International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Evaluating Machine Learning and Deep Learning for Sentiment Analysis of Customer Feedback

Numan Ali[1], Sadia Ramzan[1], Tasawar Ali[3*] and Muhammad Usman[3]

[1,3]Department of Computer Science, University of Agriculture Faisalabad, Pakistan
[1]Department of Computer Science, Government College University Faisalabad, Pakistan
[3]Department of Computer Science, Barani Institute of Sciences, Pakistan

Received 10 Nov 2024, Accepted 02 Dec 2024, Available online 03 Dec 2024, **Vol.14, No.6 (Nov/Dec 2024)**

Abstract

In recent years, customers have increasingly provided essential feedback, opinions, and recommendations for internet retailers. This article aims to develop an automated comment analyzer. We present an automated solution for analyzing and classifying customer comments derived from Amazon data domains, capable of managing a substantial volume of reviews. Supervised learning classifiers, specifically Naive Bayes (NB), Support Vector Machine (SVM), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM), are employed to categorize comments as positive or negative. This study utilizes three variations of Naive Bayes models, including Support Vector Machine, Multinomial Naive Bayes, and Complement Naive Bayes, for sentiment analysis of e-commerce reviews. The system is tested and evaluated using real-time data, including product reviews from Amazon's website, specifically analyzing 10,000 customer reviews spanning various items. Data preprocessing techniques, such as lowercase processing, stop word removal, punctuation removal, and tokenization, enhance the usability of the collected data for analysis. The models were trained on this cleaned dataset to identify and classify customer sentiment as positive or negative. The machine learning algorithms CNB, MNB, BNB, and SVM achieved accuracies of 80.00%, 79.90%, 79.35%, and 81.25%, respectively, while the deep learning algorithms GRU and LSTM obtained accuracies of 80.6097% and 76.2619%, respectively. Although the SVM model demonstrated greater accuracy than the deep learning models, it exhibited significantly slower execution times. Our findings indicate that deep learning approaches yield superior results for categorizing consumer attitudes toward products.

Keywords: *Sentiment Analysis; Machine Learning; Customer Feedback; E-commerce Reviews; Deep Learning Models*

Introduction

Web 3.0 enables individuals to express and exchange their views on current events through social media, driven by its core features such as the semantic web, artificial intelligence, and improved connectivity. Opinion mining is essential for analyzing reviews and discussions. As a result, companies are increasingly utilizing this information to improve their products' quality and performance, allowing them to stay competitive in a challenging market [1].The Internet generates much data, yet critical information gets buried in the avalanche. Text mining, computational linguistics, and natural language processing are all used for "sentiment analysis"[2].

To accurately analyze emotion, it is necessary to consider morphological negativity development. Sentiment analysis and many text-processing applications require automated negation detection in news articles. Here, we used sentiment analysis to examine the impact of user reviews on a product's selection. We have shown that sentiment analysis works well for this purpose[3].As smartphone usage has skyrocketed in recent years, so has the number of individuals who use social networking sites like Instagram, Twitter, and Facebook. Scholars have recently found the structural and semantic properties of the material as new techniques to account for them. Computational approaches are used in this study to identify document-level negation [4].

For sentiment analysis, various studies utilize Twitter data in real-time to uncover patterns using the Twitter streaming API. Positive and negative ratings are used in sentiment analysis to categorize people's thoughts. Twitter streaming API was used to collect data on Indonesia's presidential elections. Analysis of the structure's correlation with vote results was devised to forecast European election results [5].Artificial Intelligence (AI) and Machine learning (ML) have been extensively used to analyze the mood of tweets. Fake positive or negative reviews may be

*Corresponding author's ORCID ID: 0009-0008-5743-7517
DOI: https://doi.org/10.14741/ijcet/v.14.6.5

detected using deep neural networks and convolution models trained on an Amazon dataset[6].

The use of advanced machine learning algorithms is on the rise. According to the findings, CNN and RNN perform better when tested against a single dataset in a given location. Based on AdaBoost's mix of CNN models for sentiment analysis in the user-generated text, deep learning models could overcome the issue of brief messages [7].With machine learning, certain old methods rely on the language in which they are applied. They were able to get an accuracy rate of 82.9% by utilizing SVM with unigrams. Sentiment classifiers often employ NLP to extract information from text. A bag-of-words technique is also a frequent NLP strategy; however, most NLP strategies are based on n-grams [8].

Deep learning models were proven accurate in detecting sentiment in various settings.

In the evaluations, in the e-commerce business, the availability of false reviews that urge customers to buy products they don't desire is the largest difficulty with sentiment analysis. A hybrid neural system is an example of a hierarchical bidirectional RNN[9].The following are the primary contribution of this research:

- A lexicon-based method is used for each product evaluation to create a sentiment score.
- We have labeled the review texts as positive or negative when the computed sentiment score is 2 or 1.
- Combining all product reviews into one data frame may gather more sentiment-related phrases.
- A deep learning model GRU + LSTM will increase accuracy for classifying product-related sentiment.
- Comparison the NB + SVM and GRU + LSTM models for classification performance.

Literature Review

The Sentiment Analysis (SA) Application

To do sentiment analysis, one must look at how a writer approaches a certain issue or the overall polarity of a piece. Texts are categorized by their attitude or viewpoint, not their subject matter. Data mining and knowledge management methods, including sentiment analysis, data mining, natural language processing (NLP), and information retrieval, are all used in sentiment analysis. Sentiment analysis is a sophisticated method with five stages for assessing the sentiment in source materials. This process has four stages: collecting data, preparing the text, and detecting and classifying sentiment. Unsupervised learning and supervised learning both use sentiment analysis as a tool. Text-based patterns may be created by sorting the training data into categories. This unsupervised learning method does not use a database; instead of relying on a list of words that includes negative and positive phrases. Because of this, the document may be labeled depending on the frequency with which negative and positive phrases appear.

