

Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews

Mohammad Abu Kausar

Department of Information Systems, University of Nizwa, Oman
kausar@unizwa.edu.om (corresponding author)

Sallam Osman Fageeri

Department of Information Systems, University of Nizwa, Oman
sallam@unizwa.edu.om

Arockiasamy Soosaimanickam

Department of Information Systems, University of Nizwa, Oman
arockiasamy@unizwa.edu.om

Received: 15 March 2023 | Revised: 14 April 2023 | Accepted: 23 April 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.5854>

ABSTRACT

Online retailers and merchants increasingly request feedback from their clients on the products they purchase. This has led to a significant increase in the number of product reviews posted online, as more people are making purchases online. The opinions expressed in these customer reviews have a significant impact on other customers' purchase decisions, as they are influenced by other customers' recommendations or complaints. This study used Amazon, a well-known and widely used e-commerce platform, to examine sentiment categorization using several machine learning techniques while analyzing an Amazon Reviews dataset. At first, the reviews were transformed into vector representations using the Bag-of-Words approach. Word cloud was used to illustrate the text data in terms of the frequency they appear in the review. Subsequently, the machine learning methods decision trees and logistic regression were used. The two models used in this study achieved high levels of accuracy in analyzing the dataset. Specifically, the Decision Tree model outperformed the Logistic Regression one, achieving an impressive accuracy of 99% compared to the 94% of the latter.

Keywords-sentiment analysis; Amazon customer reviews; dataset; feature extraction; text classification; machine learning

I. INTRODUCTION

Over the past decades online marketplaces have grown in popularity, leading to the trend of asking customers for feedback on purchased products to improve the overall customer experience. Every day, millions of reviews about different goods, services, and locations are created online. As a result, the Internet has become the key resource for knowledge and opinions regarding products and services. However, the vast amount of product reviews and the multiple perspectives to consider may further complicate the decision-making process, leading to confusion and uncertainty, as it is difficult for customers to choose wisely when there are several viewpoints on the same product and different ratings. Therefore, e-commerce businesses should evaluate this content to get feedback on their products and help customers decide.

Sentiment analysis and classification is a field of computer science that attempts to answer this problem by sifting through given natural language texts to extract attitudes and views, using numerous approaches, such as biometrics, computational linguistics, natural language processing, and text analysis. Due to their accuracy and effectiveness, Machine Learning (ML) approaches have lately grown in favor of the semantic and review analysis fields. This study considered Amazon as one of the most popular e-commerce vendors. Potential customers may read hundreds of reviews posted by previous customers regarding the products they're interested in [1-2]. These reviews provide valuable information on a product, helping shoppers understand most of its features. Customers benefit from this, but it also helps retailers or producers to get a greater knowledge of customers' needs.

This study used supervised algorithms to address the problem of sentiment classification for online reviews and

determine the overall significance of customer evaluations and characterize them as positive or negative. The data used were reviews of Amazon Titan Men Watches, collected from amazon.in. The study employed supervised algorithms to address the issue of sentiment classification for online reviews. The goal was to determine the overall importance of customer evaluations by categorizing them as either positive or negative.

II. RELATED WORKS

Many studies used data collected from numerous sources, such as Twitter [3-4], product reviews, consumer feedback, etc. Customer engagement programs help businesses to strengthen their emotional relationships with clients from all over the world. Customer involvement is also influenced by product reviews [5]. The study of product reviews allows a company to learn how customers feel about its products. According to [6], online consumer reviews have an important influence in molding customers' online shopping decisions. The term "review" refers to the process of determining whether a product or service is suitable for a certain purpose. Every day, a massive quantity of fresh information is added to the Web [7], hence a method was proposed based on simultaneous web crawling using mobile agents.

Sentiment analysis is used by businesses to increase their competitiveness in the market, as it enables them to comprehend the opinions and experiences of their clients on their products and services [8]. Sentiment analysis is also used to capture market sentiments to design a better forecasting model for the stock market [9]. Modern organizations use sentiment analysis to improve their word-of-mouth marketing approach. Competitive companies use text-mining tools to better understand client experiences and extract valuable information from social networking platforms, newspapers, and other sources. In [10], k-Nearest Neighbors (k-NN), Decision Trees (DT), and Artificial Neural Network (ANN) models were used to identify client behavior in the banking industry. As a result, a mix of sentiment analysis, text mining, and other methods is critical in ensuring that businesses understand and capitalize on online consumer evaluations. Sentiment extraction is an excellent method of understanding customer assumptions online [11], as the data collected from internet platforms and product review sites allow businesses to improve their marketing methods. Product reviews also influence client purchase decisions. Competitive organizations, such as Amazon, use such information to make decisions [12], and leading merchants in the United States rely on Internet product reviews to improve their marketing efforts and company procedures [13]. For example, if a given product receives a large number of unfavorable reviews for its cost or quality, the company examines the problem and tries to address it as soon as possible.

