



ARTICLE

An Ensemble-Based Hotel Reviews System Using Naive Bayes Classifier

Joseph Bamidele Awotunde¹, Sanjay Misra^{2,*}, Vikash Katta² and Oluwafemi Charles Adebayo¹

¹Department of Computer Science, Faculty of Information and Communication Sciences, University of Ilorin, Ilorin, 240003, Nigeria

²Department of Computer Science and Communication, Østfold University College, Halden, Norway

*Corresponding Author: Sanjay Misra. Email: sanjay.misra@hiof.no

Received: 27 September 2022 Accepted: 07 December 2022

ABSTRACT

The task of classifying opinions conveyed in any form of text online is referred to as sentiment analysis. The emergence of social media usage and its spread has given room for sentiment analysis in our daily lives. Social media applications and websites have become the foremost spring of data recycled for reviews for sentimentality in various fields. Various subject matter can be encountered on social media platforms, such as movie product reviews, consumer opinions, and testimonies, among others, which can be used for sentiment analysis. The rapid uncovering of these web contents contains divergence of many benefits like profit-making, which is one of the most vital of them all. According to a recent study, 81% of consumers conduct online research prior to making a purchase. But the reviews available online are too huge and numerous for human brains to process and analyze. Hence, machine learning classifiers are one of the prominent tools used to classify sentiment in order to get valuable information for use in companies like hotels, game companies, and so on. Understanding the sentiments of people towards different commodities helps to improve the services for contextual promotions, referral systems, and market research. Therefore, this study proposes a sentiment-based framework detection to enable the rapid uncovering of opinionated contents of hotel reviews. A Naive Bayes classifier was used to process and analyze the dataset for the detection of the polarity of the words. The dataset from Datafiniti's Business Database obtained from Kaggle was used for the experiments in this study. The performance evaluation of the model shows a test accuracy of 96.08%, an F1-score of 96.00%, a precision of 96.00%, and a recall of 96.00%. The results were compared with state-of-the-art classifiers and showed a promising performance and much better in terms of performance metrics.

KEYWORDS

Sentiment analysis; hotel reviews; Naive Bayes algorithm; consumer opinions; web 2.0; machine learning

1 Introduction

Whenever people take a vacation or travel to a particular place for business or pleasure, one major question that comes to mind is where they would sleep for the night(s). Throughout the years, so many hotels have been built for public use. They are also of different classes. Thus, everyone looking for a hotel tries to find what suits his class. The best way to know if a hotel is right for you or not is to find out what people who have stayed there before are saying about the hotel. Nine in ten resort managers



stated that online tourist reviews are vital to the future of their businesses [1]. Opinions found on the web were found to be germane, real, and informative compared to the ones in hotel booklets [2]. The reviews on the web are too many for them to comprehend and get the polarity. This is where the task of hotel reviews sentiment analysis comes into play. It can help you decide whether or not a hotel is suitable for your trip. This is because it helps to extract the feelings from the opinion of the reviewers and classifies them thereby providing valuable information for hotel guests to use when making their choice of hotel. Hotel reviews are a fundamental prerequisite by which guests choose a hotel to stay in.

Various ensemble-based models have been used by several authors for the classification and prediction of sentiment datasets in various fields, even in hotel reviews, but many are characterized by lower accuracy with a slower training process. Naive Bayes ensemble classifiers have been proven to be used by ensuring a faster training process capacity when compared with other classifiers. The classifier still proved better in terms of performance when compared with TPU v3-8 in the training process of various data. The positive review of rooms provided by a hotel and even the category of the food served become part of the sentiment information a user often wants before settling for any hotel of their choice. Hence, it becomes necessary to categorize the hotel review sentiment information about a particular aspect or category belonging to a particular hotel. For example, a positive review of the food served by a particular hotel will surely attract customers to that hotel. Hence, this leads to the need to label each hotel based on a wide range of criteria.

The kind of reviews previously given by people about a hotel determines the rate at which people will flock to a hotel, hence determining the amount of money they will make. Therefore, Sentiment Analysis has been developed as a tool whose main objective is to derive the opinion of customers on the hotels from a database of different customer reviews. An efficient Sentiment Analysis is one that optimally derives vital data from the database in order to formulate meaningful information, which is useful in helping customers make informed decisions on the hotels to lodge in. The authors in [3] worked on sentiment analysis of hotel reviews using K-Nearest Neighbour (KNN) classifier, but it was not accurate enough. Several approaches have been applied to sentiment analysis. The authors in [4] worked on Sentiment Analysis Using Naive Bayes Classifier. They created a model that performed sentiment analysis on Twitter data using the Machine Learning (ML) method. The framework developed in this research was built with the application of the Natural Language Tool Kit (NLTK) on the tweets. A bag of words was also employed, containing both positive and negative words distinctly. Naive Bayes algorithm was employed in tweets categorisation. Nevertheless, they chose an efficient Twitter feature dataset which improved the efficiency and accuracy of the algorithm.

1.1 Motivation

The lack of a correctly labelled dataset is one of the main issues that various researchers have to deal with while building hotel reviews, sentiment, and recommendation systems. There are hardly any datasets that include reviews of hotels. Even such datasets are very difficult to be used for the purpose of hotel recommender and many cannot even be used for classification or sentiment analysis tasks. Some of the dataset's sentiment labels are missing, which is necessary to train the dataset's sentiment categorization algorithm [1]. This highlights the need for creating a dataset with hotel reviews that are appropriately categorized and tagged. Based on the literature review, it has been noted that the majority of studies concentrate on categorizing sentiments into only two groups, positive and negative. There might be some reviews, though, that are purely neutral [1]. Therefore, this proposed study extended the publicly available dataset to accommodate a 3-class problem by introducing the neutral label. It is important to note in this perspective that for a 3-class problem of this nature, baseline approaches exhibit low accuracy performance. The main cause of this reduced accuracy is the imbalance in the

data brought on by the dataset's lack of a reasonable number of neutral label data and did not even present the neutral class within the dataset.

Therefore, this paper develops a framework for sentiment analysis for hotel reviews using the Naive Bayes algorithm. The developed framework will be used to analyze sentiments in hotel review access to students based on the Naive Bayes algorithm. The introduction of neutral sentiment gives a detailed analysis of various hotels thereby helping customers make better choices from the available ones. The key contributions of this study are:

- i. A new sentiment tag was derived from the existing hotel reviews dataset using Datafiniti's Business Dataset for hotel reviews and made up of English-language critiques. Hence, design a framework for analyzing the sentiment of hotel reviews based on the redefined dataset.
- ii. Implement the proposed system using the Naive Bayes algorithm worked on a variety of linguistic aspects for assessing feelings in hotel reviews.
- iii. The proposed model was evaluated with the recent state-of-the-art model of the hotel's reviews system.

The remainder of the paper is organized as follows: [Section 2](#) presents a review of related work by reviewing some studies that are related to this study. [Section 3](#) discusses the methods and materials used in this study, and the algorithm employed with the dataset conversed in detail. [Section 4](#) discusses the experimental results and gives a detailed discussion of the proposed study, and finally, [Section 5](#) concludes the paper with a given future direction for the study.

