STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 12, Issue 2, 2025
May–August
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Optimizing Sentiment Analysis with Naïve Bayes and Random Forest Techniques: A Result-based Approach

Mintu Nagar[1], Chetan Kumar[2], Shalini Puri[3]*

Abstract

In the increased digitalization, the sentiment analysis and classification have evolved as an eminent area to determine the polarity of positive, negative, and neutral reviews of the customers and users on products. It is an integral application field that employs supervised learning, Machine Learning, and Natural Language Processing concepts. The proposed Semantic Analysis and Classification using Naive Bayes and Random Forest system accomplishes the sentiment polarity by classifying the user reviews into three categories. It extracted the features from the structured form of Amazon reviews and then employed the training and testing stages through Naive Bayes and Random Forest techniques. Several experiments were performed on the datasets of several products from an online shopping website, Amazon. The review dataset for the proposed system was collected from eight popular categories: mobile phones, scientific supplies, musical instruments, cameras, televisions, books, kitchen items, and grocery items. This system was tested on approximately 26,000 reviews. The highest number of positive and negative reviews were obtained for the "Mobile Phones" and "Cameras" categories. The best ratings of 4.5 and 4.3 were achieved using Naive Bayes and Random Forest classifiers for Amazon products, respectively. The system's predicted ratings differed only slightly from the actual ratings on Amazon, with the maximum differences being –0.20 for "Scientific Supplies" and +0.25 for "Kitchen Items". The proposed system obtained 97% precision, 92% recall, and 95% F1-score using the Naive Bayes classifier. It obtained 96% precision, 91% recall, and 93.75% F1-score using the random forest classifier. The overall 94.1% system accuracy was achieved, using the Naive Bayes classifier.

Keywords: Online products, sentiment analysis, classification, Random Forest, Naive Bayes, Amazon, reviews

***Author for Correspondence**
Shalini Puri
E-mail: shalini.puri@jaipur.manipal.edu

[1]Research Scholar, Department of Computer Science and Engineering, Kautilya Institute of Technology and Engineering, Jaipur, Rajasthan, India
[2]Associate Professor, Department of Computer Science and Engineering, Kautilya Institute of Technology and Engineering, Jaipur, Rajasthan, India
[3]Associate Professor, Department of Information Technology, Manipal University, Jaipur, Rajasthan, India

Received Date: March 05, 2025
Accepted Date: March 21, 2025
Published Date: April 19, 2025

Citation: Mintu Nagar, Chetan Kumar, Shalini Puri. Optimizing Sentiment Analysis with Naïve Bayes and Random Forest Techniques: A Result-based Approach. Journal of Artificial Intelligence Research & Advances. 2025; 12(2): 46–57p.
DOI: https://doi.org/10.37591/JOAIRA.v12i02.204465

INTRODUCTION

A subfield of Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) called sentiment analysis examines assessments of various online goods from many data sources [1–5]. In recent years, review identification and detection have received good attention from various researchers. Customer comments, reviews, and recommendations are the main sources for analyzing online products. Therefore, it is an important job for the assessment of logically linked information incorporated in review texts. This condition can be attained if the review contents are recognized accurately and efficiently. Sentiment or perception is an idea or opinion, especially a structure based on emotional concern rather than imagination. Sentiments represent the customer's point of view, such as positive (good), negative (bad), or neutral. They can be expressed in whole or in one form, or all their characteristics [6–8].

They represent the views, judgments, analyses, and beliefs of a person on a particular subject. Nowadays, the semantic analysis and classification models are gaining the attention of many NLP researchers. The purpose of this proposed research work is to design an automatic Semantic Analysis and Classification (SAC) system using Naïve Bayes (NB) and Random Forest (RF) that can identify a wide range of Amazon reviews very effectively. Although many SAC algorithms have been developed, they are limited in the review and polarity usage, classifier usage, and use of standard data domains [9–18].

These existing methods and algorithms have worked on a limited set of reviews of Amazon products. Most of them have considered a limited set of products for implementation. In addition, polarity was tested only for positive and negative reviews. Therefore, these limitations have been the inspiration for new research work using NB and RF [2–15].

The next section provides a detailed and systematic journey of existing SAC methods using several product-based and service-based data domains. The main focus here is to consider the English reviews collected through Amazon using NB and RF algorithms. Based on a few important metrics and their accuracy outcomes, these methods are contrasted with others. All of these algorithms have many critical issues, leading to the necessity of a new, automated, and functional semantic analyzer. The following section describes the design structure of the proposed SAC-NR system. The SAC-NR system design shows its framework and describes its steps in detail. The next section provides a discussion of the various test results of the SAC-NR system. It illustrates the review database of Amazon and the experiments performed on the SAC-NR system. Further, it includes a performance evaluation of existing methods based on important key parameters. Further, it explains the experimental results obtained for online products available in the Amazon data domain. All these results are analyzed and the effectiveness of the proposed system is compared with the existing system. The last section provides a conclusion and further suggestions for the future. It includes a discussion on the contributions of each chapter and lastly, the scope of the work.

