RP Current Trends In Engineering And Technology (e-ISSN: 2583-5491)
Vol. 4, No. 3 (July – September 2025) pp. 53–57

RESEARCH PLATEAU PUBLISHERS

Cite this article: A. Kumar, P. Rawat, P. Aswal, M. Nath, Sentiment analysis of Amazon reviews: Comparative performance of machine learning algorithms, *RP Cur. Tr. Eng. Tech*. **4** (2025) 53–57.

Original Research Article

Sentiment analysis of Amazon reviews: Comparative performance of machine learning algorithms

Anuj Kumar[1], Prachi Rawat[2,*], Priyank Aswal[3], Manjeet Nath[4]

[1,3,4]Department of Computer Science, School of Technology, Doon University, Dehradun, Uttrakhand, India
[2]Computer Science and Engineering Department, G.B. Pant Institute of Engineering and Technology, Pauri Garhwal, Uttrakhand, India
*Corresponding author, E-mail: prachi.rawat03@gmail.com

ARTICLE HISTORY
Received: 5 July 2025
Revised: 8 Sept. 2025
Accepted: 10 Sept. 2025
Published: 11 Sept. 2025

KEYWORDS
Natural Language Processing (NLP); Machine learning; Sentiment analysis; Text classification; Supervised learning; Consumer behaviour.

ABSTRACT
Collecting and analyzing people's opinions, emotions, perceptions, attitudes, etc. toward several entities, such as subjects, products, and services, is known as sentiment analysis (SA), sometimes known as opinion mining (OM). People are producing a tonne of thoughts and reviews about goods, services, and daily activities as a result of the quick development of Internet-based applications such blogs, social networks, and webpages. To understand human psychology, data must be digested at the same speed as it is produced. Sentiment analysis determines the author's attitude toward a particular object, organization, person, or place — whether it be favorable, negative, or neutral. This study compares many supervised Machine Learning (ML) classification algorithms using Amazon product evaluations to identify the most trustworthy methodology for sentiment analysis results. After pre-processing the product reviews, the retrieved sentiments are classified as either positive or negative. Naïve Bayes (NB), Logistic Regression (LR), K Nearest Neighbour (KNN), and Support Vector Machine (SVM) are the methods used to analyse the sentiments. F1-score, precision, recall, and confusion matrix were used to assess the models' performance. The LR performed better than other versions and offered an accuracy of 99%.

1. Introduction

E-commerce now a days become more popular because people are more interested in buying products on e-commerce websites rather than going to the physical market to save the time and get the product at home. Notably, reading product reviews before making a purchase is a typical scenario [1]. Therefore, the user has been prejudiced either favourably or unfavourably by the product reviews. However, it is an inhuman achievement to read thousands of reviews. In the age of growing machine learning-based algorithms, it takes a lot of effort to sift through hundreds of evaluations to find a product, where a review of a particular product may be differentiated to determine its uniqueness among customers globally [2]. It would thus be beneficial to extract some helpful information from these evaluations. This is where machine learning is useful. AI and machine learning have completely changed everything. Applications of machine learning have been thoroughly investigated in areas such as sentiment analysis, business, medical, and so on.

Sentiment analysis is also known as opinion mining, a branch of NLP that focuses on identifying the author's point of view on a given topic by distinguishing between distinct points of view. Using computational techniques, this approach gathers subjective data from a range of sources, such as blogs, forums, social media, and reviews. Theoretically, it is a method of classification that highlights the neutral, negative, or positive emotions expressed in a particular review [3].

Sentiment analysis is typically undertaken at a variety of levels, ranging from coarse to fine. Coarse sentiment analysis determines the sentiment of an entire document, whereas fine sentiment analysis examines individual attributes. Sentence-level sentiment analysis falls somewhere between these two [4]. Various studies performance on sentiment analysis of user reviews, which shows that sentiment classifiers' performance is dependent on the type data. Therefore, we cannot say that one classifier is better than another on every issue because one classifier does not always perform better than the other. The effects of domain evidence in choosing a feature vector can be ascertained by doing sentiment analysis on a particular domain. With this particular feature vector, many classifiers are used to do classification in order to ascertain their influence in this particular domain [5]. The quantity of sentiment analysis studies has increased in recent years. Many efforts have recently been started to handle emotion in literary texts.

