BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 12, No. 1, March 2024
Research Article

Advancing Sentiment Analysis during the Era of Data-Driven Exploration via the Implementation of Machine Learning Principles

Ali A. H. Karah Bash and Ergun Ercelebi

***Abstract*— As a result of the rapid expansion in digital communication, information technology has seamlessly integrated into our everyday existence. It's nearly inconceivable to envision life without the presence of social media. The modern era of communications and networks encompasses not just entertainment instruments, but also contemporary means for users to share crucial data, viewpoints, and concepts. Certain data and information are of such significance that they are vital for analysis and the extraction of essential data that can subsequently be employed in decision support systems. This study examines sentiment polarity analysis through the utilization of the Naive Bayes approach. Naive Bayes is a supervised machine learning model employed for the prediction and analysis of data obtained from external sources. In the training phase, the dataset is categorized into three different groups: small, medium, and large. Additionally, both positive and negative dictionaries are obtained. As for the testing phase, two dataset categories are employed. To gauge the performance of the Naive Bayes algorithm in sentiment analysis, evaluation metrics like accuracy, precision, recall, and the F1 score are utilized. These assessment metrics are computed across three varied categories of positive and negative reviews. The experimental outcomes proved that the Naive Bayes approach is superior and the most effective technique for sentiment analysis.**

Based on the findings, it can be stated that the Naive Bayes classifier delivers a high level of accuracy when analyzing the positive and negative polarity of the data. Additionally, this method requires less time to generate high-quality results.

Index Terms— **Sentiment Analysis, Favorable Polarity, Unfavorable Polarity, Naive Bayes Technique, Machine Learning, Guided Training.**

I. INTRODUCTION

THE sentiment analysis is not a recent domain; rather, its examination and evolution commenced during the early 1990s and persists into the present [1,2]. The ongoing scholarly attention dedicated to this realm underscores its significance within our societal landscape, where sentiment analysis has emerged as a prevalent tool within networks and communication systems. Its application extends to the scrutiny of data, facilitating the extraction of novel insights utilized for analytical and statistical purposes [3].

Recent investigations have extended beyond the boundaries of individual sentiment analysis methodologies, combining machine learning and artificial intelligence to examine a variety of data sources. These investigations delve into the field of predictive analysis, leveraging the subject matter of artificial intelligence and machine learning models to predict outcomes with high accuracy [4].

Furthermore, sentiment analysis techniques are used to unpack the complexities of textual content sourced from communication systems and the broad reach of the Internet. This results in new data containing texts rich in intrinsic informational value, which then becomes raw material for machine learning algorithms [5]. It is worth noting that sentiment analysis should focus on revealing hidden emotional secrets, dividing feelings into positive and negative feelings, and including feelings that extend from anger and joy to sadness, fear, and vitality [6, 7].

Nevertheless, it is important to emphasize that the purpose of employing sentiment analysis techniques is not direct decision-making, but rather data analysis, thereby reinforcing the foundation upon which machine learning methodologies foresee emotional outcomes [8].

The main focal point of this study revolves around the execution of a classification algorithm aimed at discerning positive and negative sentiments within reviews. To fulfil this objective, we employed the Naive Bayes methodology, utilizing the bag-of-words to train the classifier.

Subsequently, the aim of this research is to conduct a series of experiments to meticulously assess the effectiveness of the Naive Bayes approach in identifying positive and negative sentiments within reviews.

Furthermore, an integral aspect of this investigation entails an exploration into the impact of data pre-processing, feature selection, and data curation on the accuracy of the Naive Bayes algorithm.

The organization of this manuscript is outlined as follows: Section 1 introduces the subject matter. Section 2 introduces

Ali A. H. Karah Bash, Department of Electric and Electronic Engineering, Hasan kalyoncu University, Gaziantep 27310, Turkey, (e-mail: ali_karabash2016@yahoo.com).
https://orcid.org/0000-0002-6513-9180

Ergun Ercelebi, Department of Electric and Electronic Engineering, Gaziantep University, Gaziantep 27310, Turkey, (e-mail: ergun.erceleb@gmail.com).
https://orcid.org/0000-0002-4289-7026

Manuscript received August 09, 2023; accepted October 28, 2023.
DOI: 10.17694/bajece.1340321

Copyright © BAJECE

the background of sentiment analysis and the Naive Bayes method. Section 3 details the materials and methodologies employed in the research. Section 4 comprehensively presents the findings and subsequent discourse. Ultimately, Section 5 encapsulates the concluding remarks.

II. Background

The study published in 2021 by Manitosh, employed an ensemble learning technique for sentiment analysis using textual data. However, it is found unsuitable for sentiment classification within the domain of textual sentiment analysis [9].

In a publication from 2021, Ayushi and Sanjukta introduced a rule-based methodology for sentiment analysis, employing Natural Language Processing (NLP) techniques that encompass stemming, tokenization, part-of-speech tagging, and machine learning-based parsing for text mining [10].

In the paper [11], the researchers conducted a meta-analysis concerning the immune system's reaction to perturbations in the presence of artificial intelligence and extensive data frameworks. The study focused on analysing the sentiments expressed by individuals through social media data, while also undertaking a comparative evaluation of diverse machine learning methodologies.

