



Sentiment Analysis using various Machine Learning and Deep Learning Techniques

V. Umarani^{a,*}, A. Julian^a, J. Deepa^b

^aDepartment of Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India

^bDepartment of Information Technology, VelTech Institute of Technology, Chennai, Tamil Nadu, India

Abstract

Sentiment analysis has gained a lot of attention from researchers in the last year because it has been widely applied to a variety of application domains such as business, government, education, sports, tourism, biomedicine, and telecommunication services. Sentiment analysis is an automated computational method for studying or evaluating sentiments, feelings, and emotions expressed as comments, feedbacks, or critiques. The sentiment analysis process is automated by using machine learning techniques, which analyses text patterns faster. The supervised machine learning technique is the most used mechanism for sentiment analysis. The proposed work discusses the flow of sentiment analysis process and investigates the common supervised machine learning techniques such as Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, Support Vector Machine, Random Forest, K-nearest neighbor, Decision tree, and deep learning techniques such as Long Short-Term Memory and Convolution Neural Network. The work examines such learning methods using standard data set and the experimental results of sentiment analysis demonstrate the performance of various classifiers taken in terms of the precision, recall, F1-score, RoC-Curve, accuracy, running time and k-fold cross validation and helps in appreciating the novelty of the several deep learning techniques and also giving the user an overview of choosing the right technique for their application.

DOI:10.46481/jnsps.2021.308

Keywords: Convolution neural network, Long short term memory, Sentiment analysis, Supervised machine Learning, Deep learning

Article History :

Received: 21 July 2021

Received in revised form: 04 October 2021

Accepted for publication: 05 October 2021

Published: 29 November 2021

©2021 Journal of the Nigerian Society of Physical Sciences. All rights reserved.

Communicated by: W. A. Yahya

1. Introduction

In the present days, social media is a popular technology that uses micro blogging platforms to connect millions of people. People can freely express their thoughts, ideas, and views as short messages called tweets on many micro blogging platforms in social networks (like Twitter) and business websites or web forums [1]. Researchers gather these unstructured tweets and use a variety of methods to extract information from them.

This analysis of tweets or opinions provides predictions or measures in a variety of application domains such as business, government, education, sports, tourism, biomedicine and telecommunication services [2].

Sentiment analysis or opinion mining is the study of opinions and prediction. Sentiment analysis (SA) is one of the text mining approach that use natural language processing for binary text classification. Sentiment analysis can be performed in four levels based on the scope of text. They are document-level, sentence-level, aspect-level, and word level sentiment analysis [6]. In Document level SA, overall opinion of the document

*Corresponding author tel. no: +918610352527

Email address: umaranibharathy@gmail.com (V. Umarani)

about single entity is grouped into positive or negative. In sentence level SA, the opinion expressed in a sentence is classified as either positive or negative. In aspect level SA, opinions about entities are grouped based on specific entity elements. At word level SA, opinions about entities are grouped based on a specific word. In the proposed work, word level sentiment analysis model is developed for restaurant review data set based on machine-learning and deep learning algorithms to classify sentiments as positive or negative automatically.

The sections in this paper are organized as follows: Section 2 examines the various sentiment analysis models that have been discussed in the literature. Section 3 discusses the various machine learning techniques. Section 4 analyzes the performance of machine learning techniques used in sentiment analysis and finally discusses the conclusion and future work that need to be carried out in sentiment analysis.

2. Related Work

The researchers create several sentiment analysis models based on data collected from social networks, business websites, and dataset providers. Examples of these analysis models built from data include Amazon reviews, Tweets from Twitter micro blogging sites, Yelp travel recommendations, and Movie Reviews from IMDB, Kaggle, and others [3]. This section discusses some existing work as well as the benefits of their proposed model.

Endsuy [4] used Twitter Datasets to conduct exploratory data analysis on the US Presidential Election 2020. They compare the sentiment of location-based tweets to that of on-the-ground public opinion. They collect features such as latitude, longitude, city, country, continent, and state code using Open Cage API and sciSpacy NER. They use two datasets from Kaggle about Donald Trump and Joe Biden, both dated November 18, 2020. For lexicon-based feature extraction, they use a valence aware dictionary for sentiment reasoning (VADAR), and for classification, they use logistic regression machine learning approaches.

Bibi et al. [5] created a Cooperative Binary-Clustering Framework for sentiment analysis on Indigenous data sets using Twitter. They use majority voting to partition the data and combine single linkage, complete linkage, and average linkage approaches. Based on the confusion matrix, they divide the cluster into positive and negative. For feature selection, they use unigram, TF-IDF, and word polarity mechanisms. According to this analysis, the cooperative clustering approach outperforms the other individual partitioning techniques (75 %).

Cekik et al. [6] use a filter-based feature selection method called Proportional Rough Feature Selector (PRFS) for feature selection and test it with various classifiers such as SVM,DT,KNN, and Naive Bayes. PRFS uses rough set theory to determine whether documents belong to a specific class or not. It improves classifier performance at a 95% confidence level.

Peng et al. [7] proposed a sentiment model adversarial learning method. It is made up of three parts: a generator, a discriminator, and a sentiment classifier. To obtain efficient semantic

and sentiment information, the generator uses a multi-head attention mechanism. The Discriminator measures up the similarity of sentiment polarity in the generative vector and global vector produced by the Generator and Classifier, respectively. It uses gradients to update the weights in the generator, resulting in high-quality word embeddings. Word vectors with opposing sentiment contexts will be classified as fake vectors in this context. It avoids the issue of similar context words with polar opposite sentiments.