The study [6] utilized four classifiers for text classification: Support Vector Machine, Logistic Regression, Naïve Bayes, and K-NN. Classifiers are simply machine learning algorithms, usually referred to as models. The textual data set must be used to train these classifiers to provide likelihoods. Allocating tags and subsequently creating relationships between textual segments, and these methods are used to train classifiers. The statistical significance of each independent variable in terms of probability is evaluated using logistic regression. The nearest neighbor classifier aids in group maintenance by keeping related items together. SVM is popular for its performance in classification and regression tasks, particularly when dealing with complex data with clear margins of separation. Naive Bayes is also utilized for its simplicity, efficiency, and efficacy in text categorization and other applications that require independence. The four methods are then compared to determine which is the best classifier.

2. Literature review

Sentiment analysis is a fundamental component of NLP. Its primary objective is to extract arbitrary data from text. This is crucial for e-commerce, as knowing what customers have to say about a product or service can give businesses insightful information. With millions of product reviews, Amazon is a well-known e-commerce site and an excellent case study for sentiment analysis [7]. Modern advances in deep learning and machine learning have resulted in sophisticated models capable of efficiently extracting sentiments from massive volumes of unstructured data. The newest studies on sentiment analysis of product evaluations on Amazon are compiled and assessed in this review of the literature. It presents some methods and strategies from research that have been published in the recent four years.

According to recent studies, sentiment analysis employing both complex deep learning algorithms and conventional machine learning is becoming more and more popular. For example, Xing Fang and Justin Zhan's study gathered over 5.1 million consumer reviews across a variety of Amazon product categories, including home goods, electronics, books, and cosmetics. It calculated sentiment scores using a variety of sentiment indicators, which revealed an unequal distribution of reviews and the need for normalizing techniques to get accurate sentiment analysis results [8]. Saridena (2023) also used supervised machine learning methods to analyze sentiment in the Amazon reviews dataset from Kaggle. The study stressed the importance of feature selection and data pre-processing in improving the predictive model's functionality [9].

In sentiment analysis tasks, deep learning models—in particular, transformer-based architectures like BERT and RoBERTa—have shown improved performance and also show the capability of these models to achieve high accuracy and capture contextual subtleties in customer reviews [10].

This study examined several deep learning architectures, including convolutional neural networks (CNN) and long short-term memory networks (LSTM), to analyze the sentiment in Amazon product reviews. It stressed the benefits of using LSTM-CNN hybrids to manage large datasets and obtain contextual data [5]. The study focused on aspect-based sentiment analysis of eco-friendly products using the BERT and T5 models. The study's high accuracy (BERT at 92% and T5 at 91%) demonstrates how well these models understand user feedback and enable adjustments to product designs [11].

This study [2] compares the sentiment analysis of Amazon product evaluations using SVM and Naive Bayes classifiers. It uses TF-IDF to extract features and assesses how well the models can distinguish between reviews' positive and negative sentiments. SVM is more accurate than Naive Bayes. In addition, Naive Bayes is less computationally expensive and more efficient than SVM.

In this paper, we will conduct a comprehensive comparison of various methodologies for implementing Sentiment Analysis and assess their respective accuracies.

3. Methodology

The expected method for carrying out SA on Amazon reviews is outlined in this section. Figure 1 shows the several steps of the procedure. Gathering and preparing data is the first step. Following that, feature extraction techniques are used, and machine learning models are then trained. Lastly, an assessment is made of these models' efficacy [10]. Classifying evaluations as either good or negative and evaluating the performance of different classification method.