2 Review of Related Work

Authors in [5] suggested an approach to carry out opinion mining on a dataset with a size between two hundred and four thousand. The remaining one-fourth of the data was used for validation, leaving just three-quarters for training. Multinomial Naive Bayes and Decision trees were the two methods employed. Feature extraction was used to do tweet pre-processing. Apache Spark framework was employed because of its scalability and gave faster correct outcomes. The decision tree produced an accuracy of 100%, precision of 100%, recall of 100%, and F1-score of 100%, respectively. The authors in [6] worked on the multimodal analysis of memes for the extraction of sentiment, they used the IMGTEXT, IMGSEN, and CAPSEN models for humour and sentiment detection. It was discovered that the average testing accuracy was 62.77%, while the average F1-score was 59.05% which was a big improvement over the baseline observed. The authors of [7] also worked on the sentiment analysis of the microblogging dataset related to the coronavirus outbreak. A model was created to distinguish between the sentiments in various labels (positive, negative, and neutral) employing the collected dataset. The study first removed ambiguity from the dataset by pre-processing it, and then they partitioned the data into two, using 80% to train the models and the 20% left for testing the models. The text was then subjected to feature extraction using a bag of words and TF-IDF techniques. The LR, XGboost, SVM, NB, and DT all failed to achieve accuracy levels of 93%. The RF revealed the highest accuracy performance for the bag-of-words and TF-IDF models. Authors in [8] suggested an aspect-focused approach for mining opinions on hotels. He used different class-balanced methods and data models. He got a good prediction accuracy on hotel review sentiments. The author got about 70% to 75% and 85% to 90% for both prediction accuracy and polarity accuracy, respectively.

For an essential prediction of mortality risk in sepsis patients and the control techniques, the authors in [9] concentrated on sentiment classification. They studied 1,844 cases of sepsis with major monthly deaths at about thirty-eight per cent. Multivariate Cox's study implied that impartiality

scores and emotion polarization were practically considerable, even with the availability of known determinants for monthly death in sufferers. These observations implied that the sentiment scores obtained were helpful in predicting the likelihood of mortality risk in symptomatic patients. The authors in [10] proposed the LSTM model for the classification and analysis of online customers' reviews. The study assesses the likelihood that customers will feel a certain way about the airline's offerings. The results gained to support the crucial fundamental contribution to the literature on service evaluation, online reviews, and suggestions. The authors in [11] presented general reviews work on the application of ML based on hotel and tourism sentiment analysis. The study worked on the Aspect based Sentiment Oriented Summarization of Hotel Reviews. This was because this part was not well talked about in their review sentences. Another work by authors in [12], proposed a rough set approach (RST) for online reviews classification and analysis. The implementation of RST for predictive suggestions is the best alternative, according to experimental results, because of its increased accuracy (96.70%) and remarkably quick classification process with a large amount of data.

The authors in [13] proposed a model for the Joint Aspect and Polarity Classification for Aspect-based Sentiment Analysis with End-to-End Neural Networks. In this study, they proposed a novel framework for area-based opinion mining. They jointly modeled the identification of aspects and the categorization of their polarity in an end-to-end trainable neural network. The combination of a convolutional neural network and fast text implants did better than the best submission of the shared task in 2017, thereby designing a cutting-edge model. In [14], the authors proposed a hierarchical framework for classifying reviews using a facet-focused approach. They employed the hierarchical bidirectional Long Short-Term Memory (LSTM) architecture model for the study. Their model achieved cutting-edge outcomes on 5 out of 11 datasets for aspect-focused opinion mining. A Corpus of Basque and Catalan Hotel Reviews Annotated for Sentiment Classification was suggested by the work in [15]. For automated aspect-level sentiment classification in Basque and Catalan, they used two datasets, each of which has a small number of grammatical options. To evaluate, the paper performed 10-fold cross-validation with each fold retaining 80% of the training data. From the work, it was observed that the weighted F1-score was higher in Basque than in Catalan. Authors in [16] worked on a Targeted Aspect Based Sentiment Analysis Dataset for Urban Neighbourhoods which they termed SentiHood. It is based on the text taken from the question-answering platform of Yahoo! they aimed to deduce useful knowledge from the things users commented on. To test the experiment, they introduced their custom data, gotten from a user discussion website in which users talk about cities. The proposed study revealed a cutting-edge result from their performance evaluation.

The authors in [17] worked on a comparative study of sentiment analysis using NLP and various ML-based models on US Airline Twitter Data. Their most effective methods gave a score of seventy-seven percent for SVM and Logistic Regression while using the BOW model. The authors in [18] worked on the Span Detection for Aspect-Based Sentiment Analysis in Vietnamese. They also suggested a new framework by employing Bidirectional Long Short-Term Memory (BiLSTM) with a Conditional Random Field (CRF) layer (BiLSM-CRF) to perform span sensing in Vietnamese facet-focused opinion mining. The most impressive performance was about 63.0% F1macro for spam detection. In [19], the authors proposed subject-focused resort opinion mining. They suggested a system that can either be used to examine a specific hotel or to differentiate between many hotels. Their system gives exact and elaborate accounts that would help the public select the right hotels for their scheduled travels. Therefore, it has the ability to study and deal with reviews in multiple languages and it gives more accurate reports. The authors in [20] worked on a multilingual Twitter sentiment analysis using contributors. They studied close to two million tweets, in thirteen dialects in Europe, tagged as polarity by contributors. Those tagged tweets were employed as data for polarity algorithms to learn

various dialects. The algorithms were ordered by their mean performance level across thirteen data files. The 95.0% confidence range was approximated from 10-fold cross-validations. In [21], the authors worked on the socio-dynamics of profanity and its consequences on opinion mining on social networks. They proved that using profane attributes gives a better opinion of the mining system's performance. In [22], the authors used deep learning techniques to categorize reviews. Eight ML-based algorithms such as Naive Bayesian, Adaboost, K-NNs, RF, LR, DT, NN, and SVM, and five DL-based like CNN, LSTM, BiLSTM, RNN, and GRU were deployed with the aim of discovering polarities on fake news on the pandemic. By comparing various metrics, they discovered that CNN and BiLSTM gave the best results for designing the categorization framework due to the fact that they had correctness of 97.2%.

In [23], the authors proposed an approach for studying the polarity of coronavirus Twitter comments in Nepal. They used multiple CNN frameworks in categorizing the comments. The frameworks proved to be balanced and rugged. They proved their proposed features' extraction techniques to be efficient by applying orthodox ML classifiers, which proved that the suggested attributes can omit the composite COVID-19 tweets in many cases. In [24], the authors proposed a field-based sentiment analysis model by employing various ML-based methods. Their work is based on football tweets and labels the feature for the sentiment classification as fouls, penalties, and goal-scoring, among others. The outcomes prove that their method is efficient in identifying fans' feelings about football matches applying concepts from ordinal regression, the authors in [25] worked on an elaborate study of tweets using different ML methods. The whole procedure involved dataset preprocessing, feature extraction, and then, deploying ML methods to categorize the tweets. The observed outcomes proved that ML classifiers may identify ordinal regression with excellent outcomes. The DT had the topmost correctness of 91.81%. In [26], the authors described a sentiment analysis by performed on a dataset of tweets pertaining to COVID-19 immunizations. To categorize the tweets, they used NLP and the KNN classifiers. They observed a positive sentiment analysis of 47.3% for Pfizer, 46.2% for Moderna, and 40.1% for AstraZeneca, respectively from their experiment.