To improve the performance of aspect-based sentiment analysis, Tan et al. [8] devised an Aligning Aspect Embedding method (AAE). Using the cosine measure metric, the AAE method discovered the relationship between aspect-categories and aspect-terms. The AAE method effectively solves the misalignment problem in aspect-based sentiment analysis and improves sentiment analysis performance.

Jain et al. [9] uses the Apriori algorithm to minimize the feature set for sentiment analysis and developed a feature selection approach based on Association rule mining. For experiments, they used supervised classification methods such as naive bayes, support vector machine, random forest, and logistic regression. Kalaivani et al. [10] developed a hybrid feature selection method for sentiment analysis. For feature extraction, they use the Unigram and Bigram models, as well as the TF-IDF weighting technique. They use information gain to shrink the subset of features. To select optimal features, this method employs a genetic algorithm. Then, using various classifiers such as naive bayes and logistic regression, support vector machine (SVM), this hybrid model is put into practice. Based on this analysis, the SVM classifier outperformed the other classifiers.

Using statistical feature selection methods, Ghosh et al. [11] proposed an ensemble feature selection technique to improve the performance of the sentiment analysis process. To find the best feature set, they use information gain, the Gini index, and the Chi-square method. They utilized five distinct classifiers: Multinomial Nave Bayes, KNN, Maximum Entropy, Decision Tree, and Support Vector Machine.

Rodrigues et al. [12] formed a pattern-based method for extracting aspects and analyzing sentiment. Pattern analysis is used in this case to extract the explicit aspect syntactic pattern from product sentiments. It extracts the bigram features and uses Senti-Wordnet to determine the sentence's sentiment polarity. According to this analysis, the multi node clustering approach outperforms the single node clustering approach.

For tweet sentiment classification, Jianqiang et al. [13] created a GloVe-DCNN (Global Vectors for Word Representation – Deep Convolution Neural Network). They presented a method of word embedding that combines unigram and bigram feature vectors. A subset of sentiment features is formed by combining the twitter specific features vector and word sentiment polarity features. This feature set is used to train and predict sentiment classification labels in a deep convolution neural network. It resolves the data sparseness issue.

Imran et al. [14] use tweets from Twitter and the sentiment 140 dataset. They utilize the LSTM model to estimate sentiment polarity and emotion. For sentiment analysis, Li et al.

[15] developed lexicon integrated CNN family models. They implemented a sentiment padding approach to ensure that input data sizes are consistent and that the percentage of sentiment information in each row is increased. During neural network learning, the gradient vanishing problem between the input layer and the first hidden layer is solved using the sentiment padding method.

The detailed literature survey on machine learning methods, illustrates that the suggested methods help in analyzing the text patterns faster and they can be used to automate the sentiment analysis process [16, 17]. Deep learning handles large volume of data and it uses artificial neural network to analyze the text pattern faster and it can solve the misalignment problems by extracting local features from a sentiment. As a result, deep learning algorithm takes word embedding as input and provides high accuracy on these kinds of tasks. In the proposed work, the most widely used machine learning algorithms have been deployed to analyze the performance of deep learning models LSTM and CNN.

3. Methodology

Sentiment analysis process is carried out by several steps. First step is to collect data and perform preprocessing. In preprocessing step, data set is structured by normalization, stop word removal, stemming, and tokenization process [18]. Next step in sentiment analysis is to extract and select most relevant text features from the opinion. Feature extraction and feature selection methods are used for this purpose. These methods are used to reduce the number of input variables, avoid over fitting, decrease computational complexity or training time, and improve model accuracy [19]. Vectorization and word embedding methods are used for feature extraction [20]. Finally, machine learning techniques are used to classify or categories text as positive, negative, or neutral based on sentiment polarity of opinions. Machine learning techniques classify the sentiments based on training and test data set [21].

In this section, the most widely used machine learning classifiers namely Naïve Bayes, Logistic Regression, Random forest, Linear SVC (Support Vector Classifier) , K-nearest neighbor and Decision tree and deep learning technique such as Convolution Neural Network and Long Short term Memory have been investigated The basics of these methods are discussed in detail.

Theoretical Background

Naive Bayes

It is an intuitive approach and has good ability to work with small data with lower computation time for training and prediction [22]. It uses Bayes theorem for finding the probability of the event. Naive Bayes classifier is based on equation (1).

$$P(Y/X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n/Y) \times P(Y)}{P(X_1, X_2, \dots, X_n)} \quad (1)$$

Here $P(Y/X_1, X_2, \dots, X_n)$ is called the posterior probability of an output class Y given input features X_1, X_2, \dots, X_n .