LITERATURE REVIEW

Puri and Kaushik introduced a novel approach to text classification that leveraged an enhanced fuzzy similarity measure to accurately identify and cluster relevant features from textual data [1]. By utilizing feature clustering, it could effectively mine concepts and improve the classification performance of text documents. It addressed the challenges of traditional text classification methods, offering a more robust and precise solution for categorizing textual information. Dey and Noor first acquired the reviews and then pre-processed and cleaned them using case conversion, stemming, and stop word removal [2]. It used a bag-of-words model, and then extracted the features to predict the sentiments. It found 10 attributes from review records such as Clothing ID, Age, Title, Review Text, Rating, Recommendation, Positive Feedback Count, Division Name, Department Name, and Class Name. This research implemented the system using women's clothing. The sentiment analysis approach by Alrehili and Albalawi used the WEKA tool to convert the unbalanced dataset of customer reviews into a balanced dataset through pre-processing [3]. The pre-processing step performed case conversion, stemming, stop word removal, and considered six N-gram scenarios. Snowball, Iterated Lovins, and Lovins stemmers were used for the stemming process. Each unigram, bigram, and trigram were described in the N-gram situations, either with or without the elimination of stop words. After this, it extracted 21 features, and classified through several single classifiers and an ensemble (Bagging + Boosting + NB + RF + Support Vector Machine (SVM)) approach. Finally, it evaluated the performance using 10-fold cross-validation.

Another method by Shaheen *et al*. predicted the product rating by fetching the English reviews of Samsung, Apple, and LG products, pre-processing them through stop word removal, stemming, Part of Speech tagging, and extracting the features [4]. It focused on six key mobile attributes: Product Title, Brand, Price, Rating, Review Text, and Number of People who started the appraisal. Finally, it used seven different classifiers: Gradient Boosting, Multinomial NB, Long Short-Term Memory, RF, NB and NB-SVM, and Convolutional Neural Networks (CNN) to classify the reviews after training the classifiers for 50 epochs. It demonstrated several connections between reviews and ratings, as well as between product price and rating, review length and rating, and invention rating. It took into account the 1–5 rating system. Another such analysis method by Rintyarna *et al*. extracted the sentence level features using word sense disambiguation-based method and the domain sensitive features (DSF) using Senti-Circle-based method [5]. It achieved the results using the WEKA tool along with 10-fold cross validation. Jadhav and Jadhav illustrated a systematic and detailed study, and the general workflow of sentiment analysis [6]. It discussed different methods and solutions for the opinion spam-based problem. It emphasized the requirement of a dynamic sentiment analysis-based system with real-time data monitoring.

Kunaefi and Aritsugi followed the steps of data acquisition, cleaning, feature extraction, modelling, training, and evaluation to determine the reasons behind user purchases [7]. This method separated the non-argumentative reviews from those that contained both arguments and conclusions. Four classifiers were trained after the structural, lexical, and contextual features were retrieved. It computed the F1-score for three tasks, such as task 1 as argumentative review detection, task 2 as argumentative and decision review classification, and task 3 as decision review classification. In the study by Chopra *et al*. [8], the brand reviews were predicted by the sentiment analysis method of [8] using classifiers such as NB, RF, Decision Tree (DT), SVM, K-Nearest Neighbor (KNN), and multilayer perceptron. Dataset loading, filtering, pre-processing, feature extraction, modeling, and prediction were the phases it went through. It also assessed, predicted, and recommended the brand. Khan *et al*. worked on the sentiment classification method [9]. It first collected and compiled the data of three products, pre-processed it, and extracted the features, then trained the seven ML classifiers, compared the efficiency of different classifiers, and lastly recommended the classifier with the best results. It implemented the system for POSitive (POS), NEGative (NEG), and NEUTral (NEUT) sentiment polarities.

Elfaik and Nfaoui used word embeddings to compare nine different ML techniques for Arabic Sentiment Analysis which were Gaussian NB (GNB), Nu-SVM, Linear SVM (LSVM), Logistic Regression (LR), Stochastic Gradient Descent (SGD), RF, KNN, DT, and ADABoost (ADAB) [10]. In the opinion prediction method presented by Mandhula *et al.*, the review data was initially acquired and then pre-processed [11]. The pre-processing steps included lemmatization, detection of review spam, and removal of stop words and Uniform Resource Locators (URLs). Next, the system extracted keywords using Latent Dirichlet Allocation (LDA) and fuzzy C-Means, and subsequently classified them using a Selective Memory Architecture-Based Convolutional Neural Network (SMA-CNN). The review dataset included product ratings, user and product data, and plain text reviews. The execution involved eight product categories: Amazon instant video, books, electronics, home and kitchen, movie reviews, media, Kindle, and cameras. Mahadevan and Arock demonstrated a detailed study for the prediction of review ratings by using a combination of latent topics and associated sentiments along with multi-class recurrent neural network classifier [12]. The study assessed the effectiveness of these methods by employing non-negative matrix factorization, Latent Dirichlet Allocation (LDA), and term frequency-inverse document frequency (TF-IDF). The algorithm was tested on various datasets, including the Yelp restaurant review dataset, the 100k-Coursera course reviews dataset (covering electronics, cell phones and accessories), and the Amazon instant videos and video games.

