

# Future prospects on E-Commerce Product Experience based on Natural Language Processing and Machine Learning algorithms

<sup>1</sup>Ambika Talawar, <sup>2</sup>Akshatha M

<sup>1</sup>Student, <sup>2</sup>Assistant Professor  
Computer Science & Engineering,  
Maharaja Institute of Technology  
Mysore, India

**Abstract-** In the era of competition among the E-commerce brands, the customer feedback and product experience plays a vital role. An emotion intensity scanner is a natural language processing (NLP) tool used to assess the level of emotion or level of feeling within an item of text. The sentiment can be favorable, adverse, neutral, or it may be mixed. The analyzer provides an integer number or rating to the feelings, reflecting its intensity. A favorable sentiment, for instance, might receive an increased positive score, while an adverse feeling may have a higher negative rating. Such analyzers can be employed for a variety of purposes, including social media monitoring, market research, client feedback evaluation, and general opinion analytics. The intensity scores can be displayed in any number of methods, contingent upon the method of execution or methodology. E-commerce experience with a product relates to the complete customer experience or satisfaction when engaging with goods or services on a digital retail marketplace. It encompasses numerous aspects of the consumer background, spanning the moment a customer first comes across an item to their final choice of purchase and post-purchase acquaintances. A positive customer experience is a vital part of any profitable online shopping operation. It encompasses each phase of the consumer's experience; at the moment they find a product to the last post-purchase interactions. In addition, customized recommendations and incentive schemes adapt to a person's preferences, promoting relationships that endure. By continually enhancing and honing the product experience, e-commerce companies may make an eternal mark on the consumers they serve, building confidence and receiving an edge over their competitors in the constantly changing online marketplace. A favorable e-commerce product experience can lead to increased consumer satisfaction, sales, and brand loyalty. E-commerce enterprises should regularly monitor and evaluate customer comments, user behavior, and sales data to find areas for improvement and improve the overall product experience. Businesses may stand out in the competitive online market by providing a flawless and pleasurable e-commerce product experience. We have developed a Sentiment analysis we application using flask using VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment Intensity analyzer to, where both customer and buyer can get the benefits of the business needs. This program also displays the mood of each prior customer's remarks or feedback when they purchase a new product. The program has been tested on several E-Commerce platforms' real-time platforms. The score table provides the sentiment such as Positive, Negative, Neutral and Compound.

**Index Terms – E-Commerce, Sentiment Analysis, Natural Language Processing, Product Experience**

## I. INTRODUCTION

The swift growth of the World Wide Web along with shifts in consumer buying habits led to the development of eBay, Amazon, and a number of numerous platforms for e-commerce. In the meantime, a growing variety of goods and services can be found on the website via online comments. The reviews not only demonstrate the efficacy and standard of internet-based goods or services, but also display the buying experiences in a precise and comprehensive manner. As it turned out, review sites are seen as a trustworthy source for knowledge by customers as well as sellers. Reviews, especially of fresh and unproven products, provide customers with essential recommendations for selecting a product, and this is vital to lower shopping hazards. Furthermore, assessments assist merchants in producing products that meet consumer expectations by offering an appreciation of consumer attitudes such as motivations, satisfaction, and so on. To be sure, it is vital for corporate success to gain insights into product experience and swiftly comprehend consumers' practical needs from assessments. How to identify and gather consumer opinions and mindsets from immense informal remarks has grown into a highly contentious problem in this environment. In answer to this require, the present research presents an integrated sentiment analysis method with the objective of digging customers' perceptions of operational goods and monitoring their requirements. The study's potential utilization values are mainly represented as following. In a particular manner, it lets consumers make better choices about the product's quality and use. Additionally, it allows merchants to successfully understand consumers' practical desires, enabling for future product enhancements and marketing strategy optimization.

The Natural Language Processing (NLP) is as opinion mining, is used to locate, extract, and analyze the subjective information in texts. The Internet has permeated every facet of public's everyday lives since the information era began. People now share their ideas and leave comments on the network platform more and more frequently. On the Internet, there has been a recent increase in useful reviews of famous people, significant occasions, and commercial goods. Online evaluations convey a wealth of sentimental data, such as ideas, attitudes, and emotional inclinations like delight, annoyance, grief, criticism, and admiration, allowing potential

users to learn what the general public thinks about particular individuals, occasions, or things. In order to effectively investigate public opinion, formulate policies, and make commercial decisions, it is necessary to recognize and analyze the subjective data that comes from social network platforms. This data often represents personal thoughts and opinions.