The authors in [27] used real news data for the study that are gathered from a variety of multimedia services, including the New York Times, Health Harvard, Centers for Disease Control and Prevention (CDC), World Health Organization (WHO), and Global Health Now, using the information fusion technique. While information for fake news is gathered from Facebook, YouTube, and other social media platforms. It was found that the volume of tweets on distant healthcare services skyrocketed after the pandemic began. In [28], the authors proposed A ML model for Sentiment Analysis for Distance Education. The feedback got from the eCampus system by using ML techniques and modelled them by employing seven classifiers using Python programming language in Jupyter Notebook with the aid of a supervised learning method. The best outcomes from the analysis were gotten with 77.5% correctness of the LR classifier model. The performance level of the analysis was between 42% and 85%. [Table 1](#) gives a summary of the related work discussed.

Table 1: The summary of literature review

Reference	Year	Models	Results/Contributions	Research gap
[4]	2019	NB Classifier	NB classifier performed better on movie reviews after studying the hotel reviews dataset both classifiers showed similar outcomes. Thus, the Naive Bayes classifier was found to be better for movie reviews classification	The results show that there need for improvement on the hotel reviews classification
[5]	2016	Multinomial NB and DT classifiers	The DT classifier revealed 100% accuracy, precision, recall, and F1-score	The models was applied on another dataset apart from hotel reviews
[6]	2021	They used the IMGTX, IMGSEN, and CAPSEN models for humor and sentiment detection	The average testing accuracy and F1-score were 62.77%, and 59.05% respectively denoting a major improvement over the standard	The accuracy of the proposed models still need improvement
[7]	2021	They used NB, DT, xgboost, SVM and LR classifiers for training bag-of-words and TF-IDF models	The random forest gave the highest accuracy for both bags-of-words and TF-IDF	Their performance metrics still need improvement, and the issue of imbalance dataset still need to be properly handle
[8]	2020	They used two models; multi-class classification and regression, and three algorithms; Naive Bayes, SVM, and linear regression to analyze the reviews	He got about 70% to 75% and 85% to 90% for both prediction accuracy and polarity accuracy, respectively	Performance metrics still need improvement, and the issue of imbalance dataset still need to be properly handle
[9]	2021	They used multivariate Cox analysis to check for sentiment polarity and subjectivity	These observations implied that the sentiment scores measured were conclusive in observing the dangers of 30-day mortality in sepsis sufferers	They recommended the use of their model on other datasets

(Continued)

Table 1 (continued)

Reference	Year	Models	Results/Contributions	Research gap
[10]	2022	LSTM based model	The results gained support the crucial fundamental contribution to the literature on service evaluation, online reviews, and suggestions	The problem of imbalance dataset is still an open issues, and the dataset used is very small especially when expect want to use deep learning models
[11]	2021	ML-based techniques	The study had a significant influence on future research directions for researchers	There is still open research issues during their study
[12]	2022	They performed sentiment analysis using RST approach	The experimental results, provided and accuracy of 96.70% and there is reduction in the processing time when compared with other models	The issue of imbalance dataset still need to be properly handle
[13]	2018	In an end-to-end trainable neural network, they concurrently modeled the detection of aspects and the categorization of their polarity	It performed better than models that only depended on sentence information and gave a performance on par with models which use lots of external resources and hand-engineered attributes. Their model achieved cutting-edge outcomes on 5 out of 11 datasets for aspect-focused opinion mining	Performance metrics still need improvement, and the issue of imbalance dataset still need to be properly handle
[14]	2016	They used the hierarchical bidirectional long short-term memory (LSTM) architecture model for aspect-based sentiment analysis	It performed better than models that only depended on sentence information and gave a performance on par with models which use lots of external resources and hand-engineered attributes	There is still room for improvement in term of performance accuracy

(Continued)

Table 1 (continued)

Reference	Year	Models	Results/Contributions	Research gap
[15]	2018	Conditional random Field4 (CRF), Bag-of-words approach, and linear SVM classifier	It was observed that the weighted F1-score was higher in Basque than in Catalan	There is still room for improvement in term of performance accuracy
[16]	2016	Logistic regression & long short-term memory (LSTM)	LSTM model with final state implants did better than the one with implants at the location index	
[17]	2021	SVM, multinomial NB, RF, and LR classifiers	The most performing classifier revealed and accuracy of 77%, both SVM and LR using the Bag-of-Words classifier for feature selection	There is still room for improvement in term of performance accuracy
[18]	2021	BiLSTM and BiLSTM-CRF	The best performance was F1-macro with 62.76% accuracy for span detection using BiLSTM-CRF enabled with imbedding syllable, character and contextual implants from XLM-RoBERTa.	There is still room for improvement in terms of performance metrics
[19]	2021	Pandas, Textblob, Langdetect, mTranslate, Matplotlib	They approximated a 95% confidence range for Alpha from 10-fold cross-validations	Performance metrics still need improvement, and the issue of an imbalance dataset still needs to be properly handled
[21]	2018	Bidirectional long-short-term memory network	Gender outcomes show that females are less likely to post profane tweets than males, older users are way less likely than younger ones to upload profane things, and faith is not related to the use of profane words with the second-highest effect	

(Continued)

Table 1 (continued)

Reference	Year	Models	Results/Contributions	Research gap
[20]	2016	They gathered a set of about 1.6 million manually tagged tweets. They applied four evaluation measures and show that two of them are more appropriate to evaluate sentiment classifiers	They demonstrated that the model performance approaches the correspondence between annotators when the training set is big enough	Performance metrics still need improvement, and the issue of an imbalance dataset still needs to be properly handled
[22]	2021	Six Algorithms: Naive Bayesian, Adaboost, K-nearest neighbors, random forest, logistic regression, decision tree, neural networks, and support vector machine and five deep learning models: CNN, LSTM, BiLSTM, RNN, and GRU	They discovered that CNN and BiLSTM were the most effective to build a model for classifying fake news on COVID-19 sentiment analysis because they gave a very high accuracy of 97%	The problem of an imbalanced dataset is still an open issue, and the dataset used is very small especially when expecting to want to use deep learning models
[23]	2021	The SVM with both linear and RBF kernels, XGBoost, ANN, RF, NB, LR, and K-NN was used in the said study	They observed that fastText-based embedding (ft) does better than the other embedding types (“ds” and “da”) using performance metrics for the majority of the classifiers	Applied on a different dataset, future work suggested that the models can still be applied to other datasets to see it performed
[24]	2018	The study employed three models namely: SVM, Multinomial NB, and RF classifiers	SVM gave rugged and uniform outcomes in comparison with MNB and RF	There is still room for improvement in terms of performance accuracy
[25]	2019	The study used three classifiers like SVM, DT, and RF for the sentiment analysis	The DT had the highest performance accuracy of 91.8%	

