International Journal of Innovative Information Systems & Technology Research 14(3):247-253, July-Sept, 2026

© SEAHI PUBLICATIONS, 2026 www.seahipublications.org **ISSN: 2467-8562**

DOI:10.5281/zenodo.22135618

Sentiment Analysis Of Nigerian Hotel Reviews Using Deep Learning Techniques

Rinmiko Tanfa Zakaria; Dr. Bilkisu Maijama'a & Dr. S. E Chaku

Centre for Cyberspace Studies, Nasarawa State University Keffi, Nasarawa State, Nigeria.

ABSTRACT

Sentiment analysis is now a key method for understanding customer opinions in the hospitality industry. In this study, we tested how well deep learning techniques can analyze Nigerian hotel reviews. We collected and prepared a dataset of online hotel reviews, then classified each review as positive or negative. We built and tested three deep learning models: Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). The reviews were tokenized and converted into numerical data using word embeddings before training the models. We measured performance using accuracy, precision, recall, F1 score, and confusion matrix. The results show that the CNN model performed best, with an accuracy of 97%. The FNN also gave good results, while the LSTM had the lowest performance. These findings suggest that convolutional models work especially well for sentiment analysis of Nigerian hotel reviews and can help hospitality businesses better understand customer feedback and improve their services.

Keywords: deep learning techniques, sentiment analysis, hospitality industry

1. INTRODUCTION

Over the past decade, Nigeria's hospitality industry has grown steadily, thanks to more domestic tourism, business travel, and the rise of online travel platforms. As more people use sites like Booking.com, TripAdvisor, and Google Reviews to find and book hotels, there has also been an increase in user-generated content about hotel experiences.

Sentiment analysis, or opinion mining, has become a popular way to evaluate customer feedback by identifying whether a review is positive, negative, or neutral. Advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) have helped this field grow, making it possible for automated systems to efficiently analyze large amounts of text data (Liu, 2012)

Traditional approaches to sentiment analysis mostly used lexicon-based approaches and traditional machine learning algorithms like Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression. These methods worked fairly well for structured text but often missed context, sarcasm, negation, and word relationships. Deep learning techniques, including Long Short-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs), have shown better results in understanding these complexities, making them a good fit for analyzing Nigerian hotel reviews (Devlin, Chang, Lee, & Toutanova, 2019).

Neural networks like Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory Networks (LSTM) have improved accuracy in sentiment analysis tasks (LeCun, 2015). While FNN processes input in a straightforward manner without considering sequential dependencies, CNN is designed to detect spatial hierarchies in data, making it effective for feature extraction. In contrast, LSTM, a specialized type of Recurrent Neural Network (RNN), can retain long-

term dependencies, making it particularly suitable for analysing sequential text data in order (Hochreiter & Schmidhuber, 1997).

This study aims to design and use deep learning models (FNN, CNN, and LSTM) to classify review sentiment as positive or negative. The performance of these models will be measured and compared using standard metrics like accuracy, precision, recall, F1-score, and confusion matrix.

2. RELATED WORK

Sentiment analysis is now a common tool for understanding customer opinions in the hospitality industry, where online reviews can greatly affect booking choices and hotel reputation. (Ameur, Hamdi, & Yahia, 2023) reviewed 90 studies on hotel review sentiment analysis and found that research differs widely in pre-processing methods, feature representation, sentiment classification levels (document, sentence, or aspect), and model selection. Few studies have tried to bring these findings together or suggest future research directions. Their review also showed that the field is moving from traditional methods to deep learning approaches that can better capture complex language patterns in reviews.