In the referenced study, the scholars employed machine learning techniques to conduct sentiment analysis, operating at both sentence and perspective levels. Their findings revealed variations in the efficacy of the methodologies employed, emphasizing the challenges in drawing definitive conclusions based on the existing state-of-the-art approaches [12].

In a publication from the year 2021, the researchers employed text-mining methodologies for the generation and manipulation of variables. Subsequently, a supervised probabilistic machine learning algorithm was utilized to categorize tweets into positive and negative sentiments. The authors then conducted two distinct experiments to comprehensively assess the effectiveness of their model [13].

The researchers delved into the progression of sentiment analysis, considering the emergence of text processing techniques and the shift from rule-based to statistical text comprehension. This exploration included a comparative analysis of benchmark performance across diverse applications and datasets, utilizing state-of-the-art models [14].

Within the study documented in [15], a comparative assessment was conducted on the efficacy of five distinct machine-learning techniques for sentiment analysis. These techniques encompass Support Vector Machine (SVM), Logistic Regression, Naive Bayes, Random Forest, and K-Nearest Neighbor. The evaluation was carried out utilizing a publicly accessible dataset sourced from kaggle.com.

In the publication referenced as [16], the authors comprehensively addressed the progression of the sentiment analysis workflow. Their inquiry encompassed an exploration into prevalent supervised machine learning methodologies, including multinomial naive Bayes, Bernoulli naive Bayes, logistic regression, support vector machine, random forest, K-nearest neighbour, decision tree, and deep learning techniques.

In a distinct publication denoted as [17], the authors directed their attention toward the scrutiny of online application reviews. Their primary objective revolved around discerning the polarity – positive or negative – of these reviews. To achieve this, the authors commenced their analysis by subjecting the data to preliminary processing, involving both data cleaning and the elimination of stop words.

In prior research endeavours, techniques were employed to enhance comprehension of work content. This involved the amalgamation of rule-based strategies and publicly available machine learning models, empowering analysts to efficiently pinpoint significant elements within extensive document collections [18].

Furthermore, in the study outlined in [19], traditional classifiers and deep neural networks, along with hybrid amalgamations thereof, underwent experimentation. The purpose was to fine-tune pertinent parameters with the aim of attaining optimal classification accuracy on a labelled movie review corpus.

III. Materials and Methods

In contemporary society, access to the Internet has become a basic necessity, performing a myriad of roles ranging from information retrieval to engaging with social media platforms. Prominent companies such as Amazon, Adidas, and Nike are now proactively soliciting consumer input, with the goal of anticipating demands and forecasting upcoming purchasing trends. This sector delves into a carefully designed methodology aimed at extracting valuable consumer insights, a practice that some companies leverage by developing software to discern consumer preferences and thus monetize this data. This section provides a comprehensive explanation of the proposed methodology, its systematic implementation process, and a detailed presentation of the project functions. The overall breakdown of sentiment analysis implementation includes distinct phases including preprocessing, training, and testing.

A. System Prerequisites

Table 1 presents identifying details of the software and hardware used in this work. The software is compatible with a variety of operating systems, including all versions of Windows as well as the PC software platform. The application runs on a personal computer (PC) equipped with a computer software platform, appropriately dimensioned random access memory (RAM), a capable hard disk (HD), and central processing unit (CPU).

B. Dataset Compilation and Pre-processing

User feedback stands as a pivotal metric driving the enhancement of service quality and the refinement of deliverables. Blogs, review platforms, data repositories, and microblogs collectively offer valuable insights into the reception of products and services. Within this study, the dataset pertinent to sentiment analysis encompassing both positive and negative sentiments has been sourced from [20–22]. Moreover, we employed the dataset accessible via the following online link: https://www.cs.uic.edu/~liub/FBS.

This dataset comprises two distinct classes representing (positive and negative) polarities and incorporates expansive dictionaries for both positive and negative sentiments (comprising more than 2000 words each). The primary aim of the pre-processing phase is to meticulously extract the dataset's distinctive features, which subsequently serve as fundamental elements during the training phase.

TABLE I
SOFTWARE AND HARDWARE REQUIREMENTS.

SOFTWARE	Operational system	compatibility includes Windows 7, Windows Vista, and Windows 8
	Programming Language	computer software platform
HARDWARE	CPU	Any Intel processor is suitable.
	RAM	At least 4 GB or higher.
	Processing Speed	Core i5 or higher
	Hard Disk	Requires a minimum of 100 GB or more of storage capacity.

1.1. *pre-processing*

The comprehensive depiction of the entire pre-processing process is presented in Figure 1. The sequence commences with the integration of a dataset containing both affirmative and negative reviews, subsequently reorganized into a singular array. Subsequent stages involve the elimination of special characters and numerical elements through a structured elimination process. These parcels are systematically organized into word structures within distinct classes, with each word being allocated an independent array cell. To exemplify, consider the sentence "I like you," which is organized as depicted in Figure 2.