(Continued)

Table 1 (continued)

Reference	Year	Models	Results/Contributions	Research gap
[26]	2021	NLP and supervised KNN classification algorithm	The results of the study revealed that Pfizer has 47.3%, Moderna has 46.2%, and AstraZeneca with 40.1% of positive sentiments, respectively	Performance metrics still need improvement, and the issue of an imbalance dataset still needs to be properly handled
[27]	2021	NVivo software	It was found that the number of tweets on distant care delivery skyrocketed after the coronavirus pandemic outbreak	Performance metrics still need improvement, and the issue of an imbalance dataset still needs to be properly handled
[28]	2020	The study used various models like DT, MLP Classifier, XGB, SVM, Multinomial LR, Gaussian NB, and K-NN Classifiers	The best outcomes from the analysis were gotten with 77.5% correctness of the LR model. The performance level of the analysis was between 42% and 85%	Performance metrics still need improvement
[29]	2021	A hybrid CNN-LSTM	The study revealed an accuracy of 91.3% and outperformed ML-based models in sentiment analysis	Performance metrics still need improvement using other models
[30]	2017	Gated multimodal embedding LSTM with temporal attention	GME-LSTM (A) better modeled the multimodal organization of speech through time and carried out better sentiment analysis	The issue of an imbalance dataset still needs to be properly handled
[31]	2016	GME-LSTM(A) Recursive deep learning	Outcomes show that AROMA achieved remarkable performance boosts compared to the standard RAE model. It also performed better than many well-known methods used in the research	

3 Methodology

3.1 Dataset Specification

The dataset used in this study was obtained from [kaggle.com](https://www.kaggle.com) [32]. This dataset includes 1493 luxury hotels scored in the CSV format and 515,000 customer reviews of English sentences. For further investigation, the exact location of the hotels is also provided. Negative Review, Positive-Review, and Reviewer Score are the dataset attributes that were employed in this study, and for a proper analysis, a new score called was introduced in the study called Neutral-Review. Table 2 provides a summary of the dataset utilized in the study along with details for each attribute. The dataset can be downloaded from <https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe>

Table 2: Details of the experiment-related datasets

Features	Description
Hotel_Address	Address of hotel.
Review_Date	The date that the reviewer published the relevant review.
Average_Score	Based on the most recent comment from the previous year, we determined the hotel's average score.
Hotel_Name	Hotel name.
Reviewer_Nationality	Nationality of the reviewer.
Negative_Review	Negative review the reviewer gave to the hotel. If the reviewer does not give a negative review, then it should be: 'No Negative'.
Review total negative word counts	Total number of words in the negative review.
Positive_Review	Positive review the reviewer gave to the hotel. If the reviewer does not give a negative review, then it should be: 'No Positive'.
Review Total Positive Word Counts	Total number of words in the positive review.
Reviewer_Score	Score the reviewer has given to the hotel, based on his/her experience.
Total number of reviews reviewer has given	Number of reviews the reviewers has given in the past.
Total number of reviews	The total number of valid reviews the hotel has.
days since review	The duration between the review date and the scrape date.
Tags	Tags the reviewer gave the hotel.
Additional number of scoring	There are also some guests who just made a scoring on the service rather than a review. This number indicates how many valid scores without review in there.
Lat	Latitude of the hotel.
Lng	Longitude of the hotel.

3.2 The Proposed Model

Over the past few years, the usage of a Naive Bayes-based classifier has drawn a lot of attention [33]. The classification accuracy of Naive Bayes may be higher than that of conventional classifiers. Because of this, this study seeks to classify hotel reviews using Naive Bayes. Fig. 1 depicts the proposed model framework. The Naive Bayes classifier and word embedding rank as the proposed method's two most crucial elements.

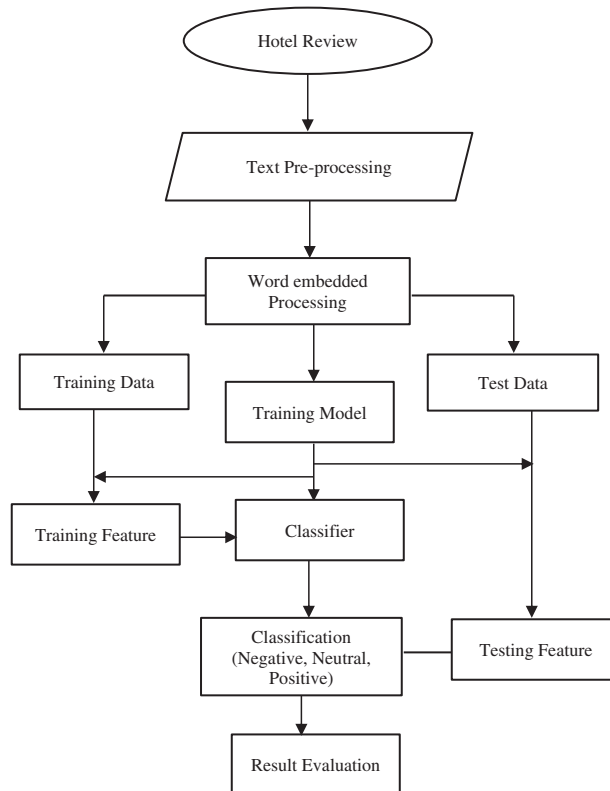


Figure 1: The proposed flowchart

3.2.1 Naive Bayes' Classifier

The Naive Bayes classifier has been shown to be effective in a number of situations despite its relative simplicity and solid assumptions. With the use of Bayesian classification, one can combine observable data, prior knowledge, and useful learning methods [34]. The primary idea of the Naive Bayes technique is to identify the potential categories from a text document. Using the combined probabilities of the terms and classes. The philosophy of the word sovereignty serves as its cornerstone. Algorithm 1 gives the detailed Naive Bayes algorithm.

The conditional probability of sentiment is given as:

$$p(\text{sentiment/sentence}) = \frac{p(\text{sentence/sentiment}) p(\text{sentiment})}{p(\text{sentence})}$$

Algorithm 1: Naive Bayes Algorithm

```

1: train Naive Bayes (D, C) returns log P(c) and logP(w|c);
2: each class c ∈ C # calculate P(c) terms
   Ndoc = number of document in D
   Nc = number of document from D in class c logprior[c] ← log
   V ← vocabulary of D;
   bigdoc[c] ← append(d);
For d ∈ D with class c
3: For each word w in V
   # Calculate P(w|c) terms
   count(w,c) ← # of occurrences of w in bigdoc[c]
4: loglikelihood[w,c] ← log  $\frac{\text{count}(w,c) + 1}{\sum_{w_2 \text{ in } v} (\text{count}(w_2,c) + 1)}$ 
   return logprior, loglikelihood, V
5: function TEST NAIVE BAYES (testdoc, logprior, loglikelihood, C, V) returns best c
6: for each class c ∈ C
   sum[c] ← logprior[c]
7: for each position i in testdoc
   word ← testdoc[i]
8:   if word ∈ V sum[c] ← sum[c] + loglikelihood[word,c]
   return argmaxc sum[c]
9:   else
   D with class c
10: compute the P(c) terms;
11: return P

```

3.2.2 Pre-Processing

User reviews are typically written informally, with no regard for grammar or other writing conventions. Before using the data to evaluate the suggested technique, we do a pre-processing procedure to remove such problems from the data. The pre-processing stage includes tokenization, hash tag removal, extra white space removal, special symbol removal, and customized stop word removal [34]. While it is customary to eliminate “Not” as a stop word from evaluations, we have chosen to leave it in because it is an essential component of determining a sentence’s polarity value. For sentiment analysis and pre-processing procedures, Python programming language is employed.

