# Hotel review analysis on machine learning and deep learning based on pre-trained Glove Embedding

Md Fardin Rahman Ami, A S M Nasim Khan, Mohammad Nasif Sadique Khan, Ehsanur Rahman Rhythm, Md Sabbir Hossain and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
Brac University
{md.fardin.rahman.ami,a.s.m.nasim.khan, mohammad.nasif.sadique.khan,ehsanur.rahman.rhythm, md.sabbir.hossain1}@g.bracu.ac.bd,
annajiat@gmail.com

***Abstract*—The development of technology has revolutionized the decision-making process, including the selection of lodging. User-generated feedback and reviews play a crucial role in influencing the decisions of prospective visitors in the hospitality industry. These feedbacks are utilized by our system to generate appropriate recommendations for future customers. The foundation of the system is sentiment analysis, which involves categorizing reviews into positive and negative sentiments. A comprehensive preprocessing phase including data cleaning and feature extraction is implemented prior to training. The proposed system leverages a diverse set of models including k-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB), Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier (SVC), AdaBoost, Multi-Layer Perceptron (MLP), Attention-based Long Short-Term Memory (ATTENTION-LSTM), Recurrent Neural Network (RNN) with LSTM and Convolutional Neural Network (CNN) with (LSTM). Each model was individually trained on a scrupulously preprocessed dataset, which included extensive data modification and feature extraction. The RNN-LSTM model emerges as the leader, achieving the maximum level of accuracy among the models. This demonstrates the capability of deep learning techniques to capture complex dependencies in textual data. The results of this initiative will have a substantial impact on the hospitality and tourism industries.**

***Index Terms*—Sentiment Analysis, Tourism, CNN-LSTM, Hotel reviews, Machine-Learning, RNN-LSTM, Deep-Learning**

## I. Introduction

In the current digital planet, online evaluations exert an unprecedented amount of impact over consumer decisions, thereby rewriting the hospitality industry. Individuals readily share their experiences on platforms such as hotel websites, Facebook, and YouTube due to ubiquitous internet access. The effect of these feelings on online reservations is substantial. Sentiment Analysis (SA) with Machine Learning (ML) is used to generate a pattern of the deluge of written criticism. Our custom-tailored Natural Language Processing (NLP) model for SA decodes positive and negative sentiments from reviews. Our method utilizes a pipeline that transforms unprocessed text via data cleansing, tokenization, and vectorization. Then, Machine Learning and Deep Learning fuel classifications based on particular requirements. By automating sentiment extraction, our method bridges language nuances and fosters a robust framework for sentiment analysis by analysing with different models in total of 11 models including both neural network models like, CNN-LSTM, RNN-LSTM also non-neural models likewise, decisin-tree, KNN, MNB etc. This inspires the hotel industry to strategically respond to consumer perceptions in the dynamic digital era, improving growth.

## II. Related Works

With Advancement of NLP techniques, we can now analyze the sentiment using various models. Machines can now sense the sentiment of any text using different NLP techniques. To do so, we train these models using text data with predefined sentiment labels. But some text data are complicated to categorize into predefined labels. These types of data are usually removed in data preprocessing, but the removal of controversial or disputed data from sentiment analysis datasets can be problematic as it obstructs the performance of automated sentiment analysis systems. In this paper by Kenyon-Dean et al. proposes the inclusion of a ”complicated” class of sentiment to capture such disputed text and argues that its inclusion in sentiment analysis frameworks can improve the quality of automated sentiment analysis systems in real-world settings [1].The paper proposes the construction of a new publicly available Twitter sentiment analysis (TSA) dataset named MTSA, consisting of over 7,000 tweets annotated with 5x coverage, to analyze classifier performance and gain insights into sentiment analysis dataset and model design. The preprocessing and feature extraction involve representing the textual data in a vector space using N-Grams (unigrams and bigrams), mean word embedding (GLoVE embeddings built from Twitter data), and SentiWordNet scores. The experimental design includes agglomerating tweets based on majority labeling and comparing different models (SVM with linear or RBF kernel, Random Forests, Naive Bayes, and K-Nearest-Neighbors) to determine the best performing model variant. Cross-validation is used to account for possible variance in the results. The results show that the accuracy improvement in sentiment analysis is not solely due to a change in the distribution of classes, but rather there must be a qualitative difference between high-agreement and low-agreement tweets. The model poorly classifies the ”complicated” tweets, indicating the need for a separate category to handle such challenging data.The paper does not provide specific numerical results or statistical analyses. The evaluation of sentiment analysis is done using weighted and macro F1-scores, but the paper does not provide detailed information on the performance of the classifiers . The author focuses on the construction and analysis of the MTSA dataset, but it does not compare the performance of the proposed ”complicated” class with existing sentiment analysis methods .It also does not discuss the generalizability of the findings beyond the specific dataset and Twitter sentiment analysis.