The pre-processing procedure comprises several key stages, as detailed below:

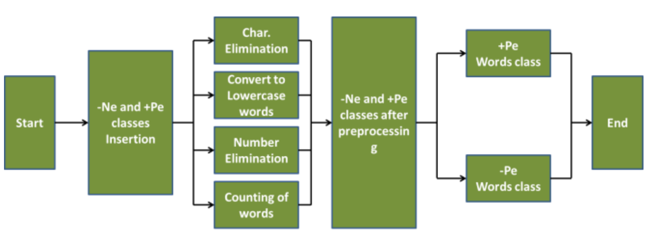

Fig.1. pre-processing steps

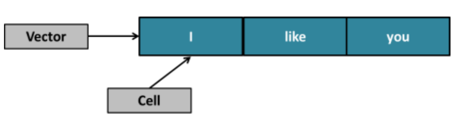

Fig. 2. Sentences representation as vector and cell in this work.

1.1.1 *Removal of Special Characters*

The process of removing special characters involves examining both the positive and negative class datasets. It initiates by scanning for specific characters like (@, #, $, % ^, &, (,), _, <, >, =!) within each word, sentence, and paragraph. Subsequently, these special characters are removed from both the negative and positive dataset classes, as demonstrated in the following example:

Before applying the special characters elimination process: "I @@ like to * e ^^ at && some %$ food &". After applying the special characters elimination process: "I like to eat some food".

1.1.2 *Lowercase Conversion*

The conversion of all letters within the negative and positive classes to lowercase is undertaken upon completion of the special character elimination. This conversion is facilitated by invoking a dedicated computer software platform function, illustrated as follows:

- Before Lowercase Conversion:
  "He like gOing TO The School In thE Afternoon. He is solving H.W."
- After Lowercase Conversion:
  "he like going to the school in the afternoon. he is solving hw."

1.1.3 *Removal of Numerical Characters*

The inclusion of numerical values in the dataset reviews does not convey positive or negative emotions, and it merely adds to the dataset's length without providing any advantages. As a result, it is necessary to remove these numbers to reduce the dataset's size, as illustrated in the following example:

- Before Number Elimination:
  "He is 0 going 77 to 78 the 32 school 84 for studying789."
- After Number Elimination:
  "He is going to the school for studying."

1.1.4 *Word Count*

The word count process involves comparing words from the positive/negative classes within the dataset against those existing in the dictionary. Repeated word occurrences are counted and stored in a new package for testing. Subsequently, probabilities (prior probability, conditional probability, and posterior probability) are computed. The following illustrates the word count procedure:

- Before Word Count:
  "He loves some foods. He loves playing tennis. He likes swimming. He likes me."
- After Word Count:
  [4, 2, 0, 0, 0, 2, 0, 0…]

1.2. *Training Stage*

A comprehensive dataset containing positive and negative words is harnessed to train the system. The system then seeks to create a frequency distribution function, with the aim of identifying the most relevant terms.

Following this, the system proceeds to evaluate the terms to determine the polarity of statements, discerning whether they convey a negative or positive connotation. Terms accumulated from the preceding stages are enlisted for system training.

For this training endeavour, the Naive Bayes approach is adopted, facilitating the system's ability to discern sentiments—whether they are positive or negative—markedly illustrated in Figure 3.

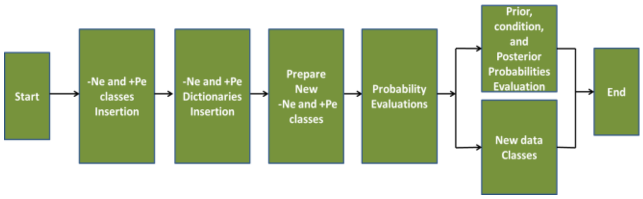

Fig.3. Training stages in this work.

The dataset is partitioned into three distinct categories: small, medium, and large. Moreover, both positive and negative dictionaries are procured, serving as a basis for comparison against the words within the positive and negative classes. This comparison is employed to construct respective positive and negative bags of words.

In the testing phase, a bag of words is utilized, comprising the shared terms between the positive and negative classes within the dataset and dictionaries. Consequently, residual words exclusive to individual classes are omitted from this bag.

1.3. The Proposed Algorithm

Naive Bayes represents a machine learning algorithm suitable for a range of classification tasks, including document classification and sentiment prediction, among others. Its label "Naive" originates from the inherent indirect relationship between the features, where altering the value of one feature does not directly impact the others, a characteristic ingrained in the Naive Bayes algorithm. This method is rooted in Bayes' Theorem [19], underpinning its classification approach.

Notably, the Naive Bayes approach offers simplicity in construction and remarkable efficiency, especially when handling extensive datasets. The underlying formula employed within the Naïve Bayes algorithm is provided below.

$$P_{(C|X)} = \frac{P_{(X|C)}P_{(C)}}{P_{(X)}} \qquad (1)$$

Where $P_{(C|X)}$ is conditional probability, $P_{(X|C)}$ is the class prior probability $P_{(C)}$ and $P_{(X)}$ are predictor prior probability.

Figure 4 illustrates the overarching schematic of the envisaged system. The developed system is implemented using computer software platform and is bifurcated into two principal components: the training segment and the testing segment. Further elaboration of the training component can be garnered through examination of the algorithmic code.

```
Procedure:  Training NB (L, F)
            V ⟵ Extract Vocabulary (F)
            V ⟵ Count All Features (F)
For each label ϵ L do
            Nlabel ⟵ Count Features Belonging To The Label (F)
            priorProb[label] ⟵ N_label/N
            text_label ⟵ Combine Text Of All Features Belonging To The Label (F, label)
 for each f ϵ V do
            F_ilabel ⟵ Count Frequency Of Features (text_label, f)
 for each f ϵ V do
            condProb[f][label] ⟵ F_ilabel / (∑ i′∈V F_i′label)
                  end for
              end for
      end for
Return V, priorProb, condProb
End Procedure
```

Fig.4. The code schematic of the proposed system in this work.