According to [6], most of the product evaluations on the Internet are less polar and more positively balanced. Also, the distribution of product reviews for similar items differs by platform [14]. The diversity of reviews is driven by several factors, including the rating system, the business strategy of the online platform, and the review frequency [6]. In [15], a Web Crawling Model based on Java Applets was presented. Furthermore, it is important to mention that firms use sentiment

analysis to improve company operations and client retention, as analyzing product evaluations allows a company to understand client experiences [16]. A client can leave a review to indicate whether he is happy or unhappy with a particular product or service. However, the majority of product reviews do not represent the level of client happiness. In [16], a study was conducted based on reviews on the Internet to classify consumer happiness, according to auditory and language characteristics, into four classes: extremely positive, positive, neutral, or highly unfavorable. The findings of this study on the impact of review length on online sentiments were consistent with [17]. Customers prefer to rely on comprehensive and insightful product reviews before making a final purchase [18]. As a result, companies must discover the true degree of customer satisfaction with their products and services in order to make an informed decision based on online product reviews.

Authors in [19] evaluated the way internet reviews influence Amazon book sales. According to this study, customers view online reviews as a reliable source of information and prefer more accessible and detailed reviews. According to this study, internet reviews have a major impact on user experiences and product costs. These findings were in agreement with those of [20] on online reviews and feelings. The study also looked at the valence of internet reviews. In [19], the influence of online reviews on purchases was found to be contradictory. Some studies concluded that the valence of online reviews has a major influence on sales, while others found that it has no impact. The impact is also affected by variables such as product categories and qualitative text qualities. In [21], an analysis was conducted on 142.8 million Amazon user reviews, focusing on determining the usefulness and unhelpfulness of each review by examining the summary headline, the product remark, and helpfulness information, filtering out blank and non-English product evaluations to improve the accuracy of the results, and choosing only those with the most votes. The study concluded that an investigation of online product evaluations on Amazon plays an important part in today's e-commerce. Helpful reviews provide thorough information on specific products or services based on client feedback [22]. Customers rely on reviews with the most votes to make a purchase decision. The results are congruent with the findings of [23] on how customer reviews impact internet purchases. Positive product reviews enable clients to build greater trust in the things they want to buy online.

Sentiment analysis has gaps due to issues with the accuracy and reliability of the utilized methods, as well as the lack of standardization in sentiment interpretation and classification. More research is needed to optimize customer engagement programs and improve the understanding of customer sentiments. This study aims to address the potential gaps in the sentiment analysis based on Machine Learning Approaches in Amazon Product Reviews using Decision Trees and Regression models. Specifically, this study investigated the effectiveness of different customer engagement strategies in influencing customer sentiments and explored the impact of these strategies on sentiment classification accuracy.

III. METHODOLOGY

Sentiment analysis is considered a study of people's sentiments, feelings, and views as conveyed via writing, and is useful in understanding other peoples' points of view on any subject. Opinions can be positive, negative, or neutral. The suggested method was based on an ML prediction model, to analyze both positive and negative reviews by binary classification. Figure 1 shows the steps of this method, beginning with data collection and ending with the evaluation of each classification model.