In a variety of disciplines, sentiment analysis is applied. The government uses sentiment analysis to understand public perceptions of various issues better. For turning disgruntled consumers into advocates, sentiment analysis is used in online commerce to analyze their shopping experience and thoughts about product quality. Customers' feedback and opinions about goods and services may be assessed using sentiment analysis. Tweetfeel is a standout example of a real-time tweet analysis program. Blogger-centric contextual advertising, which focuses on creating personalized adverts on blogs based on the interests of the businesses, uses sentiment analysis. As a result of these discoveries, sentiment analysis is frequently used in various disciplines to detect and analyze certain behavioral patterns and sentiments.

In-Text Classification Deep Learning (DL) Approaches

Machine learning algorithms have fallen out of favor in favor of deep learning alternatives. For text categorization, deep learning algorithms get the most trustworthy results. Nonlinear and complicated data interactions are largely responsible for their success. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs) are the three primary deep learning algorithms used to categorize text and documents.

In artificial intelligence, RNNs may be used to forecast journal citation counts. The investigation used bidirectional LSTM on paper metadata text to explicitly forecast the citation count. In estimating the number of times, a publication will be cited; the research does an excellent job.

Text categorization using a deep graph-LSTM algorithm. The results of the experiment were confirmed in court proceedings in India. The research accurately identified a new instance into a related category with 99% accuracy.

The Kalman filter decreases data noise and errors in an accelerated gradient LSTM model. The research was used to anticipate the stock market using Twitter and Yahoo data. The Kalman filter improved the model's performance, obtaining an accuracy rate of 90.42%.

A neural network-based deep learning algorithm was utilized to identify COVID-19 contaminated areas. The algorithm analyzed tweets from the United Kingdom and the United States. Using bidirectional LSTM improves geolocation accuracy, according to the trial.

Datasets and Knowledge-Based

Movies and Twitter use different approaches when it comes to corpora. Movie review sentences were employed as a corpus for the movie domain, while tweets were used for the Twitter domain. Customer evaluations from e-commerce and internet platforms like WebKB, industrial sector, newsgroups, and Yahoo

dataset are the most frequent sources for Sentiment Analysis[10]. These text or blog analyses are most likely to point to a single structure. As a result, other researchers used random news items from the Giga word database to train the target sentiment classifier. Research on educational and movie review systems has improved thanks to previous studies that published over 1000 publications in various fields, including education, movies, and the home environment. The researchers also produced a large-scale hotel review dataset withnegative and positive ratings.

The NN models (LSTM-GRNN and Conv-GRNN) for sentiment classification at the document level[11]. Document-level sentiment analysis was achieved on various IMDB, and Yelp Dataset Challenges datasets, including several large review datasets[12].

Video games, Amazon product evaluations, smartphone product reviews, blogs and tweets are just a few industries that have experienced system improvements in recent years. Unrestricted access to a variety of datasets.

- ChnSentiCorp-Hou, ChnSentiCorpMov, Chn-SentiCorpEdu,and ChnSentiCorp[13]
- Amazon review dataset[14]
- Dataset of cross-language sentiment classification[15]
- Amazon reviews, Yahoo answers, IMDB reviews, and Yelp reviews [16]
- IMDB, Yelp 2014, and Yelp 2013 Datasets[17]
- Yelp 2013, and IMDB datasets [18]
- Yelp 2014 and IMDB datasets[19]
- IMDB, Amazon and RCV1[20]
- IMDB Dataset[21]
- The information is accessible from the website https://www.cs.jhu.edu/[22]
- IMDB and Stanford sentiment treebank dataset and dataset[23]
- IMDB Dataset, (http://ww17.ripadvisor.com/) and Yatra (https://www.yatra.com/) Dataset[13]
- Yelp 2015, and IMDB datasets[24]
- French Articles Dataset[25]
- IMDB and Yelp Dataset Challenge[12]
- Movie Dataset Standard[26]
- Stanford Sentiment Treebank (SSTB)[27]
- Movie Review Dataset. (Moraes et al., 2013).

Sentiment Analysis of Different Methods

Text Preprocessing, Stemming, and other NLP techniques rely on Sentiment Analysis. Multiple approaches are used to determine a text's emotion. This article examines the relationship between Amazon product reviews and the ratings provided by consumers. LSTM with Word2Vec provides the most accurate results[28]. The Internet has become the primary place for people to express their views on products and administrations, as well as on social concerns and the executive plan. By using machine learning, cross-domain sentiment classification is the goal of this study. If successful, it will be a major step forward in solving domain-dependent tasks[29]. Deep learning and natural language processing produced highly accurate sentiment estimates for 12 different categories of Amazon user reviews. There is a high degree of generalizability to predictions made within and across categories. Without the need for additional approaches, deep learning and conformal prediction can correct class imbalances [30]. The polarity of Amazon and Flipkart customer comments may be better understood using a system for automatically assessing and categorizing these comments. Different lexicons and supervised algorithms were used to classify assessment procedures.Mobile phone, amazon online source, positive/negative sentiment, and feature extraction were the most effective uses of the existing algorithms for many reviews [31].The polarity of Amazon and Flipkart customer comments may be better understood using a system for automatically assessing and categorizing these comments[32].Using Business Intelligence to help firms streamline their operations and increase customer satisfaction. Online purchasing, particularly electrical items, has seen a significant uptick in the last few years. Use these evaluations to assist customers in making an informed purchase and help firms better understand how consumers received their products[33].Several deep learning algorithms are being evaluated by academics using Amazon.com reviews. It was possible to create and test four different types of RNN: LRNN, GLRNN, GRNN, and URN. The LRNN algorithm has the best accuracy of 88.39% on the balanced dataset [34].The field of Sentiment Analysis (SA) is one of the most rapid and active expanding in academia today. Amazon is an example of an online shop that enables consumers to rate and review its items. Various Amazon product review categories have been studied to determine the best machine learning classification approach [35].Textual data generated by online company websites include user evaluations, comments, recommendations, and messages. Sentiment analysis is a popular method for analyzing text data and extracting sentiment from it. 90% of customers are exploring several internet channels to assess the quality of their purchase[1].This studyinvestigates sentiment categorization using several machine learning algorithms. Various metrics were assessed, including cross-entropy loss function, recall, precision, and accuracy. The top-performing model was picked and retrained for binary classification by the author[36].Vaccination-related tweets were automatically categorized as either anti-, pro-, or neutral by machine-learning algorithms. 60% of tweets were classified as supporting, 23% vaccinated, and 17% vaccinated. Vaccine-related occurrences impact the number and polarity of tweets. [37].Snippext is a language model constructed using semester-supervised learning. Using a two-pronged SOTA strategy, Snipext has made significant strides in improving its labeled training data. Snippext performs

similarly to past SOTA findings with about half the required training data; experiments show [38].