Figure 1 Importance of E-Commerce



Figure 1 shows the reviews of online goods are crucial for marketplace rankings as well as a fantastic form of social proof. Customer reviews offer a thorough overview in addition to descriptions and images. Reviews foster trust and increase the willingness of potential buyers to try a novel brand. Product reviews are a potent source of social evidence that help people believe in and trust your brand. According to Gorgias' data on over 10,000 ecommerce sites, adding reviews to your product pages can increase revenue by an estimated 1.5%. This is made possible by the fact that customer reviews serve as testimonies for skeptic buyers and are perhaps more persuasive than any sales or marketing you can develop on your own.

## II. RELATED WORKS

Many authors have contributed the research on E-Commerce product experience field. *Digvijay mali et al.* described many techniques for sentiment analysis and presents a productive framework. Additionally, it emphasizes the superiority of the Naive Bayes classifier over alternative classification techniques. The methods that have been found to be the most effective by survey respondents are presented in this paper along with a list of other methods that are available for performing sentiment analysis on review data. Instead of relying solely on supervised or unsupervised learning algorithms improved by heuristic methods like ANN to increase accuracy, a semi-supervised strategy is suggested in which focus is placed on significant opinion words discovered using WordNet. NLTK with the Naive Bayes probabilistic model will be used for sentence-level sentiment analysis. Graphical and statistical representations of the results will be used. *Thilageswari et al.* explained about an examination of sentiment examination techniques, procedural obstacles and trends to offer researchers with an overview and evaluation of the already available information. The subsequent part of the research will go via the extraction of features and categorization technique for sentiment analysis of consumer feedback so as give readers an in-depth understanding of the approaches. The comprehension acquired from sentiment analysis trials serves as the foundation to establish the next steps. *P. Sudhakaran et al.* has contributed towards popular machine learning algorithms as Multiclass Support Machine algorithm for classifying sentiment polarity with detailed description. The proposed method is implemented in Python software and experimented on online product-reviews data taken from Amazon.com. Sentence level and opinion level classification is obtained with promised outcomes. From the results it is noted that the proposed method outperforms than the existing method such as Naïve Bayes and Random Forest algorithms.

The main objective of this paper is to design and implement a ML algorithm for analyzing the online product reviews data. To do that a multiclass SVM classifier is designed, implemented and experimented on product reviews data taken on amazon.com. *Jenny Yow Bee Yin et al.* extended the research on how to figure out people on Twitter feel regarding the online retailer and the topic at hand. The data indicates that Twitter users had favorable sentiments toward Lazada and Shopee, with favorable emotions outweighing negative emotions for both businesses. For data extraction from Twitter using the Twitter API, data cleansing, and sentiment analysis, RapidMiner was selected as the tool. Positive, Negative, or Neutral sentiments are the three categories into which the sentiment is divided. Shopee and Lazada will be able to learn more about the audience's opinions of the experience and recommendations thanks to these results. The findings advance our knowledge of e-commerce business services.

Social media is used to build relationships and to exchange all types of emotions, grievances, and opinions. Social media platforms like Twitter have developed as crucial marketing sensing channels. The behaviors and attitudes of customers toward the business or brand are substantially influenced by those internet reviews. *Siti Fidyanti Nurfadila Hadju et al.* the purpose of this study is to