3.2.3 Word Embedding

Any learning model cannot directly comprehend the raw text. In order to train the learning model on it, it should be transformed into some numerical form. According to authors in [35], word embedding is currently the best method for converting unstructured text into numbers. In addition to vectorizing the raw text, it also creates connections between the words. The first layer in the suggested framework is the embedding layer, which accepts user reviews and transforms them into vector form for training. The maximum sequence length is determined in accordance with the length of the maximum word review, and the embedding dimension is set to 100.

- (i) Collection of data to be used: The data were obtained and downloaded from kaggle.com
- (ii) Prepare the text for analysis: The text obtained from the dataset will then be pre-processed in preparation for analysis. The majority of documents are created and stored in order to enhance comprehension. Machines sometimes find document analysis difficult. Before beginning text analysis research, there is often the need to clean and parse the text to ensure it is in a format that a computer can use (machine-readable).
- (iii) Detecting the sentiment in the text: The polarities in the text are then identified by the training process: This involves the framework learning to an input (for example, a word) with its matching result (label) as per the specimen used for training. The feature extractor moves the words into a feature vector. Braces of feature vectors, as well as labels, are then put into the classifier with the aim of developing a framework.
- (iv) Classifying sentiment: This has to do with the prediction process which involves using the feature extractor for changing unobserved texts into feature vectors. These feature vectors are then put into the framework, that produces previsioned labels.
- (v) Presenting the output in a simple manner. This is where data visualization comes in. This involves converting data into an ocular format like maps, charts, or graphs, to enhance comprehension for people. This would enable people to identify patterns, trends, and anomalies in the data.

The proposed flowchart and sentiment analysis framework is shown in [Figs. 1 and 2](#).

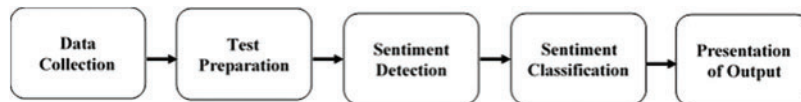


Figure 2: Sentiment analysis framework

3.3 Steps Taken in Analysis

First, the python libraries are imported as they have useful functions in helping with the analysis of our dataset. Importing python libraries into the Jupyter notebook. Then, data is read into the jupyter notebook as seen in [Fig. 3](#).

The number of negative reviews is shown in [Fig. 4](#).

Thereafter, the ratings are shown in [Fig. 5](#), three are assigned as negative with the aim of training the algorithm.

Then, reviews with ratings of five are assigned as positive with the aim of training the algorithm as shown in [Fig. 6](#).

The length of the dataset is calculated as shown in [Fig. 7](#).

Then, the positive and negative responses are combined as shown in [Fig. 8](#).

In addition, a new column is added for sentiment [Fig. 9](#).

Then, the positive and negative responses are mixed together in the notebook and displayed as shown in [Fig. 10](#).

Thereafter, the data were grouped as training as well as the validation set, and the reviews were vectorized (converted to numbers) using a pipeline.


```
In [2]: data = pd.read_csv('tripadvisor_hotel_reviews.csv')
data.head(20)
```

Out[2]:

	Review	Rating
0	nice hotel expensive parking got good deal sta...	4
1	ok nothing special charge diamond member hillo...	2
2	nice rooms not 4* experience hotel monaco seat...	3
3	unique, great stay, wonderful time hotel monac...	5
4	great stay great stay, went seahawk game aweso...	5
5	love monaco staff husband stayed hotel crazy w...	5
6	cozy stay rainy city, husband spent 7 nights m...	5
7	excellent staff, housekeeping quality hotel ch...	4
8	hotel stayed hotel monaco cruise, rooms genero...	5
9	excellent stayed hotel monaco past w/e delight...	5
10	poor value stayed monaco seattle july, nice ho...	2
11	nice value seattle stayed 4 nights late 2007. ...	4
12	nice hotel good location hotel kimpton design ...	4
13	nice hotel not nice staff hotel lovely staff q...	3
14	great hotel night quick business trip, loved l...	4
15	horrible customer service hotel stay february ...	1
16	disappointed say anticipating stay hotel monac...	2

Figure 3: Sample data reading from the dataset into the Jupyter notebook

```
In [6]: positive = five.loc[:len(negative)]
positive
```

Out[6]:

	Review	Rating
0	unique, great stay, wonderful time hotel monac...	5
1	great stay great stay, went seahawk game aweso...	5
2	love monaco staff husband stayed hotel crazy w...	5
3	cozy stay rainy city, husband spent 7 nights m...	5
4	hotel stayed hotel monaco cruise, rooms genero...	5
...
3210	perfect hotel hotel does not really need glowi...	5
3211	perfect hotel small hotel comfortable perfect ...	5
3212	ordinary location extraordinary hotel know lov...	5
3213	classy indulgence awesome experience, staff n'...	5
3214	first-rate experience stay library hotel wife ...	5

3215 rows × 2 columns

Figure 4: Number of negative reviews

```
In [4]: negative = data.loc[data['Rating']<3]
negative = negative.reset_index(drop=True)
negative
```

Out[4]:

	Review	Rating
0	ok nothing special charge diamond member hilt...	2
1	poor value stayed monaco seattle july, nice ho...	2
2	horrible customer service hotel stay february ...	1
3	disappointed say anticipating stay hotel monac...	2
4	great location need internally upgrade advanta...	2
...
3209	deceptive staff deceptive desk staff claiming ...	2
3210	not impressed unfriendly staff checked asked h...	2
3211	ok just looks nice modern outside, desk staff ...	2
3212	hotel theft ruined vacation hotel opened sept ...	1
3213	people talking, ca n't believe excellent ratin...	2

3214 rows × 2 columns

Figure 5: Negative reviews

```
In [5]: five = data.loc[data['Rating']==5]
five = five.reset_index(drop=True)
five
```

Out[5]:

	Review	Rating
0	unique, great stay, wonderful time hotel monac...	5
1	great stay great stay, went seahawk game aweso...	5
2	love monaco staff husband stayed hotel crazy w...	5
3	cozy stay rainy city, husband spent 7 nights m...	5
4	hotel stayed hotel monaco cruise, rooms genero...	5
...
9049	best best reason affordable better hotels city...	5
9050	night not trust not disappointed stay hotel, s...	5
9051	great play stay stay loyal inn package deal ha...	5
9052	great choice wife chose best western quite bit...	5
9053	best kept secret 3rd time staying charm, not 5...	5