Data Collection

The study's dataset was a tab-separated values (TSV) file including Amazon product reviews. The data put into a Pandas Data Frame consists of two columns: Sentiment (the desired variable) and Review (text data).

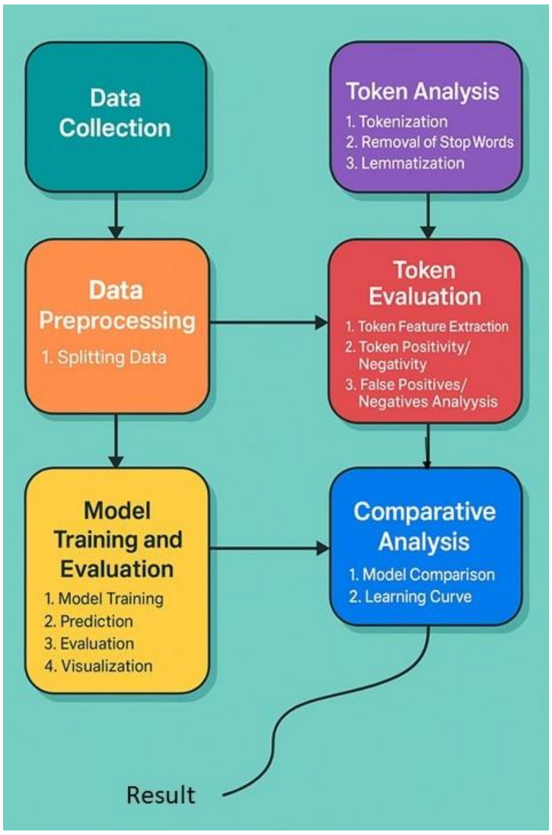

Figure 1: Sentiment analysis work flow for Amazon reviews.

Data Preprocessing

In NLP operations, data preparation is a crucial step that increases the effectiveness of knowledge discovery. In order to prepare and enhance the dataset for more efficient analysis, it incorporates methods such as data cleansing, integration, transformation, and reduction [11]. 3942 positive and 2975 negative reviews make up our dataset's total of 6917 reviews. A 90-10 split ratio was then used to separate the dataset into training and testing sets.

The text input is converted into numerical features with Count Vectorizer. This generates a matrix of the token counts from the text, avoiding frequent English stop words and focusing on tokens that are present in at least four papers but not more than 80% of them.

Model Training and Evaluation

Four distinct machine learning models are developed and tested using the test data. The models are as follows:

Multinomial Naive Bayes: The MNB algorithm employ Bayesian learning for predicting the tags of a text utilizing Bayes' theorem. MNB determines the probability of each independent label in a given sample. This method provides the label with the highest probability. The MNB employed a default set of settings. First, the alpha parameter has a value of 1.0. Next, fit_prior is True, and class_prior is None, that is the default value [13].

Logistic Regression: The supervised machine learning technique known as logistic regression uses multinomial logistic regression to categorize sentiment into two groups with positive and negative labels or into several classes [14]. In the classification process, real-valued features are extracted from the input, multiplied by a weight, added together, and then run through a sigmoid function to get a probability. To make a decision, a threshold value is utilized.

Support Vector Machine: It is a non-probabilistic classifier that can handle both discrete and continuous variables and segregate data either linearly or nonlinearly. In many cases, it outperforms most other algorithms in classification and has a strong theoretical basis. SVM is frequently used in sentiment classification because it works well for text categorization. Finding the best hyperplane to divide classes is the primary objective of the SVM classifier. Since a bigger margin lowers the classifier's generalization error, the hyperplane with the largest margin to the closest training point from either class is said to have an effective separation.

K-Nearest Neighbours (KNN): Text is represented as a spatial vector in K-NN), which is represented by S. = S (T_1, W_1; T_2, W_2; … T_n, W_n). The texts with the highest similarity are chosen once similarity between each text and the training text is identified. In addition to feature vectors in a dataset, KNN operates on class labels. KNN keeps track of every case and uses similarity metrics to help classify fresh cases. The following steps are taken for every model:

Model Training: The training data matrix is utilized to train each model.