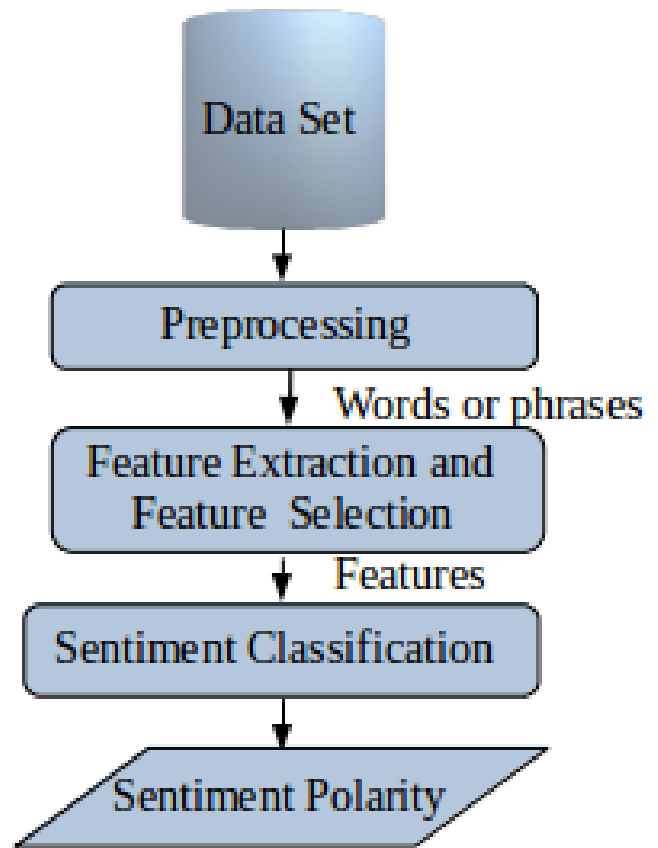


Figure 1: Sentiment analysis Process

$P(X_1, X_2, \dots, X_n/Y)$ is the likelihood of input features X_1, X_2, \dots, X_n given their class Y. $P(Y)$ is the prior probability of class Y. $P(X_1, X_2, \dots, X_n)$ is the marginal probability. The common distributions are the normal (Gaussian), multinomial, and Bernoulli distributions used in Naïve Bayes Classifier.

Logistic Regression (LR)

LR is a most widely used binary classifier which uses logistic function to predict the probability of observation output value (Y_i) with input X_i [23]. If $P(Y_i/X_i)$ is greater than 0.5 then predict class 1 otherwise considered it as 0. The logistic function is defined by equation (2).

$$P(Y_1/X) = \frac{1}{1 + e^{-(1+2)x}} \quad (2)$$

where $1+2$ are learning parameters, x is the training data, Y is observation output value and e is the Euler's number

Support Vector Classifier (SVC)

SVC classifies the data by finding hyper plane that maximize the margin between the predicted classes in the training data [24]. Support vector classifier is represented by equation (3).

$$f(x) = 0 + \sum_{j \in S} (j \times K(x_j, x_j)) \quad (3)$$

Where K is the kernel function which is used to compare the similarity of (x_j, x_j) observations and j is the learning parameter, S is the set of support vector observations. SVC uses linear kernel or radial basis function kernel to create a hyper plane decision boundary.

Decision tree classifier

Decision tree classification is a non-parametric supervised method suitable for both classification and regression problem. Decision tree contain set of decision nodes, each node specifies decision rule which create split to decrease the impurity recursively until all leaf nodes are belonging to specific class. Decision tree classifier uses various split quality measures like Gini impurity or Entropy or MAE (Mean Absolute Error) to decrease the impurity at a node. By default, It uses Gini impurity which is calculated by equation (4).

$$G(n) = 1 - \sum_{i=1}^c P_i^2 \tag{4}$$

Where $G(n)$ is the Gini- impurity at node n and P_i specifies the proportion of observed class c at node n .

Random Forest

Random Forest is a kind of ensemble learning method in which many decision trees are trained and each tree carries bootstrapped sample of observations called out of bag observations. For every observation, Random Forest learning algorithm calculates the overall score by comparing the observation’s true value with the prediction from a subset of trees not trained using that observation. This overall score is taken as a measure of random forest performance.

K-Nearest Neighbor

KNN is simple and most widely used classifier in supervised machine learning [24,25]. KNN algorithm first identifies the K closest neighbors based on the distance metrics and K -neighbors predict their classes based on k observations. Most widely used distance metrics as Euclidean or Manhattan distance or Minkowski distance. They are defined by equation (5), (6) and (7).

$$d_{euclid} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

$$d_{manhat} = \sum_{i=1}^n |(x_i - y_i)| \tag{6}$$

$$d_{minkowski} = \left(\sum_{i=1}^n |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{7}$$

where x_i and y_i are the observations and p is the hyper parameter

Convolution Neural Network

It is a type of feed forward neural network called ConvNet (Convolution Network) which uses hierarchical structure for fast feature extraction and classification [26,27]. Most important layers form a CNN is shown in Figure 2.

1. Convolution Layer - It extracts the features and creates feature maps. Convolution is a mathematical operation which is a measure of two overlapping function. It uses filter or kernels to detect important feature in the input. The number of output features (N_o) is calculated by equation (8) based on number of input features (N_i) and convolution properties like kernel size k , padding size p and stride size s .

$$N_o = \left\lfloor \frac{N_i + 2p - k}{s} \right\rfloor + 1 \tag{8}$$

Here, kernel size specifies the size of filter, stride specifies the rate at kernel increases and padding adjust the size of input according to the requirement of input matrix. When kernel detects the important features then it is stored in a feature map. Zero padding adds many zeros to fit the input matrix with kernel size. If Zero padding is used in convolution layer called wide convolution otherwise it is called narrow convolution. Other variants are zero padding called valid padding which keeps only valid parts and ignore other parts.