Many studies have shown that deep learning models, especially Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, work well for classifying hotel review sentiment. (Kusumaningrum, Jayanto, Nawangsari, Nisa, & Wibowo, 2023) developed Sentilytics 1.0, a web tool for analyzing Indonesian hotel reviews. It uses a CNN for document-level classification and an improved LSTM for aspect-level classification. Their system reached F1-scores of about 0.95 for document-level analysis and 0.92 for aspect-polarity detection. This shows that hybrid deep learning models can provide detailed, aspect-specific insights, such as separate evaluations of food, room, service, and location. However, their work also highlighted a common problem: model performance drops for aspects or sentiment classes that are unclear or not well represented in the training data.

Most Nigerian-focused sentiment analysis research looks at social media data, which makes this study different. (Oyewusi, Adekanmbi, & Akinsande, 2020) pointed out that Nigerian Pidgin, while sharing much of its vocabulary with Standard English, has changed a lot due to code-switching and code-mixing. As a result, common English words can have very different meanings in Nigeria. They argued that models trained only on Standard English do not work well for Nigerian text. Their study provided a Pidgin-specific sentiment lexicon and a dataset of 14,000 annotated tweets to help fill this gap.

(Venkata Srinivas, Ch.Satyanarayana, Ch.Divakar, & Sirisha, 2021) used 1.6 million tweets from Kaggle, splitting them into 80% training and 20% testing data. They trained three models with Python's Keras library: LSTM, CNN, and a simple neural network. The LSTM performed best with an accuracy of 87%, followed by CNN at 82%.

(Deriu & Cieliebak, 2016) introduced a multi-task training method for CNN. Their model classified tweets as subjective or objective, negative or positive, and ironic or not, all at once, instead of using separate models for each task. When compared to single-task models, the multi-task model performed slightly better.

Each model has its own features. FNN is a basic neural network that processes inputs in one direction and does not have memory. CNN is built to find spatial patterns in text. LSTM, a type of recurrent neural network (RNN), is good at capturing long-term dependencies in sequences, but it can be harder to train because of the gradient vanishing problem when target information is not used enough (Bengio, Simard, & Frasconi, 1994).

By comparing these models, this research hopes to show which approach works best for sentiment analysis in the hospitality sector, especially for Nigerian data.

Even though deep learning models work well for sentiment classification, there is not much research on using them to analyze hotel reviews in Nigeria. Most studies look at Western markets, so we do not know much about how these models perform in Nigeria's hospitality sector. Abuja, Nigeria's capital, is a major business and travel center with a growing hotel industry. Knowing what customers think in this area is important for hotel operators to stay competitive and meet guest expectations.

This study also helps travelers by allowing hotels to improve their services based on customer feedback. Better service quality leads to better guest experiences, which benefits the tourism and hospitality sector in Abuja. The research findings can also help policymakers and industry regulators make data-driven decisions to improve service quality in the hospitality industry.

3 METHODOLOGY

This study uses a quantitative approach to examine deep learning models with Nigerian hotel reviews. The research starts by evaluating FNN, CNN, and LSTM models quantitatively, followed by qualitative analysis for further explanation. Customer reviews were collected from booking.com, where users share their hotel experiences. The reviews were labeled using the sentiment intensity analyser method in Python, assigning each a binary classification of positive or negative. Reviews were converted to numerical data with GloVe word vectors to help interpret the models. The hotel review data was gathered from booking.com using Python's Scrapy library and Chrome driver. In total, 20,000 reviews were collected. The dataset was split into 80% for training and 20% for testing.

Data pre-processing was done to prepare the reviews for deep learning models. The steps included removing punctuation, special characters, emojis, and images, as well as removing stopwords, tokenizing, lemmatizing, and padding or truncating sequences. All experiments used the Python programming language with libraries such as TensorFlow, Keras, PyTorch, and Scikit-learn.

The FFN model starts with an embedding layer that turns input tokens into 100-dimensional dense vectors. Next, a global max pooling layer extracts the most important features from the review sequence. Two fully connected dense layers with ReLU activation help the model learn non-linear sentiment patterns. Dropout layers are added to prevent overfitting. Finally, a sigmoid output layer gives a probability score for binary classification.