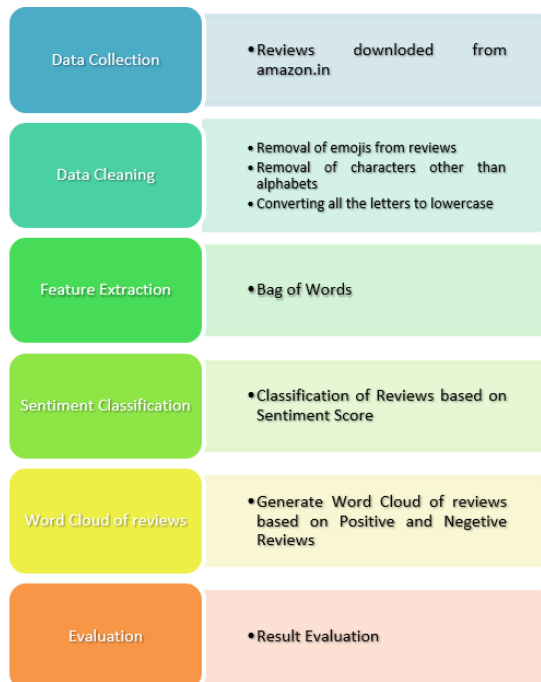


Fig. 1. Overall sentiment analysis approach for Amazon reviews.

A. Programming Environment

Python is one of the most used programming languages in data science and ML, as it offers a large library collection for solving various ML problems. Python was chosen due to its extensive libraries and ease of use. Scikit-learn is a Python package that provides supervised ML algorithms [24], including many classification algorithms, such as SVM and Naive Bayes, and feature extraction techniques.

B. The Dataset

Figure 2 presents an example of an Amazon review to better grasp the dataset's structure and format. An Amazon user review consists of the following four key components that assist in comprehending and analyzing the reviews:

- Summary: The title of the review.
- Review text: The review's actual content.
- Rating: The product's user rating on a scale of 1 to 5.
- Helpfulness: The percentage of persons who considered the review beneficial.



Fig. 2. Actual Amazon customer review sample.

Data collection was the initial stage of the study. Raw data were obtained from the website amazon.in and converted to Comma Separated Values (CSV) format. The data set contained 4960 Titan Men Watches reviews.

C. Data Cleaning

This is a critical stage in examining any type of data and has a significant influence on the success of ML models. There are numerous types of pre-processing procedures and the appropriate ways must be selected. This study employed four distinct data preparation steps in the reviews:

1. Remove Emojis.
2. Remove HTML tags.
3. Lowercase all letters.
4. Filter numbers and special characters.

1) Emoji Removal

Although people use emojis to express their feelings, they were removed since they did not affect the identification of the polarity of the review.

2) HTML Tag Removal

The HTML tags of the retrieved reviews were removed because they did not affect the determination of the polarity of the review.

3) Converting all Letters to Lowercase

In different reviews, there is a good chance that identical words will appear in different situations and the system will recognize them as distinct words. Converting all letters to lowercase was used to avoid such problems.

4) Filtering Numbers and Special Characters

All the unnecessary elements in determining the review's polarity were eliminated to make the data tidy and clean. The special letters and digits were eliminated.

D. Feature Extraction

ML algorithms interpret data in specified formats. The text data were turned into numerical feature vectors, a process known as vectorization. Bag of Words is one such approach that involves tokenization, normalization, and counting. This study used CountVectorizer to represent words in terms of Bag of Words. CountVectorizer requires specifying an N-gram range, which is a tuple consisting of the lowest and maximum length of the sequence of words to be regarded as features.

E. Word Cloud of Reviews

The word cloud is a method of displaying text data where each word's size corresponds to its frequency or importance. A word cloud can be used to highlight significant textual information, and they are widely used in social network data analysis. This study used word clouds to represent the most often-used terms in reviews.

IV. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is a method of displaying and evaluating data concealed in rows and columns, using several charts to provide maximum insight into the dataset. This study investigated the Positive, Negative, and Neutral review categories of Amazon reviews based on the sentiment score. Figure 3 shows the Amazon product review categories. This study scrapped 4960 reviews of a specific product and stored them, after cleaning, in a CSV file.

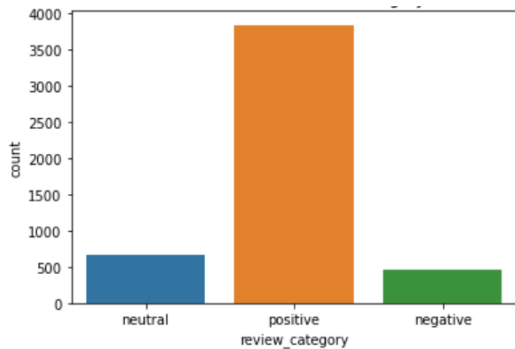


Fig. 3. Distribution of reviews' categories.

