Check for updates

OPEN ACCESS

Volume: 4

Issue: 2

Month: June

Year: 2025

ISSN: 2583-7117

Published: 30.06.2025

Citation:

Sameeksha Dwivedi, Dr. Pankaj Richhariya "Sentiment Analysis of Hotel Reviews: Identifying Key Factors for Customer Satisfaction" International Journal of Innovations in Science Engineering and Management, vol. 4, no. 2, 2025, pp. 416–423.

DOI:

10.69968/ijisem.2025v4i2416-423

Sentiment Analysis of Hotel Reviews: Identifying Key Factors for Customer Satisfaction

Sameeksha Dwivedi[1], Dr. Pankaj Richhariya[2]

[1]*Research Scholar, Department of Computer Science, Bhopal Institute of Technology & Science.*
[2]*HOD, Department of Computer Science, Bhopal Institute of Technology & Science.*

Abstract

Deep learning is a subfield of machine learning which has achieved great success in certain sentiment analysis and other natural language processing applications. Specifically, state-of-the-art outcomes in sentiment analysis tasks have been shown when deep neural networks are used. This study aimed to perform sentiment analysis on hotel reviews to improve the services based on customer feedback. After completing data cleansing and feature engineering, the study employed dataset of the 36,000 reviews of a variety of hotels and locales. Reviews with ratings below 2.0 were classified as negative, while those with ratings over 4.5 were classified as positive, thereby creating the desired variable. Random forest, Logistic regression, support vector machine, and Naive Bayes were among the machine learning techniques tested in this study. When compared to these other methods, nevertheless, suggested bidirectional LSTM model fared better, with an accuracy of 97.44%. Using the confusion matrix, ROC curve, and accuracy measures, the study assessed the model. The findings demonstrated that proposed model outperformed the baseline paper model by a wide margin, with an improvement in accuracy from 92% to 97.44% when using Bi-LSTM.

Keywords; sentiment analysis, machine learning, deep learning, LSTM.

INTRODUCTION

There has been a meteoric rise in the use of online reviews as a means through which consumers may learn more about a product before purchasing it. As e-commerce expands and more people go online, consumers have more opportunities than ever to research purchases and voice their own opinions on products and services. In today's information-driven economy, both consumers and businesses rely heavily on online reviews. Customers now heavily rely on internet reviews when making purchasing decisions. Customers may use reviews to gauge the standard of product/service, and companies can utilize the input to enhance their offerings. Opinion mining, or sentiment analysis, is a method for automatically determining whether a given piece of text is positive, negative, or neutral. Financial markets, government, and business all make use of sentiment analysis. The hotel sector has come to value sentiment analysis as a tool for better understanding client feedback and enhancing service quality in recent years. "Sentiment analysis of hotel reviews" is a crucial job for hotel management since hotel reviews, in particular, have grown to be a valuable source of information for future customers. Sentiment analysis can be performed at three different levels, including the document level, sentence level, and phrase level [1]. Sentiment Analysis is not limited to just reviews or Twitter data, it also applies to stock markets [2] [3], news articles [4], political debates [5], and other areas. It can play a crucial role in enhancing consumer products-related businesses [6].

Deep learning is a subfield of machine learning which has achieved great success in certain sentiment analysis and other natural language processing applications. Specifically, state-of-the-art outcomes in sentiment analysis tasks have been shown when deep neural networks are used. These networks include "Convolutional Neural Networks" and "Long Short-Term Memory" networks. The implementation of deep learning methods for this sentiment analysis of the hotel reviews will be the main emphasis of this study.

A lot of prior research has been done in this field where words and phrases have been classified with prior positive or negative polarity [7].The purpose of this study is to develop a robust model that can accurately sort hotel reviews into positive and negative categories. To do this, we will utilize the well-known deep learning framework Tensor Flow to train a model on a dataset consisting of hotel reviews. Due to its ability to acquire high-level features from raw data, methods based on deep learning have attracted substantial interest in the area of sentiment analysis. Deep learning algorithms may gather sophisticated representation of "input data" with no the need of human-crafted features, in contrast to typical machine learning approaches. This is especially helpful in sentiment analysis tasks, where nuanced and context-dependent verbal clues are commonly used to determine sentiment. When compared to rule-based or standard machine learning approaches, the potential for improved accuracy in deep learning methods for "sentiment analysis" is a major benefit. It has been shown that deep learning algorithms perform better on large and varied datasets because they are able to learn complicated and non-linear correlations between input features as well as sentiment labels.