To divide the data corpus into numerous sections, the aspect model described by Wang *et al.* first retrieved implicit aspects using matching algorithms, and then it extracted aspect words [13]. Then it built the data-based features, combined three features, trained the classifier, and finally tested it. The model of Srikanth *et al.* used NB to classify the online reviews by including the top-level end-to-end process flow [14]. Sindhu *et al*. first acquired sales data and predicted the sales of each product, and then analyzed the customer behavior to determine the most-demanded and least-demanded products

[15]. Finally, it evaluated the precision, Recall, and F1 measures. Chauhan *et al.* involved systematically collecting and evaluating citizens' opinions and experiences with these initiatives [16]. By leveraging various feedback channels, such as surveys, social media, and public forums, researchers could gain valuable insights into the effectiveness, strengths, and areas for improvement of these programs. This approach not only enhanced transparency and accountability but also ensured that the programs are better aligned with the needs and expectations of the public, ultimately leading to more informed policy decisions and improved service delivery.

The study by Meena *et al*. focused on leveraging machine learning techniques to segment customers based on their behavior and satisfaction levels [17]. By analyzing various influential factors that impact e-satisfaction, the research aimed to identify distinct customer segments and understand their specific needs and preferences. This approach allowed businesses to tailor their marketing strategies and service offerings to better meet the expectations of different customer groups, ultimately enhancing overall customer satisfaction and loyalty. The study by Ahmad and Mishra delved into the application of cutting-edge ML algorithms to analyze sentiments expressed in both handwritten and electronic text documents [18]. By employing advanced techniques, this research aimed to accurately identify and classify emotions, opinions, and attitudes present in various forms of written communication. The novel approach not only enhanced the understanding of sentiment dynamics but also bridged the gap between traditional handwritten documents and modern digital texts, providing a comprehensive solution for sentiment analysis across different mediums.

Existing systems used criteria like F1 score, recall, accuracy, and precision to assess their performance. The performance outcomes of various systems are compared based on the use of multiple or hybrid classifiers. Furthermore, analyzing these systems' future requirements, difficulties, and constraints highlights important problems and the reach of current algorithms. Although the majority of current systems and methodologies have successfully implemented classification models with an extensive collection of product reviews, there remains a need for an effective sentiment analysis classification system for numerous online products from platforms like Amazon and Flipkart that span various data domains and sources.

DETAILED DESIGN OF THE PROPOSED SAC-NR SYSTEM

The proposed Semantic Analysis and Classification (SAC-NR) model framework is a robust and time-efficient model. At the initial level, the customer registers itself and uses the URL or the website address for further processing. It categorizes the English reviews that are gathered online from Amazon shopping websites. This model uses two SL classifiers called NB and RF. The SAC-NR system is framed in two phases. These phases are the training and testing phases. The SAC-NR model initiates customer registration in the training phase through which the customer registers himself/herself. The customer uses his/her credentials for further processing. He/she enters several addresses of products and items on the Amazon website. It pre-processes these reviews, partitions them, and extracts their features. After that, it classifies them into pre-defined categories using any one classification technique. These categories are positive, negative, and neutral categories, and they are mutually exclusive from each other.

The system is trained using both classifiers. It is framed to categorize the reviews of several product items by employing one technique at a time. After that, this SAC-NR system is tested using a new set of reviews of products and items from Amazon. It follows all the procedural steps. Lastly, the selected classifier categorizes these reviews accurately.

Step I: Review Acquisition and Collection

The SAC-NR system design begins with a homepage. This homepage renders the first page. It asks the customer to enter his/her login credentials. First, the customer needs to register himself/herself. After registration, his/her account is created. Then the customer enters the link number and the website address links. After entering the website address in the browser, the SAC-NR checks the presence or absence of the website address on the server. A sizable collection of reviews for eight product categories was gathered from the Amazon shopping website to perform the SAC-NR system's experiments.

Step II: Review Pre-processing and Cleaning

In the pre-processing step, the SAC-NR model gathers the entire set of reviews of the product items from the webpage. It pre-processes all the reviews to get the important words only. As the undesired words increase the size of the database and space complexity, they are discarded. That is why the SAC-NR system uses an NLTK-based corpus. All of the review statements are separated, and only the written content is taken out. In this process, the system removes all stop words and applies the Porter stemmer for stemming. Additionally, it performs lemmatization. The significant terms that are extracted, specifically verbs, adverbs, nouns, and adjectives, are then sent for further processing.