2. Pooling layer - It progressively reduces size of input representation and the number of weights needed. Hence it reduces overall number of input features and computation in the network as well as controls the over fitting. Pooling layer reduce the dimensionality of feature map without reducing important information. It uses max pooling, average pooling or sum pooling technique to reduce dimensionality. Max pooling takes the maximum value from the feature map. Average pooling finds the average value from the feature map and Sum pooling finds the sum of all values from the feature map. The new values form a matrix called pooled feature map. The pooling process is also called sub sampling or down sampling.
3. Flattening Layer – It convert the multidimensional data to single dimensional data. Flattening Layer is connected to fully connected layer for classification.
4. Fully Connected Layer – In Fully connected Layer, neurons have full connections to all activations in neurons of previous layer. Fully connected layers classify the feature which is obtained from the convolution and pooling layers into various classes based on the training set. It uses soft max activation function in output layer for text classification [33].Soft max function [24] is defined by equation 9. It takes input vector x_j of n real numbers and normalize the output into a range $[0,1]$.

$$softmax(x_j) = \frac{e^{x_j}}{\sum_{k=1}^n e^{x_k}} \tag{8}$$

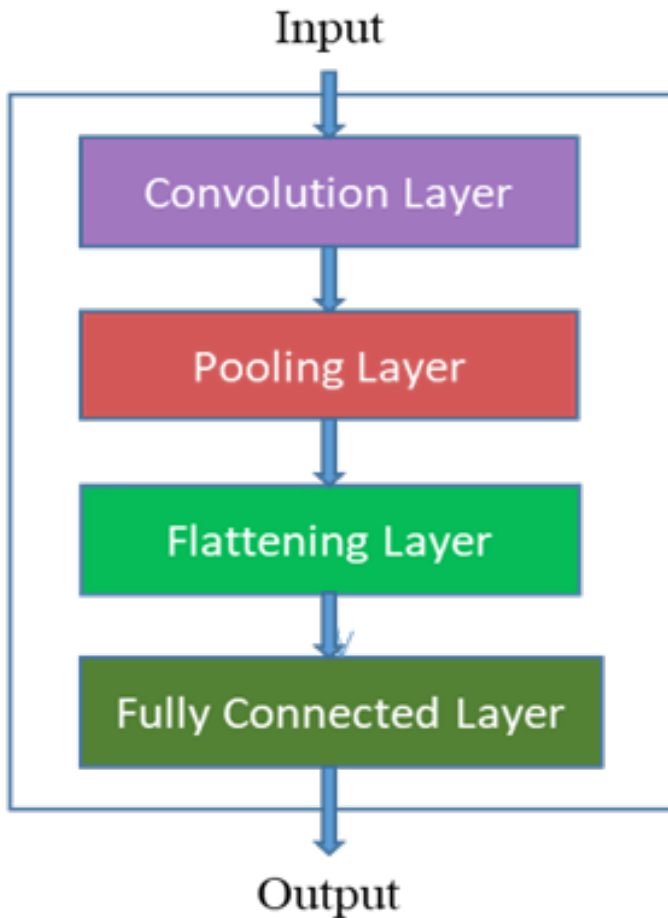


Figure 2: Layers in Convolution Neural Network

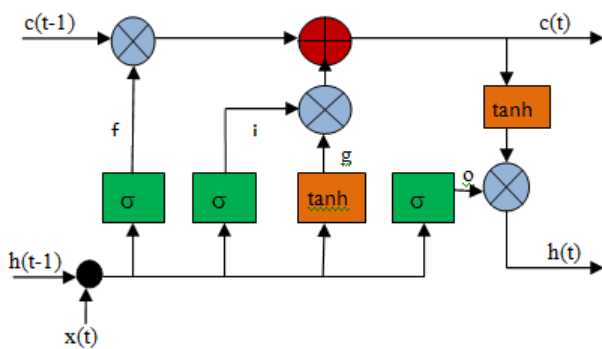


Figure 3: Hidden state Transformation in LSTM cell

LSTM control the data flow passes through cell with the help of gates. Input gate describe how much of newly calculated state for present input, Forget gate describe how much of previous state passes through it and output gate describe how much of internal state passes to next layer. These gates help to adjust the LSTM hidden state and every time, cell state c is calculated based on hidden state g , previous cell state, forget gate and input gate. Finally hidden state at time t is calculated by cell state with output gate. The transformations applied for hidden state is shown in Figure 3. The calculations done by the hidden state transformation are described by the following equations

$$i = ((W_i \times h_{t-1}) + (U_i \times x_t)) \tag{9}$$

$$f = ((W_f \times h_{t-1}) + (U_f \times x_t)) \tag{10}$$

$$o = ((W_o \times h_{t-1}) + (U_o \times x_t)) \tag{11}$$

$$g = \tanh((W_g \times h_{t-1}) + (U_g \times x_t)) \tag{12}$$

$$c_t = ((C_{t-1} \circ f) + (g * i)) \tag{13}$$

$$h_t = \tanh(c_t) \circ o \tag{14}$$

Here i, f, o, g, c_t and h_t refer input gate, forget gate, output gate and hidden state at time $t-1$, cell state at time t and hidden state at time t . Here σ is the sigmoid function which helps to adjust the output of gates between value 0 and 1. W, U are the weight matrix and transition matrix which helps to reduce the number of parameters learned by LSTM. x_t refer input at time t . The present input x_t and hidden state h_{t-1} are used to calculate internal hidden state g . Finally hidden state h_t is calculated from c_t and output gate value o .

4. Results and Discussion

Restaurant reviews from Kaggle were taken to examine the various supervised learning techniques for assessing the sentiment analysis process. This data set contains restaurant review text containing 1000 text reviews. Here Anaconda Python platform is used to evaluate and pre-process the restaurant reviews. 70% of the reviews are used for training, while 30% are used to test the supervised learning technique. Nltk is used for pre-processing, and Keras, Tensor flow (backend) is used to create LSTM (RNN with memory) and CNN neural network models[30]. Experiments are carried out by Google Collab which provide python development environment and run code in Google cloud. The Figure 4 shows the five rows of initial configuration of restaurant data set which contain 1000 reviews and its opinion classification positive or negative. The statistical summary of restaurant data set is shown in Figure 5.