The Bayes algorithm is recognized for its directness and efficiency, consistently yielding precise outcomes. Thus, the probabilistic supervised technique, Naive Bayes, is harnessed to prognosticate positive and negative emotions. This is achieved by computing the probability associated with dataset classes.

Three distinct probabilities (prior, conditional, and posterior) necessitate computation. The prior probability is assessed by incorporating prior knowledge pertaining to positive/negative dataset classes, employing the subsequent formula:

$$\text{Positive prior prob.} = \frac{No\_of\_positive\_reviews}{No\_of\_classes} \qquad (2)$$

$$\text{Negative prior prob.} = \frac{No\_of\_negative\_reviews}{No\_of\_classes} \qquad (3)$$

The conditional probability is construed as the likelihood probability value assigned to each word within the dataset's classes, specifically within the confines of negative and positive classes, expressed as the ensuing formula (1).

The posterior probability entails a connection between the prior and conditional probabilities, and its computation is determined through the subsequent relationship:

$$\text{Posterior prob.} = \text{prior prob. X conditional prob.} \qquad (4)$$

1.4. *Testing Phase*

The process of testing is illustrated in Figure 5. The algorithm initiates by incorporating both positive and negative reviews into the testing segment of the simulation. Subsequently, pre-processing procedures are executed on these reviews, involving a comparison of all words within the reviews with the respective positive and negative word repositories.

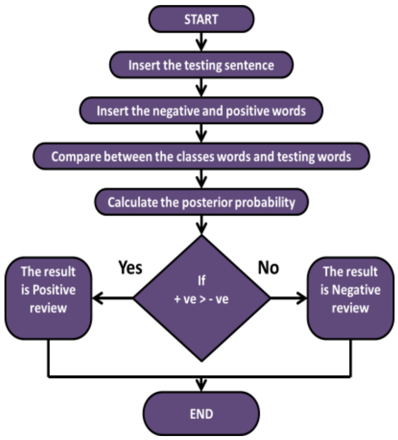

Fig.5. The Testing procedure in this work.

Specific linguistic elements such as negations and adverbs are gathered and employed for juxtaposition with evaluation reviews, thereby augmenting the likelihood of categorizing them as either negative or positive. This augmentation contributes to an overall enhancement in the effectiveness and precision of the process. The resultant lexicon is subsequently harnessed to compute three distinct probabilities: prior, conditional, and posterior.

The procedure encompasses the computation of prior probabilities using Equations 2 and 3, leveraging historical data and their impact on testing reviews. Further, conditional probabilities (defined by Equation 1) are calculated, expressing the likelihood of a variable's value given knowledge of another variable's outcome. These probabilities collectively determine the positive and negative posterior probabilities through the application of Equation 4.

These resultant values are stored within novel packages, utilized for the purpose of making comparative assessments that culminate in the ultimate determination of sentiment polarity (positive/negative). The outcomes of these comparisons find their place within the dataset package, serving as the foundational framework for sentiment classification. Should the positive probability eclipses the negative counterpart, the sentiment is classified as positive; conversely, it is categorized as negative. In instances where the positive and negative probabilities are equal, a neutral sentiment is indicated. The procedural depiction is provided through the algorithmic representation, facilitating the categorization of document sentiment, as depicted in Figure 6.

```
Procedure:   Testing NB (L, V, priorProb, condProb)
             W ⟵ Extract Features (V)
for each label ϵ L do
             Score_label ⟵ LogpriorProb[label]
for each  f ϵ W do
             Score_label += logcondProb [f][label]
                   end for
             end for
Return argmaxlabel ϵ L score[label]
End Procedure
```

Fig.6. Code algorithm of testing procedure in this work.

IV. Result and Discussion

This section provides an elaborate account of the comprehensive experimental outcomes derived from the implementation of the Naive Bayes algorithm. The experiments encompass three distinct datasets of reviews, comprising 50, 100, and 200 reviews, respectively. Notably, each dataset encompasses an equal distribution of positive and negative reviews.

A. *Performance Metric*

The evaluation metrics commonly employed to assess the efficacy of classification algorithms encompass precision metrics such as accuracy, precision, recall, and the F1 score. Moreover, it is imperative to consider computational resource costs, a crucial aspect often taken into account during classifier construction.

Table 2 elucidates the metrics utilized for calculating precision, encompassing accuracy, and the F1 score. The confusion matrix presents the actual and predicted label distributions, with each row indicating the predicted label and each column denoting the actual label of the sentence. The True Positive (TP) represents the count of sentences that are correctly classified as positive (both actual and predicted labels are positive). The True Negative (TN) signifies the count of sentences that are accurately classified as negative (both actual and predicted labels are negative).