Prediction: The test data's sentiments are predicted based on the training model.

Evaluation: The performance of the model is evaluated by computing the accuracy, precision, recall, F1-score, and confusion matrix.

Visualization: A heat map is employed for displaying the confusion matrix.

Comparative Analysis

Model Comparison*:* The performance metrics of all models are compared, and their accuracies are plotted for better visualization.

Learning Curve*:* A learning curve is plotted for the chosen model to evaluate its performance across different training sizes.

Token Analysis

Token Feature Extraction*:* The tokens (words) used in the Naive Bayes model are extracted and analysed for their contribution to positive and negative sentiments.

Token Positivity/Negativity*:* The number of positive and negative tokens are calculated, and specific tokens can be searched to analyse their sentiment.

False Positives/Negatives Analysis*:* Analysing incorrect predictions to understand the limitations of the models.

4. Results

Using comprehensive experimental findings, this section illustrates how well our suggested methodology for sentiment analysis performs [15]. The desired output was acquired when the code was successfully executed. Their predicted efficiency is determined and compared after testing and training the model with the dataset to reveal which approach is more effective in classifying [3]. Five parameters—accuracy, precision, F1-score, support, and confusion matrix—are used to compare the methods, as was previously indicated [7]. Our dataset has 6917 reviews in total, of which 2975 are unfavourable and 3942 are in the positive category. All four models' accuracy, precision, recall, and F1-score values are displayed in the table 1 below. Figure 2 displays the comparison of models evaluated on various metrics.

Table 1: Comparison of all the four Models on the basis of four evaluation metrics.

Models	Accuracies	Precision	Recall	F1score
Naïve Bayes	98.91%	0.98	1.00	0.99
Logistic Regression	99.34%	0.99	0.99	0.99
SVM	99.06%	0.99	0.99	0.99
KNN	98.69%	0.99	0.99	0.99

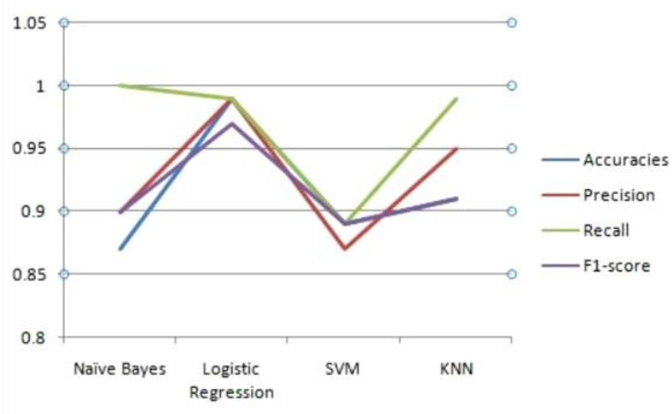

Figure 2: Graph of comparison of models on the basis of evaluation metrics.

Finally, a variety of criteria are used to compare the model to standard models. Model assessment metrics are required to evaluate model performance. Using known real facts, a confusion matrix is created that shows the percentage of accurate and inaccurate assessments or forecasts. For data fitting based on positive and negative classes, this matrix shows true positive (TP), false negative (FN), false positive (FP), and true negative (TN) values. Researchers used metrics including accuracy, precision, recall, F1 score, and others to assess their model based on these numbers.

Given below the confusion matrix of all the four models in Figures 3, 4, 5 and 6.

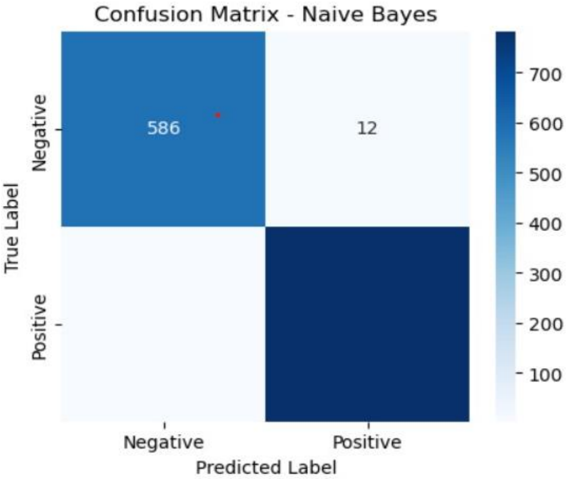

Figure 3: Confusion Matrix of Naïve Bayes.