Evaluation Parameters

Precision, recall, F1-score, accuracy, AUC score, and training time were used to assess the classifier's performance [31]. They are calculated by

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

Long Short Term Memory

LSTM is a type of feedback or recurrent neural network designed to solve various sequential and time series problems [28,29]. LSTM takes current observation and previous observation as input and store information in a gated cell or memory cell. Input data is propagated to LSTM layers to make a prediction. LSTM layer contain set of recurrently connected blocks, each one carries memory cell and three gates namely input, output and forget gate. Each memory cell has input weight, output weight and hidden state are used to process the input data.

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday off Rick Steve recommendation and loved it.	1
4	The selection on the menu was great and so were the prices.	1

Figure 4: First five rows of Restaurant Review Data set

```

Summary statistics of numerical features :
      Liked      Pos      Neg
count 1000.00000 1000.00000 1000.00000
mean   0.50000    0.50000    0.50000
std    0.50025    0.50025    0.50025
min    0.00000    0.00000    0.00000
25%    0.00000    0.00000    0.00000
50%    0.50000    0.50000    0.50000
75%    1.00000    1.00000    1.00000
max    1.00000    1.00000    1.00000
    
```

Total number of reviews: 1000

Figure 5: Statistical summary of dataset

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1\ score = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \tag{17}$$

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{18}$$

Here True positive (TP) refers to restaurant reviews that were initially classified as positive and are also expected to be positive. False Positive (FP) reviews are those that are initially identified as negative but are predicted to be positive. True Negative (TN) refers to restaurant reviews that were initially categorized as negative and that the classifier predicted to be negative. False Negative (FN) applies to restaurant reviews that are initially positive but are expected to be negative. AUC score specifies the area under ROC curve from prediction score.

Result analysis of Machine learning Technique for Sentiment Analysis

Initially, pre-processing is carried out in sentiment reviews by removing non-character data such as digits and symbols, as well as punctuation and converts the sentence into lowercase. After preprocessing, cleaned text reviews are converted to numerical data that contain sentiment tokens and sentiment score called feature vector. The feature vector is formed TF-IDF vectorizer. The TF-IDF stands for Term-Frequency Times Inverse Document-frequency and it assign weight to each word based on how often it appears in the review text. The TF-IDF refer term-frequency times inverse document-frequency which assign weight to each word based on the frequency of that word appeared in review text. After extracting features with a vectorizer, six machine learning classifiers are used for sentiment

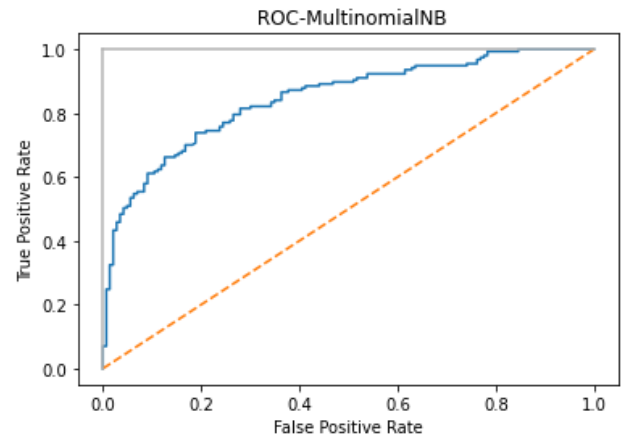


Figure 6a: ROC curve of Multinomial Naïve Bayes

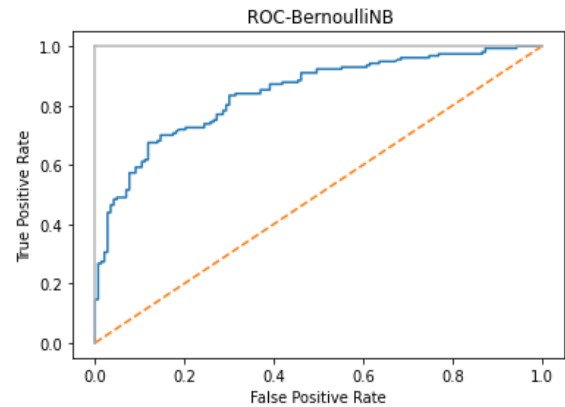


Figure 6b: ROC curve of Bernouli Naïve Bayes

analysis: naive bayes, logistic regression, random forest, linear SVC (Support Vector Classifier), K-nearest neighbor and decision tree. The performance of classifier is assessed by precision, recall, F1-score, accuracy, AUC score, ROC curve and training time. The classification report of classifier is shown in Table 1. From this table, the highest AUC score obtained for Naïve Bayes Classifier is 0.7642. So Naive Bayes model provide better prediction compared to other machine learning classifier for this restaurant data set. The table also shows that time taken for Naive bayes model is low compared to other machine learning algorithms. ROC curve of machine learning classifier is depicted in Figures 6a to 6f.