The paper titled ”Convolutional Neural Networks for Sentence Classification” by Yoon Kim at New York University offers a thorough investigation into the utilization of convolutional neural networks (CNNs) for tasks involving the classification of sentences in the field of natural language processing (NLP) [2]. The study commences by acknowledging the remarkable achievements of deep learning models in domains like computer vision and speech recognition, highlighting the vital role played by word vector representations in NLP. These word vectors, which encapsulate the semantic attributes of words, are derived from pretrained models like word2vec, which underwent training on extensive text datasets. The paper introduces an elementary CNN architecture founded on these word vectors, presenting a slight modification of the CNN architecture originally devised for NLP applications by Collobert et al. (2011). Within this architecture, sentences are portrayed as sequences of word vectors, and convolutional operations are employed with diverse window sizes to capture local features.

The paper provides an intricate exposition of the CNN model, detailing the incorporation of convolutional filters, nonlinear functions, and max-over-time pooling operations to discern pertinent features from the input sentences. It underscores the benefits of CNNs, initially devised for computer vision but adept at effectively tackling NLP tasks, such as semantic parsing, search query retrieval, and sentence modeling. Various model adaptations are explored, encompassing static and non-static representations, as well as a multi-channel architecture that amalgamates both static and non-static word vectors. Techniques for regularization, such as dropout and l2-norm constraints, are employed to enhance the model’s ability to generalize.

The experimental configuration entails the assessment of the CNN model on several benchmark datasets, encompassing movie reviews, sentiment analysis, question classification, and customer reviews. The paper provides a meticulous delineation of the hyperparameters employed in the experiments, with an emphasis on the pivotal role of dropout regularization. Pre-trained word vectors sourced from word2vec, trained on an extensive corpus of Google News data, are harnessed. In cases where words are not present in the pre-trained set, they are initialized randomly. The study scrutinizes different model variations, including CNN-rand (random initialization of word vectors), CNN-static (use of static word vectors), CNN-non-static (fine-tuning of pre-trained vectors), and CNN-multichannel (integration of both static and non-static vectors). The results are comprehensive and encompass comparisons with alternative methods. They demonstrate that even the elementary CNN model with static vectors delivers exceptional performance, surpassing more intricate models across a spectrum of NLP benchmarks. The paper concludes by underlining the significance of pre-trained word vectors and the potential of CNNs as potent tools for sentence-level classification tasks in NLP, even with minimal adjustment of hyperparameters. In essence, the paper furnishes an exhaustive and detailed exploration of CNNs in the realm of NLP sentence classification, offering insights into model structure, techniques for regularization, and empirical findings.

Lukas Christ et al. authored a paper that conducts a comprehensive investigation into the MuSe 2022 challenge, aimed at advancing the field of multimodal sentiment analysis [3]. The primary objective of this research is to provide valuable insights into this domain by furnishing baseline results and a detailed overview of the MuSe 2022 competition, comprising three distinct sub-challenges, each addressing diverse facets of sentiment analysis.

Within the MuSe-Humor sub-challenge, the authors employ various features and modalities, encompassing audio, video, and textual elements, to discern humor in press conference recordings. Their findings indicate that video-based features, particularly VGGish, outperform audio-based counterparts, implying that humor is frequently conveyed through facial expressions and visual cues. Textual features also demonstrate the capability to detect humor, with sentence-level BERT features surpassing token-level variants. Intriguingly, the amalgamation of modalities does not consistently enhance performance.