Summary

In sentiment analysis, machine learning algorithms analyze how artificially intelligent computer systems conclude the text. The polarity may be positive and negative, or it might be completely neutral. Customers' judgments and utterances about a product or service indicate their sentiments and attitudes about it.

Methodology and Data Collection

Overview

The examination of online social networks has been the subject of some research. It is possible to categorize them in three distinct ways: geometric, statistical, and topological. In the past, most analysis systems employed detection, extraction, selection, and classification as their primary analysis processes. They can correctly identify the analysis or visualization of Online Social Networks (OSNs) by utilizing the mutual exchange of methods and methodologies. Python, R-Studio, MATLAB, and Weka are used in social network analysis to generate various graphs using various methods and methodologies. A summary of social network analysis using graphs is also provided in this study. First, a sampling of social networking websites, such as Amazon, will be examined. Different tests on social networks like Amazon will outline the common aspects of social network sites. By using visualization tools, it is possible to discover many connections and qualities of social network participants. Visualization and analysis of network graphs may be done using a variety of open-source applications.

Methodology For the Amazon Sentiment Analysis (SA)

The study of online social networks has been the subject of specific research. It is possible to categorize them in three distinct ways: geometric, statistical, and topological. It's common for analytic systems to go through these four steps: detection, extraction and selection. They use various methods and strategies to jointly share their work to correctly estimate the OSN network analysis or graph visualization. A summary of social network analysis using graphs is also provided in this study.

First, a sampling of social networking websites, such as Amazon, will be examined. Different tests on social networks like Amazon will describe the typical aspects of social network sites. In social network visualization approaches, distinct elements and properties of social network members may be found by looking at the interconnections and connections between them. There are a variety of open-source programs that may be used to visualize and analyze network graphs. We leverage Amazon customer reviews to do sentiment analysis. Let's rapidly study the CSV file to proceed with the various steps in order. Stop words are the first thing we delete while working on NLP problems. Thenumber of stop words may estimate how much information we've been missing. Stop words have been incorporated into the Natural Language Toolkit library. So far, we've learned a lot about how reviews may be used to extract fundamental attributes.

A thorough cleaning of our dataset is necessary before extracting text and features. The training dataset was pre-processed to provide these characteristics. For spelling correction, we utilize the "textblob" library since this step is more beneficial in pre-processing to decrease the copies of words to grasp certain words, which are completely unintelligible in reading. Tokenization was utilized to categorize the user evaluations into various words and phrases. User reviews were turned into a "textblob" library, which was subsequently turned into a list of keywords. "ing,""ly,""s," and so on are all suffices that stemming refers to the removal of by employing a simple rule-based technique. As seen in the following illustration, we used a porter stemmer from the NLTK library, as shown in Figure.1

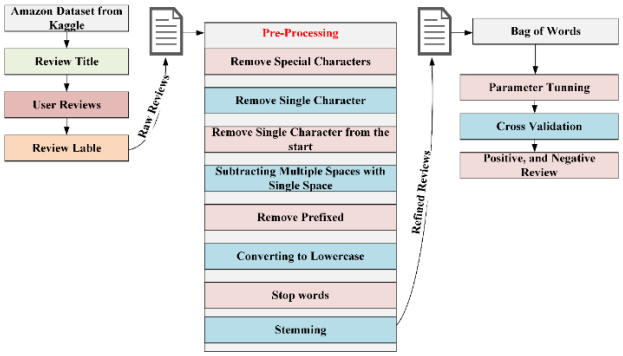

Figure 1 Amazon customer review sentiment analysis is shown in this flowchart

Dataset Collection

To increase the quality and efficiency of reviews for keyword trends, a model for better sentiment analysis was built leveraging an ensemble approach. Identifying the weather in a statement is the first step. The SSA determines whether or not a statement is positive or negative if it is subjective. Because we may state that sentences are simply little documents, researchers have not examined any fundamental differences between the assessments of sentences and document levels. The stream of data has been gathered from the data itself. There are approximately 400,000 customer reviews of Amazon goods in the database. For our study, we utilized 10,000 customer reviews from Amazon; for this, we used 80% training data and 20% testing and validation, as shown in Figure 2.

	Lable	Title	Review
2	2	Great CD	My lovely Pat has one of the GREAT voices of her generation. I have listened to this CD for YEARS and I still LOVE IT. '
3	2	One of the best game music soundtra	Despite the fact that I have only played a small portion of the game, the music I heard (plus the connection to Chron
4	1	Batteries died within a year ...	I bought this charger in Jul 2003 and it worked OK for a while. The design is nice and convenient. However, after abou
5	2	works fine, but Maha Energy is better	Check out Maha Energy's website. Their Powerex MH-C204F charger works in 100 minutes for rapid charge, with opti
6	2	Great for the non-audiophile	Reviewed quite a bit of the combo players and was hesitant due to unfavorable reviews and size of machines. I am w
7	1	DVD Player crapped out after one yea	I also began having the incorrect disc problems that I've read about on here. The VCR still works, but hte DVD side is u
8	1	Incorrect Disc	I love the style of this, but after a couple years, the DVD is giving me problems. It doesn't even work anymore and I us
9	1	DVD menu select problems	I cannot scroll through a DVD menu that is set up vertically. The triangle keys will only select horizontally. So I cannot
10	2	Unique Weird Orientalia from the 193	Exotic tales of the Orient from the 1930's. "Dr Shen Fu", a Weird Tales magazine reprint, is about the elixir of life that
11	1	Not an "ultimate guide"	Firstly,I enjoyed the format and tone of the book (how the author addressed the reader). However, I did not feel that
12	2	Great book for travelling Europe	I currently live in Europe, and this is the book I recommend for my visitors. It covers many countries, colour pictures,
13	1	Not!	If you want to listen to El Duke , then it is better if you have access to his shower,this is not him, it is a gimmick,very v
14	1	A complete Bust	This game requires quicktime 5.0 to work...if you have a better version of quicktime (I have 7.5), it will ask you to inst
15	2	TRULY MADE A DIFFERENCE!	I have been using this product for a couple years now. I started using it because my hair had gotten so dry from all th
16	1	didn't run off of USB bus power	Was hoping that this drive would run off of bus power, but it required the adapter to actually work. :(I sent it back.
17	1	Don't buy!	First of all, the company took my money and sent me an email telling me the product was shipped. A week and a half
18	2	Simple, Durable, Fun game for all ages	This is an AWESOME game! Almost everyone know tic-tac-toe so it is EASY to learn and quick to play. You can't play j
19	2	Review of Kelly Club for Toddlers	For the price of 7.99, this PC game is WELL worth it, great graphics, colorful and lots to do! My four year old daughte
20	2	SOY UN APASIONADO DEL BOX	Y ESTE LIBRO ESTÃ ESPLÃ‰NDIDO !Lo disfrutas, lo puedes usar como obra de consultaNos trae LAS HISTORIA DE LO