RELATED WORK

Many different methods are used for processing of text in emotional analysis. [8] In this study, the authors offer a technique for automatically sorting reviews into positive, neutral, and negative categories using the Multinomial Nave Bayes Classifier. The researchers tested several models based on preliminary processing, extraction of features, as well as feature selection techniques in order to evaluate the efficacy of this strategy. The researchers used a 10-fold cross-validation procedure, which entails dividing the data into Ten equal parts and employing nine parts for training along with one part for testing, to assess the effectiveness of the suggested strategy. Each step was performed ten times, and every component was tested only once. The trials indicated that the highest performance, having an average F1-Score greater than 91%, was achieved by combining preprocessing & feature selection techniques

Khare, Arpit, et al [9].uses Twitter data from the 2019 Lok Sabha election in India to analyze public opinion during the campaign. In order to deal with the unsupervised character of this challenge, the authors have developed an automated tweet analyzer utilizing the Transfer Learning approach. The authors' Machine Learning model makes use of Linear Support Vector Classifiers and the Term "Frequency Inverse Document Frequency" approach to dealing with tweets' textual input. In addition, the authors have improved the model's capacity to deal with caustic tweets written by certain users, a factor that has not been previously taken into account by academics in this field.

The use of social media is becoming more important in political campaigns. T Alashri, Saud, et al [10] the authors of this research examine a dataset consisting of over 22,000 potential Facebook posts and over 48,000,000 comments in order to draw conclusions on the topic of online debate. In this article, the author focuses on how the 2016 U.S. presidential contenders communicated with voters. The author outlines a fresh approach to categorizing commenters as either diehard believers, moderates, moderate dissenters, or diehard dissidents. Sentiment analysis is then used to each group's commentary on policies and their implementation. Finally, the author discusses potential directions for further research into the interplay between political campaigns and social media.

Through unsupervised learning on huge Twitter corpora, the Author [11] introduces a word embedding approach that makes use of latent contextual semantic links and co-occurrence statistical properties between words in tweets. Twitter's sentiment characteristics are constructed using word embeddings, n-grams, as well as word sentiment polarity scores. A deep convolutional neural network is fed the feature set to train and make predictions about sentiment labels. After putting our model through its paces on five different Twitter data sets, the author found that it outperformed the baseline model—a word n-grams model—on accuracy & F1-measure for sentiment classification.

Zhiyong Luo [12] examine the lack of background information in brief paragraphs makes sentiment analysis difficult. Convolutional neural network as well as recurrent neural networks (RNNs) are examples of deep learning models that have been used to text sentiment analysis with rather impressive results in recent years. This research describes a jointed convolutional neural network (CNN) and recurrent neural network (RNN) architecture for short-text sentiment analysis, which makes use of both the "coarse-grained local features" created by CNN and the long-distance relationships learnt by RNN. The experimental findings on the three benchmark corpora (MR, SST1, and SST2) indicate a significant improvement above the state-of-the-art, with accuracies of 82.28 percent, 51.50 percent, and 89.95 percent, respectively.

Chang, Victor, et al [13] provides a heuristic approach for visual & multimedia analytics to do sentiment analysis on evaluations of high-end hotels. The experiment comprises collecting data from Booking.com, preprocessing it, designing features to be included in the model, classifying it using a Random Forest algorithm, analyzing it, and visualizing the results. The findings highlight the importance of location, cleanliness, and employee training for luxury hotel pleasure. The study includes a list of crucial properties for future changes, and the "sentiment analysis model" works well.

