig. 2 Flow of Search Products

Table 3. Searching of Products

Actors	User.
Description	After login, the user can search the products in which he/she is interested. After searching user could be logout.
Pre-Condition	Users have login in web portal.
Steps	Login into the system. Type the name of product. See reviews sentiments. Logout.
Post-Condition	Search product successfully.

B. *Flow Diagram for the Web Development*

The interaction between the Web Application and the User will be shown in fig. 3. It depicts the flow of information between activities such as product, product reviews, feelings, login, and other application settings. All the reviews come from the System API on the backend.

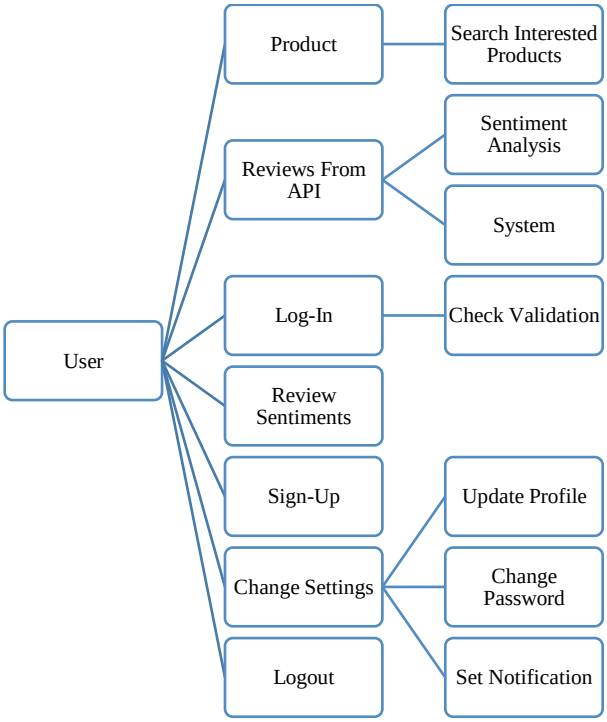

Fig. 3 Case Diagram of Web Application

III. METHODOLOGY

A methodology is a thorough document that outlines a product's or process's features. The design specification might, for example, contain required dimensions, environmental factors, ergonomic aspects, aesthetic elements, and maintenance requirements. It might also provide examples of how the design should be implemented, assisting others in their efforts. The following composite viewpoint diagram is following in our research.

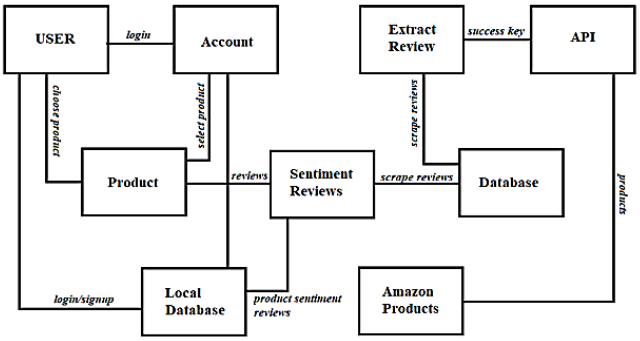

Fig. 4 Composite Viewpoint

It represents a user interacting with the application, and then the application sending user interests to API, which API uses to extract reviews of products related to user's interests, and then the database will send back a response to the user containing the product that was selected by users, and on the backend, API are crawling the web pages and extracting reviews, performing sentiment analysis on reviews and classification, and then displaying the sentiment analysis.

A. *Sequence Diagram of Login*

This is a representation of the product reviews application's sequence diagram, Fig 5, which demonstrates how objects communicate with one another. In this sequence diagram, we can see the login process: the user hits the login button, the application replies by navigating to the login page, the user enters their user's name and password, the program checks the database for valid usernames and passwords, and finally returns a response.

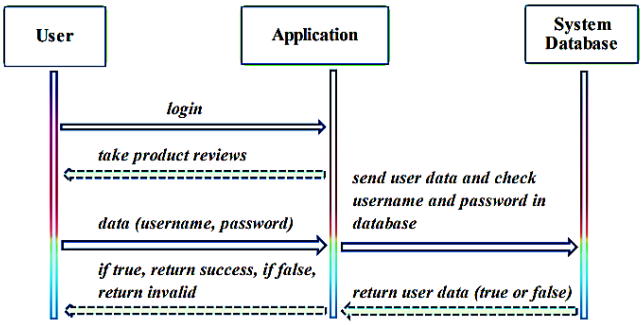

Fig. 5 Sequence diagram of Login

B. *Sequence Diagram of Product*

The process of seeing a product is depicted in this sequence diagram. The user clicks on the product button, and the program retrieves the product description from the database and displays it to the users.