False Positive (FP) indicates the quantity of sentences falsely predicted as positive, when in reality they are negative. Conversely, False Negative (FN) quantifies the number of sentences falsely predicted as negative, while they are actually positive.

TABLE 2
CONFUSION MATRIX FOR THE BINARY CLASSIFIER.

			Actual
		Positive	**Negative**
Estimated	**Positive**	**TP**	**FP**
	Negative	**FN**	**TN**

The accuracy formula, as represented by Equation 5, showcases the ratio of correctly predicted answers, encompassing both True Positives (TP) and True Negatives (TN), in relation to the total number of answers.

In addition, precision and recall are utilized as alternative performance metrics. Precision, captured by Equation 6, quantifies the count of accurately classified positive answers derived from the classifier.

$$Accuracy = \frac{(T_N + T_P)}{(T_N + T_P + F_N + F_P)} \quad (5)$$

$$\Pr ecision = \frac{T_P}{(F_P + T_P)} \quad (6)$$

Recall quantifies a classifier's proficiency in correctly identifying the potential positive answers within the anticipated responses. This computation is expressed mathematically through Equation 7.

While a higher precision value corresponds to improved recall, attaining elevated levels of both precision and recall concurrently in practical contexts is exceedingly challenging. Thus, striking a balance between these two metrics becomes imperative. The F1 score, constituting the harmonic mean of precision and recall, offers a means to assess this equilibrium. Mathematically, the F1 score is determined as follows:

$$\mathrm{Re}\,call = \frac{T_P}{(F_N + T_P)} \tag{7}$$

$$F_1 = \frac{2x\,\mathrm{Re}\,callx\,\mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \tag{8}$$

B. Dataset and Pre-processing

In this study, the dataset utilized is sourced from [23]. This dataset encompasses distinct classes of positive and negative polarities and is presented as a unified package, devoid of partitioning into separate positive and negative subsets [23–25].

The preparation of this dataset entails several stages. The initial phase involves the incorporation of both positive and negative classes into the dataset. Subsequently, these positive and negative classes are consolidated within a single array, a process constituting the second step.

In the third step, the dataset undergoes a treatment wherein special characters and numerical values, such as (#, $, %, &, ^, *, (,), -,:, 9, 5, 654), are removed. Following this, the array is organized such that each word within the classes occupies an individual cell, denoting the fourth step. The final step is executed to eliminate any vacant cells present within the array [26].

For the development of the network, a set of 200 negative and 200 positive reviews was employed as input data. The negative word classes exhibit a minimum size of 204 and a maximum size of 1644, whereas the positive word classes range from a minimum of 219 to a maximum of 1975 words. The initial phase of the project involved importing these classes into computer software platform and subsequently initiating pre-processing on both positive and negative categories. The cumulative word count across all classes is tabulated in Table 3. Furthermore, the count of negative words, both with and without repetition, is provided in the same table. Similarly, the count of positive words, considering both repetition and non-repetition, is also pivotal for sentiment analysis and is outlined in Table 3.

TABLE 3
THE TERMS WITHIN THE CATEGORIES, BOTH THE NEGATIVE AND POSITIVE WORDS CONSIDERING REPETITION AND NON-REPETITION.

Total number of words in the classes	Negative words		positive words	
	With repetition	Without repetition	With repetition	Without repetition
163322	78737	9090	84585	9470

C. Training Stage

To establish the proficiency of the system in discerning positive and negative emotions, the Naive Bayes method was iterated ten times during the learning process. In this endeavor, the word dictionaries for positive and negative sentiment encompassed a total of 3004 words, with 1502 words allocated to each polarity. Prior probability calculations, achieved through Equations 2 and 3, were performed for all instances of negative and positive polarities. The outcomes of these calculations are meticulously documented in Table 4. Similarly, posterior and conditional probabilities, computed using equations (4 and 1) respectively, were determined for all instances of negative and positive polarity, with detailed results featured in Table 4.

TABLE 4
ASSESSMENT OF PRIOR, CONDITIONAL, AND POSTERIOR PROBABILITIES.

	Negative Polarity	Positive Polarity
prior probability value	0.500	0.500
conditional probabilities value	0.500	6.11688e-05
posterior probabilities value	0.0003368	3.05844e-05

D. Testing Stage

In the testing phase, testing reviews are initially input as one or more paragraphs. Subsequently, these reviews undergo preprocessing, involving the removal of special characters and numbers, followed by the conversion of all words to lowercase. The processed reviews are then organized into sentences within a testing array through sentence segmentation. A comparison of these sentences with the bag of words is performed, and outcomes are stored in the testing database. Binary values (zeros and ones) are assigned based on word matches between the testing reviews and the bag of words. To evaluate the efficiency of the proposed method, experiments are conducted three times with 50 and 100 reviews.

For the testing results involving 50 reviews, the system's performance is analyzed using a smaller dataset comprising 25 positive and 25 negative reviews. The first dictionary group correctly identifies 75 positive and 80 negative reviews out of 100 each. Subsequent dictionary groups also achieve progressively better results, culminating in 90 positive and 89 negative reviews correctly identified. These outcomes are summarized in Table 5.

Similarly, testing results are conducted for a dataset of 100 reviews (50 positive and 50 negative). Employing the first dictionary group, 90 positive and 82 negative reviews are correctly identified. Successive dictionary groups yield improved results, with the final group correctly identifying 95 positive and 92 negative reviews. These results are presented in Table 5.