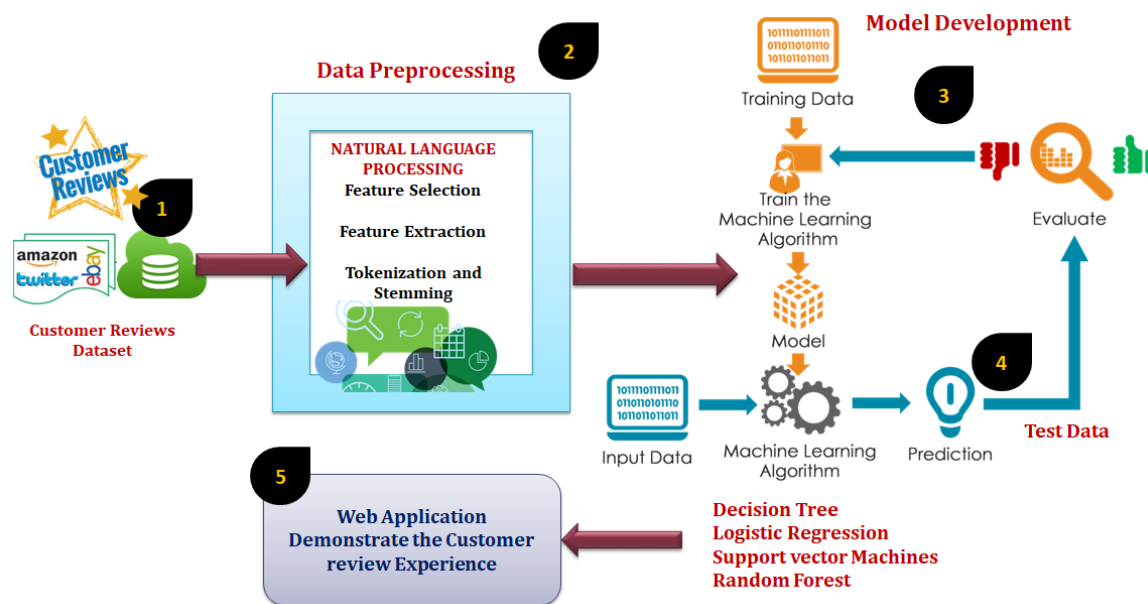
offer e-commerce management a new angle by showing how user-reviewed product reviews can be criteria for selecting management tactics. The CRISP-DM approach can be used to gather existing reviews, analyse them, and turn them into data for product segmentation. The authors examine multiple machine learning algorithms to find the most effective method for sentiment analysis. Data were gathered from reviews of products in the following three categories: fashion, grocery, and phones. Then, using comparisons between four algorithms (Decisions Tree, Random Forest, Gradient Boost, and Super Vector Machine), this data is positioned. With a 95.875% accuracy rate, the SVM method can't be argued against as the best one. Once the prediction model has been developed, it must be connected to the dashboard software for E-commerce management and associated merchants to see it.

**III. PROPOSED METHODOLOGY**

Recent years have seen a rapid growth in e-commerce. Online product reviews have increased as a result of the growth in online shopping. Because the client's perception of the products is affected by the recommendations or concerns of other buyers, inbuilt comments in customer reviews has an important effect on the consumer's choice to purchase them. The project's objective is to establish a platform for the effective creation of online products utilizing machine learning and opinion mining methods based on customer reviews. Based on the aim of the project, we have designed the objectives as:

- Initially reviews data will be collected from the online E-Commerce Platform
- Text processing is applied using Natural Language Processing Techniques
- Split the dataset into Train and test.
- Develop the Machine Learning Models for the classification of the reviews.
- Performance of each model is evaluated separately and comparison study can be given.
- Finally, we propose an effective approach of following customer needs and evaluating online merchant experiences, greatly helping subsequent product innovation and marketing strategy improvement.

Figure 2: Proposed System of Ecommerce System



The figure 2 shows the four significant components that constitute the recommended technique's method are reviewing text gathering, preprocessing of text, mood feature extraction, and sentiment evaluation. Figure 1 illustrates the extensive investigation environment.

**Step 1: Data Gathering Phase**

To start with, a substantial amount of internet-based review texts are collected through online sources such as Flipkart, Amazon, eBay, or any other Electronic commerce Portal. In addition, given input is organized into collections of word arrays through preprocessing of text to facilitate the next statistical calculations.

**Step 2: Data reviewing Phase**

Review writings, as an example of unorganized data, aren't allowed to be handled and assessed immediately. We need to apply Text Processing techniques to balance and structure the data. Text Processing methods Feature Selection, Feature Extraction, Tokenization and Stemming are applied.