Figure 2 This is a snippet of data from Amazon's product reviewer database

Pre-processing of Dataset

Cleansing data and eliminating stop-words is one of the most important steps in increasing outcomes performance.

Basic Pre-Processing

So far, we've learned a lot about how reviews may be used to extract fundamental attributes. A thorough cleaning of our dataset is necessary before extracting text and features. The training dataset was pre-processed to provide these characteristics.

Lower Case

As a first step, we lowered the case of our dataset. We can prevent the same terms from appearing more than once in our dataset. When determining the word count, we distinguish between "Analytics" and "analytics.".

Removal of Stop Words

The section on basic feature extraction previously covered the elimination of "stop words" from user evaluations. We have used the same pre-processing procedure that we used in the past. We've utilized preexisting libraries and a list of placeholder terms as a workaround.

Common Word Removal

After removing stop words in the previous stage, we've also deleted frequent terms in this one. A decision will be made on whether or not to keep or eliminate the ten most commonly appearing terms. We've deleted any terms that aren't relevant to the way user reviews are classified.

Removal of Rare Words

We removed the most frequent terms from the user review and eliminated the rarest words. Because of their rarity, noise dominates the associations people make between them and other words. We may simply drop the are words and use the general word form instead to expand the number of words.

Spelling Correction

To reduce the number of copied words, we utilize the "textblob" library for spelling correction. In addition, for those words that are utterly unintelligible when read aloud.

Tokenization

Tokenization was employed to break down the user evaluations into a series of words or phrases. We used the "textblob" package to turn user evaluations into a blob, which we translated into a series of words.

Stemming

In stemming, suffixes like "ing,""ly,""s," and so on are removed through the application of a simple rule. The NLTK library's porter stemmer was used.

Lemmatization

Instead of removing sufficiency, lemmatization turns the word into its root word. The terminology is used for the lemmatization source term, and the morphological study is carried out. Therefore, lemmatization is typically preferred to lemmatization.

Proposed Models

Naïve Bayes (NB)

Naive Bayes classifiers categorize problem occurrences based on vectorized feature values. All naive Bayes classifiers assume that a feature's value is unaffected by its connection to other features, given its class variable. Fruit is a red, round, 10-centimeter apple. A naïve Bayes classifier has no association between color,

roundness, diameter and apple probability. Maximum likelihood is often used to estimate Naive Bayes model parameters[39].

Support Vector Machine (SVM)

SVM may be used for both classification and regression. Even if we assert regression, classification is its finest use. The objective of the SVM approach is to locate a hyperplane in an N-dimensional space that classifies the data points. The size of the hyperplane is determined by the number of features. If there are two input features, the hyperplane is effectively a straight line. As the number of input characteristics hits three, the hyperplane changes into a two-dimensional plane. When there are more than three qualities, visualisation becomes difficult [40].

Gated Recurrent Units (GRUs)

Recurrent neural networks employ GRUs to gate. The GRU is like an LSTM with a forget gate but lacks an output gate, making it simpler. GRU and LSTM perform equally in modeling polyphonic music, speech signals, and spoken language. GRUs perform better with smaller, less frequent datasets[41].

Long Short-Term Memory (LSTM)

A recurrent neural network example is the Short-Term Long Memory. In RNNs, the output of the previous step serves as the input for the subsequent step. Hochreiter and Schmidhuber were the LSTM's designers. Long-term RNN dependencies were addressed since the RNN was incapable of predicting words stored in long-term memory, but was able to generate more accurate predictions based on more recent input. RNN is less effective as the gap length increases. By default, LSTM may retain data for an extended period of time. Using this tool, time-series data is analysed, forecasted, and categorized [42].

Results and Discussion

Overview

Sentiment Analysis research is one of the scientific community's most current and challenging study topics. This paper addresses one of the most difficult aspects of sentiment analysis in bipolar words. Firms that want to maximize sentiment analysis must employ cutting-edge technologies and techniques. The definition in the presence of the backdrop, its impact on the product's total rating, and the fundamental characteristic of the study were evaluated, and the findings were astounding. Python's Platform has been used to analyze the work done so far. We've used Google co-lab as an integrated development environment (IDE) for these tests. The average semantical analysis and rating for each Amazon product were evaluated in this study. In addition, we looked at all of the reviews on Amazon.

Sentiment Analysis using Naïve Bayes (NB)

The Naive Bayes classifier is a simple yet effective tool in Machine Learning (ML). Classification based on the Bayes' formula is based on a strong assumption of independence between features. Natural Language Processing, for example, benefits greatly from the Naive Bayes classification when applied to textual data. Naive Bayes (NB) classifiers have been employed in our study of a large Amazon review dataset of [1:10000] reviews. Polarity 1 and 2 are used for negative and positive, respectively. Classifiers such as the Complement NB model (CNB), the Multinominal NB model (MNB), and the Bernoulli NB model (BNB) have all been used. If you're new to machine learning, you may utilize the cheat sheet provided by sklearn to figure out which model is appropriate for a specific situation. Use the NB classifier, it says. Meanwhile, we'd want to understand more about the model known as the "NB.".