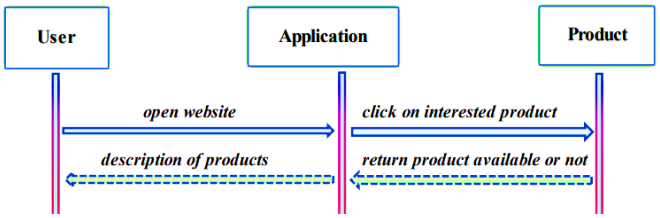

Fig. 6 Sequence diagram of Product

C. *Sequence Diagram of Reviews*

We can see how a user may see product reviews in this sequence diagram. Users will click on the product feed to access sentiment reviews from a database, which will then be shown in the activity.

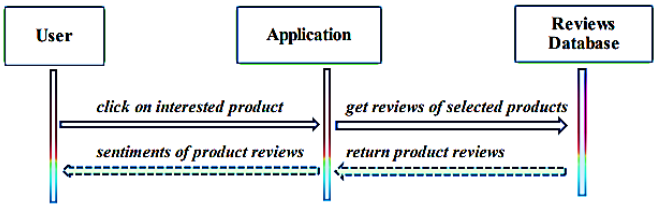

Fig. 7 Sequence diagram of Reviews

IV. RESULTS AND DISCSSUION

The Vader sentiment analysis is 82.9% accurate, compared to TextBlob's accuracy of 81.2%. Which unmistakably demonstrates that TextBlob's accuracy is higher than Vader's sentiment analysis. Therefore, we employed the Vader analysis to examine the product reviews before further website implementation. In this, we display the outcomes of several Amazon items. The website allows customers to simply write reviews and check

ratings. The following images also display the website's outcomes.

The Vader sentiment analysis is 82.9% accurate, compared to TextBlob's accuracy of 81.2%. Which unmistakably demonstrates that TextBlob's accuracy is higher than Vader's sentiment analysis. Therefore, we employed the Vader analysis to examine the product reviews before further website implementation. In this, we display the outcomes of several Amazon items. The website allows customers to simply write reviews and check ratings. The following images also display the website's outcomes.