PRELIMINARIES

LSTM

Long short-term memory (LSTM) networks are an offshoot of RNNs that were developed to fill in the gaps where RNNs fell short. The LSTM model has been developed and it has been observed to have a higher precision and recall compared to the traditional autoregressive approach [14]. RNN is a network that processes the current input by remembering previous output & keeping it in its memory for a short time. It's most well-known uses are in the areas of voice recognition, non-Markovian control, and musical creation. However, RNNs are not without their flaws. It can't keep data for a very long time, for starters. In order to forecast the present output at times, it is required to refer to information that was stored a long time ago. However, RNNs are unable to deal with "long-term dependencies" of any kind. Second, there's is no precise control over the amount of the historical background should be "forgotten" and how much should be "carried forward." One key distinction between RNNs & LSTMs is the fact that the LSTM's buried layer is a gated unit, also known as a gated cell. It has four layers that work together to generate the cell's output, which is a long distance from the cell's current state. These two elements are then sent to the next secret level. LSTMs contain "three logistic sigmoid gates" & one tanh layer, whereas RNNs only have one tanh layer. There are now gates installed to regulate the flow of data into and out of the cell. They decide what information the following cell will require and what may be disregarded. The output is often a number between 0 and 1, where 0 indicates complete exclusion and 1 complete inclusion.

Bi-LSTM

In this research paper, we propose using Bidirectional Long Short-Term Memory (Bi-LSTM) to analyze text data for the purpose of sentiment analysis. A bidirectional LSTM is an extension of the standard LSTM that processes the input sequence in both forward and backward directions. In the context of sentiment analysis, Bi-LSTMs can be used to process a sequence of words in a review and make a prediction about the sentiment of the review [15]. The standard LSTM processes the input sequence in one direction, which can be problematic for sequential data such as text. This is because the context of a word can depend on both the words that come before it and the words that come after it. By processing the input sequence in both forward and backward directions, the Bi-LSTM is able to capture the context of a word from both directions. When processing sequential input, the bidirectional long short-term memory (BiLSTM) RNN architecture considers both historical and prospective contexts. It uses 2 LSTM layers, one for forward reading and one for reverse reading of the sequence. When the results of the two layers are combined, a more accurate depiction of input sequence is obtained. Making a neural network capable of storing sequence data in both forward and backward orientations is called "bidirectional long-short term memory" (bi-lstm). A bi-lstm is distinct from a standard LSTM since its input goes in both directions. The standard LSTM only allows us to choose one way for the input to go, either backwards or forwards. In contrast, bi-directional input allows us to keep track of both the present and the past. Let's look at an example to help clarify.

METHODOLGY

The proposed methodology has been illustrated in **Figure 1**

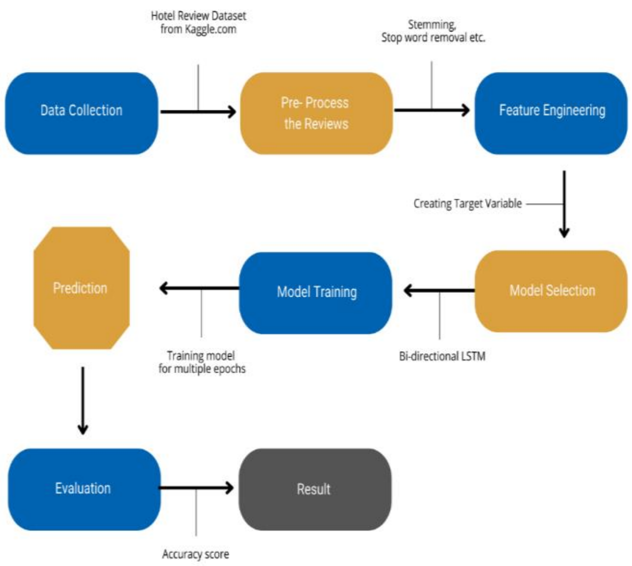

Figure 1 Flow Chart of Proposed Model

Dataset

In this research, hotel reviews data was collected from Kaggle. Kaggle is a popular platform for machine learning and data science projects, where users can access a wide variety of datasets to use for their own research. The dataset used in this research contained around 36,000 hotel reviews from various hotels and locations. The dataset was collected by scraping reviews from various hotel booking websites and was made publicly available on Kaggle.

```
df.sample(10)
```

	reviews.rating	reviews.text
26739	2.0	For the price go stay somewhere nicer! Looks a...
18750	5.0	stayed 1 nite on road trip. paid $40 on pricel...
1839	4.0	I stayed at the Super 8 for one night. The sin...
13122	3.0	Beds were awesome!! To warm in the pool/hottub...
18991	4.0	The hotel staff was great very courteous. Clos...
2866	4.0	Great front desk , very friendly and helpful, ...
34592	5.0	Our daughters go to Camp Monterey. We've staye...
19847	1.0	i had a not so pleasant experience the room wa...
16420	5.0	Was driving forever. I failed to book early an...
8923	3.0	Room was nice, the but the air conditioning wa...