Performance Metrics

Calculate each class's accuracy, recall, F-measure, and support.

- False positives are split by true positives in the ratio tp/(tp + fp): tp denotes true positives. Precision is the classifier's ability to avoid incorrectly classifying a negative sample as positive.
- Positive and negative results may be genuine or false, but the recall is equal to the difference between the two. The classifier's capacity to find all positive samples is called recall.
- The F-1 score, ranging from 1 to 0, may be used to describe precision and recall as a weighted harmonic mean.
- The F-1 score is weighted more heavily on recall than accuracy by a beta factor. As long as beta is equal to or larger than 1.0, it implies that both accuracy and recall are equally important.
- To estimate the degree of support, the number of different classes in y true is counted.

Because we compared three different classifiers and found that the F1 score in CNB and the MNB classifier is equal to the precision, we can conclude that both recall and precision are equally critical. The BNB classifier F1 score is higher than precision, but this isn't a big deal, as seen in Table 1

Table 1 Classification Report of the different NB Classifiers

	CNB			MNB			BNB			
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Support
1	0.80	0.80	0.80	0.80	0.80	0.80	0.85	0.72	0.78	1010
2	0.80	0.79	0.80	0.80	0.79	0.80	0.75	0.87	0.81	990
Accuracy			0.80			0.80			0.79	2000
Macro avg	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.79	0.79	2000
Weighted avg	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.79	0.79	2000

In the NB analysis, we have evaluated that the Complement NB model (CNB) gains the maximum accuracy, 80.00%, compared to the other two classifiers, as shown in Table Tabl.

Table 2 The different NB classifier's accuracy comparison

Complement NB model (CNB)	80.00%
Multinominal NB model (MNB)	79.90%
Bernoulli NB model (BNB)	79.35%

Plot Confusion Matrices

A classification algorithm's performance may be summarized using a confusion matrix. Just looking at classification accuracy might be deceptive if your dataset has an uneven number of observations in each class or has more than two classifications. Calculating a confusion matrix may give you a better sense of your classification model's accuracy.

CNB, MNB, and BNB classifications have been displayed in the confusion matrix. We looked at the diagonal and determined that the relationship was 1-to-1. Otherwise, we used words like 1-to-2 to describe uncorrected categorized situations, as shown in Figure.

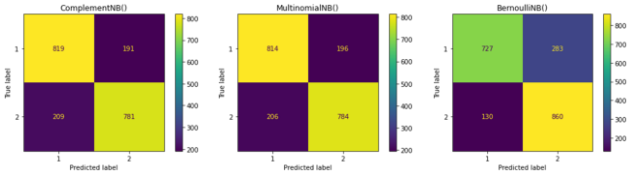

Figure 3 Confusion Matrices of the different NB Classifier

Receiver Operating Characteristic (ROC)

It analyzes the quality of classifier output using receiver operating characteristic (ROC) metrics. The Y-axis of a ROC plot normally shows the true positive rate, while the X-axis shows the false positive rate. As a result, the "ideal" point on the plot is located in the lower-left corner, where the false positive rate is 0, and the actual positive rate is 1. The bigger the area under the curve (AUC), the better. However, this isn't particularly practical. To increase the true positive rate while lowering the false positive rate, the "steepness" of ROC curves is critical.

ROC curves were drawn for three different classifiers, and we found that the "optimal" point for CNB had a false-positive rate of zero and a true-positive rate of 0.8808, as seen in the comparison to the other two classifiers as shown in Table3 and Figure4.

Table 3 The Ratio for the Receiver Operating Characteristic (ROC)

CNB	0.8808
MNB	0.8718
BNB	0.8718

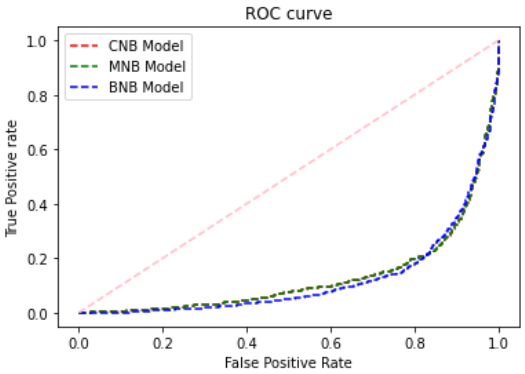

Figure 4 The Graph the Receiver Operating Characteristic (ROC)

Term Frequency-Inverse Document Frequency (TF-IDF)

It is possible to calculate a word's relative importance in a collection of documents using an algorithm known as TF-IDF. The inverse document frequency of each word may be calculated by multiplying it by the document's total number of instances of that term.

Automatic text analysis is its primary purpose, although it may also be used to score words in machine learning algorithms for NLP (NLP). We tested three NB classifiers, CNB, MNB, and BNB, and found the CNB classifier to be the most accurate, as shown in Table 4.

Table 4 The accuracy of TF-IDF for different NB Classifiers

CNB	81.05%
MNB	80.90%
BNB	78.65%

Pandas Data Frame

The numerical data index includes count, mean, standard deviation, minimum, maximum, and lower, 50, and higher%iles. The lower%ile is 25, while the higher is 75. Median and 50th%ile equal. Object data

count, uniqueness, topness, and frequency are indexed (e.g., strings or timestamps). Top of scale is more often. Freq indicates the most common value. Timestamps comprise the first and last list items. A random one is chosen if several object values have the greatest count. When analyzing a Data Frame with several data types, the numeric columns are analyzed by default. Include = 'all' includes all types of features. Include and exclude parameters limit which Data Frame columns are assessed in output. As stated in Table 5 e 5, the parameters are not considered while studying a series.

Table 5 The Description for Pandas Data Frame

	Sentiments
Count	10000
Mean	1.5124
Std	0.499871
Min	1
25%	1
50%	2
75%	2
Max	2

Data Visualization in Our Analysis

Plotting Histograms of Data frame columns might be useful for in-depth analysis in certain cases. It helps a lot if you use the dataframe.hist() method This function allows us to create histograms with whatever number of columns we choose.