Step III: Feature Extraction with Polarity Detection

After completing the pre-processing and classification steps, the SAC-NR system finds relevant features from the reviews. These extracted features are used to train the NB and RF models. For the testing phase, 30% of the product subgroups were retained, while the remaining 70% were dedicated to the training phase. These outcomes are derived for rating, accuracy computation, polarity computation, and categorization reviews. The factor-dependent analytical results are shown including the assessment of the current algorithms' performance metric in terms of utilization percentage. The training data is used, and the CSV file is used for the updated review process. This system finds out the polarity of the reviews. The polarity can be neutral, positive, and negative. These detection and identification are important for categorization and classification purposes.

Step IV: Sentiment Classification

The reviews are divided into three groups per the SAC-NR system. For this kind of classification, it makes use of the class corpus. Precision, recall, and F1 score are among the system performance metrics determined by this method. These metrics utilize parameters such as TR_POS (True Positive), TR_NEG (True Negative), FAL_POS (False Positive), and FAL_NEG (False Negative). True positive, false positive, true negative, and false negative are represented by TR_POS, FAL_POS, TR_NEG, and FAL_NEG, respectively.

FACTOR-DEPENDENT ANALYTICAL RESULTS OF EXISTING SYSTEMS

Factor-dependent analysis measures the performance of the existing systems based on two parameters, such as polarity and classifier usage. It shows the analytical results graphically. By contrasting the various classifier types and sentiment polarity methods now in use, these findings and observations were produced.

Factor I: Analyzing Polarity Factor

Figure 1 illustrates the percentage usage of the three types of sentiment polarity. They are POS_NEG (POSitive NEGative), POS_NEG_NEUT (POSitive NEGative NEUTral), and a combination of both. It is observed that most of the researchers preferred POS_NEG sentiment analysis systems. So, the existing methods preferred the POS_NEG, POS_NEG_NEUT, and POS_NEG with POS_NEG_NEUT groups of polarity with 81.33, 15.37, and 3.3% contributions, respectively.

Factor II: Analyzing Classification Factor

Numerous classifiers, including single and hybrid classifiers, were employed in the SAC algorithms that were in use at the time. SVM, RF, NB, DT, KNN, LR, Linear Regression (LiR), Genetic Algorithms (GA) and Artificial Neural Networks (ANN) are among them. Several systems opted to employ a hybrid set of classifiers, rather than relying on a single classifier in their implementation. The percentage employment of different classifiers in the current systems is shown in Figure 2. NB, SVM, RF, ANN, DT, KNN, LR, LiR, and GA classifiers were utilized by the majority of these contributors, with utilization rates of 86.78, 72.16, 66.34, 33.5, 19.5, 8.8, 4.4, 2.22, and 2.22%, respectively.

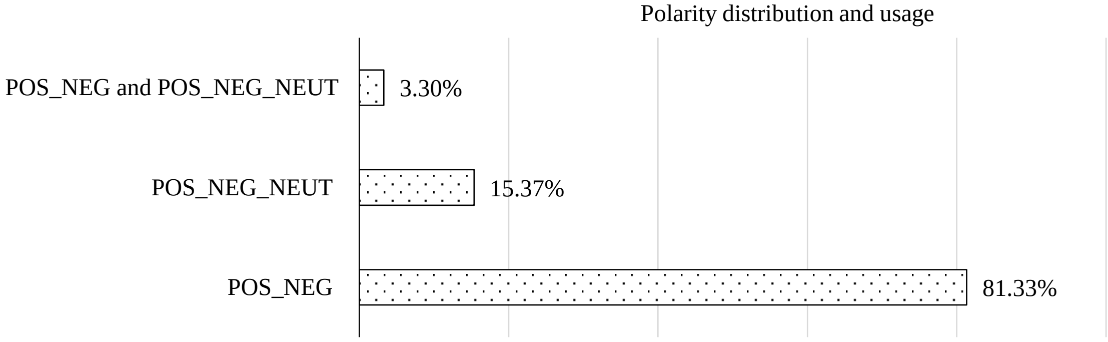

Figure 1. Depicting the sentiment polarity distribution with % usage.