Transitioning to the MuSe-Reaction sub-challenge, the authors delve into the prediction of emotional reactions to stimuli. The outcomes reveal that audio-based results generally lag behind those obtained from the video modality, with features such as Facial Action Units (FAUs) proving effective in capturing emotional expressions. Remarkably, the Amusement category consistently outperforms others, likely attributable to its association with non-verbal signals like laughter. However, combining the most successful features from each modality does not yield significant enhancements in this context.

In the MuSe-Stress sub-challenge, the authors pivot their focus toward forecasting continuous valence and arousal values in stressful scenarios. In this case, video features encounter challenges in effectively generalizing to the test data, displaying noteworthy disparities between development and test results. Conversely, audio features, notably DeepSpectrum features, demonstrate superior generalization, with fewer performance discrepancies between the development and test datasets. Interestingly, audio features exhibit better performance for valence compared to arousal, contrary to prior findings in multimodal emotion recognition.

In summary, the paper underscores the contributions of the MuSe 2022 challenge to the field of multimodal sentiment analysis. It emphasizes the potential for further enhancements through more refined methodologies and serves as a valuable resource for researchers and participants involved in the MuSe 2022 challenge. Ultimately, the study's significance lies in its transparent methodology, realistic baseline results, and the provision of open-source resources, all of which facilitate the development and evaluation of innovative approaches in multimodal sentiment analysis.

When we try to apply Sentiment analysis in short texts, it becomes very challenging due to limited contextual information. This paper by dos Santos et al. proposes a deep convolutional neural network that combines character-to-sentence level information to achieve state-of-the-art results in sentiment prediction for bothTwitter messages and movie reviews [4]. The proposed network architecture uses word-level, character-level, and sentence-level representations, to achieve high accuracy by leveraging convolutional layers to extract features. It also uses a softmax operation to compute conditional probability distributions of sentiment labels. For movie review they got an accuracy of 85.7% and 86.4% for Twitter messages.

The network architecture includes convolutional layers to extract relevant features from character-level inputs . The network was trained using stochastic gradient descent to minimize negative log-likelihood over the training set . It generates a score for each sentiment label using a softmax operation, transforming the scores into a conditional probability distribution of labels given the sentence and network parameters.This proposed approach achieves state-of-the-art results for sentiment prediction in both binary positive/negative classification and fine-grained classification for both datasets. A feed-forward neural network instead of a recursive one and it does not require input about the syntactic structure of the sentence. The addition of a convolutional layer allows the extraction of character-level features, providing flexibility in feature extraction. The architecture combines character-level, word-level, and sentence-level representations to perform sentiment analysis effectively.

This study by Cook et al. plans to anticipate self-destructive ideation and uplifted mental side effects in grown-ups who have as of late been released from mental ongoing or trauma center settings utilizing AI and normal language handling (NLP) in Madrid, Spain [5]. Collapse is an essential in general success concern, and early unquestionable check of people in danger is key for persuading mediation and balance attempts. The review gathered information from members who answered organized mental and actual well-being instruments at numerous subsequent places. Notwithstanding the organized information, members were asked to answer an unstructured inquiry, "how would you feel today?". NLP-based models were utilized to dissect the text reactions to this genuine inquiry, while calculated relapse expectation models were assembled utilizing the organized information. The outcomes showed that the NLP-based models, which used unstructured text information, accomplished moderately high prescient qualities for recognizing people in danger of self-destructive ideation and mental pain. In any case, it is significant that the exhibition of the NLP-based models was somewhat lower contrasted with the organized information-based models, which approached more unambiguous and distinct data. Predictions based on NLP could have significant effects on mental health. The review's discoveries recommend that NLP can be used to distinguish people in danger of self-destruction or mental misery quickly. In situations where lengthy structured surveys are impractical, this automated approach may offer a cost-effective screening alternative. The utilization of NLP for anticipating ailments has shown guarantee in different biomedical applications, and this exploration stretches out its pertinence to emotional well-being evaluation utilizing instant messages, explicitly short message administration (SMS) messages, happening beyond clinical settings. Continuous alarms created from SMS texts hold the possibility to illuminate ideal clinical intercessions, in this manner forestalling self-destruction and tending to elevated mental pain.