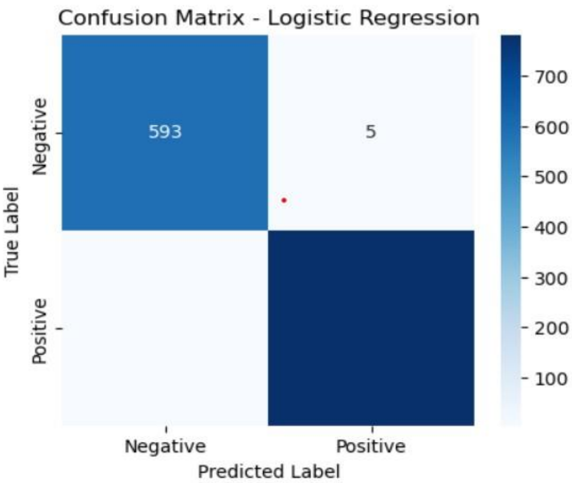

Figure 4: Confusion Matrix of Logistic Regression.

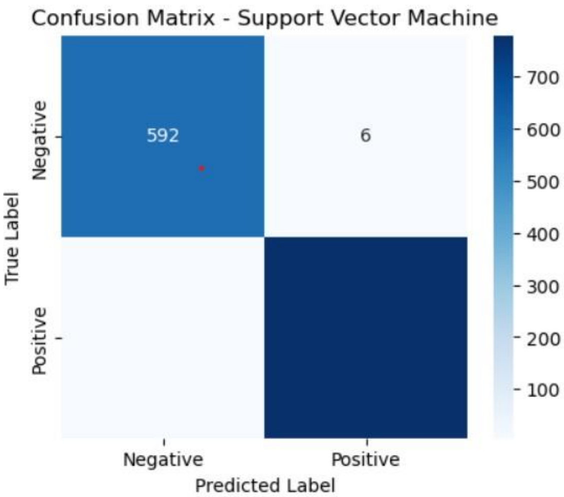

Figure 5: Confusion Matrix of Support Vector Machine.

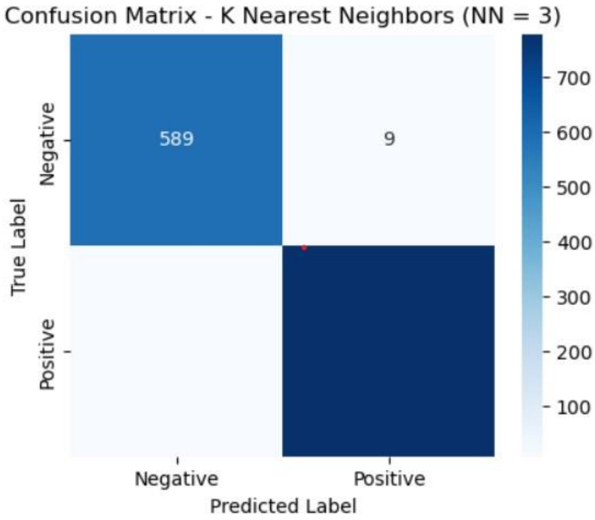

Figure 6: Confusion Matrix of K-Nearest Neighbour.

Also, the learning rate curve of all the four models are given below in Figures 7, 8, 9 and 10.