9054 rows × 2 columns

Figure 6: Positive reviews

```
In [3]: len(data)
```

Out[3]: 20491

Figure 7: Length of the dataset

Combine negative and positive responses

```
In [7]: all_data = pd.concat([negative, positive], axis=0)
all_data
```

```
Out[7]:
```

	Review	Rating
0	ok nothing special charge diamond member hilito...	2
1	poor value stayed monaco seattle july, nice ho...	2
2	horrible customer service hotel stay february ...	1
3	disappointed say anticipating stay hotel monac...	2
4	great location need internally upgrade advanta...	2
...
3210	perfect hotel hotel does not really need glowi...	5
3211	perfect hotel small hotel comfortable perfect ...	5
3212	ordinary location extraordinary hotel know lov...	5
3213	classy indulgence awesome experience, staff n'...	5
3214	first-rate experience stay library hotel wife ...	5

6429 rows × 2 columns

Figure 8: Combination of positive and negative reviews

```
In [8]: all_data['Sentiment'] = np.where(all_data['Rating'] == 5, 'Positive', 'Negative')
all_data
```

```
Out[8]:
```

	Review	Rating	Sentiment
0	ok nothing special charge diamond member hilito...	2	Negative
1	poor value stayed monaco seattle july, nice ho...	2	Negative
2	horrible customer service hotel stay february ...	1	Negative
3	disappointed say anticipating stay hotel monac...	2	Negative
4	great location need internally upgrade advanta...	2	Negative
...
3210	perfect hotel hotel does not really need glowi...	5	Positive
3211	perfect hotel small hotel comfortable perfect ...	5	Positive
3212	ordinary location extraordinary hotel know lov...	5	Positive
3213	classy indulgence awesome experience, staff n'...	5	Positive
3214	first-rate experience stay library hotel wife ...	5	Positive

6429 rows × 3 columns

Figure 9: New column added for sentiment

```
In [9]: all_data = all_data.sample(frac=1)
all_data=all_data.reset_index(drop=True)
all_data.head(12)
```

```
Out[9]:
```

	Review	Rating	Sentiment
0	entertainment team visited excellence punta ca...	5	Positive
1	good location problems hotel good location clo...	2	Negative
2	heaven earth stayed different hotels inns far ...	5	Positive
3	no wonder hotel 1, lowdown, stay hotel ruin ho...	5	Positive
4	comfortable nice location summer months stayed...	5	Positive
5	pay short place chosen stay prepared, cabbies ...	1	Negative
6	nice soooo quiet comfortable quiet room, block...	5	Positive
7	start, just got meant dream holiday upgraded r...	1	Negative
8	absolute best, fantastic week, husband enjoyin...	5	Positive
9	average resort stayed occidental grand flamenc...	2	Negative
10	not expected, excited come puerto rico stay wo...	2	Negative
11	great hotel lovely area recently stayed night ...	5	Positive

Figure 10: Mixture of positive and negative reviews

The Performance metrics

These metrics were applied in evaluating the classifier:

- (i) Accuracy: measures the ratio of correct previsions to the total previsions.
- (ii) Precision: measures the ratio of true positives to the total positives predictions.
- (iii) Recall measures the ratio of useful outcomes gotten to the total useful outcomes.

Performance:

$$Precision = \frac{TP}{TP + FP}$$

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

$$Recall = \frac{TP}{TP + FN}$$

where

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

4 Experimental Results

Simulations of the sentiment analysis algorithm were performed using Jupyter Notebook, an open-source, open-standards data analysis software that offers services for interactive computing across dozens of programming languages including the ones mostly used for sentiment analysis, Python, and R. The Naive Bayes algorithm was implemented on the sentiment analysis dataset. The results were then examined. The work was carried out on a Jupyter Notebook, a computer with a processor speed of 1.99 GHz, RAM size of 4 GB and hard disk of size 500 GB, and a Windows 8.1

(64 bits) Operating system. The results of the performance evaluation metrics used in this study are stated based on accuracy, precision, and recall.

4.1 Result of Proposed System

With an accuracy of 96.08%, the model is said to classify hotel reviews with a high level of accuracy. Recall, precision, and F1-score are also calculated as shown in [Table 3](#).

Table 3: The performance matric results

	Accuracy	Precision	Recall	F1-score	Support
Negative	0.96	0.97	0.95	0.96	790
Positive	0.96	0.95	0.97	0.96	818
Macro average	0.96	0.96	0.96	0.96	1608
Weighted average	0.96	0.96	0.96	0.96	1608

4.2 The Confusion Matrix

From [Fig. 11](#), which presents hotel review dataset on Naive Bayes model, the classification performance for class negative represented by 0.0 was 100% because all instances were correctly classified. For class label neural represented by 1.0, 98 instances were correctly classified out of 96, 1 instance were misclassified as positive and 1 instance were misclassified as negative. For positive, the classification performance was 100%.

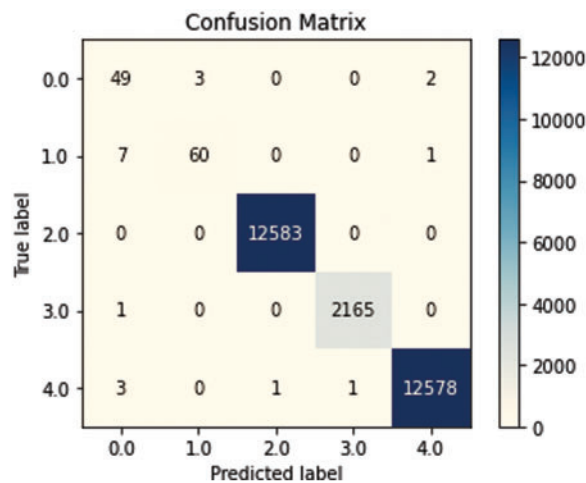


Figure 11: Confusion matrix

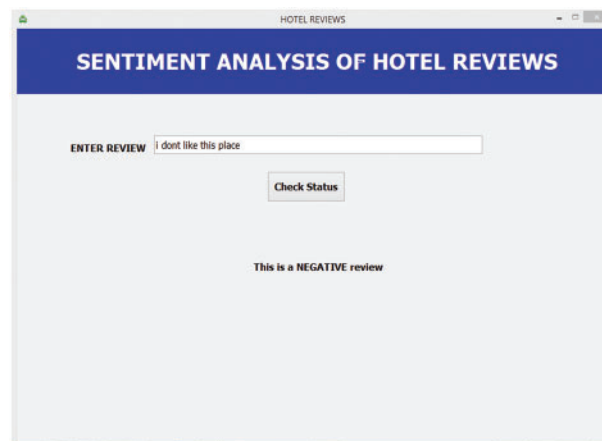
The comparison of the proposed model with other state-of-the-art classifiers is given in [Table 4](#).