## III. Reasearch Methodology

### A. Dataset Collection

The dataset comprises diverse reviews encompassing a broad spectrum of guest perspectives. The data was scraped from Kaggle which contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe taken from Booking.com [**?**]. Written in a unique blend of formal and informal tone, it mirrors the contemporary linguistic practices of native speakers. Our dataset comprises reviews both positive and negative, where positive review holds 70% of the total and 30% being negative. We have split our dataset in 80-20 ratio for train-test purpose. This eclectic collection of text-based data forms the foundation for training and evaluating our sentiment analysis models, enabling accurate classification of sentiments as positive or negative, ultimately driving deep insights for effective decision-making in the hotel industry.

### B. Data Preprocessing

The ambition of data Preprocessing is to make the data more understanable by the models. The splited a sentence into list of words. For data preprocessing, there are many useful techniques. Specifically in this paper, tokenization is the first step in preprocessing. Subsequently, the removal of stop words, commonly used across languages, is implemented as they hold minimal significance in natural language processing. For English, words like "is", 'the", and "and" fall under the category of stop words as well as punctuation marks, emoticons, pictorial icons, random words, single alphabets, and HTML tags which are removed from the review text which help in noise reduction of the data.Lemmatization ia a NLP technique by which we can convert a word into its base form to enchnace the semantic consistency, converting words like *walking* to *wak*, *am/is/are* to *be* and *ducks* to *duck*. To facilitate these operations, the system leverages *SpaCy*, a versatile open-source Python toolkit designed for advanced natural language processing tasks.*SpaCy* efficiently executes tokenization, lemmatization, and stop word removal, streamlining preprocessing activities, particularly within the domain of English language processing.

TABLE I
DATA PREPROCESSING

| **Before Processing** | ***After Processing*** |
|---|---|
| Well located place in the area, The bathroom was clean and fresh . The rooms are well ventilated size also pretty spacious and. | well located place area bathroom clean fresh room ventilated size spacious |
| All staffs are welcoming, well behaved. food is above average.The bedrooms are furnished well. | staffs welcoming well behaved food above average bedrooms furnished well |

### C. Feature Extraction

In the initial phase, all features except the text and label columns were retained. The text data, being natural language, often lacked proper formatting. It contained elements like HTML tags, punctuation marks, diverse symbols, and stopwords. These elements hold negligible semantic significance within sentences and have limited impact on stress and nonstress classification. To address this, a preprocessing step was undertaken to normalize the text and eliminate unnecessary symbols, resulting in a simplified form containing only words. Lemmatization was employed to obtain word root forms, enriching the text's consistency. To integrate this processed data into the model, the words needed to be transformed into a format that the model could comprehend. Various word embedding techniques exist, including word2vec, TF-IDF, countervectorizer, and rapid text. Among these, we explored the use of word2vec and countervectorizer methods. However, word2vec's negative value vector conversion posed challenges, making it less adaptable for certain models. Consequently, word2vec embedding yielded less satisfactory outcomes. As an alternative, countvectorizer was selected for word embedding assignments, facilitating subsequent comparisons.

### D. Word representation

Since computers do not understand words or their context, it is needed to convert text into the appropriate, machine-interpretable form. The aforementioned representation refers to vectors that are spatially arranged in a manner that vectors in close proximity exhibit greater semantic similarity by mathematical representation.

*1) Count Vectorizer:* Count Vectorizer is a method that generates a matrix in which each distinct word is denoted by a column, and each text sample from the document is represented as a row. The purpose of this method is to convert a given text into a vector representation based on the frequency of each word in the entirety of the text.

TABLE II
WORD PREPROCESSED

| **Reviewed text** | ***Preprocessed text*** |
|---|---|
| Sad people. Do not smile, and No matter how cool thier life is they bore themselves in any place they are the unluckiest human on earth | "sad", "people", "smile", "matter", "cool","life","bore", "them", "place", |
| The service of the hotel was best Location, but costs are highest in the area, enough Sunlight and air . Recommend | "service", "good", "location", "area", "cost", "high","sunlight", "air", "recommend" |
| Great location, neat but ugliest view, Darkest room I have ever seen. Restaurants in near distance, erves nice and fresh breakfast | "ugly", "view", "dark", "great", "location", "neat","restaurant", "near", "distance", "breakfast", "serve", "fresh" |

The present study employs the vector representation as a characteristic. Count vectors were retrieved from the dataset and subsequently utilized as input for many models.