Matplotlib Axes

In Figure 5 (a), we can see that our number of total reviews is frequently zero and that most of the data is between 0 and 1000. Let's scope our data down and then plot again.

In Figure 5 (b), we can see that our number of positive and negative reviews is frequently zero and that most data is between 0 and 1000. Let's scope our data down and then plot again.

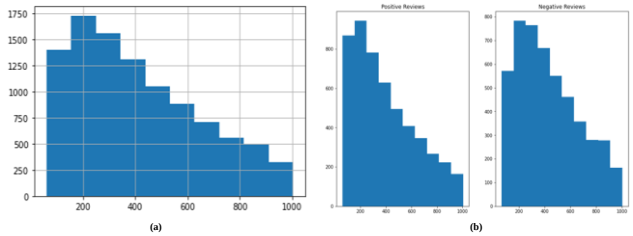

Figure 5 Error! No text of specified style in document.The Histogram Visualization by Using Matplotlib Axes

Working With the Most Frequent Words

In our dataset, we have used two review classes (positive and negative); we can find some frequently used words, as shown in Figure 6 and Table 6.

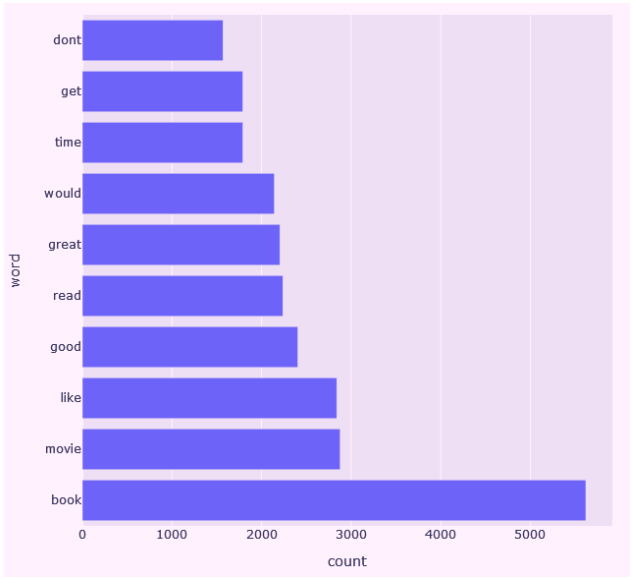

Figure 1 The graph of Frequent Words

Table.1 The actual values for Frequent Words

	Word	count
0	book	5620
1	movie	2878
2	like	2839
3	good	2405
4	read	2237
5	great	2204
6	would	2143
7	time	1791
8	get	1790
9	dont	1569

Support Vector Machine (SVM)

A method known as the Support Vector Machine (SVM) may be utilized for both classification and regression problems. However, categorization is the most common usage for it. In the SVM method, data points are represented by n-dimensional coordinates (n is the number of features you have), where a specific coordinate value represents each feature. After then, classification is carried out by identifying the hyperplane that most separates the two groups. In our analysis, we have used 10000 user reviews for the product on amazon.

In Table 7, we have analyzed that the F1 score is 0.81, which is greater than class 1 and less than class. That means that class 2 is classified more accurately in our analysis. Class 2 is related to the positive class in our dataset.

Table 2 Classification Report of the SVM Classifier

	SVM			
	Precision	recall	f1-score	Support
1	0.80	0.82	0.81	968
2	0.83	0.81	0.82	1032
Accuracy			0.81	2000
Macro avg	0.81	0.81	0.81	2000
Weighted avg	0.81	0.81	0.81	2000

Table 8 shows the overall evaluated results for the SVM classifier; from 2000 support reviews of user reviews of Amazon, there are 1625 reviews correctly predicted, and 375 are wrong predicted with an accuracy of 81.25%.

Table 8 The overall statistics of SVM

	SVM
Correct Prediction	1625
Wrong Prediction	375
Accuracy	81.25%

The SVM classifier's confusion matrix has been shown. We looked at the diagonal and determined that the relationship was 1-to-1. Otherwise, we used words like 1-to-2 to describe uncorrected categorized situations, as shown in Figure 7.

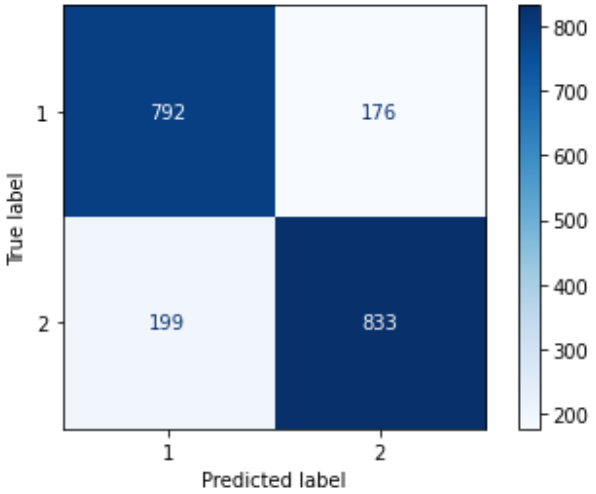

Figure.2 Confusion Matrices of SVM Classifier

Sentiment Analysis using GRU + LSTM

Activation and input data are inputs to a block of function in this network, providing an output. The output is transferred to a new block containing the next batch when all the input data have been processed. A recurrent neural network is an apt term for this network. The following
Figure **3** may help clarify things.

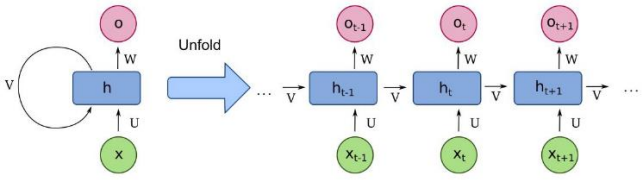

Figure 3 The RNN Model Process Diagram

Using RNN in real-time would be impossible due to its inability to remember the context of a discussion, as seen above. As a result, the Gated Recurrent Unit (GRU) was shown as a solution. It features a memory cell unit to retain the context of prior sequences.

"Short-Term Long Memory" is the abbreviation for LSTM. Even more sophisticated than GRU is LSTM. Though LSTM was developed long before GRU, it is more complicated. It contains a variety of gates for dealing with various input parameters. However, this raises the model's computation burden and makes training more time-consuming than GRU.