The text data were presented graphically using Python's Word Cloud package. Figures 4-6 show the most common words used in reviews, as positive reviews, and as negative reviews, respectively.



Fig. 4. Most common words used in the reviews.



Fig. 5. Most common words used in positive reviews



Fig. 6. Most common words in negative reviews.

A. Unigrams

Unigrams determine the frequency of single words in the reviews. Figures 7-8 show the unigrams of the top 20 positive and negative reviews, respectively.

B. Bigrams

Unigrams do not offer a clear understanding of a consumer's intended message. Bigrams can capture the meaning behind consecutive pairs of words in the reviews. Combining two adjacent words into a single unit can offer valuable insight into the context and meaning of a text. Bigrams show the frequency of two-word combinations in a text review. For instance, in the sentence "I love this product", the bigram "love this" would be generated, while "I hate this product" would produce the bigram "hate this". Analyzing the most prevalent bigrams in a collection of reviews can help to understand what consumers are attempting to convey regarding their encounter with a product or service.

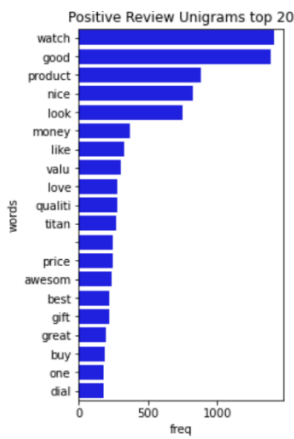


Fig. 7. Top 20 positive review unigram.

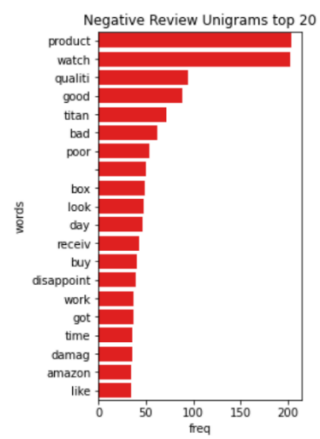


Fig. 8. Top 20 negative review unigram.

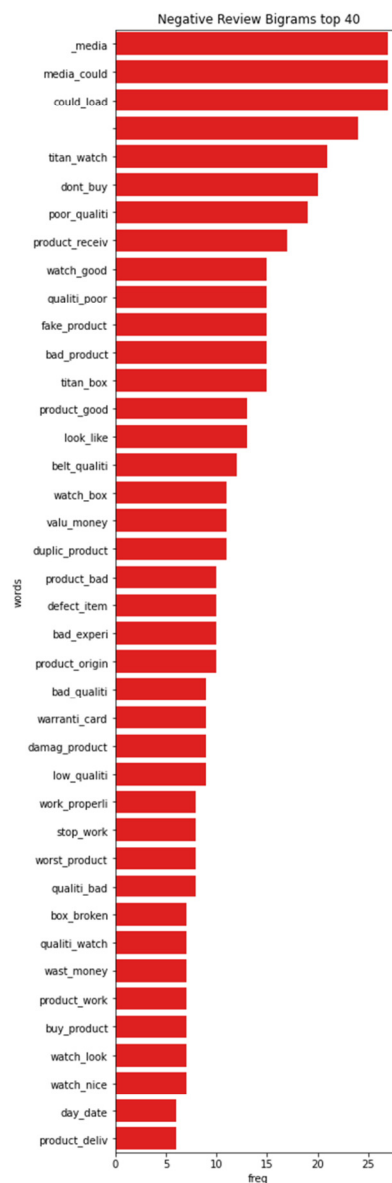
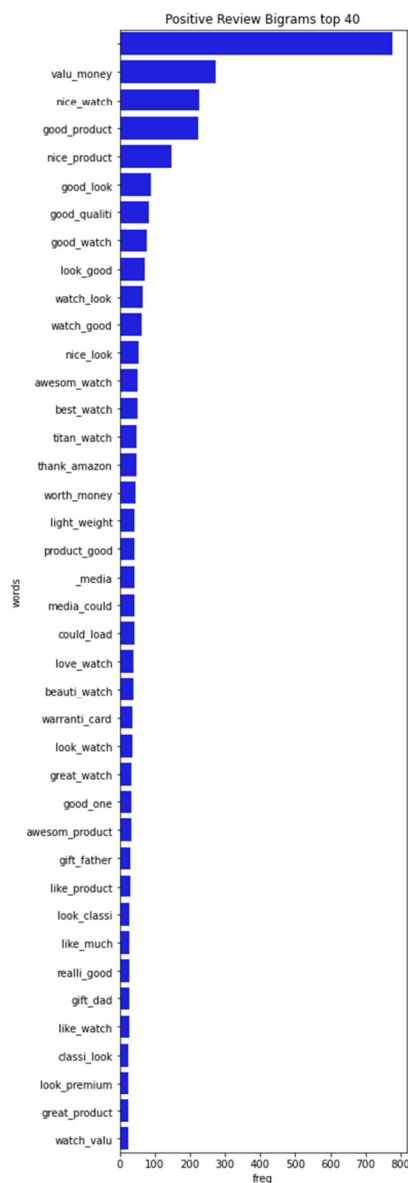