Result analysis of deep learning techniques in Sentiment Analysis

Sentiment analysis is carried out by deep learning techniques CNN and LSTM. After pre-processing the review text are converted into tokens. The result of tokenizer is shown in Figure 7. Following tokenization, the sentiment text is passed to word2vec model which converts word to vectors. The total number of words obtained for training data is 5118 words total, with a vocabulary size of 1839 and maximum sentence length is 18, while the total number of words obtained for test data is

Table 1: Performance Analysis of Machine Learning Classifiers for Sentiment analysis

Classifier	Class	Classification Report				Accuracy		AUC Score	Time to train (Sec-onds)
		Precision	Recall	F1-score	Support	Train	Test		
Naive Bayes Multinomial	0	0.74	0.78	0.76	256	0.76	0.7633	0.7642	0.009278
	1	0.79	0.75	0.77	244				
Naive Bayes Bernouli	0	0.72	0.79	0.76	256	0.76	0.756	0.758	0.020505
	1	0.79	0.73	0.76	244				
Logistic Re- gression	0	0.67	0.87	0.75	256	0.73	0.73	0.736	0.06633
	1	0.83	0.61	0.70	244				
Random Forest	0	0.65	0.93	0.76	256	0.72	0.7233	0.732	0.57832
	1	0.89	0.54	0.67	244				
Linear SVC	0	0.69	0.82	0.75	256	0.74	0.74	0.743	0.014712
	1	0.80	0.67	0.73	244				
KNN	0	0.70	0.72	0.71	256	0.72	0.716	0.716	0.079166
	1	0.74	0.71	0.72	244				
Decision Tree	0	0.65	0.80	0.72	256	0.70	0.70	0.704	0.1707437
	1	0.77	0.61	0.68	244				

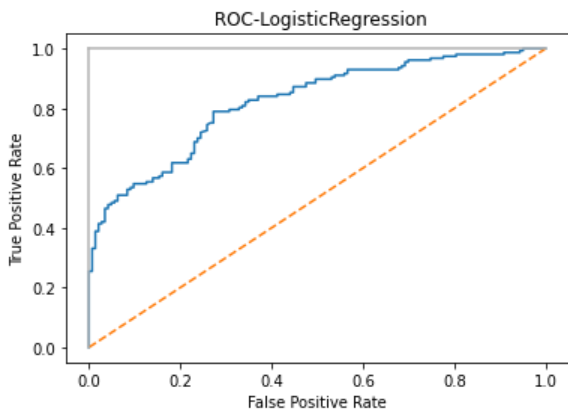


Figure 6c: ROC curve of Logistic Regression

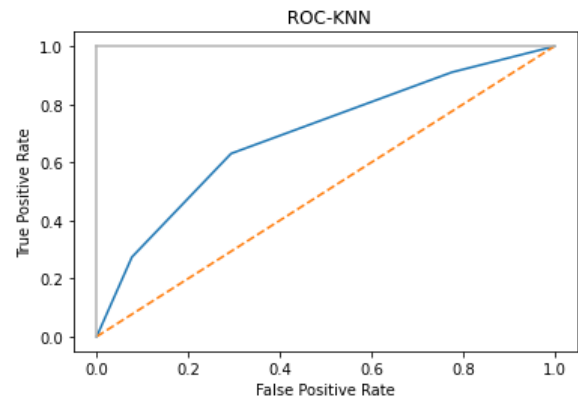


Figure 6e: ROC curve of KNN

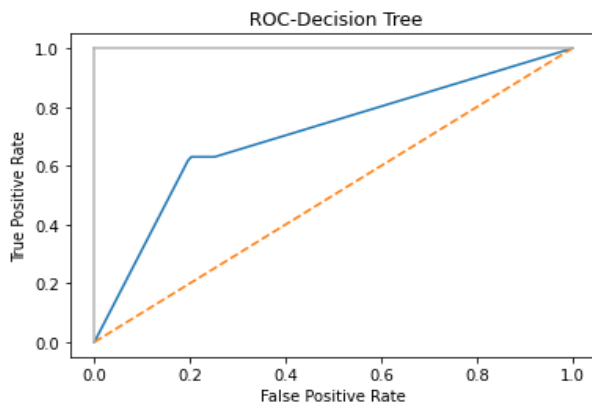


Figure 6d: ROC curve of Decision Trees

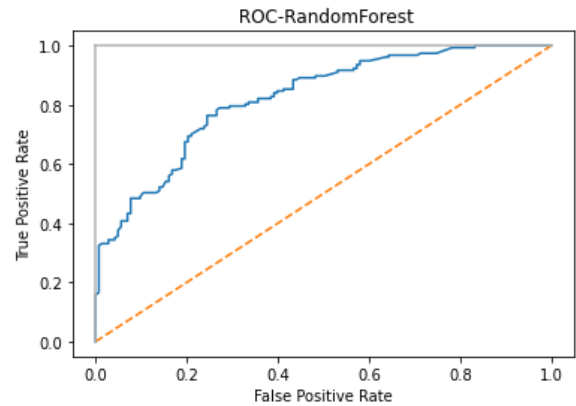


Figure 6f: ROC curve of Random Forest

574 words, with a vocabulary size of 415 and maximum sentence length of review text is 15.

The CNN model passes the input data to series of layers. It

	Text_Final	Tokens	Liked	Positive	Negative
0	wow loved place	[wow, loved, place]	1	1	0
1	crust not good	[crust, not, good]	0	0	1
2	not tasty texture nasty	[not, tasty, texture, nasty]	0	0	1
3	stopped late may bank holiday rick steve recommendation loved	[stopped, late, may, bank, holiday, rick, steve, recommendation, loved]	1	1	0
4	selection menu great prices	[selection, menu, great, prices]	1	1	0