Table 4: The comparison of the proposed model with other classifiers

Model	Accuracy	Precision	Recall	F1-score
RNN	86.00	0.73	0.63	0.60
GRU	90.00	0.79	0.74	0.76
LSTM	89.00	0.78	0.70	0.70
Bi-LSTM	89.00	0.81	0.70	0.74
BERT-RF	92.00	0.86	0.82	0.84
Proposed model	96.08	0.96	0.96	0.96

Table 4 shows the comparison of the proposed model with some state-of-the-art models used to review hotel sentiment. The results of the proposed model revealed that the classifier performed better across all the performance metrics than the models used for comparison. From Table 3, the results produce an accuracy of 96.08%, precision of 0.96, F1-score of 0.96, and recall of 0.96. The second better model is BERT-RT with an accuracy of 92.36%, precision of 0.86, F1-score of 0.82, and recall of 0.84. The worst of all the classifiers is the RNN with an accuracy of 86.00%, precision of 0.73, F1-score of 0.60, and recall of 0.63. the performance of the proposed model is due to the positive sentiment polarity of the richness of reviews (sentiment), and the positive reviews are much greater when compared with other classes under reviews for various metrics. The bottommost of all is the neutral class because of its scarceness of reviews in the dataset.

As shown in Fig. 12, the user will input the review and click on the ‘check status’, then the application would detect the polarity and display the kind of review it is whether positive or negative. If it is a positive review, the application will display “This is a POSITIVE review” “NEUTRAL” or “This is a NEGATIVE review” as shown in Fig. 12. After a number of tests, it was discovered that the application detects the polarity of the reviews with a high level of accuracy.

**Figure 12:** GUI application for sentiment analysis of hotel reviews

5 Discussion

The hotel recommendation system provides us with an interesting research area by assisting a user to select a suitable hotel based on his/her requirements and availability from the online hotel reviews given by the customer. This guarantees that clients may choose the optimal option depending on the submitted inquiry. Therefore, this study presents a model for a user inquiry-based decision support system that provides information on hotels and ratings related to them, if necessary, as output in response to user inquiries. Unlike earlier research that employed aspect-based Sentiment Analysis, the suggested model concentrates on categorizing reviews' initial perceptions and then classifying the reviews based on various aspect criteria. Finally, a suitable hotel is chosen, and its reviews are displayed in accordance with the user inquiry input.

The proposed model experiments are based on a hotel review dataset crawled from kaggle.com. The data comes from Booking.com scraping. The entire content of the file is already freely accessible to everyone. Please be aware that Booking.com is the original owner of the data. In the task of sentiment analysis, the proposed model revealed an accuracy of 96.08%, and 96% across other performance evaluations for precision, recall, and F1-score, respectively, which are significantly greater than comparable cutting-edge models. There are numerous ways the proposed system might be applied to the tourism systems. Instead of exhaustively searching for hotels with superior reviews that are available on different online portals, the suggested model can be used by tourists to locate hotels in a specific area, with specific elements like improved personnel, value, reviews, and evaluations among others.

Notwithstanding the proposed model's positive results, there are not many restrictions that can be fixed by future research. Well first of all, because recently published reviews are more frequently read, we could learn more by tracking changes in consumer acceptance of reviews over time. Secondly, approaches to addressing class disparity and imbalance like Bootstrap Aggregating (Bagging), or Random Under-sampling and Over-sampling should be considered on the hotel reviews datasets, the classification of the reviews using positive, neutral, and negative alone may not be enough to get a better classification accuracy [1]. The application of deep learning could be considered for better classification results, especially on any huge hotel review datasets. The results of the categorization could only be verified mechanically, which could be subject to human errors. Hence, approaches like the Dempster-Shafer method or the fuzzy ensemble method might be tested in order to create method ensembles. The existence of labeled data with more qualifications might aid in lowering such human-level errors. Additionally, if suitable multilingual datasets are discovered, the vector encoding mode can then be changed to apply the suggested approach to such datasets.

5.1 Contribution in Terms of Managerial and Theoretical Perspective

This study proposed a method for a user inquiry-based classification model that provides information on hotel reviews and provided a recommendation as output in response to user inquiries. The study determined the varying significance of the inquiry found in user input while defining traveler sentiment as an attitude construct and describing sentiment classification. This study proposed a Naive Bayes model in sentiment classification as positive, neutral, or negative. The model has been implemented on the hotel reviews dataset collected from a publicly online available source. The hotel qualitative review class was used as input for the proposed model. The proposed method results were evaluated based on various performance metrics from the reviewed literature. The results revealed that the proposed model performed better than other state-of-the-art models compared with hotel reviews sentiment analysis. A managerial orientation is provided by this study, which is important for the provider of

hotel services to better serve their clients and travelers that may want to lounge. This study finds that managers have to concentrate on consumers' welfare, food service, and rooms to be lounged through their online comments to provide high-quality information. The purpose of this study investigation is also to point toward the effectiveness of qualitative material in fostering customer attitude. This study also suggests ways to gather feedback and data to assess client opinions, perceptions, and satisfaction. Any hotel service provider organization can take into account the usefulness of the suggested model, which can help categorize hotel reviews for business growth based on profitability, enlargement, and positive responses based on favorable feedback.

6 Conclusion

Choosing a proper hotel with a good environment and affordable price is necessary for hotel users and helping them to choose a proper hotel from online hotel reviews becomes important, and this gives an interesting research field called hotel endorsement system. Customers will be able to make the best travel options possible depending on the submitted inquiry. Therefore, this study presented a framework for a hotel recommendation model that gives hotel users and reviewers the choice of the correct and appropriate hotel when traveling. This proposed model provided recommendations as an output based on the user queries. Sentiment analysis was performed on a dataset that includes 1493 premium hotels from across Europe and 515,000 customer reviews and ratings. The performance of the model is measured using accuracy, precision F1-score, and recall. The reviews were pre-processed; Naive Bayes algorithm was used to train the dataset. Then, it classified the reviews in the dataset with high accuracy of 0.96. From the experimental results, it can be concluded that the proposed model can detect sentiment polarity in hotel reviews using the Naive Bayes classifier with a good accuracy result. It also gave a good precision and recall score. In the future, this algorithm will be applied in the cryptocurrency space to see how accurate this classifier is going to be in classifying cryptocurrency tweets to ascertain the perception of the public on cryptocurrencies.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ray, B., Garain, A., Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, 98, 106935. <https://doi.org/10.1016/j.asoc.2020.106935>
2. Ponnappureddy, S., Priskin, J., Vinzenz, F., Wirth, W., Ohnmacht, T. (2020). The mediating role of perceived benefits on intentions to book a sustainable hotel: A multi-group comparison of the Swiss, German and USA travel markets. *Journal of Sustainable Tourism*, 28(9), 1290–1309.
3. Dey, L., Chakraborty, S., Biswas, A., Bose, B., Tiwari, S. (2016). Sentiment analysis of review datasets using Naive Bayes and K-NN classifier. arXiv preprint arXiv:1610.09982.
4. Surya, P. P., Subbulakshmi, B. (2019). Sentimental analysis using Naive Bayes classifier. *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, pp. 1–5. Vellore, India.
5. Jain, A. P., Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pp. 628–632. Bangalore, India, IEEE.