TABLE 5
RESULT OF TESTING STAGE FOR 50 AND 100 REVIEWS

Total no. of testing reviews for positive and negative	Dictionary size	No. of positive reviews correctly identified in entirety	No. of negative reviews correctly identified in entirety.
50 reviews	1000	75	80
	2000	80	83
	3000	80	84
	4000	90	89
100 reviews	1000	90	82
	2000	89	85
	3000	93	90
	4000	95	92

E. Evaluations of Accuracy, Precision, Recall and F1

The evaluation of accuracy entails assessing the closeness between calculated values and their true counterparts, while the percent error quantifies the error ratio relative to the actual value, presented as a scaled factor of 100. This measure contributes to appraising work efficiency and enables the comparison of results in experimental contexts.

The assessment of accuracy spans three distinct review groups, engaged in both training and testing phases, illuminating the impact of varying review quantities on sentiment analysis. Additionally, diverse dictionary sizes are investigated to showcase the effectiveness and efficiency of the Naive Bayes algorithm in sentiment analysis as depicted in Figure 7.

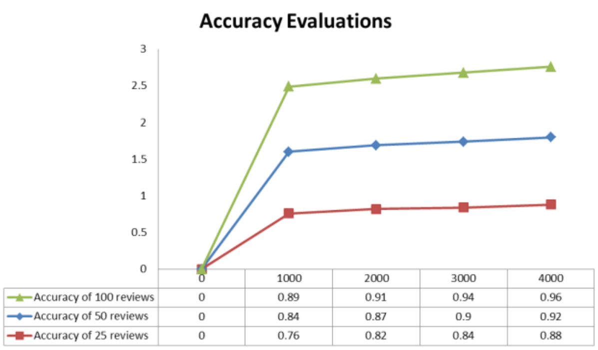

Fig. 7. Accuracy evaluation for 25, 50, and 100 reviews.

Precision, a metric gauging measurement consistency, is determined through the calculation of the standard deviation within a dataset. It provides insight into the precision of positive or negative responses within reviews of the corresponding polarity. The maximum precision value signifies the minimal occurrence of false estimations as depicted in Figure 8.

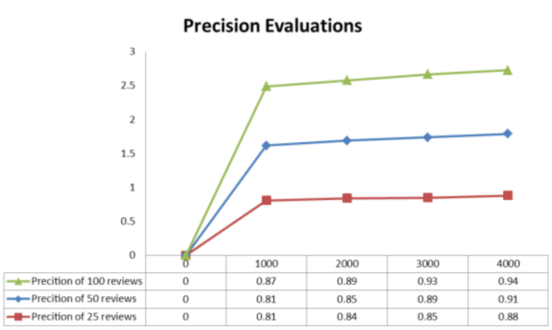

Fig. 8. Precision evaluation for 25, 50, and 100 reviews.

Figure 9 illustrates the Recall values corresponding to review quantities of 25, 50, and 100. Recall, also known as sensitivity, reveals the ratio of correctly predicted positive or negative reviews to the total reviews within the original dataset. This metric underscores the Naive Bayes method's capability to anticipate the maximum achievable count of positive or negative reviews within the anticipated set.

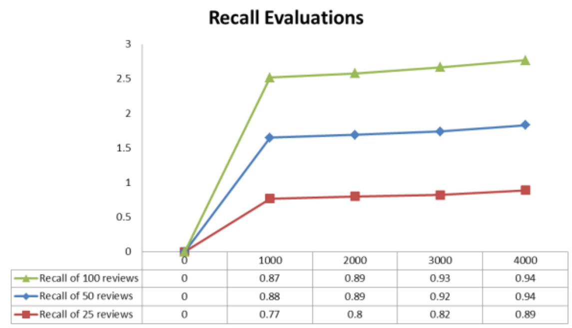

Fig. 9. Recall evaluation for 25, 50, and 100 reviews.

Recognizing the inherent difficulty of simultaneously optimizing recall and precision within real-world contexts, the F1 score is computed to achieve equilibrium between these two metrics. Functioning as a combination of precision and recall, the F1 score accommodates both false positive and false negative reviews, as illustrated in Figure 10.

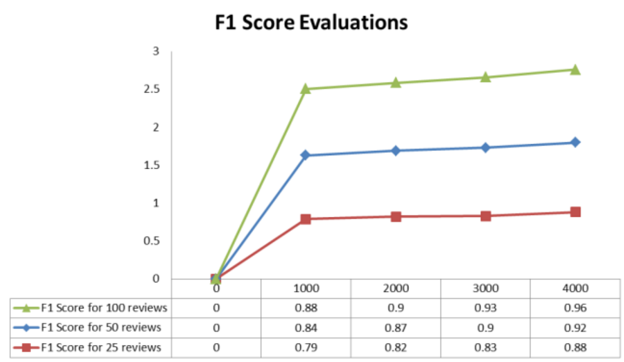

Fig. 10. F1 Score evaluation for 25, 50, and 100 reviews.