There are now two basic graphical representation methods: one-hot representation and distributed representation. An M-bit of state flip flop is used to encode M states in one-hot representation, where each bit of state is stored in a distinct flip flop, with only one

bit true at any given time. One-hot representation encodes a single word into a vector expressing a specific feature according to the criteria mentioned above.

Nonetheless, it presupposes that every phrase in the sentence has a distinct personality, which fails to convey the sequence and connection that exists between words. In addition, if the amount of an eigenvalue is incredibly excellent, it will build gigantic redundant sparse matrices, culminating in the dimension curse. Word2vec, a shallow neural network-based word embedding algorithm. Word2vec, a single of the distributed representations, allows an unsupervised model to be learned from an enormous corpus of manuscripts that incorporates words into a low-dimensional vector space.

**Step 3: Development of the model.**

Initially we divide the dataset into train and test groups. Create machine learning models for the classification of the reviews.

**Step 4: The performance evaluation.**

When each model's effectiveness is assessed independently and a comparison study is possible.

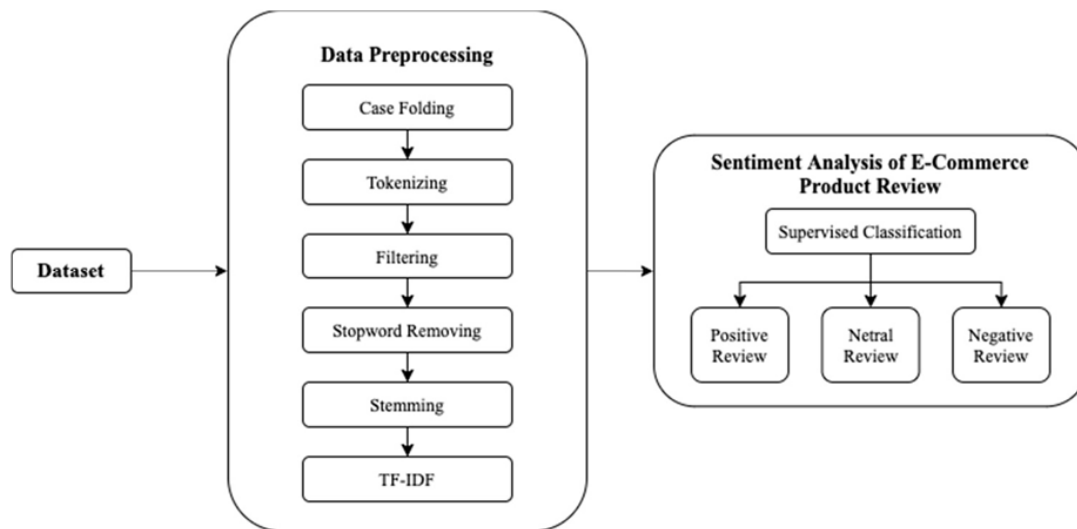
**Step 5: Online Product Experience System Demonstration.**

Finally, we offer a practical method for tracking client needs and online product experiences, which is crucial for future product development and marketing plan optimization.

**IV. SYSTEM IMPLEMENTATION**

The method's viability and validity have been confirmed by research into customers' experiences reading books they purchased from Amazon.com online. The outcomes show that the method correctly identifies the emotional trends of reviews and captures factors influencing reading experiences from reviews. The study offers a practical method for tracking client wants and mining online product experience, which is crucial for future product development and marketing plan optimization.

Figure 2: Architectural Diagram



**Pre-processing of Data**

Figure 3 shows the pre-processing strategies used on the real-time E-Commerce dataset for this study. Text pre-processing is a critical component in NLP for improving data textual quality. As a component of the review pre-processing, for example, "Good" and "GrEat" are converted to "good" and "great," instead of mixed uppercase and lowercase letters. Furthermore, any punctuation and stop words that are frequently employed but have little impact on meaning have been eliminated, including "-", "/", "the," and "a." In addition, the reviews underwent a tokenization procedure, which involves breaking down a text into a collection of terms known as "tokens." For performing word separations, the tokenizing methodology depends upon the space character since each token can frequently be distinguished from another token or separated from it by a space character. After that, the lemmatization approach was employed to revert each token to its lexicon or base form.