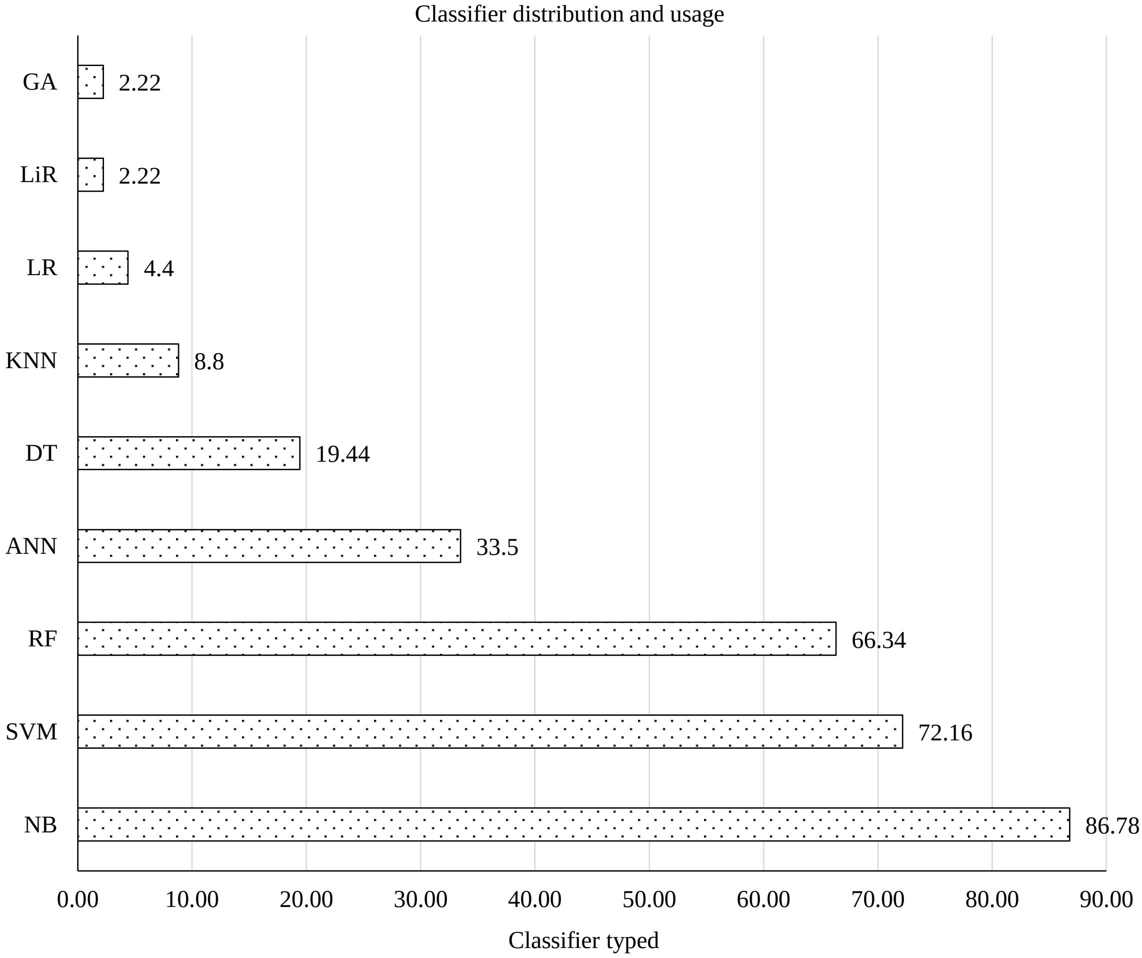

Figure 2. Depicting the classifiers with % Usage.

NB served as a significant classifier for the majority of them. Likewise, the remaining classifiers have other percentage usages ranging from 2.22 to 72.16%. Therefore, it is found that the feature extraction method emerged as the most highly recommended approach utilized in the sentiment analysis model of the research works. Conversely, these works primarily used bi-polar sentiment analysis techniques, which were used 81.33% of the time. It includes the positive and negative review group in sentiment polarity. Most of these contributors used Amazon as their primary source. Other findings stated that many researchers preferred "mobile phones" for analysis. Lastly, the NB classifier was used by 86.78% of contributors. The observations and findings obtained from these analytical results addressed and reviewed the concepts of automatic identification of sentiments.

SYSTEM CONFIGURATION AND DATASET USED

This section depicts the SAC-NR system's experimental configuration and SPECS. Further, it illustrates the required system configuration, the Amazon data set used to describe the training and testing groups. The computer system set up for executing the proposed SAC-NR system includes the following specifications: a Windows-10 operating system, an Intel(R) Core(TM) i5 processor, 6.00 GB of RAM, a 500 GB hard disk, and a 64-bit operating system. The SAC-NR system used Python 3.9.1 (64-bit), PostgreSQL 13, Django 1.11.26, Beautiful Soap, Natural Language Tool Kit, Selenium WebDriver, Chrome 32-bit driver, and SK-learn. It also used Microsoft Visual C++ 2015 Redistributable (x15) and Microsoft Visual Studio Code.

The SAC system is run on the Google Chrome browser by running the address link "http://127.0.0.1:8000/". Amazon is a worldwide online shopping E-commerce website, which primarily provides many offers to sell or resell various products to its customers. It provides a large-scale shopping platform for electronics, hardware, and software products; books and reading material; fashion and makeup products; groceries; furniture; household things; home essentials, kitchen items; wearables; construction materials, and many more life-based products. The SAC-NR system works on eight product groups. This data is used to form all group and sub-group sets and the review dataset for testing. Here 70 and 30% of groups are kept for the training and testing phases respectively.