*2) Global Vector Model (GloVe):* The Global Vector model is a computational framework that represents words in a distributed manner. The algorithm under consideration is an unsupervised learning technique that integrates the characteristics of two distinct model families, specifically the global matrix factorization and local context window approaches. The GloVe model is a log-bilinear model that utilizes a weighted least-squares goal function during training. This objective function is applied to a matrix of global word-word co-occurrence . The matrix shown illustrates the probability of co-occurrence between terms within a certain corpus.

### E. Word Embedding

Converting raw text into a numerical form is crucial for training learning models. Word embedding is the contemporary choice for this purpose, transforming text into vectors while capturing word relationships [6]. In our sequential model, the first layer is an embedding layer that converts user reviews into trainable vector format. The embedding dimension is 100, and the maximum sequence length is set according to the longest review.

## IV. MODEL IMPLEMENTATION

### A. Sentiment analysis using machine learning

*Classifiers:*
For our proposed system we have used several Machine Learning models and some deep learning models to classify the data. As for the machine learning models we have opted to use k-Nearest Neighbors (KNN), Multinomial Naive Bayes (MNB), Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier (SVC), AdaBoost, Multi-Layer Perceptron (MLP) and for the models for deep learning we have used neural network models like Attention-based Long Short-Term

Memory (ATTENTION-LSTM), Recurrent Neural Network (RNN) with LSTM and Convolutional Neural Network (CNN) with (LSTM).

### B. Model Analysis

After preprocessing the data by tokenizing and lemmatization, we used the count vector feature to modify the data used in the machine learning models and gloVe for the data to be fed in the deep learning models. We have found quite comparatively well accuracy and results from the Logical Regression and Random Forest model from the machine learning side but overall, the best results were from the neural network side which used the RNN-LSTM model. We have trained the model 15 epochs with Adam (learning rate 0.001) as a optimizers and loss function as binary crossentropy. Then we have achieved 96.02% accuracy. The table 1 presents performance metrics for various machine learning and deep learning models used in the study. These metrics evaluate the effectiveness of each model in categorizing reviews into positive and negative sentiments.

TABLE III
MODEL PERFORMANCE METRICS

| Model Names | F1 Score Positive (%) | F1 Score Negative (%) | Accuracy (%) |
|---|---|---|---|
| KNN | 94 | 74 | 89.77 |
| MNB | 96 | 85 | 93.69 |
| Logistic Regression | 96 | 87 | 94.3 |
| Decision Tree | 94 | 77 | 89.94 |
| Random Forest | 96 | 86 | 93.92 |
| SVC | 96 | 86 | 94.11 |
| AdaBoost | 93 | 69 | 88.14 |
| MLP | 96 | 84 | 93.16 |
| Attention-LSTM | 97 | 89 | 94.89 |
| RNN-LSTM | 97 | 91 | 96.02 |
| CNN-LSTM | 97 | 90 | 95.5 |

### C. Model Training

To start our study, we carefully chose several traditional machine learning models known for their distinct abilities.
Our aim in using these models was two-fold. Initially, we expected these models to provide useful baseline results that could act as reference points for assessing more advanced models. Secondly, we aimed to identify any inherent patterns or tendencies in the data that could guide us toward the most effective modeling approach. The unique feature of attention mechanisms, allowing the model to focus on specific parts of the input data, could be crucial in identifying subtle cues related to human stress levels. Additionally, the use of pre-trained GloVe embeddings to align the dataset meticulously within the framework of deep learning models. Both the RNN and CNN used the Long Short-Term Memory (LSTM) architecture, which included components of the Convolutional Neural Network (CNN) into the LSTM model. The training methodology used for each of these deep learning models included 15 epochs, with a batch size of 32.