**Feature Extraction**

Feature extraction is a process in machine learning and data analysis in which raw data is turned into a set of relevant features that can be used to represent and characterize the underlying patterns in the data. The purpose of feature extraction is to lower the dimensionality of the data while maintaining as much important information as possible, making it easier and more efficient for machine learning algorithms to deal with the data. Feature extraction is critical in improving the performance of machine learning models in several domains, including computer vision, natural language processing, and signal processing.

**Models for Classification**

Data is classified using a technique known as classification. It is used in sentiment analysis to categorize data into binary classification (for example, "positive" and "negative") and ternary classification (for example, "positive," "negative," and "neutral"),

and then the sentiment analysis process is completed on the basis of those classifications. In the sentiment analysis of customer evaluations, lexicon-based and machine learning techniques are commonly used.

The effectiveness of categorization algorithms can be quantified using Accuracy, F-Score, Cross-entropy, Recall, and Precision. These metrics are useful for assessing the effectiveness of supervised machine learning algorithms since they are based on one element of a matrix known as the confusion matrix or contingency table. A confusion matrix is widely used to assess the performance of an algorithm. When evaluating the performance of machine learning models, several metrics and parameters are used to assess how well the model performs on a given task. The choice of evaluation metrics depends on the type of problem (classification, regression, clustering, etc.) and the specific requirements of the application. Here are some commonly used performance evaluation parameters:

**Precision:** Precision is calculated by dividing the number of true positive predictions by the total number of positive predictions (true positive + false positive). It measures how well the model avoids false positives.

$$Precision = \frac{TTP}{TTP+FP}$$

**Recall:** Recall is defined as the number of correct positive predictions divided by the total number of correct positive cases (true positive + false negative). It measures the model's ability to detect positive cases.

$$Recall = \frac{TTP}{TTP+FN}$$

**Accuracy:** Accuracy is one of the most basic and extensively used criteria for classification jobs. It calculates the fraction of accurately predicted instances to the total number of instances in the dataset. However, it may not be appropriate for datasets with imbalances.

$$Accuracy = \frac{TP+TFTP+FP+TN+FN}{Total}$$

**F-score:** The harmonic mean of precision and recall is used to get the F1 score. It delivers a good combination of precision and recall, which is essential when dealing with skewed datasets.

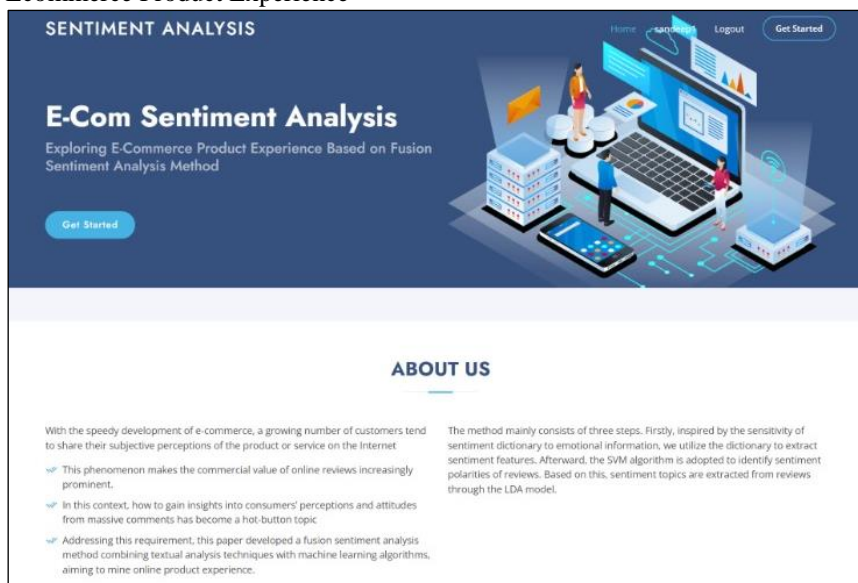
$$F-score = 2 \left( \frac{Precision * Recall}{Precision + Recall} \right)$$

**Cross-entropy:** Cross-entropy or log loss is used to assess the performance of classification models. The log loss function generates a probability between 0 and 1.