Fig. 9. Top 40 positive and negative review bigrams.

Figure 9 shows the top 40 positive and negative review bigrams. The bar plot shows that the most common bigrams in positive reviews were "value for money", "lovely watch", "good product", etc. These bigrams suggest that customers were satisfied with the quality and value of the product they purchased. In contrast, the most commonly used bigrams in negative reviews were "media could", "could not", "titan watch", etc., implying that consumers experienced problems with the product, such as difficulty with media playback or loading times. Overall, bigrams are a useful tool in natural language processing and sentiment analysis, providing valuable insight into the opinions and experiences of consumers. By identifying and analyzing the most common bigrams in a set of reviews, a better understanding of the strengths and weaknesses of a product or service can be gained. This information can then be used to make more informed decisions.

V. RESULTS AND ANALYSIS

This study used Python and Jupyter, in conjunction with supporting libraries, to perform data purification, visualization, pre-processing, and ML modeling. Supervised ML was used to create sentiment classification models, by building training and testing sets. A collection of features was then taken from the training and testing data and supplied into a classifier model, such as Logistic Regression (LR) and Decision Tree (DT). The complete review dataset was separated into two parts: 75% of the data was used to train the models, while 25% was used to assess their performance. The performance of the models was validated using accuracy and confusion matrix.

A. Accuracy (A)

Accuracy (A) is defined as the proportion of correctly predicted occurrences to the total number of occurrences by:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative. Table I shows the accuracy results of the models used in this study.

TABLE I. ACCURACY OF MODELS

Model	Accuracy
Logistic Regression	94%
Decision Tree	99%

The results show that both Logistic Regression and Decision Tree were effective. However, the two models had different levels of accuracy, as DT achieved 99% outperforming LR, which achieved 94%, as shown in Table I and Figure 10. This could be attributed to DT's ability to handle complex decision-making processes that may not be captured by LR. The difference in accuracy between the two models is significant, and it may impact the model selection decision depending on the importance of the classification task. For example, in a high-stakes classification task, such as in medical diagnosis, the higher accuracy of DTs could be crucial to make accurate and reliable predictions.

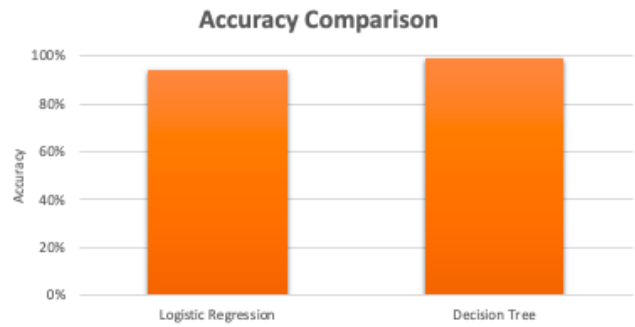


Fig. 10. Accuracy comparison for the methods used.

B. Confusion Matrix

It is worth noting that the Accuracy values reported are not the only performance metrics to consider when evaluating the models. Other metrics such as Precision, Recall, and F1-score can provide a more nuanced view of the models' performance and should be considered alongside Accuracy. They can be used to assess the performance of the categorization methods. These parameters are important for measuring the efficacy of supervised ML algorithms since they are based on the confusion matrix [25]. A confusion matrix is widely used to visualize an algorithm's performance. Classification terms TP, TN, FN, and FP are used to compare class labels in this matrix, as shown in Table II and Figure 11. Based on the data of the confusion matrix, Precision, Recall, F-measure, and Accuracy can be used to evaluate the performance of the classifier.