Result of Tokenizer

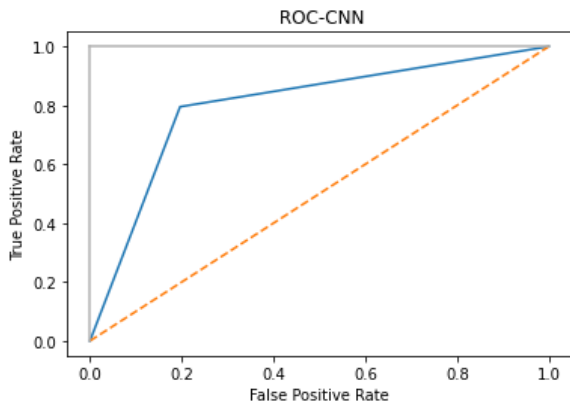


Figure 7a: ROC curve of CNN

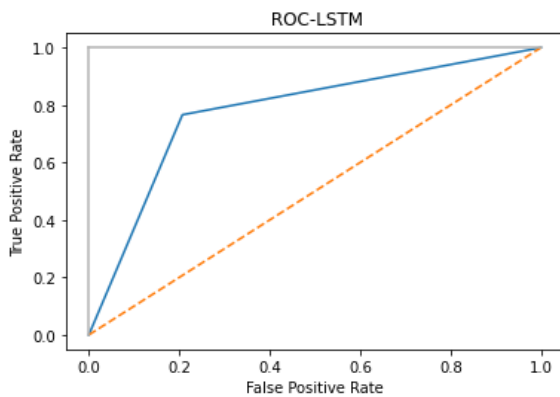


Figure 7b: ROC curve of LSTM

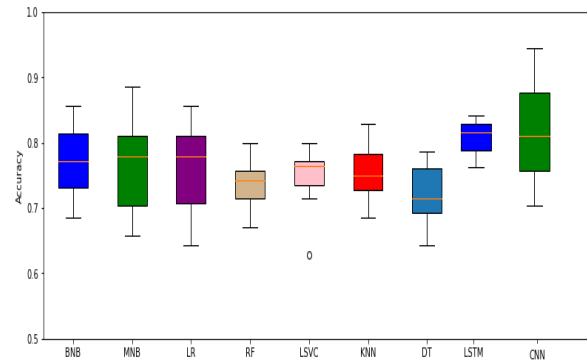


Figure 8: Accuracy score of Machine learning and Deep Learning algorithms

dom seed = 20. The results of accuracy score for each algorithm obtained is shown in Figure 8.

This box plot shows the spread of accuracy score across each cross validation fold of these classifiers. From this box plot, the mean value of accuracy score obtained for Bernoulli Naive ayes (BNB) is 0.7714, Multinomial Naïve Bayes (MNB) is 0.767, Logistic regression (LR) is 0.762, Random forest (RF) is 0.738, Linear Support Vector Machine (LSVC) is 0.748 K nearest Neighbor (KNN) is 0.752 and Decision tree (DT) is 0.72. The mean score value of accuracy score for LSTM is 0.823 and CNN is 0.828. Hence deep learning algorithm (LSTM and CNN) provide high prediction compared to machine learning classifier algorithms.

According to this experimental study, machine learning classifiers such as naive bayes, logistic regression, SVC, and KNN are faster to train than deep learning models such as CNN and LSTM. The results of the experiment show that LSTM and CNN take longer to train but have higher accuracy in both training data (98%) and test data (84%) for this restaurant dataset.

uses convolution layer to extract features from input data, max pooling layer to reduce the dimensionality of trainable parameters and a sigmoid activation function in output dense layer. Similarly, the input data in the LSTM model is passed to embedding layer, spatial dropout, LSTM, and output layer. The spatial dropout rate is taken as 0.2 for avoid over fitting. For compilation, both the CNN and LSTM model use the Adam optimizer. The classification report of CNN and LSTM classifier is shown in Table 2. The ROC curve of CNN and LSTM is shown in Figure 7a and 7b.

Finally K-fold cross validation (K=10) is carried out to evaluate the machine learning and deep learning algorithm with ran-

Table 2: Performance Analysis of Deep learning techniques for Sentiment Analysis

Classifier	Class	Classification Report				Accuracy		AUC Score	Time to train (Sec-onds)
		Precision	Recall	F1-score	Support	Train	Test		
CNN	0	0.83	0.79	0.81	256	0.98	0.845	0.84	7.3
	1	0.76	0.80	0.78	244				
LSTM	0	0.80	0.67	0.73	256	0.96	0.77	0.748	9.08
	1	0.73	0.76	0.74	244				

5. conclusion

Sentiment analysis on restaurant data set is carried out to classify the opinion as positive or negative. Initially preprocessing is carried out to reduce the features and speed up the classification task. The classification task is performed by machine learning methods (Multinomial Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, Random Forest, SVC, KNN, Decision Tree) and deep learning methods (LSTM and CNN). Machine learning methods use bag of model approach (TFID-Vectorizer) to convert text into vector form and Deep learning methods use word embedding method for converting text into vector form. The classifier performance is evaluated by various metrics such as precision, recall, f1 score, AUC score and time taken for training. The accuracy score of each classifier is tested by k-fold cross validation technique. From these findings of the experiments, neural network-based learning has a higher training accuracy and a longer running time compared with machine learning classifier. For future work, improve the performance of classifier by adopting various feature selection technique and analyze the text prediction on multilingual texts.

Acknowledgments

The authors gratefully appreciate the anonymous reviewers for their constructive contributions to this work.