As Table 4 reveals, the visual representations elucidate that smaller review sets and a limited dictionary vocabulary result in diminished accuracy values, whereas larger review sets and a comprehensive dictionary lead to enhanced accuracy (approximately ranging from 76% to 96%). The highest level of accuracy is achieved with 4000 informative dictionary words and 100 training reviews. This trend is mirrored in the precision figure, wherein smaller review sets and datasets yield an 81% precision rate, which progressively rises to around 94% with augmented review quantities and expanded dictionary sizes. The pinnacle precision is realized with a dictionary comprising 4000 words and 100 training reviews. Additionally, heightened review quantities and the inclusion of informative dictionary words correspondingly contribute to elevated recall and F1 scores in the figures. With 4000 informative dictionary words and 100 training reviews, recall reaches approximately 94%, and the F1 score approaches 96%.

F. Analysis of Time

The analysis of time, as succinctly presented in Table 6, offers valuable insights into the temporal aspects of the training phase within this study. Particularly noteworthy is the observation that a reduced volume of reviews coupled with a diminished count of informative words in the dictionary or bag led to a modest requirement of approximately 50 minutes for the Naive Bayes algorithm to complete its training. Conversely, an elevation in the count of informative words contained within the dictionary engendered a commensurate augmentation in the duration of the training process. To illustrate, the utilization of a dictionary encompassing four thousand words necessitated an extended training period of roughly one and a half hours for the proposed systems. In a similar vein, when contending with a corpus of 100 reviews and a quota of 1000 informative words held within the bag, the proposed classifier expended a comparable time span of approximately one and a half hours; this temporal demand increased proportionally as the word count within the bag expanded.

Of notable significance is the circumstance where the management of a collection comprising 200 reviews, interwoven with a dictionary harboring 4000 informative words, prompted the proposed Naive Bayes classifier to commandeer a training interval of 5 hours and 10 minutes. This underscores the discernible reality that while an augmentation in the number of reviews and a proliferation of dictionary size can bestow a boon upon enhanced accuracy and precision, this marked improvement is coupled with the trade-off of heightened temporal investment. Hence, the imperative of striking equilibrium between optimal performance and judicious time efficiency takes center stage as a pivotal deliberation.

TABLE 6
ANALYSIS OF THE TIME IN THIS WORK IN DIFFERENT DICTIONARIES.

Reviews No.	Dictionary word			
	1000	2000	3000	4000
25	33 min	40 min	50 min	1 h
50	1h.30min	2h.12min	2h.55min	3h.15min
100	3h.25min	3h.50min	4h.20min	5h.20min

In the comparison between the Naive Bayes method employed in this article and the Support Vector Machine (SVM) method as utilized in reference [1], it was observed that the method used in this study demonstrates high efficiency even when utilizing relatively small positive and negative dictionaries. Furthermore, when comparing the method applied in Reference [27], it was observed that the approach employed in our article operates with high efficiency and expeditiously performs sentiment analysis.

V. CONCLUSION

Due to its wide range of applications, sentiment analysis remains a highly active research area for numerous scholars. In our study, we employed the Naive Bayes approach for sentiment analysis across various dataset categories because of is a simple and effective method. The system classifies and determines the polarity of the text into positive and negative classes. This classifier utilizes a dataset that is integrated into a machine-learning framework. The experimental findings demonstrate that the Naive Bayes classifier yields highly satisfactory results. In this work, the Naive Bayes classifier underwent training and testing with three various numbers of review sets and dictionary sizes. In the first scenario, an accuracy of 0.88 % was attained using 50 training reviews and a 4000-word dictionary. In the second scenario, where 100 training reviews and the same 4000-word dictionary were used, the accuracy increased to 0.92 %. In the third scenario, with 200 training reviews and the same 4000-word dictionary, an accuracy of 0.96 % was achieved. Furthermore, this approach is also characterized by its efficiency in reaching decisions within a shorter timeframe.

ACKNOWLEDGMENT

The authors express their gratitude to the science lab personnel within the Department of Electrical and Electronic Engineering for their valuable assistance and support throughout the duration of the study.