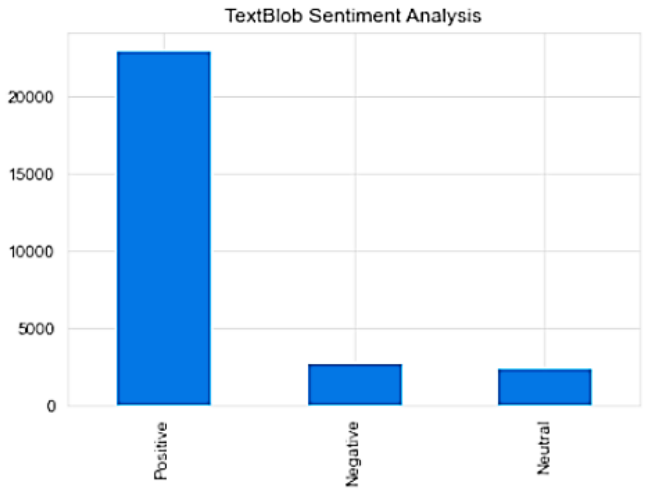

Fig. 8 Results of TextBlob

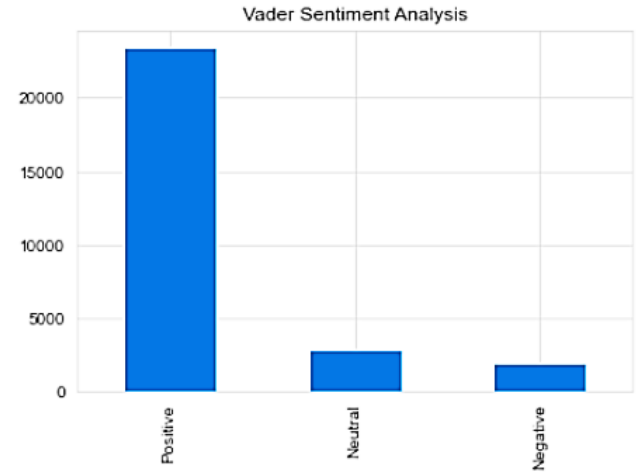

Fig. 9 Results of VADER

Fig. 10 Website Home Page

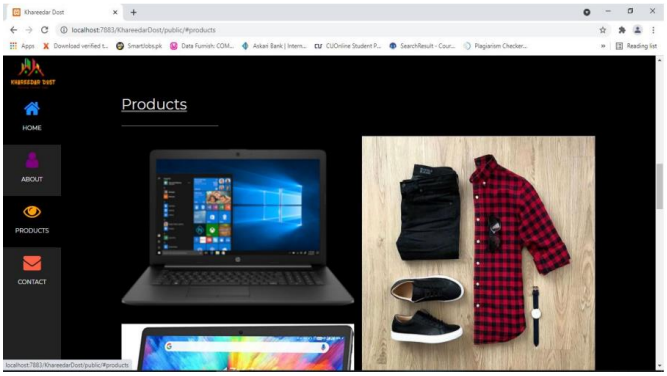

Fig. 11 Amazons Products

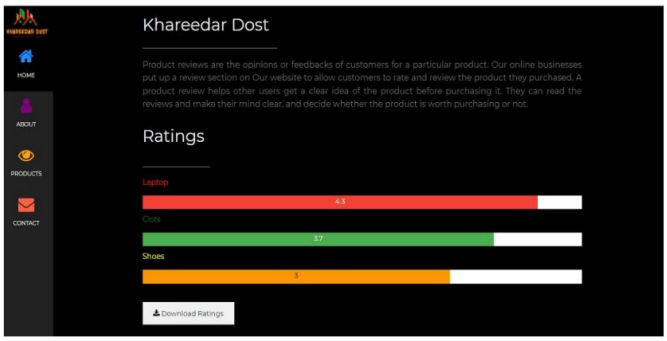

Fig. 12 Rating of Some Products

V. Conclusion and Future Work

The analyst and the NLP programmer were able to analyse 1000 textual reviews and scores to learn what customers liked and didn't like about the product without having to sort through millions of reviews. Many analysts and purchasers may analyse and purchase the greatest goods with the aid of this technique.

Even while all our tests indicate that the system is accurate, we cannot claim that it is mistake-free since there is still room for error and more work must be done to make the system even more accurate. This technology can also be of utilized along with Big Data as well. For improved performance, the system may be moved to Hadoop clusters and Amazon NLP engines.

REFERENCES

[1] Ö. Aktaş, B. Coşkuner, and İ. Soner, "Turkish Sentiment Analysis Using Machine Learning Methods: Application on Online Food Order Site Reviews," *arXiv Prepr. arXiv2201.03848*, vol. 1, no. 1, pp. 1–10, 2022, [Online]. Available:https://arxiv.org/abs/2201.03848

[2] N. Hossain, M. R. Bhuiyan, Z. N. Tumpa, and S. A. Hossain, "Sentiment Analysis of Restaurant Reviews using Combined CNN-LSTM," *2020 11th Int. Conf. Comput. Commun. Netw. Technol.ICCCNT 2020*, 2020, doi: 10.1109/ICCCNT49239.2020.9225328.

[3] A. S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews," *Int. J. Comput. Sci. Inf. Technol.*, vol. 13, no. 3, pp. 15–30, 2021, doi: 10.5121/ijcsit.2021.13302.

[4] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0015-2.

[5] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015, doi:10.1016/j.procs.2015.07.523.

[6] S. Haider, M. Tanvir Afzal, M. Asif, H. Maurer, A. Ahmad, and A. Abuarqoub, "Impact analysis of adverbs for sentiment classification on Twitter product reviews," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 4, pp. 1–15, 2021, doi: 10.1002/cpe.4956.

[7] M. Aarti and A. Patil, "Sentiment Analysis for Product Reviews," *Int. J. Adv. Res. Comput. Sci.*, vol. 5, no. 5, pp. 202–204, 2014, doi: 10.21917/ijsc.2019.0266.

[8] V. Bonta, N. Kumaresh, and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," *Asian J. Comput. Sci. Technol.*, vol. 8, no. S2, pp. 1–6, 2019, doi: 10.51983/ajcst-2019.8.s2.2037.

[9] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 23, pp. 1–12, 2021, doi: 10.1002/cpe.5909.