## V. EXPERIMENTAL RESULTS

Customer information can be utilized by e-commerce sites to offer tailored product recommendations based on browsing history, purchase behavior, and preferences. This tailoring of product suggestions can enhance the customer's shopping experience by showcasing relevant items of interest. A seamless and straightforward checkout process is crucial for minimizing cart abandonment. It should have clear calls-to-action, a variety of payment options, and a secure and reliable payment gateway. Reducing the number of steps and forms required during checkout can also help streamline the process. Offering responsive and helpful customer support channels, such as live chat, email, or phone, is essential for addressing customer queries, concerns, or issues promptly. A dedicated customer support team that is knowledgeable about the products and provides timely assistance can significantly improve the overall experience. Ecommerce businesses should actively seek customer feedback and suggestions to identify areas of improvement in the product experience. Regularly analyzing customer data and implementing changes based on user feedback can help enhance the overall ecommerce experience.

Figure 3: Home page of Ecommerce Product Experience



Home Page is depicted in Figure 3 and 4 depict the details on Signing and registration details. First, using the dictionary to extract sentiment features, we are motivated by the sentiment dictionary's sensitivity to emotional information. The SVM technique is then

used to determine the sentiment polarization of the reviews. Using the LDA model, sentiment topics are derived from reviews based on this.

Figure 4: Registration of users

Sample result 1:

**The Sentiment of**

"If you're looking for a phenomenal dining experience, look no further than [restaurantname]. their menu is a culinary masterpiece, offering an incredible selection of dishes that cater to diverse tastes and preferences. the atmosphere is warm and inviting, and the staff is friendly and attentive, making every visit a true pleasure. I've dined at [restaurantname] several times and have never been disappointed – it's truly a must-visit destination for any food lover."

**is 98.0% positive !**

**Score table**

SENTIMENT METRIC	SCORE
Positive	0.394
Neutral	0.606
Negative	0.0
Compound	0.98

Sample result 2:

**The Sentiment of**

"When I visited the hotel, it was very dirty, I am unhappy with service from you. very bad and negative staff"

**is 4.0% positive !**

**Score table**

SENTIMENT METRIC	SCORE
Positive	0.0
Neutral	0.317
Negative	0.683
Compound	0.04

## VI. CONCLUSION

Product experience can be enhanced by utilizing sentiment analysis to understand and analyze customer feedback and opinions. Sentiment analysis is the procedure used to identify the sentiment or emotional tone indicated in a text, such as customer reviews, comments on social media, or survey results. By applying sentiment analysis to product-related data, industries can gain valuable insights into consumer happiness, pinpoint areas for development, and make data-driven decisions to better the product experience. Here's how sentiment analysis can contribute to product experience. Sentiment analysis can be useful to customer reviews, comments, and feedback to determine the overall sentiment expressed towards a product. By automatically categorizing reviews as positive, negative, or neutral, businesses can gauge customer satisfaction levels and identify specific pain points or areas of delight. This information can be used to prioritize product improvements or highlight positive aspects to potential customers. In conclusion, sentiment analysis, particularly utilizing the VADER (Valence Aware Dictionary for sentiment Reasoning) sentiment analysis tool, can significantly enhance the ecommerce product experience. By analyzing customer feedback, reviews, and other textual data, businesses can gain insights into customer sentiment and emotions towards their products. Here's how VADER sentiment analysis can contribute to improving the ecommerce product experience. By classifying customer sentiment as positive, negative, or neutral, VADER sentiment analysis offers a quantitative assessment of consumer sentiment. This assists companies in determining the general degree of client happiness and pinpointing areas for product improvement.