6. Alluri, N. V., Krishna, N. D. (2021). Multi modal analysis of memes for sentiment extraction. *2021 Sixth International Conference on Image Information Processing (ICIIP)*, pp. 213–217. Shimla, India.
7. Mahbub, N. I., Islam, M. R., Al Amin, M., Islam, M. K., Singh, B. C. et al. (2021). Sentiment analysis of microblogging dataset on coronavirus pandemic. *2021 5th International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1–6. Khulna, Bangladesh.
8. Abro, S., Shaikh, S., Abro, R. A., Soomro, S. F., Malik, H. M. (2020). Aspect based sentimental analysis of hotel reviews: A comparative study. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 4(1), 11–20. <https://doi.org/10.30537/sjcms>
9. Zou, Y., Wang, J., Lei, Z., Zhang, Y., Wang, W. (2021). Sentiment analysis for necessary preview of 30-day mortality in sepsis patients and the control strategies. *Journal of Healthcare Engineering*, 2021. <https://doi.org/10.1155/2021/1713363>
10. Jain, P. P. K., Srivastava, G., Lin, J. C. W., Pamula, R. (2022). Unscrambling customer recommendations: A novel LSTM ensemble approach in airline recommendation prediction using online reviews. *IEEE Transactions on Computational Social Systems*, 9(6), 1777–1784. <https://doi.org/10.1109/TCSS.2022.3200890>
11. Jain, P. K., Pamula, R., Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41, 100413. <https://doi.org/10.1016/j.cosrev.2021.100413>
12. Jain, P. K., Prakash, A. (2022). Predicting consumer recommendation decisions from online reviews: A rough set approach. *Wireless Personal Communications*, 1–18. <https://doi.org/10.1007/s11277-022-09719-6>
13. Schmitt, M., Steinheber, S., Schreiber, K., Roth, B. (2018). Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. arXiv preprint arXiv:1808.09238.
14. Ruder, S., Ghaffari, P., Breslin, J. G. (2016). A hierarchical model of reviews for aspect-based sentiment analysis. arXiv preprint arXiv:1609.02745.
15. Barnes, J., Lambert, P., Badia, T. (2018). Multibooked: A corpus of basque and Catalan hotel reviews annotated for aspect-level sentiment classification. arXiv preprint arXiv:1803.08614.
16. Yuan, J., Wu, Y., Lu, X., Zhao, Y., Qin, B. et al. (2020). Recent advances in deep learning based sentiment analysis. *Science China Technological Sciences*, 63(10), 1947–1970. <https://doi.org/10.1007/s11431-020-1634-3>
17. Tusar, M. T. H. K., Islam, M. T. (2021). A comparative study of sentiment analysis using NLP and different machine learning techniques on US airline twitter data. *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, pp. 1–4. Khulna, Bangladesh.
18. Nguyen, K. T. T., Huynh, S. K., Phan, L. L., Pham, P. H., Nguyen, D. V. et al. (2021). Span detection for aspect-based sentiment analysis in Vietnamese. arXiv preprint arXiv:2110.07833.
19. Gharzouli, M., Hamama, A. K., Khattabi, Z. (2022). Topic-based sentiment analysis of hotel reviews. *Current Issues in Tourism*, 25(9), 1368–1375. <https://doi.org/10.1080/13683500.2021.1940107>
20. Mozetič, I., Grčar, M., Smailović, J. (2016). Multilingual twitter sentiment classification: The role of human annotators. *PLoS one*, 11(5), e0155036. <https://doi.org/10.1371/journal.pone.0155036>
21. Cachola, I., Holgate, E., Preoțiu-Pietro, D., Li, J. J. (2018). Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media. *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 2927–2938. Santa Fe, New Mexico, USA.
22. Bangyal, W. H., Qasim, R., Ahmad, Z., Dar, H., Rukhsar, L. et al. (2021). Detection of fake news text classification on COVID-19 using deep learning approaches. *Computational and Mathematical Methods in Medicine*, 2021. <https://doi.org/10.1155/2021/5514220>
23. Sitaula, C., Basnet, A., Mainali, A., Shahi, T. B. (2021). Deep learning-based methods for sentiment analysis on Nepali COVID-19-related tweets. *Computational Intelligence and Neuroscience*, 2021. <https://doi.org/10.1155/2021/2158184>

24. Aloufi, S., El Saddik, A. (2018). Sentiment identification in football-specific tweets. *IEEE Access*, 6, 78609–78621. <https://doi.org/10.1109/ACCESS.2018.2885117>
25. Saad, S. E., Yang, J. (2019). Twitter sentiment analysis based on ordinal regression. *IEEE Access*, 7, 163677–163685. <https://doi.org/10.1109/Access.6287639>
26. Shamrat, F. M. J. M., Chakraborty, S., Imran, M. M., Muna, J. N., Billah, M. M. et al. (2021). Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(1), 463–470. <https://doi.org/10.11591/ijeecs.v23.i1.pp463-470>
27. Iwendi, C., Mohan, S., Ibeke, E., Ahmadian, A., Ciano, T. (2022). COVID-19 fake news sentiment analysis. *Computers and Electrical Engineering*, 101, 107967. <https://doi.org/10.1016/j.compeleceng.2022.107967>
28. Osmanoğlu, U. Ö., Atak, O. N., Çağlar, K., Kayhan, H., Talat, C. A. N. (2020). Sentiment analysis for distance education course materials: A machine learning approach. *Journal of Educational Technology and Online Learning*, 3(1), 31–48. <https://doi.org/10.31681/jetol.663733>
29. Jain, P. K., Saravanan, V., Pamula, R. (2021). A hybrid CNN-LSTM: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5), 1–15. <https://doi.org/10.1145/3457206>
30. Chen, M., Wang, S., Liang, P. P., Baltrušaitis, T., Zadeh, A. et al. (2017). Multimodal sentiment analysis with word-level fusion and reinforcement learning. *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 163–171. Glasgow, UK.
31. Al-Sallab, A., Baly, R., Hajj, H., Shaban, K. B., El-Hajj, W. et al. (2017). Aroma: A recursive deep learning model for opinion mining in arabic as a low resource language. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 16(4), 1–20. <https://doi.org/10.1145/3086575>
32. Jason L, L. (2021). 515K hotel reviews data in Europe. <https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe>
33. Jain, K., Kaushal, S. (2018). A comparative study of machine learning and deep learning techniques for sentiment analysis. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 483–487. Noida, India. <https://doi.org/10.1109/ICRITO.2018.8748793>
34. Awotunde, J. B., Misra, S., Adeniyi, A. E., Abiodun, M. K., Kaushik, M. et al. (2022). A feature selection-based K-NN model for fast software defect prediction. *International Conference on Computational Science and Its Applications*, pp. 49–61. Malaga, Spain.
35. Folorunso, S., Ogundepo, E., Basajja, M., Awotunde, J., Kawu, A. et al. (2022). FAIR machine learning model pipeline implementation of COVID-19 data. *Data Intelligence*, 4(4), 971–990. https://doi.org/10.1162/dint_a_00182