[10] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce ProductReviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[11] I. Awajan, M. Mohamad, and A. Al-Quran, "Sentiment Analysis Technique and Neutrosophic Set Theory for Mining and Ranking Big Data from Online Reviews," *IEEE Access*, vol. 9, pp. 47338–47353, 2021, doi: 10.1109/ACCESS.2021.3067844.

[12] N. B. Van Le and J. H. Huh, "Applying sentiment product reviews and visualization for bi systems in vietnamese e-commerce website: Focusing on vietnamese context," *Electron.*, vol. 10, no. 20, 2021, doi: 10.3390/electronics10202481.

[13] A. Chinnalagu and A. K. Durairaj, "Context-based sentiment analysis on customer reviews using machine learning linear models," *PeerJ Comput. Sci.*, vol. 7, 2021, doi: 10.7717/PEERJ-CS.813.

[14] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[15] Y. Su and Y. Shen, "A Deep Learning-Based Sentiment Classification Model for Real Online Consumption," *Front. Psychol.*, vol. 13, no. April, pp. 1–9, 2022, doi: 10.3389/fpsyg.2022.886982.

[16] Y. Qin, X. Wang, and Zeshui X., "Ranking Tourist Attractions through Online Reviews: A Novel Method with Intuitionistic and Hesitant Fuzzy Information Based on Sentiment Analysis," *International Journal of Fuzzy Systems*, Springer, vol. 24, no. 2, pp. 755-777, 2022.

[17] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product reviews: A review," *2017 International conference on Inventive Communication and Computational Technologies (ICICCT)*, IEEE, 2017, pp. 298-301.

[18] R. S. Jagdale, V. S. Shirat, and S. N. Deshmukh, "Sentiment Analysis on Product Using Machine Learning Techniques," *Cognitive Informatics and Soft Computing: Proceeding of CISC 2017*, Springer, 2019, pp. 639-647.

[19] A. N. Jebaseeli and E. Kirubakaran, "A Survey on Sentiment Analysis of (Products) Reviews," *International Journal of Computer Applications*, Citeseer, vol. 47, no. 11, 2012.

[20] V. Vyas and V. Uma, "Approaches to Sentiment Analysis on Product Reviews," *Sentiment Analysis and Knowledge Discovery in Contemporary Business*, IGI Global, pp. 15-30, 2019.

[21] S. Smetanin and M. Komarov, "Sentiment Analysis of Product Reviews in Russian using Convolutional Neural Network," *2019 IEEE 21st conference on Business Informatics (CBI)*, IEEE, vol. 1, pp. 482-486, 2019.

[22] M. Ali Fauzi, "Word2Vec model for sentiment analysis of product reviews in Indonesian language," *International Journal of Electrical and Computer Engineering*, IAES institute of Advanced Engineering and Science, vol. 9, no. 1, pp. 525, 2019.

[23] M. Z. Asghar, A. Khan, S. Ahmed, and F. M. Kundi, "A Review of Feature Extraction in Sentiment Analysis," *Journal of Basic and Applied Scientific Research*, vol. 4, no. 3, pp. 181-186, 2014.

[24] N. Sultana, P. Kumar, and M. R. Patra et. al., "Sentiment Analysis for Product Review," *ICTACT J Soft Comput*, vol. 9, no. 3, 2019.

[25] A. Cernian, V. Sgarciu, B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," *2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, IEEE, 2015, pp. WE-15.

[26] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, Elsevier, vol. 3, no. 2, pp. 143-157, 2009.

[27] Y. Basani, H. V. Sibuea, S. I. P. Sianipar, and J. P. Samosir, "Application of Sentiment Analysis on Product Review E-Commerce," *Journal of Physics: Conference Series*, IOP Publishing, vol. 1175, no. 1, pp. 012103, 2019.

[28] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon Review Classification and Sentiment Analysis," *International Journal of Computer Science and Information Technologies*, Citeseer, vol. 6, no. 6, pp. 5107-5110, 2015.
[29] R. Bose, R. K. Dey, S. Roy, and D. Sarddar, "Sentiment Analysis on Online Product Reviews," *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018*, Springer, 2020, pp. 559-569.
[30] S. Zad, M. Heidari, J. H. Jones, and O. Uzuner, "A Survey on Concept-Level Sentiment Analysis Techniques of Textual Data," *2021 IEEE World AI IoT Congress (AIToT)*, IEEE, 2021, pp. 0285-0291.