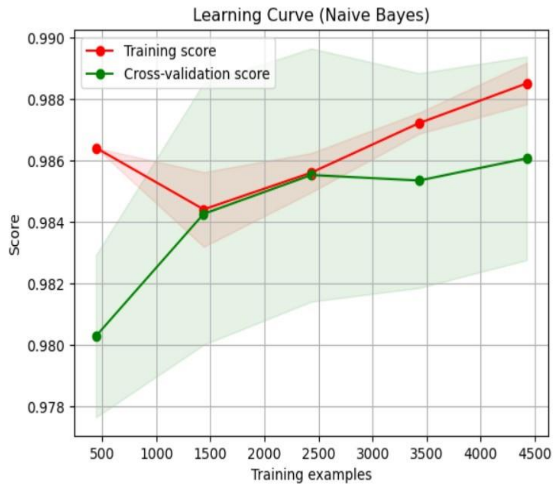

Figure 7: Learning Curve of Naïve Bayes.

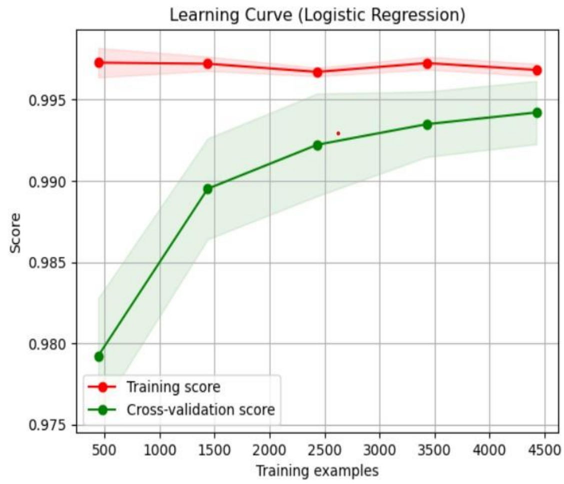

Figure 8: Learning Curve of Logistic Regression

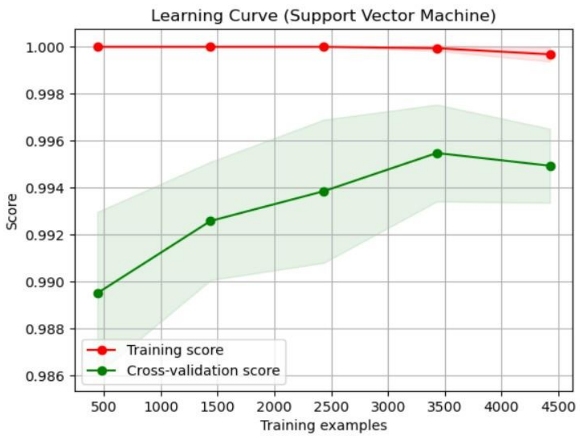

Figure 9: Learning Curve of SVM.

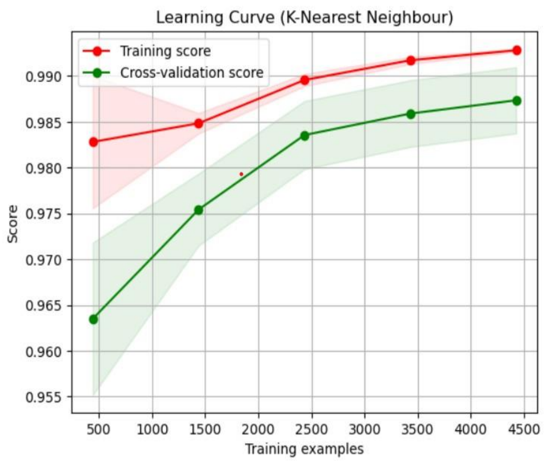

Figure 10: Learning Curve of KNN.

5. Conclusions

This work considers conducted research on various machine learning-based tools and approaches for categorization and sentiment analysis. This includes discussion of well-known algorithms and strategies that are often employed for sentiment classification [16]. One of the key findings is the identification of trends in customer sentiments. Models with varying degrees of accuracy, precision, recall, and F1-score include Naive Bayes, Logistic Regression, SVM, and KNN. The sentiment distribution research identified similar motifs in both positive and negative evaluations, offering organizations actionable information.