Figure 2 Dataset Sample

The dataset was checked for any duplicates, missing values or other anomalies. Once the data was cleaned, only 12,000 reviews were selected for sentiment analysis. The reviews were selected based on their relevance to the research question and their representativeness of different hotels and locations. The selected reviews were then utilized to train as well as test the "machine learning models" for sentiment analysis.

To ensure the accuracy and reliability of the data, we conducted a thorough quality check before including it in our study.

Data Pre Processing

Sentiment analysis relies heavily on data preparation, which involves transforming raw textual data into a more manageable organized format. In this research, a number of methods were used to prepare the data for sentiment analysis, including normalizing case, eliminating punctuation along with tokenizing text, stemming the text, lemmatizing the text, and deleting stop words. The quality of the classification is directly impacted by the preprocessing operations carried out. [16]

First, the dataset was cleaned up by removing any unnecessary columns in order to retrieve the data required for sentiment analysis. As punctuation and capitalization are irrelevant to sentiment analysis, they were removed from the cleaned text data. The tokenized text was next subjected to a process of stemming and lemmatization, which stripped the words of their inflected forms. In contrast to lemmatization, which reduces words back to their root form, stemming retains the full meaning of the original word. For instance, stemming would reduce "running," "ran," and "runs" to "run," whereas lemmatization would do the same.

In the end, stop words were filtered out of the textual information. A, an, the, is, etc. are examples of stop words that do not add much meaning to a sentence and should be avoided while doing sentiment analysis. By excluding these, we may lessen the amount of background noise in data and direct the algorithm's attention to the key terms.

	reviews.rating	reviews.text	processed_text
29034	1.0	try to get room . it's got a great view of san...	try get room got great view santana row
27679	1.0	Roaches in room, Carpet soiled and worn. Tile ...	Roaches room Carpet soiled worn Tile room grea...
12754	2.0	jUST OK.The hotel was very quiet. Rooms were v...	jUST hotel quiet Rooms warm breakfast room col...
15174	2.0	Reserved a room with a single bed, but those r...	Reserved room single bed room sold I stay one ...
4364	5.0	The gentleman at the counter was super friendl...	The gentleman counter super friendly welcoming...
26999	9.6	arriving from Qubec, we were really tired and ...	arriving Qubec really tired received u big smi...
34804	5.0	After driving hours all you want is a bath an...	After driving hour want bath bed And hotel wen...
23919	2.0	Free coffee refills and budget rates make this...	Free coffee refill budget rate make motel top ...
30530	1.0	The hotel was disappointing. We did not receiv...	The hotel disappointing We receive room cleani...
23292	1.0	A poor excuse for a BB	A poor excuse BB

Figure 3 Data after pre-processing

Feature engineering was used in this research to develop a sentiment analysis dependent variable. Reviews were coded as favorable or negative to construct the dependent variable. The following procedures were carried out in order to produce the requisite variable:

Relevant columns were extracted: The original dataset contained several columns such as hotel name, review title, review text, and ratings. Only the review text and ratings columns were selected as relevant for sentiment analysis.

Ratings were used to label reviews: The ratings column contained a numerical rating between 1 and 10 for each review. Reviews with ratings below 2.0 were labeled as negative, while those with ratings above 4.5 were labeled as positive. Reviews with ratings between 2.0 and 4.5 were not used in the analysis, as they were considered neutral.

Proposed Model

In this study, various machine learning algorithms such as decision tree, random forest, multinomial naive Bayes, gradient boosting, logistic regression, and support vector machine were tested with two word vectorization techniques: CountVectorizer and TfidfVectorizer. However, none of these models provided satisfactory results in predicting the sentiments. Therefore, we tried bidirectional LSTM and trained it for multiple epochs, which yielded higher accuracy than the machine learning algorithms.