## V. RESULT ANALYSIS AND DISCUSSIONS

Non-neural models were first selected due to their adaptability, since they include a broad spectrum of machine learning methods, which include both linear and non-linear approaches. This enabled us to examine several approaches for predicting reviews and get a holistic understanding of the issue. Moreover, these models have gained significant recognition in the respective discipline, providing a robust basis for our preliminary inquiry. However, we recognized the potential for improvement by incorporating neural models, particularly those with attention mechanisms. Neural models, known for their ability to capture intricate patterns in complex datasets, are a valuable asset. The unique feature of attention mechanisms, enabling models to focus on specific data elements, holds promise for identifying subtle cues in sentiment analysis. Neural models' adaptability, ability to prioritize relevant information, and capacity to uncover hidden features make them an essential addition to enhance our performance in predicting positive and negative reviews. From Fig1, MNB, Random Forest, Logistic Regression, SVC and MLP shows drastic high score for the F1 scores of both positiove and negative reviews among the basic 8 ML models that we have been implemented. In addition to that, in Fig2 the comparison between deep learning models has been plotted with F1 scores of both positive and negative reviews.

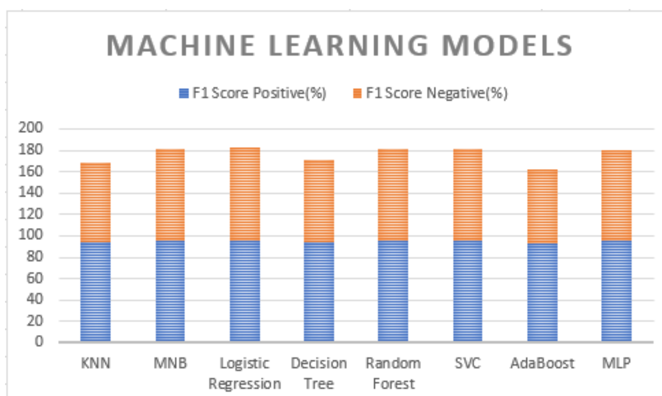


Fig. 1. Comparisons of outcomes among Machine Learning Models

### A. Comprehensive Model Comparison and Analysis

Given is comparative comparison of performance indicators for several machine learning models in three categories: F1 Score Positive (%), F1 Score Negative (%), and Accuracy (%). Initially, it is evident that the models demonstrate a notable level of proficiency, as shown by F1 scores over 90% continuously for the positive class. It is worth mentioning that both Attention-LSTM and RNN-LSTM exhibit exceptional F1 scores in the negative class, namely 97% and 91% respectively. This indicates their notable proficiency in detecting

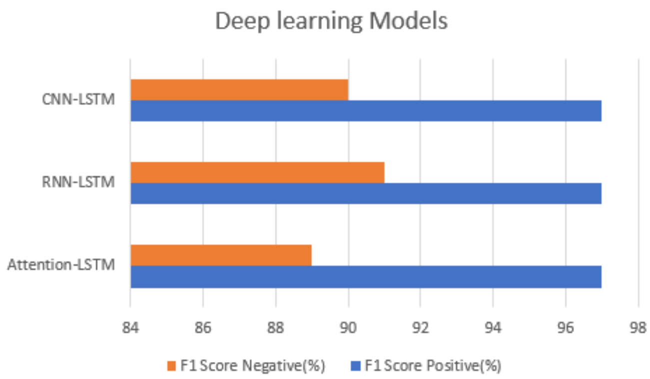


Fig. 2. Comparisons of outcomes among Deep Learning Models

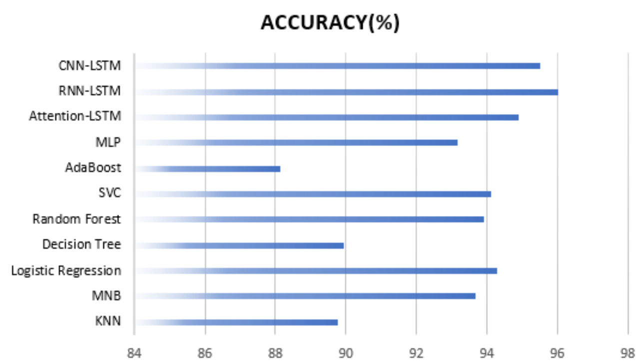


Fig. 3. Comparisons of outcomes among Deep Learning Models

negative outcomes. Regarding the overall accuracy, the models demonstrate commendable performance, as Logistic Regression, SVC, RNN-LSTM, and CNN-LSTM achieve accuracy scores beyond 94%. Once again, the RNN-LSTM model demonstrates its superiority by achieving an amazing accuracy rate of 96.02%, while the Attention-LSTM model closely trails behind with a commendable accuracy rate of 94.89%. Upon comparing the F1 scores of the positive and negative classes, it becomes apparent that the majority of models exhibit a somewhat equitable performance.