The Compiling Model

In the compiling model of the GRU and LSTM, we have used 5 epochs; for training and testing, we have to use a total of 10000 reviews from the amazon users. We have calculated both models' validated loss and accuracy and compared them based on these 2 parameters. After evaluation, we have calculated that GRU has less loss and greater accuracy than the LSTM model. The GRU has 0.6759 val_loss and 0.8044 val_accuracy, as shown in Table 3.

Table 3 The compiling model for the GRU and LSTM

	GRU		LSTM	
Epoch	Val_Loss	Val_Accuracy	Val_Loss	Val_Accuracy
1	0.4984	0.7606	0.4176	0.8131
2	0.4152	0.8131	0.4194	0.8181
3	0.5562	0.8044	0.5050	0.8181
4	0.5385	0.8025	0.5735	0.8081
5	0.6759	0.8044	0.6029	0.7625

Figure **4** shows the graphical representation of the compiling model for the GRU + LSTM models. In these graphs are the representation loss, accuracy, val_loss and val_accuracy.

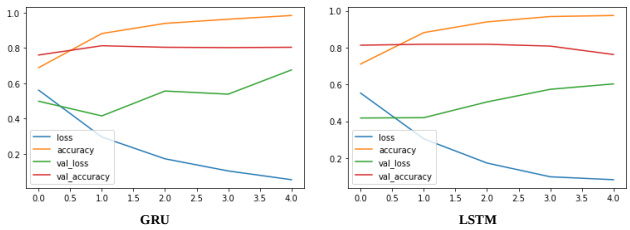

Figure 4 The Graph of Compiling Model of the GRU + LSTM

Table 10shows the overall evaluated results for the GRU and LSTM classifier; from 2000 support reviews of user reviews of Amazon,1613 reviewsare correctly predicted, and 388 are wrong predicted, with the accuracy of 81.6097% for the GRU classifier. Also, 1526 reviews are correctly predicted, and 475 are wrong, with an accuracy of 76.2619% for the GRU classifier.

Table 10 The overall statistics for GRU + LSTM Classifier

	GRU	LSTM
Correct Prediction	1613	1526
Wrong Prediction	388	475
Accuracy	80.6097	76.2619

The GRU and LSTM classifiers' confusion matrices are shown in the graphs. It was determined that we correctly classed it as negative-to-negative in the diagonal. Otherwise, we used phrases like "negative-to-positive" and "so on" to describe uncorrected categorized situations, as shown in **Error! Reference source not found.**Figure 5.

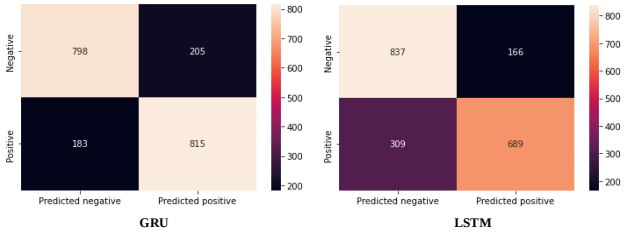

Figure 5 Confusion Matrices of the GRU and LSTM Classifier

In Table 5we have analyzed that the F1 score is equal to precision in the GRU classifier, which is 0.81, so recall and precision are equally important. The LSTM model'sF1 score is 0.76, greater than class 1 and less than class. That means that class 2 is classified more accurately in our analysis. Class 2 is related to the positive class in our dataset.

Table 5 Classification Report of the GRU + LSTM classifier

		1	2	Accuracy	Macro avg	Weighted avg
GRU	precision	0.81	0.8		0.81	0.81
	recall	0.8	0.8		0.81	0.81
	f1-score	0.8	0.81	0.81	0.81	0.81
	support	1003	998	2001	2001	2001
LSTM	precision	0.73	0.81		0.77	0.77
	recall	0.83	0.69		0.76	0.76
	f1-score	0.78	0.74	0.76	0.76	0.76
	support	1003	998	2001	2001	2001

Conclusion

Sentiment analysis is a challenging and contemporary topic in science. This paper addresses a key difficulty in bipolar-word sentiment analysis, emphasizing the need for companies to utilize existing methodologies to minimize in-person research. The findings reveal significant insights when examining the company's historical data and its impact on overall ratings. Today, various tools and techniques visualize online social networks (OSNs), enabling the detection of network properties and their influence. Sentiment analysis is vital for interpreting text data, as daily streams of customer feedback, comments, and tweets provide marketers with valuable insights for effective campaigns. While numerous algorithms assess emotions, some specifically consider bipolar keywords, which shift their meaning based on context. The introduction of social networking sites has revolutionized online interactions, facilitating the exchange of ideas and information. Using social network analysis (SNA) techniques, we can analyze graphs with extensive nodes and connections. This study proposes a novel technique incorporating nostalgic elements based on product characteristics, using feedback from Amazon customers. After preprocessing 10,000 reviews from a dataset of 400,000, we compared two machine learning algorithms (CNB, MNB, BNB, SVM) and two deep learning models (GRU, LSTM). The results showed that GRU and LSTM delivered superior accuracy with lower loss rates, achieving 80.61% and 76.26%, respectively. In contrast, machine learning models obtained accuracies of 80.00%, 79.90%, 79.35%, and 81.25%. Notably, SVM required significantly more execution time than the deep learning models.

References

[1] S. Wassan, X. Chen, T. Shen, M. Waqar, and N. Jhanjhi, "Amazon product sentiment analysis using machine learning techniques," *Revista Argentina de Clínica Psicológica,* vol. 30, no. 1, p. 695, 2021.

[2] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in *2018 IEEE international conference on innovative research and development (ICIRD)*, 2018, pp. 1-6: IEEE.

[3] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon review classification and sentiment analysis," *International Journal of Computer Science and Information Technologies,* vol. 6, no. 6, pp. 5107-5110, 2015.

[4] P. Pandey and N. Soni, "Sentiment analysis on customer feedback data: Amazon product reviews," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 320-322: IEEE.

[5] C. Rain, "Sentiment analysis in amazon reviews using probabilistic machine learning," *Swarthmore College,* 2013.