To create the bidirectional LSTM model, we relied on the TensorFlow and Keras frameworks. In order to convert the phrases into dense vectors, the input is first sent via an embedding layer. The sequence's preceding and subsequent word dependencies are captured by the Bidirectional LSTM layer. After the output of the layer has been normalized by a batch-normalization layer, it is fed into a 64-unit Bidirectional LSTM layer. Two completely linked dense layers, one with 512 units and the other with 1, are employed after Dropout layer is applied to avoid overfitting. The "Sigmoid activation function" is used in the last layer to transform the output into a probability score among 0 and 1. In order to train the model, we use “binary cross-entropy loss function" also the Adam optimizer. Accuracy is used as a measure of the model's efficacy. The model architecture consists of several layers:

Embedding layer: This layer takes the input text data and maps each unique word to a high-dimensional vector representation. Here, the input vocabulary size is 10,000, and each word is mapped to a 32-dimensional vector.

Bidirectional LSTM layer: This layer is made up of the two LSTM layers which perform forward and reverse processing on the input text. Because of this, the model is better able to capture sequential dependencies included in the input text data. Each time step's output sequence is returned by the 32-unit first layer of LSTM.

Batch Normalization layer: To maintain a uniform distribution of values throughout training, this layer normalizes output of the preceding layer.

Second Bidirectional LSTM layer: This layer consists of two LSTM layers that process the input text in both forward and backward directions. The second LSTM layer has 64 units and returns the final output sequence.

Dropout layer: The synapses among neurons in layer below are randomly severed by a small percentage (0.2 in this example). As a result, the model is compelled to learn additional robust characteristics, which aids in preventing overfitting.

Dense layer: This 512-neuron layer uses rectified linear unit function of activation to transform the data. In this layer, more complex representation of the input information are learned.

Output layer: There is just one sigmoid activation function neuron in this layer. To convert the layer-previous output into probability value of 0 to 1 that indicates the projected emotion of the input text, the sigmoid function is utilized.

The model was then trained using the cleaned and prepared data, an Adam optimizer, and an accuracy metric. We also used an 80:20 split between the training and validation sets. We used a number of training epochs to improve the model's performance to our satisfaction.

After the model got trained, we used a ROC curve and a confusion matrix to determine how well it performed. The model's total effectiveness is shown by the accuracy metric. The number of accurate and wrong predictions produced by the model are shown in the confusion matrix, while the ROC curve depicts trade-off between true positive rate and false positive rate. Using these data, we were able to gauge how well our "sentiment analysis model" performed.

RESULTS AND DISCUSSION

Based on the evaluation of multiple machine learning algorithms, logistic regression and SVM models achieved accuracy of 90.60% and 90.70%, respectively, which are not up to the desired accuracy. Therefore, the bidirectional LSTM model was chosen for the sentiment analysis. The proposed Bidirectional LSTM model achieved an accuracy of 97.44%, which is significantly higher than the machine learning models. This indicates that the proposed model is capable of predicting the sentiment of hotel reviews with high accuracy. Therefore, this model can be utilized by hotels to analyze their customer feedback and improve their services.

Below is the summary of the results in tabular form:

Table 1 Model Comparison without Image Augmentation

Model	Accuracy
Logistic Regression	90.60%
SVM	90.70%
Bidirectional LSTM (Proposed Model)	97.44%