Eight categories such as, cell phones, scientific supplies, musical instruments, cameras, televisions, books, kitchenware, and groceries, are included in the SAC-NR system's review data set. G1 through G8 are the names of these groupings, respectively. Each group consists of one subgroup for testing and three subgroups for training. We tested this approach on over 26,000 reviews. The entire number of Amazon training reviews, test reviews, and reviews is shown in Table 1. For training, 7608, 712, 220, 3590, 5398, 196, 650, and 1398 reviews in total were extracted from Amazon. A total of 19772 training reviews that were gathered from the Amazon website were employed by the SAC-NR system. For Amazon, the SAC-NR system gathered 25963 total reviews and 6191 testing reviews.

EXPERIMENTAL RESULTS OF THE PROPOSED SAC-NR SYSTEM

The experimental results of the SAC-NR system are computed for review groups, polarity computation, rating, and review classification. It measures the performance in terms of SYStem PREcision (SYS_PRE), SYStem RECall (SYS_REC), and SYStem F1-measure (SYS_F1). Lastly, it evaluates the performance of the SAC-NR system with the results of the existing systems, and discusses the result findings.

The accuracy and performance of the SAC-NR system are obtained in terms of SYS_PRE, SYS_REC, and SYS_F1-measure. TR_POS, TR_NEG, FAL_POS, and FAL_NEG are the terminology used to calculate all three metrics. The most important performance metric, accuracy, is represented in Eq. (1) as:

$$\text{Accuracy}=(\text{TR\_POS}+\text{TR\_NEG})/(\text{TR\_POS}+\text{FAL\_POS}+\text{FAL\_NEG}+\text{TR\_NEG}) \quad (1)$$

Table 1. Amazon training and testing data.

Group no.	Training review count	Test review count	Final count of reviews
G1	7608	2613	10221
G2	712	223	935
G3	220	70	290
G4	3590	1077	4667
G5	5398	1683	7081
G6	196	59	255
G7	650	195	845
G8	1398	271	1669
Total reviews	19772	6191	25963

Eq. (2) defines SYS_PRE as the proportion of accurately predicted positive observations to the total number of predicted positive observations. Meanwhile, SYS_REC represents the proportion of accurately predicted positive observations to all actual positive observations, as provided by Eq. (3). The weighted average of SYS_PRE and SYS_REC is called the SYS_F1-Score, which is represented in Eq. (4).

SYS_PRE = TR_POS/(TR_POS + FAL_POS) … (2)

SYS_REC=TR_POS/(TR POS + FAL_NEG) … (3)

SYS_F1-Score = 2 * (SYS_REC * SYS_PRE) / (SYS_REC + SYS_PRE) … (4)

Various experiments have been performed on the SAC-NR system using the group-wise testing dataset of Amazon. These results are shown for the group-wise and overall reviews obtained for Amazon products. They include the experimental results on the rating of the products, results of the various polarity types, and the accuracy results in terms of SYS_PRE, SYS_REC, and SYS_F1-measure.

Experimental Results on Review Sets

Table 2 presents the positive, negative, neutral, and total count of reviews for the products within the SAC-NR system. The outcomes of both classifiers are averaged to produce each review set. There were 359 neutral ratings, 1311 bad reviews, and 4521 good reviews for the SAC-NR system. A final total of 6191 reviews were obtained.

Resultant Products with Features

The product's star ratings on Amazon are provided by the SAC-NR system, which compiles all testing review sets. The system's anticipated rating is contrasted with Amazon's actual rating for the products. Two Amazon products' actual and anticipated ratings are shown in Figure 3. These items include the Sony Bravia 108 cm (43 in) Full HD Smart LED TV KDL 43W6603 (Black) (2020_Model) and the Sony Bravia 80 cm (32 in) HD Ready LED TV KLV 32R202G (Dark Brown). Out of five stars, the Sony Bravia 80 cm has an actual rating of 4.4 and a forecasted rating of 4.4. The Sony Bravia 108 cm (43 in) has an actual rating of 4.7 and a forecasted rating of 4.5.

Polarity-Based Results

The system's expected rating for eight classes utilizing NB and RF classifiers for Amazon testing datasets is shown in Table 3. It also displays the rating discrepancies and actual rating for each case. When the actual and anticipated ratings are identical, the zero-rating difference is achieved. There is relatively little difference between the system's anticipated ratings and Amazon's actual ratings. The maximum differences found were –0.20 and +0.25 for the "Scientific Supplies" and "Kitchen Items" groups, respectively.

Table 2. Summary of final counts for positive, negative, and neutral reviews in the Amazon testing data.