## REFERENCES:

- [1] S. M. Mohammad, "Sentiment analysis: Detecting valence, emotions, and other effectual states from text," in *Emotion Measurement*. Sawston, U.K.: Woodhead, 2016, pp. 201\_237.
- [2] M. Taboada, "Sentiment analysis: An overview from linguistics," *Annu. Rev. Linguistics*, vol. 2, no. 1, pp. 325\_347, Jan. 2016.
- [3] F. Emmert-Streib, O. Yli-Harja, and M. Dehmer, "Utilizing social media data for psychoanalysis to study human personality," *Frontiers Psychol.*, vol. 10, p. 2596, Nov. 2019.
- [4] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, "Twitter as a tool for health research: A systematic review," *Amer. J. Public Health*, vol. 107, no. 1, pp. e1\_e8, Jan. 2017.
- [5] R. E. Wilson, S. D. Gosling, and L. T. Graham, "A review of Facebook research in the social sciences," *Perspect. Psychol. Sci.*, vol. 7, no. 3, pp. 203\_220, May 2012.
- [6] M. S. Shaharudin, Y. Fernando, C. J. C. Jabbour, R. Sroufe, and M. F. A. Jasmi, "Past, present, and future low carbon supply chain management: A content review using social network analysis," *J. CleanerProd.*, vol. 218, pp. 629\_643, May 2019.
- [7] B. Qi and A. M. Costin, "Investigation of the influence of Twitter user habits on sentiment of their opinions towards transportation services," in *Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience*. Reston, VA, USA: American Society of Civil Engineers, 2019, pp. 314\_321.
- [8] R. P. S. Kaurav, K. G. Suresh, and S. Narula, "New education policy: Qualitative (contents) analysis and Twitter mining (sentiment analysis)," *J. Content, Community Commun.*, vol. 12, no. 1, pp. 4\_13, 2020.
- [9] R. K. Dong, "Emotion and international business: Theorising fear of failure in the internationalisation," *Frontiers Psychol.*, vol. 13, Mar. 2022, Art. no. 850816, doi: [10.3389/fpsyg.2022.850816](https://doi.org/10.3389/fpsyg.2022.850816).
- [10] U. Warriar, P. Singh, and C. W. Jien, "Factors that lead Amazon.com to a successful online shopping platform," *Int. J. Tour. Hosp. Asia Pasific*, vol. 4, no. 1, pp. 7\_17, 2021.
- [11] C. A. Haryani, A. N. Hidayanto, N. F. A. Budi, and Herkules, "Sentiment analysis of online auction service quality on Twitter data: A case of E-bay," in *Proc. 6th Int. Conf. Cyber IT Service Manage. (CITSM)*, Aug. 2018.
- [12] H. Luo, "A review of research on identification of false reviews in Ecommerce," *J. Manage. Humanity Res.*, vol. 3, pp. 9\_15, Jul. 2020.
- [13] K. Thirumaran, H. Jang, Z. Pourabedin, and J. Wood, "The role of social media in the luxury tourism business: A research review and trajectory assessment," *Sustainability*, vol. 13, no. 3, pp. 12\_16, 2021.
- [14] S. Tao and H.-S. Kim, "Online customer reviews: Insights from the coffee shops industry and the moderating effect of business types," *Tourism Rev.*, vol. 77, no. 5, pp. 1349\_1364, Aug. 2022.
- [15] B. von Helversen, K. Abramczuk, W. Kopeć, and R. Nielek, "Influence of consumer reviews on online purchasing decisions in older and younger adults," *Decis. Support Syst.*, vol. 113, pp. 1\_10, Sep. 2018.
- [16] C. Campbell and J. R. Farrell, "More than meets the eye: The functional components underlying influencer marketing," *Bus. Horizons*, vol. 63, no. 4, pp. 469\_479, Jul. 2020.
- [17] J. M. Luo, H. Q. Vu, G. Li, and R. Law, "Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews," *Int. J. Hospitality Manage.*, vol. 98, Sep. 2021, Art. no. 103032.
- [18] J. R. Saura, D. Palacios-Marqués, and D. Ribeiro-Soriano, "Exploring the boundaries of open innovation: Evidence from social media mining," *Technovation*, Jan. 2022, Art. no. 102447.
- [19] P. Dash, J. Mishra, and S. Dara, "Sentiment analysis on social network data and its marketing strategies: A review," *ECS Trans.*, vol. 107, no. 1, pp. 7417\_7425, Apr. 2022.
- [20] S. Gupta and R. Sandhane, "Use of sentiment analysis in social media campaign design and analysis," *Cardiometry*, vol. 22, pp. 351\_363, May 2022.