Here are visual results of proposed mode –

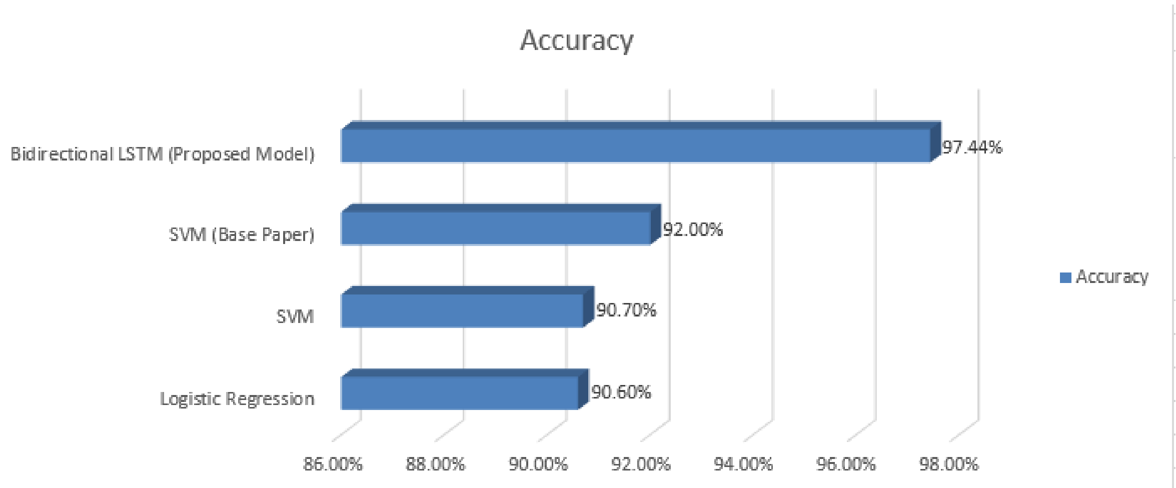

Figure 1 Result Comparison

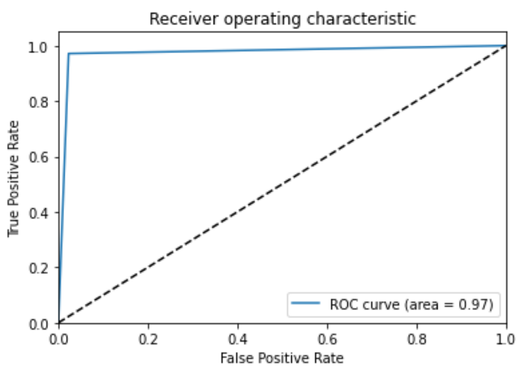

Figure 2 ROC Curve

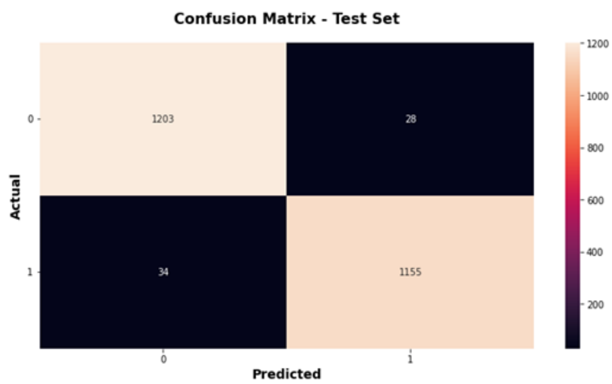

Figure 3 Confusion Metrix

The sentiment analysis of the hotel reviews using the SVM-based base paper model has a 92% success rate. Whereas proposed model improved upon the underlying paper model by using bidirectional LSTM to increase accuracy to 97.44%.

For a number of reasons, the proposed model is an improvement over the published model. When it comes to sentiment analysis and other forms of natural language processing, deep learning algorithms like bidirectional LSTM have shown to be effective. When compared, SVM is a more conventional machine learning technique that may struggle with the nuances of real language.

Since word embedding may capture more relevant and subtle connections between words, it was employed to represent the words in review text in the proposed model. The original paper model relied on a bag-of-words & TFIDF framework, both of which may have been inadequate for accurately capturing linguistic nuance. The comparison table of our results and the base paper's results is as follows:

Table 2 Base Paper Comparison Table

Existing Work	Accuracy Score
SVM [17]	92.00%
Propose Proposed work d work	Accuracy Score
Bi-Directional LSTM	97.44%

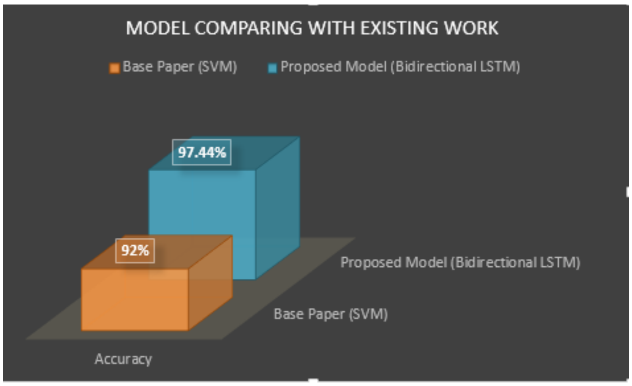

Figure 4 Graph result of proposed work

The data clearly shows that the proposed model employing Bidirectional LSTM beat the SVM model employed in the underlying research, with an accuracy of 97.44% vs 92%. This demonstrates that, in comparison to the SVM model employed in the foundational research, the proposed model is more suitable for "sentiment analysis" of hotel reviews.