References

- [1] P. Poomka, N. Kerdprasop & K. Kerdprasop, "Machine Learning Versus Deep Learning Performances on the Sentiment Analysis of Product Reviews" *International Journal of Machine Learning and Computing* **11** (2021) 103, doi: 10.18178/ijmlc.2021.11.2.1021.
- [2] K. Klimiuk, A. Czoska, K. Biernacka & L. Balwicki, "Vaccine Misinformation on Social Media—Topic-Based Content and Sentiment Analysis of Polish Vaccine-Deniers Comments on Face book", *Human Vaccines & Immunotherapeutic* **17** (2021) 2026.
- [3] H. Tsaniya, R. Rosadi & A. S. Abdullah, "Sentiment Analysis towards Jokowi's Government Using Twitter Data with Convolutional Neural Network Method", *Journal of Physics Conference Series*, **1722** (2021) 012017, doi:10.1088/1742-6596/1722/1/012017.
- [4] R. D Endsuy, "Sentiment Analysis between VADER and EDA for the US Presidential Election 2020 on Twitter Datasets", *Journal of Applied Data Sciences* **2** (2021) 8.
- [5] M. Bibi, W. Aziz, M. Almarashi, I. H. Khan, M. S. A. Nadeem & N. Habib, "A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis", *IEEE Access* **8** (2020) 68580.
- [6] R. Cekik & S. Telceken, "A New Classification Method Based on Rough Sets Theory", *Soft Computing* **6** (2018) 1881.
- [7] B. Peng, J. Wang & X. Zhang, "Adversarial Learning of Sentiment Word Representations for Sentiment Analysis", *Information Sciences* **541** (2020) 426.
- [8] X. Tan, Y. Cai, J. Xu, H. F. Leung, W. Chen & Q. Li, "Improving Aspect-Based Sentiment Analysis via Aligning Aspect Embedding", *Neuro computing* **383** (2020) 336.
- [9] A. Jain & V. Jain, "Sentiment Classification Using Hybrid Feature Selection and Ensemble Classifier" *Journal of Intelligent & Fuzzy Systems*, **4**(2021) 221.
- [10] P. Kalaivani & K. L. Shunmuganathan, "Sentiment Classification of Movie Reviews by Supervised Machine Learning Approaches", *Indian Journal of Computer Science and Engineering* **4** (2013) 285.
- [11] M. Ghosh & G. Sanyal, "Performance Assessment of Multiple Classifiers Based on Ensemble Feature Selection Scheme for Sentiment Analysis", *Applied Computational Intelligence and Soft Computing* **2018** (2018) 10.
- [12] A. P. Rodrigues & N. N. Chiplunkar, "A New Big Data Approach for Topic Classification and Sentiment Analysis of Twitter Data", *Evolutionary Intelligence* **2** (2019)11.
- [13] Z. Jianqiang, G. Xiaolin & Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis", *IEEE Access* **6** (2018) 23253.
- [14] W. Li, L. Zhu, Y. Shi, K. Guo, & E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN-LSTMfamily models", *Applied Soft Computing* **94** (2020)106435.
- [15] A. S. Imran, S. M. Daudpota, Z. Kastrati & R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets", *IEEE Access* **8** (2020) 181074.
- [16] S. Rani, N. S. Gill & P. Gulia, "Survey of Tools and Techniques for Sentiment Analysis of Social Networking Data", *International journal of Advanced computer Science and applications* **12** (2021) 222.
- [17] R. Cekik & A. K. Uysal, "A novel filter feature selection method using rough set for short text data", *Expert Systems with Applications* **160** (2020) 113691.
- [18] I. S. Ahma, A. B. Azuraliza & M. R. Yaakub, "A review of feature selection in sentiment analysis using information gain and domain specific ontology", *International Journal of Advanced Computer Research* **9** (2019) 283.
- [19] C. Albon , "Machine Learning with python cook book : Practical solutions from preprocessing to deep learning", *OReilly media* (2018) 366.
- [20] Z. Wu & S. King, "Investigating gated recurrent networks for speech synthesis", *IEEE International Conference on Acoustics, Speech and Signal Processing* (2016) 5140.
- [21] B. Peng, J. Wang & X. Zhang, "Adversarial learning of sentiment word representations for sentiment analysis", *Information Sciences* **541** (2020) 426.
- [22] Z. Jianqiang, G. Xiaolin & Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis" , *IEEE Access* **6** (2018)23253.
- [23] N. Isnaini, M. S. Mubarak & M. Y. A. Bakar, "A multi-label classification on topics of Indonesian news using K-Nearest Neighbor", *Journal of Physics: Conference Series* **1192**(2019) 012027.
- [24] T. Anuprathibha & C. S. KanimozhiSelvi, "Enhanced Medical Tweet Opinion Mining using Improved Dolphin Echolocation Algorithm Based Feature Selection", *International journal of Innovative Technology and Exploring engineering* **2**(2019)20.
- [25] H. Zikang, Y. Yong, Y. Guofeng & Z. Xinyu, "Sentiment analysis of agri-

- cultural product ecommerce review data based on deep learning”, International Conference on Internet of Things and Intelligent Applications, **27**(2020) 7.
- [26] M. M. Ali, “Arabic sentiment analysis about online learning to mitigate covid-19”, *Journal of Intelligent Systems* **30**(2021) 524.
- [27] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund & J. Kim, “Covidsent: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis”, *IEEE Transactions on Computational Social Systems* **29**(2021)175.
- [28] Z. Wang, H. Wang, Z. Liu & J. Liu, “Rolling Bearing Fault Diagnosis Using CNN-based Attention Modules and Gated Recurrent Unit”, *Global Reliability and Prognostics and Health Management* **7**(2020) 6.
- [29] A. Vieira & W. Brandao, “Evaluating Acceptance of Video Games using Convolutional Neural Networks for Sentiment Analysis of User Reviews”, *Proceedings of the 30th ACM Conference on Hypertext and Social Media* **2**(2019) 273.
- [30] K. Hirota & F. Masahiro, “Efficient Attention Mechanism by Softmax Function with Trained Coefficient”, *IEICE Technical Report* **339** (2021) 52.