TABLE II. CONFUSION MATRIX

P V \ A V	Predicted-Negative	Predicted-Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

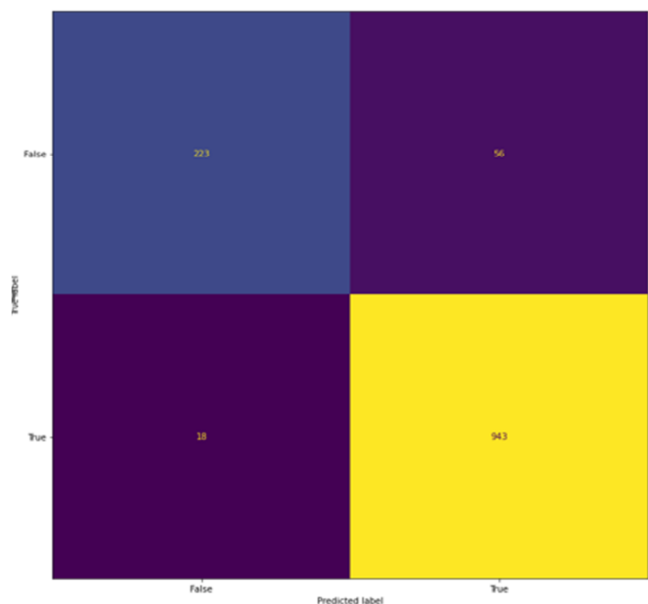


Fig. 11. The confusion Matrix of the LR model.

VI. CONCLUSION

Sentiment analysis is a necessary and popular technique for collecting information from text data on e-commerce websites. E-commerce platforms produce enormous volumes of text data every day in the form of suggestions, reviews, tweets, and comments. Additionally, emoticons, ratings, and reviews all suggest people's opinions. A customer can learn more about a product and make an informed choice by extrapolating information from reviews. This study used multiclass and binary classification for Amazon reviews of a product using supervised machine learning methods such as Logistic Regression and Decision Tree. Logistic Regression produced an excellent outcome with 94% accuracy and Decision Tree produced outstanding results with 99% accuracy. E-commerce websites should consider various feature extraction methods and machine learning techniques, examine additional product categories, analyze unstructured data, and incorporate sentiment analysis into customer experience strategies to enhance customer satisfaction and loyalty.