CONCLUSION

The proposed methodology included data collection, preprocessing, feature engineering, model building, and model evaluation. The data collection involved obtaining a dataset of around 36,000 reviews from various hotels and locations from Kaggle. However, for the purpose of sentiment analysis, only 12,000 reviews were selected after basic data cleaning and removal of irrelevant columns. The data preprocessing involved various steps like lowercasing, tokenizing, removing punctuation, stemming, lemmatization, stop words removal, and word embedding using Keras. The target variable was engineered by labeling reviews negative whose rating was below 2.0 and positive for those whose rating was above 4.5. Random forest, Logistic regression, support vector machines, and naive Bayes were only some of the machine learning techniques tested throughout the model-building process. These methods, however, failed to deliver the expected precision. For this reason, a bi long short-term memory (Bi-LSTM) model was used and trained across many iterations.

The proposed model outperformed the base paper SVM model by a wide margin, with a precision of 97.44% vs 92%. Accuracy, ROC curve, & confusion matrix assessment measures all returned positive result for proposed model.

REFERANCE

[1] Erevelles, Sunil, Nobuyuki Fukawa, and Linda Swayne. "Big Data consumer analytics and the transformation of marketing." Journal of business research 69.2 (2016): 897-904.

[2] Yu, Liang-Chih, et al. "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news." Knowledge-Based Systems 41 (2013): 89-97.

[3] Hagenau, Michael, Michael Liebmann, and Dirk Neumann. "Automated news reading: Stock price prediction based on financial news using context-capturing features." Decision Support Systems 55.3 (2013): 685-697.

[4] Xu, Tao, Qinke Peng, and Yinzhao Cheng. "Identifying the semantic orientation of terms using S-HAL for sentiment analysis." Knowledge-Based Systems 35 (2012): 279-289.

[5] Maks, Isa, and Piek Vossen. "A lexicon model for deep sentiment analysis and opinion mining applications." Decision Support Systems 53.4 (2012): 680-688.

[6] Qiu, Guang, et al. "DASA: dissatisfaction-oriented advertising based on sentiment analysis." Expert Systems with Applications 37.9 (2010): 6182-6191.

[7] Hatzivassiloglou, Vasileios, and Kathleen McKeown. "Predicting the semantic orientation of adjectives." 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics. 1997.

[8] Farisi, Arif Abdurrahman, Yuliant Sibaroni, and Said Al Faraby. "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier." *Journal of Physics: Conference Series*. Vol. 1192. No. 1. IOP Publishing, 2019.

[9] Khare, Arpit, et al. "Sentiment analysis and sarcasm detection of indian general election tweets." *arXiv preprint arXiv:2201.02127* (2022).

[10] Alashri, Saud, et al. "The 2016 US Presidential Election on Facebook: an exploratory analysis of sentiments." (2018).

[11] Jianqiang, Zhao, Gui Xiaolin, and Zhang Xuejun. "Deep convolution neural networks for twitter

sentiment analysis." *IEEE access* 6 (2018): 23253-23260.

[12] Wang, Xingyou, Weijie Jiang, and Zhiyong Luo. "Combination of convolutional and recurrent neural network for sentiment analysis of short texts." *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*. 2016.

[13] Chang, Victor, et al. "An improved model for sentiment analysis on luxury hotel review." *Expert Systems* 40.2 (2023): e12580.

[14] Huang, Xin, et al. "Lstm based sentiment analysis for cryptocurrency prediction." International Conference on Database Systems for Advanced Applications. Springer, Cham, 2021.

[15] Pronko, Rafał. "Simple bidirectional LSTM solution for text classification." Proceedings ofthePolEval2019Workshop (2019): 111.

[16] Krouska, Akrivi, Christos Troussas, and Maria Virvou. "The effect of preprocessing techniques on Twitter sentiment analysis." 2016 7th international conference on information, intelligence, systems & applications (IISA). IEEE, 2016.

[17] Hassan, Rakibul, and Md Rabiul Islam. "Impact of sentiment analysis in fake online review detection." *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2021.

[18] Shivani Wadhwani 2023. Sentiment Analysis of User Youtube Comments Using Classifier Algorithm. *International Journal of Innovations in Science, Engineering And Management*. 2, 1 (Feb. 2023), 25–32.