However, a subset of models, including AdaBoost and KNN, have a propensity towards predicting the positive class, resulting in much lower F1 scores for the negative class. When examining the trade-off between accuracy and recall, as measured by F1 scores, both Logistic Regression and Random Forest demonstrate a satisfactory equilibrium in accurately recognizing positive and negative examples. This is seen in their balanced F1 scores across both classes. Additionally, it is worth mentioning that the Multi-Layer Perceptron (MLP) has a notable F1 score of 96% for the positive class, while exhibiting a comparatively lesser score of 84% for the negative class. This observation implies that the MLP excels in accurately recognizing positive situations. When considering model complexity and interpretability, it is seen that simpler models such as Multinomial Naive Bayes and Decision Tree provide satisfactory performance, but not as remarkable as more sophisticated models like LSTMs. One example of a classification algorithm, Naive Bayes, obtains an accuracy of 80.48%, whilst another algorithm, Decision Tree, has a higher accuracy rate of 89.94%.

Models such as RNN-LSTM and CNN-LSTM demonstrate superior performance in terms of generating high F1 scores and accuracy, simpler models such as Logistic Regression and Random Forest provide a well-balanced level of performance.

### *B. Model Selection*

The CNN-LSTM and RNN-LSTM models have shown superior performance in the context of sentiment analysis tasks using textual input. These models have repeatedly shown outstanding performance across several measures, making them very suitable for this particular application. Both the CNN-LSTM and RNN-LSTM models demonstrated exceptional performance in terms of F1 scores for both positive and negative classes. Notably, the RNN-LSTM model outperformed its counterparts by achieving an amazing F1 score of 91% specifically for the negative class.

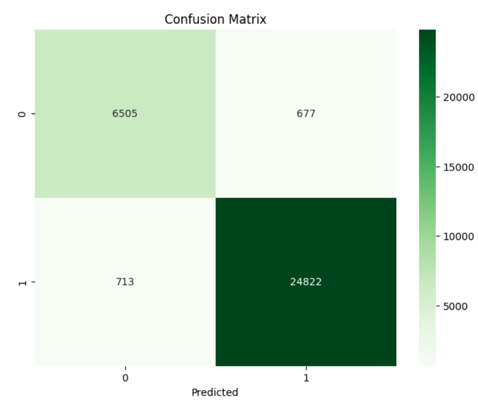


Fig. 4. RNN-LSTM Confusion Matrix

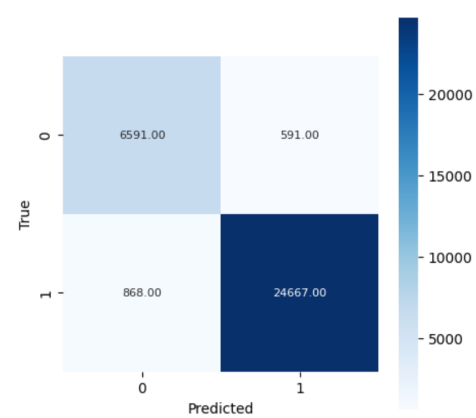


Fig. 5. CNN-LSTM Confusion Matrix

This demonstrates their aptitude in discerning positive and negative attitudes in textual data, underscoring their appropriateness for sentiment analysis tasks that need an equitable categorization of sentiments. Moreover, in relation to the overall accuracy, the RNN-LSTM model exhibited superior performance compared to all other models, with an accuracy rate of 96.02%. The CNN-LSTM model closely followed with an accuracy rate of 95.5%. The aforementioned accuracy rates demonstrate a remarkable proficiency in accurately categorizing feelings within textual material, a critical factor for the efficacy of sentiment analysis applications.The delicate trade-off between accuracy and recall has significant importance in sentiment research. These models demonstrate a commendable equilibrium, making them dependable options for collecting nuanced thoughts. The strong performance shown by the system in this aspect is especially advantageous for sentiment analysis tasks that need a significant level of precision in sentiment categorization.