Group no.	Positive review count	Negative review count	Neutral review count	Final count of reviews
G1	1853	175	47	2075
G2	329	93	5	427
G3	102	45	3	150
G4	603	240	25	868
G5	487	104	67	658
G6	270	163	82	515
G7	396	179	54	629
G8	481	312	76	869
Total polarity	4521	1311	359	6191

Figure 3. Review rating results for the product "Television" of Amazon.

Table 3. Comparison of product ratings by SAC-NR and Amazon with rating differences.

Group no.	Actual rating			Rating by Amazon	Average rating
	NB	*RF*	*Average rating*		
G1	4.1	4.2	4.15	4.3	0.15
G2	4	4.2	4.1	3.9	–0.2
G3	3.6	3.6	3.6	3.7	0.1
G4	4	4.7	4.35	4.3	–0.05
G5	4.1	4.2	4.15	4.2	0.05
G6	4.4	4.2	4.3	4.5	0.2
G7	3.9	3.8	3.85	4.1	0.25
G8	3.7	3.8	3.75	3.8	0.05
Final total	3.975	4.0875	4.03125	4.1	0.06875

Accuracy Results

The SAC-NR system efficiently obtained its performance results, demonstrating high accuracy with the overall testing dataset. This accuracy was calculated by averaging the results of SYS_PRE, SYS_REC, and SYS_F1-Score. The system's classification results employing both the NB and RF classifiers are shown in Table 4 in terms of SYS_PRE, SYS_REC, and SYS_F1-Score. It also shows the final SYS_PRE, final SYS_REC, final SYS_F1-score, and final accuracy. The SAC-NR system achieved a total of 0.97 SYS_PRE, 0.92 SYS_REC, 0.94 SYS_F1-score, and 0.94 accuracies. The proposed system obtained 97% SYS_PRE, 92% SYS_REC, and 95% SYS_F1-score with NB. It obtained 96% SYS_PRE, 91% SYS_REC, and 93.75% SYS_F1-score using an RF classifier.

Table 4. SYS_PRE, SYS_REC, and SYS_F1-Score results with NB and RF.

Group number	Naïve Bayes			Random Forest			Final SYS_PRE	Final SYS_REC	Final SYS_F1-Score	Total accuracy
	SYS_PRE	*SYS_REC*	*SYS_F1-Score*	*SYS_PRE*	*SYS_REC*	*SYS_F1-Score*				
G1	0.98	0.92	0.92	0.93	0.84	0.88	0.96	0.88	0.90	0.91
G2	0.94	0.90	0.94	0.93	0.91	0.91	0.94	0.91	0.93	0.92
G3	0.95	0.92	0.95	0.92	0.84	0.92	0.94	0.88	0.94	0.92
G4	1.00	0.92	0.87	1.00	0.90	0.94	1.00	0.91	0.91	0.94
G5	0.94	0.94	0.97	0.97	0.91	0.95	0.96	0.93	0.96	0.95
G6	0.97	0.81	0.98	0.97	0.93	0.95	0.97	0.87	0.97	0.94
G7	0.98	0.95	0.97	0.96	0.97	0.95	0.97	0.96	0.96	0.96
G8	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.99	1.00	1.00
Total	0.97	0.92	0.95	0.96	0.91	0.9375	0.97	0.92	0.94	0.94125

PERFORMANCE COMPARISON OF THE SAC-NR WITH OTHER MODELS

Many observations were found from the experimental results and their analysis. It is observed that although existing systems worked upon the sentiment analysis and classification, yet they worked upon a limited number of classification algorithms. The proposed SAC-NR system achieved high scores for the system rating, SYS_PRE, SYS_REC, and SYS_F1-score. The overall accuracy of the system is 94.1% and the best classifier is NB with 97% accuracy. Many observations were found from the experimental results and performance analysis of the SAC-NR system.

- The maximum positive and negative reviews were obtained for the groups “Mobile Phones” and “Cameras”, respectively, on Amazon.
- Additionally, the “Books” category received the highest number of neutral reviews. The best ratings were achieved with the NB classifier, where the NB and RF classifiers achieved ratings of 4.5 and 4.3 for Amazon products, respectively.
- The system’s predicted ratings differed very little from the actual ratings on Amazon. The maximum differences found were –0.20 for “Scientific Supplies” and +0.25 for “Kitchen Items”.
- The SAC-NR system obtained 97% SYS_PRE, 92% SYS_REC, and 95% SYS_F1-score using the NB classifier.
- It obtained 96% SYS_PRE, 91% SYS_REC, and 93.75% SYS_F1-score using the random forest classifier.
- The overall 94.1% system accuracy was achieved, using the NB classifier.