REFERENCES

- [1] X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, Jun. 2015, Art. no. 5, <https://doi.org/10.1186/s40537-015-0015-2>.
- [2] J. McAuley, "Amazon product data," *Recommender Systems and Personalization Datasets*. https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews.
- [3] M. A. Kausar, A. Soosaimanicka, and M. Nasar, "Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, 2021, <https://doi.org/10.14569/IJACSA.2021.0120252>.
- [4] M. Mahyoob, J. Algaraady, M. Alrahiali, and A. Alblwi, "Sentiment Analysis of Public Tweets Towards the Emergence of SARS-CoV-2 Omicron Variant: A Social Media Analytics Framework," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8525–8531, Jun. 2022, <https://doi.org/10.48084/etasr.4865>.
- [5] N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," *Spatial Information Research*, vol. 28, no. 5, pp. 601–607, Oct. 2020, <https://doi.org/10.1007/s41324-020-00320-2>.
- [6] V. Schoenmueller, O. Netzer, and F. Stahl, "The Polarity of Online Reviews: Prevalence, Drivers and Implications," *Journal of Marketing Research*, vol. 57, no. 5, pp. 853–877, Oct. 2020, <https://doi.org/10.1177/0022243720941832>.
- [7] Md. A. Kausar, V. S. Dhaka, and S. K. Singh, "An Effective Parallel Web Crawler based on Mobile Agent and Incremental Crawling," *Journal of Industrial and Intelligent Information*, vol. 1, no. 2, pp. 86–90, Jun. 2013, <https://doi.org/10.12720/jiii.1.2.86-90>.
- [8] I. Karamitsos, S. Albarhami, and C. Apostolopoulos, "Tweet Sentiment Analysis (TSA) for Cloud Providers Using Classification Algorithms and Latent Semantic Analysis," *Journal of Data Analysis and Information Processing*, vol. 7, no. 4, Nov. 2019, Art. no. 69212, <https://doi.org/10.4236/jdaip.2019.74016>.
- [9] U. P. Gurav and S. Kotrappa, "Sentiment Aware Stock Price Forecasting using an SA-RNN-LBL Learning Model," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6356–6361, Oct. 2020, <https://doi.org/10.48084/etasr.3805>.
- [10] A. Rahman and M. N. A. Khan, "A Classification Based Model to Assess Customer Behavior in Banking Sector," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2949–2953, Jun. 2018, <https://doi.org/10.48084/etasr.1917>.
- [11] V. K. Jain, S. Kumar, and P. Mahanti, "Sentiment Recognition in Customer Reviews Using Deep Learning," *International Journal of Enterprise Information Systems (IJEIS)*, vol. 14, no. 2, pp. 77–86, Apr. 2018, <https://doi.org/10.4018/IJEIS.2018040105>.
- [12] J. Lim, M. Park, S. Anitsal, M. M. Anitsal, and I. Anitsal, "Retail Customer Sentiment Analysis: Customers' Reviews of Top Ten U.S. Retailers' Performance," *Global Journal of Management and Marketing*, vol. 3, no. 1, pp. 124–150, 2019.
- [13] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques," in *Cognitive Informatics and Soft Computing*, Singapore, 2019, pp. 639–647, https://doi.org/10.1007/978-981-13-0617-4_61.
- [14] V. Vyas and V. Uma, "Approaches to Sentiment Analysis on Product Reviews," in *Sentiment Analysis and Knowledge Discovery in Contemporary Business*, IGI Global, 2019, pp. 15–30.
- [15] Md. A. Kausar, V. S. Dhaka, and S. K. Singh, "Web Crawler Based on Mobile Agent and Java Aplets," *International Journal of Information Technology and Computer Science*, vol. 5, no. 10, pp. 85–91, Sep. 2013, <https://doi.org/10.5815/ijitcs.2013.10.09>.
- [16] S. Govindaraj and K. Gopalakrishnan, "Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features," *ETRI Journal*, vol. 38, no. 3, pp. 494–501, 2016, <https://doi.org/10.4218/etrij.16.0115.0684>.
- [17] M. Ghasemaghaei, S. P. Eslami, K. Deal, and K. Hassanein, "Reviews' length and sentiment as correlates of online reviews' ratings," *Internet Research*, vol. 28, no. 3, pp. 544–563, Jan. 2018, <https://doi.org/10.1108/IntR-12-2016-0394>.
- [18] P. Sasikala and L. Mary Immaculate Sheela, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS," *Journal of Big Data*, vol. 7, no. 1, May 2020, Art. no. 33, <https://doi.org/10.1186/s40537-020-00308-7>.
- [19] S. K. Sharma, S. Chakraborti, and T. Jha, "Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach," *Information Systems and e-Business Management*, vol. 17, no. 2, pp. 261–284, Dec. 2019, <https://doi.org/10.1007/s10257-019-00438-3>.
- [20] A. Y. L. Chong, B. Li, E. W. T. Ngai, E. Ch'ng, and F. Lee, "Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach," *International Journal of Operations & Production Management*, vol. 36, no. 4, pp. 358–383, Jan. 2016, <https://doi.org/10.1108/IJOPM-03-2015-0151>.
- [21] J. Du, J. Rong, S. Michalska, H. Wang, and Y. Zhang, "Feature selection for helpfulness prediction of online product reviews: An empirical study," *PLOS ONE*, vol. 14, no. 12, 2019, Art. no. e0226902, <https://doi.org/10.1371/journal.pone.0226902>.
- [22] Meenakshi, A. Banerjee, N. Intwala, and V. Sawant, "Sentiment Analysis of Amazon Mobile Reviews," in *ICT Systems and Sustainability*, Singapore, 2020, pp. 43–52, https://doi.org/10.1007/978-981-15-0936-0_4.
- [23] K. Q. Anh, Y. Nagai, and L. M. Nguyen, "Extracting Customer Reviews from Online Shopping and Its Perspective on Product Design," *Vietnam Journal of Computer Science*, vol. 06, no. 01, pp. 43–56, Feb. 2019, <https://doi.org/10.1142/S2196888819500088>.
- [24] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [25] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Systems with Applications*, vol. 57, pp. 117–126, Sep. 2016, <https://doi.org/10.1016/j.eswa.2016.03.028>.