The use of CNN-LSTM and RNN-LSTM models in sentiment analysis is justified due to their established efficacy, particularly in cases where precise sentiment categorization has significant significance. These models possess the capability to effectively manage the complexities inherent in text input and extract significant sentiment information, making them the favored options for constructing a sentiment analysis model.

## VI. Future Works

In the realm of future endeavors, a significant avenue lies in the development of a specialized dataset that facilitates cross-lingual learning. This innovative dataset would encompass diverse languages, enabling the sentiment analysis model to transcend language barriers and extract sentiment nuances across different linguistic landscapes. By engaging in cross-lingual learning, the model could gain a more comprehensive understanding of the universal aspects of sentiment while accommodating the intricacies that arise from linguistic variations.

Furthermore, the integration of BERT (Bidirectional Encoder Representations from Transformers) stands as a promising augmentation strategy. BERT, renowned for its contextual embedding capabilities, could profoundly enhance the model's comprehension of context-dependent sentiment cues. By assimilating BERT into the existing architecture, the sentiment analysis model could capitalize on its proficiency in capturing intricate contextual relationships, potentially leading to a re- fined accuracy in sentiment classification.

The integration of cross-lingual learning and BERT has the potential to provide a novel phase of sentiment analysis, marked by a broader range of languages and an increased awareness of contextual factors. This all-encompassing strategy could enable a model to flourish in understanding and translating sentiment in a variety of languages, dialects, and socioeconomic settings, resulting in more precise and subtle sentiment analysis results on a global basis.

## VII. Conclusion

The findings of our research resulted in beneficial results. Through the utilization of advanced deep learning architectures, the system effectively processed and understood feedback from several languages, leading to enhanced levels of precision and reliability in hotel suggestions.The RNN-LSTM and CNN-LSTM models exhibited their adeptness in capturing the subtleties of customer input, hence enabling the system to effectively identify emotions and thoughts across several languages. The sentiment analysis task exhibited a notable level of accuracy, as evidenced by the constant attainment of F1 scores over 90% by both models.One notable advantage of the recommendation system is in its capacity to offer tailored hotel recommendations by leveraging sentiment analytics. Consequently, this feature aids passengers in making well-informed choices. The use of deep learning models, namely RNN-LSTM and CNN-LSTM, enhances the system's ability to comprehend the sentiment-laden characteristics of client feedback, thus enhancing the overall quality of hotel suggestions. This method essentially offers a helpful tool for tourists looking for lodgings by utilizing the strength of cross-lingual sentiment analysis.

## References


[1] K. Kenyon-Dean, E. Ahmed, S. Fujimoto, J. Georges-Filteau, C. Glasz, B. Kaur, A. Lalande, S. Bhanderi, R. Belfer, N. Kanagasabai, R. Sarrazingendron, R. Verma, and D. Ruths, "Sentiment analysis: It's complicated!" in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 1886–1895. [Online]. Available: https://aclanthology.org/N18-1171

[2] Y. Kim, "Convolutional neural networks for sentence classification," *CoRR*, vol. abs/1408.5882, 2014. [Online]. Available: http://arxiv.org/abs/1408.5882

[3] L. Christ, S. Amiriparian, A. Baird, P. Tzirakis, A. Kathan, N. Müller, L. Stappen, E.-M. Meßner, A. König, A. Cowen, E. Cambria, and B. W. Schuller, "The MuSe 2022 multimodal sentiment analysis challenge," in *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge*. ACM, oct 2022. [Online]. Available: https://doi.org/10.1145%2F3551876.3554817

[4] C. dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, Aug. 2014, pp. 69–78. [Online]. Available: https://aclanthology.org/C14-1008

[5] B. L. Cook, A. M. Progovac, P. Chen, B. Mullin, S. Hou, E. Baca-Garcia *et al.*, "Novel use of natural language processing (nlp) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in madrid," *Computational and mathematical methods in medicine*, vol. 2016, 2016.

[6] A. Mandelbaum and S. Adi, "Word embeddings and their use in sentence classification tasks," *arXiv.org*, Oct 2016. [Online]. Available: https://arxiv.org/abs/1610.08229