[6] N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," *Spatial Information Research,* vol. 28, no. 5, pp. 601-607, 2020.

[7] N. Shrestha and F. Nasoz, "Deep learning sentiment analysis of amazon. com reviews and ratings," *arXiv preprint arXiv:1904.04096,* 2019.

[8] J. Sadhasivam and R. B. Kalivaradhan, "Sentiment analysis of Amazon products using ensemble machine learning algorithm," *International Journal of Mathematical, Engineering and Management Sciences,* vol. 4, no. 2, p. 508, 2019.

[9] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques," *J. Comput. Sci.,* vol. 14, no. 5, pp. 714-726, 2018.

[10] R. Bekkerman, R. El-Yaniv, N. Tishby, and Y. Winter, "Distributional word clusters vs. words for text categorization," *Journal of Machine Learning Research,* vol. 3, no. Mar, pp. 1183-1208, 2003.

[11] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759,* 2016.

[12] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1422-1432.

[13] F. Alshuwaier, A. Areshey, and J. Poon, "Applications and Enhancement of Document-Based Sentiment Analysis in

Deep learning Methods: Systematic Literature Review," *Intelligent Systems with Applications,* p. 200090, 2022.
[14] Z. Li, Y. Zhang, Y. Wei, Y. Wu, and Q. Yang, "End-to-End Adversarial Memory Network for Cross-domain Sentiment Classification," in *IJCAI*, 2017, pp. 2237-2243.
[15] J. T. Zhou, S. J. Pan, I. W. Tsang, and S.-S. Ho, "Transfer learning for cross-language text categorization through active correspondences construction," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
[16] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480-1489.
[17] T. Chen, R. Xu, Y. He, Y. Xia, and X. Wang, "Learning user and product distributed representations using a sequence model for sentiment analysis," *IEEE Computational Intelligence Magazine,* vol. 11, no. 3, pp. 34-44, 2016.
[18] Z.-Y. Dou, "Capturing user and product information for document level sentiment analysis with deep memory network," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 521-526.
[19] Z. Wu, X.-Y. Dai, C. Yin, S. Huang, and J. Chen, "Improving review representations with user attention and product attention for sentiment classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32, no. 1.
[20] R. Johnson and T. Zhang, "Supervised and semi-supervised text categorization using LSTM for region embeddings," in *International Conference on Machine Learning*, 2016, pp. 526-534: PMLR.
[21] S. Zhai and Z. M. Zhang, "Semisupervised autoencoder for sentiment analysis," in *Thirtieth AAAI conference on artificial intelligence*, 2016.
[22] P. Liu, X. Qiu, and X. Huang, "Adversarial multi-task learning for text classification," *arXiv preprint arXiv:1704.05742,* 2017.
[23] J. Hong and M. Fang, "Sentiment analysis with deeply learned distributed representations of variable length texts," *Stanford University Report,* pp. 1-9, 2015.
[24] G. Rao, W. Huang, Z. Feng, and Q. Cong, "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing,* vol. 308, pp. 49-57, 2018.
[25] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM model for document-level sentiment analysis," *Machine Learning and Knowledge Extraction,* vol. 1, no. 3, pp. 832-847, 2019.
[26] L. L. Dhande and G. Patnaik, "Review of sentiment analysis using naive bayes and neural network classifier," *Int. J. Sci. Eng. Technol. Res,* vol. 3, no. 07, pp. 1110-1113, 2014.
[27] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, 2014, pp. 69-78.
[28] B. Darshan, P. H. Prasad, and M. D. Teja, "Product reviews classification using sentiment analysis."
[29] C. Sindhu, S. Thejaswin, S. Harikrishnaa, and C. Kavitha, "Mapping Distinct Source and Target Domains on Amazon Product Customer Critiques with Cross Domain Sentiment Analysis," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, 2022, pp. 782-786: IEEE.
[30] U. Norinder and P. Norinder, "Predicting Amazon customer reviews with deep confidence using deep learning and conformal prediction," *Journal of Management Analytics,* vol. 9, no. 1, pp. 1-16, 2022.
[31] A. Dadhich and B. Thankachan, "Sentiment analysis of amazon product reviews using hybrid rule-based approach," in *Smart Systems: Innovations in Computing*: Springer, 2022, pp. 173-193.
[32] M. E. Alzahrani, T. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.
[33] Z. Desai, K. Anklesaria, and H. Balasubramaniam, "Business Intelligence Visualization Using Deep Learning Based Sentiment Analysis on Amazon Review Data," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1-7: IEEE.
[34] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, "Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews," *Mathematical Problems in Engineering,* vol. 2021, 2021.
[35] M. Y. A. Salmony and A. R. Faridi, "Supervised Sentiment Analysis on Amazon Product Reviews: A survey," in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, 2021, pp. 132-138: IEEE.
[36] A. S. AlQahtani, "Product sentiment analysis for amazon reviews," *International Journal of Computer Science & Information Technology (IJCSIT) Vol,* vol. 13, 2021.
[37] L. Tavoschi *et al.*, "Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy," *Human Vaccines & Immunotherapeutics,* pp. 1-8, 2020.
[38] Z. Miao, Y. Li, X. Wang, and W.-C. Tan, "Snippext: Semi-supervised Opinion Mining with Augmented Data," *arXiv preprint arXiv:2002.03049,* 2020.
[39] A. Shahi, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Naive Bayesian decision model for the interoperability of heterogeneous systems in an intelligent building environment," *Automation in Construction,* vol. 54, pp. 83-92, 2015.
[40] C. M. Rocco and J. A. Moreno, "Fast Monte Carlo reliability evaluation using support vector machine," *Reliability Engineering & System Safety,* vol. 76, no. 3, pp. 237-243, 2002.
[41] Z. Chu, J. Yu, and A. Hamdulla, "LPG-model: A novel model for throughput prediction in stream processing, using a light gradient boosting machine, incremental principal component analysis, and deep gated recurrent unit network," *Information Sciences,* vol. 535, pp. 107-129, 2020.
[42] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model," *Computational Intelligence,* vol. 37, no. 1, pp. 409-434, 2021.