By showcasing the value of many machine learning models on a substantial dataset of Amazon reviews, the study advances the field's comprehension of sentiment analysis in e-commerce. It contributes to the fields of data science and NLP by showing how automated sentiment analysis may enhance business strategy by providing a deeper comprehension of customer feedback.

Authors' contributions

The author read and approved the final manuscript.

Conflicts of interest

The author declares no conflict of interest.

Funding

This research received no external funding.

Data availability

No new data were created.

References

[1] P. Nandwani, R. Verma, A review on sentiment analysis and emotion detection from text, *Soc. Netw. Anal. Mining* **11** (2021) 81.

[2] S. Dey, S. Wasif, D.S. Tonmoy, S. Sultana, J. Sarkar, M. Dey, A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews, In: *2020 International Conference on Contemporary Computing and Applications, IC3A 2020*, Institute of Electrical and Electronics Engineers Inc., (2020) pp. 217–220.

[3] B.K. Shah, A.K. Jaiswal, A. Shroff, A.K. Dixit, O.N. Kushwaha, N.K. Shah, Sentiments detection for Amazon product review, In: *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, Institute of Electrical and Electronics Engineers Inc. (2021).

[4] M.S. Neethu, R. Rajasree, Sentiment analysis in twitter using machine learning techniques, In: *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Tiruchengode, India (2013), pp. 1-5.

[5] J. Sangeetha, U. Kumuran, Comparison of sentiment analysis on online product reviews using optimised RNN-LSTM with support vector machine, *Webology* **19** (2022) 3883–3898.

[6] K. Shah, H. Patel, D. Sanghvi, M. Shah, A comparative analysis of logistic regression, random forest and KNN models for the text classification, *Augmented Human Res.* **5** (2020) 12..

[7] S. Saad, B. Saberi, Sentiment analysis or opinion mining: A review, *Int. J. Adv. Sci. Eng. Inform. Technol.* **7** (2017) 1660.

[8] X. Fang, J. Zhan, Sentiment analysis using product review data, *J. Big Data* **2** (2015) 5.

[9] A. Saridena, Sentiment analysis on Amazon product reviews, *J. Student Res.* 12 (2023) 1-5.

[10] H. Ali, E. Hashmi, S.Y. Yildirim, S. Shaikh, Analyzing Amazon products sentiment: A comparative study of machine and deep learning, and transformer-based techniques, *Electronics (Switzerland)* **13** (2024) 1305.

[11] M.K. ShaikVadla, M.A. Suresh, V.K. Viswanathan, Enhancing product design through AI-driven sentiment analysis of Amazon reviews using BERT, *Algorithms* **17** (2024) 59.

[12] M.F. Bin Harunasir, N. Palanichamy, S.C. Haw, K.W. Ng, Sentiment analysis of Amazon product reviews by supervised machine learning models, *J. Adv. Inform. Technol.* **14** (2023) 857–862.

[13] Imamah, F.H. Rachman, Twitter sentiment analysis of Covid-19 using term weighting TF-IDF and logistic regresion, In: *Proceeding - 6th Information Technology International Seminar, ITIS 2020*, Institute of Electrical and Electronics Engineers Inc., (2020) pp. 238–242.

[14] M. Birjali, M. Kasri, A. Beni-Hssane, A comprehensive survey on sentiment analysis: Approaches, challenges and trends, *Knowl. Based Syst.* **226** (2021) 107134.

[15] K. Sravana Kumari, B. Manjula, An efficient machine learning-based sentiment analysis for Amazon product reviews, *NeuroQuantology* **20** (2022) 1661-1668.

[16] S. Shah Muhammad, S. Awan, M. Ahmad, S. Aftab, S. Ahmad, Machine learning techniques for sentiment analysis: A review, *Int. J. Multidis. Sci. Eng.* **8** (2017) 27-31.