REFERENCES

[1] R. Singh and R. Singh, "Applications of sentiment analysis and machine learning techniques in disease outbreak prediction–A review," Mater. Today Proc., vol. 81, pp. 1006–1011, 2023.
[2] G. Gupta et al., "DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms," Diagnostics, vol. 13, no. 6, p. 1093, 2023.
[3] E. M. Mercha and H. Benbrahim, "Machine learning and deep learning for sentiment analysis across languages: A survey," Neurocomputing, vol. 531, pp. 195–216, 2023.
[4] H. Rahman, J. Tariq, M. A. Masood, A. F. Subahi, O. I. Khalaf, and Y. Alotaibi, "Multi-tier sentiment analysis of social media text using supervised machine learning," Comput. Mater. Contin, vol. 74, pp. 5527–5543, 2023.
[5] M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, "Machine learning sentiment analysis, COVID-19 news and stock market reactions," Res. Int. Bus. Financ., vol. 64, p. 101881, 2023.
[6] Y. Wang et al., "Supervised Gradual Machine Learning for Aspect-Term Sentiment Analysis," Trans. Assoc. Comput. Linguist., vol. 11, pp. 723–739, 2023.
[7] M. A. Jassim, D. H. Abd, and M. N. Omri, "A survey of sentiment analysis from film critics based on machine learning, lexicon and hybridization," Neural Comput. Appl., vol. 35, no. 13, pp. 9437–9461, 2023.
[8] S. Ghosh et al., "'Do we like this, or do we like like this?': Reflections on a Human-Centered Machine Learning Approach to Sentiment Analysis," in International Conference on Human-Computer Interaction, Springer, 2023, pp. 63–82.
[9] Y. N. Prajapati, U. Sesadri, T. R. Mahesh, S. Shreyanth, A. Oberoi, and K. P. Jayant, "Machine Learning Algorithms in Big Data Analytics for Social Media Data Based Sentimental Analysis," Int. J. Intell. Syst. Appl. Eng., vol. 10, no. 2s, pp. 264–267, 2022.
[10] A. Mitra and S. Mohanty, "Sentiment Analysis Using Machine Learning Approaches," in Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2, Springer, 2021, pp. 63–68.
[11] S. Qaiser, N. Yusoff, M. A. Remli, and H. K. Adli, "A comparison of machine learning techniques for sentiment analysis," Turkish J. Comput. Math. Educ., 2021.
[12] J. Srivastava and N. Katiyar, "Machine Learning Technique for Target-Based Sentiment Analysis," in Micro-Electronics and Telecommunication Engineering, Springer, 2021, pp. 161–169.
[13] M. Hayouni and S. Baccar, "Sentiment analysis using machine learning algorithms," 2021.
[14] A. A. Ansari, "Evolution of Sentiment Analysis: Methodologies and Paradigms," Trends Data Sci. Appl. Theory Pract., pp. 147–174, 2021.
[15] M. H. I. Ahmad Hapez, N. L. Adam, and Z. Ibrahim, "Performance analysis of machine learning techniques for sentiment analysis," in Advances in Visual Informatics: 7th International Visual Informatics

Conference, IVIC 2021, Kajang, Malaysia, November 23–25, 2021, Proceedings 7, Springer, 2021, pp. 205–213.

[16] V. Umarani, A. Julian, and J. Deepa, "Sentiment analysis using various machine learning and deep learning Techniques," J. Niger. Soc. Phys. Sci., pp. 385–394, 2021.

[17] A. Bhardwaj and P. Srivastava, "A machine learning approach to sentiment analysis on web based feedback," in Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020, Springer, 2021, pp. 127–139.

[18] I. I. Mestric, A. Kok, G. Valiyev, M. Street, and P. Lenk, "Aspect Level Sentiment Analysis Methods Applied to Text in Formal Military Reports," Inf. Secur., vol. 46, no. 3, pp. 227–238, 2020.

[19] A. Mehta, Y. Parekh, and S. Karamchandani, "Performance evaluation of machine learning and deep learning techniques for sentiment analysis," in Information Systems Design and Intelligent Applications: Proceedings of Fourth International Conference INDIA 2017, Springer, 2018, pp. 463–471.

[20] O. Iparraguirre-Villanueva et al., "The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities using the CNN-LSTM model," Vaccines, vol. 11, no. 2, p. 312, 2023.

[21] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and As. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in International conference on advances in ICT for emerging regions (ICTer2012), IEEE, 2012, pp. 182–188.

[22] F. Hallsmar and J. Palm, "Multi-class sentiment classification on twitter using an emoji training heuristic." 2016.

[23] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and As. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in International conference on advances in ICT for emerging regions (ICTer2012), IEEE, 2012, pp. 182–188.

[24] Z. Ke, H. Xu, and B. Liu, "Adapting BERT for continual learning of a sequence of aspect sentiment classification tasks," arXiv Prepr. arXiv2112.03271, 2021.

[25] O. Iparraguirre-Villanueva et al., "The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities using the CNN-LSTM

[26] F. Hallsmar and J. Palm, "Multi-class sentiment classification on twitter using an emoji training heuristic." 2016.

[27] Sharma, D., & Sabharwal, M, "Sentiment analysis for social media using SVM classifier of machine learning. " Int J Innov Technol Exploring Eng (IJITEE), 2019, 8(9), 39-47.

BIOGRAPHIES

Ali A. H. Karah Bash achieved his B.S. degree in Computer and Information Engineering from University of Mosul, situated in Mosul, Iraq, in 2006. Subsequently, he embarked on a scholarly trajectory, attaining both Master's and Doctoral degrees in Electrical and Electronics Engineering from Gaziantep University, Turkey, in 2014 and 2022, correspondingly. His scholarly interests revolve around embedded systems, signal and image processing, as well as artificial neural networks. Importantly, he is associated with the ORCID ID (0000-0002-6513-9180).

Ergun ERÇELEBI attained his B.S. degree in Electrical and Electronics Engineering from METU, located in Gaziantep, Turkey, in 1990. Following this, he pursued his academic journey and secured both M.S. and Ph.D. degrees in Electrical and Electronics Engineering from Gaziantep University, Turkey, in 1992 and 1999, respectively. From 2011 onward, he has held the esteemed position of Professor within the Electrical and Electronics Engineering Department at the University of Gaziantep in Turkey. Subsequently, in 2014, he assumed the role of department head. His academic pursuits center around embedded systems, signal processing, and artificial neural networks. Notably, his ORCID ID is (0000-0002-4289-7026).