CONCLUSION AND FURTHER SUGGESTIONS

The proposed efficient SAC-NR system classified customer reviews using Naive Bayes (NB) and Random Forest (RF) classifiers. These reviews were sourced from an online shopping website. The SAC-NR system determined the polarity of the reviews, categorizing them as positive, negative, or neutral. The observations of the SAC-NR stated that the customers and users preferred to give positive reviews for different products available on the Amazon website. The detailed survey showed the necessity of the SAC-NR system. This proposed research work can further be extended in multiple dimensions as given below. The SAC-NR system can further be extended with the reviews of other online shopping websites, social media, service-based applications, discussion forums, and other related data domains. The feature dimension and the complexity of the proposed system can further be reduced. The proposed system can further be extended with more reviews and more product categories. More SL or unsupervised ML-based techniques can be applied. It can further be extended to categorize the emoticons, images, and pictures present in the reviews. Another research dimension is to design a generic, integrated, and robust system to handle dynamic, real-time, and embedded applications. The system can be extended by including the ranking and recommendation aspects within it.

Declaration of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

1. Puri S, Kaushik S. An enhanced fuzzy similarity based concept mining model for text classification using feature clustering. In 2012 IEEE Students Conference on Engineering and Systems. 2012 Mar 16; 1–6.
2. Dey UK, Noor A. Comparative Exploration of Prediction Algorithms for Sentiment Analysis Using NLP. In 2019 International Conference on Computer Communication and Informatics (ICCCI). 2019 Jan 23; 1–6.
3. Alrehili A, Albalawi K. Sentiment analysis of customer reviews using ensemble method. In 2019 IEEE International conference on computer and information sciences (ICCIS). 2019 Apr 3; 1–6.
4. Shaheen M, Awan SM, Hussain N, Gondal ZA. Sentiment analysis on mobile phone reviews using supervised learning techniques. Int J Mod Educ Comput Sci. 2019 Jul 1; 10(7): 32–43.
5. Rintyarna BS, Sarno R, Fatichah C. Evaluating the performance of sentence level features and domain sensitive features of product reviews on supervised sentiment analysis tasks. J Big Data. 2019 Dec; 6: 1–9.
6. Jadhav HB, Jadhav AB. Systematic approach towards sentiment analysis in online review's. In Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2018). Cham: Springer International Publishing; 2020; 358–369.
7. Kunaefi A, Aritsugi M. Characterizing user decision based on argumentative reviews. In 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT). 2020 Dec 7; 161–170.
8. Khan S, Chopra K, Sharma P. Brand review prediction using user sentiments: Machine learning algorithm. In 2nd IEEE International Conference on Data, Engineering and Applications (IDEA) 2020 Feb 28; 1–8.
9. Khan A, Younis U, Kundi AS, Asghar MZ, Ullah I, Aslam N, Ahmed I. Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective. In Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC). Vol. 2. Cham: Springer International Publishing; 2020; 23–29.
10. Elfaik H, Nfaoui EH. A comparative evaluation of classification algorithms for sentiment analysis using word embeddings. In Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Vol. 4: Advanced Intelligent Systems for Applied Computing Sciences. Cham: Springer International Publishing; 2020; 1–11.
11. Mandhula T, Pabboju S, Gugulotu N. Predicting the customer's opinion on amazon products using selective memory architecture-based convolutional neural network. J Supercomput. 2020 Aug; 76(8): 5923–47.
12. Mahadevan A, Arock M. Review rating prediction using combined latent topics and associated sentiments: an empirical review. Serv Oriented Comput Appl. 2020 Mar; 14(1): 19–34.
13. Wang B, Wang R, Liu S, Chai Y, Xing S. Aspect-level sentiment analysis of online product reviews based on multi-features. InSemantic Technology: 9th Joint International Conference, JIST 2019, Hangzhou, China, November 25–27, 2019, Revised Selected Papers 9. Singapore: Springer; 2020; 161–169.
14. Srikanth K, Murthy NV, Prasad Reddy PV. Sentiment classification on online retailer reviews. In ICCCE 2020: Proceedings of the 3rd International Conference on Communications and Cyber Physical Engineering. Singapore: Springer Nature Singapore; 2020 Oct 12; 1557–1563.
15. Sindhu C, Rajkakati D, Shelukar C. Context-based sentiment analysis on amazon product customer feedback data. In Artificial Intelligence Techniques for Advanced Computing Applications: Proceedings of ICACT 2020. Singapore: Springer Singapore; 2020 Jul 24; 515–527.

16. Chauhan G, Puri S, Nahta R. Analysis of Government Flagship Programs using Public Feedbacks. In 2023 IEEE 2nd International Conference for Innovation in Technology (INOCON). 2023 Mar 3; 1–7.
17. Meena P, Kumar C, Puri S. Customer segmentation and behavioral systems through influential effective elements: An E-satisfaction analysis using machine learning. In AIP Conf Proc. 2023 Jun 15; 2782(1). AIP Publishing.
18. Ahamad R, Mishra KN. Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach. J Big Data. 2025 Jan 23; 12(1): 11.