This architecture is computationally efficient and performed well in this study. The results suggest that simpler feedforward models can be very effective when sentiment cues are clear in the dataset.

A one-dimensional convolutional layer with 128 filters and a kernel size of 5 is used to find local patterns like sentiment expressions (for example, "very good" or "not satisfied"). Global max pooling then keeps the most important features found by each filter, no matter where they appear in the text.

The extracted features go into a dense layer for further learning, followed by dropout regularization. The sigmoid output layer then gives the final sentiment prediction.

The model uses a pre-trained embedding layer with an embedding matrix that gives word meanings learned from external sources. This layer is set as non-trainable to keep the pre-learned relationships and lower computational cost.

A single LSTM layer with 128 units was used to learn patterns across the review text. The final sigmoid output layer gives the probability of positive sentiment for binary classification.

4. RESULTS AND DISCUSSION

A total of 20,000 customer reviews of hotels located in Abuja were collected from online travel platforms. After pre-processing, the dataset was divided into training (90%) and testing (10%) sets. The sentiment distribution chart is below.

Table 1. Sentiment class distribution

Sentiment Class	Number of Reviews	Percentage
Positive	11892	59.4%
Negative	8108	40.5%

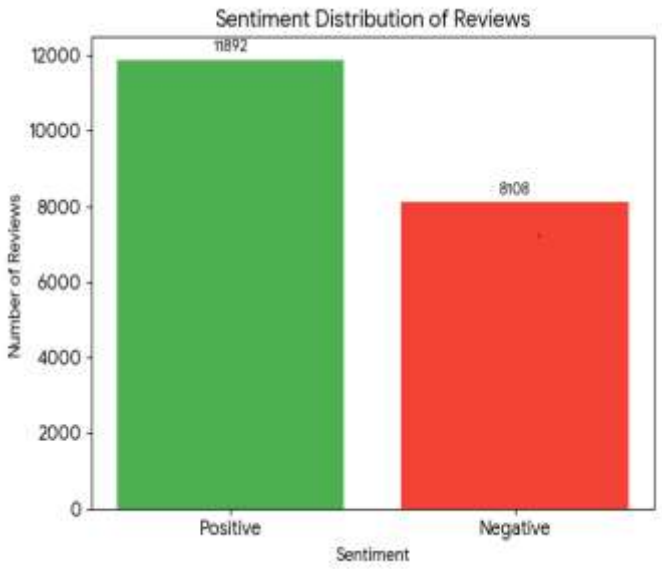

Figure 1. Bar chart showing sentiment class distribution

Most of the reviews captured were positive.

Table 2. Model's comparison for accuracy, precision, recall and f1 score

Model	Accuracy	Precision	Recall	F1-Score
CNN	97	98	97	97
FNN	93	93	93	93
LSTM	45	57	45	35

The results indicate that the CNN model outperformed the other models across all evaluation metrics. This suggests that CNN architectures provide better contextual understanding of hotel review texts and even though it is mostly used in image classification, it can also be efficient is sentiment analysis task.

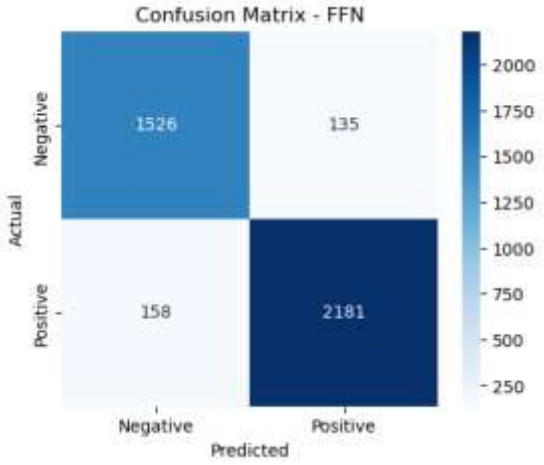

Figure 2. Confusion matrix for FFN

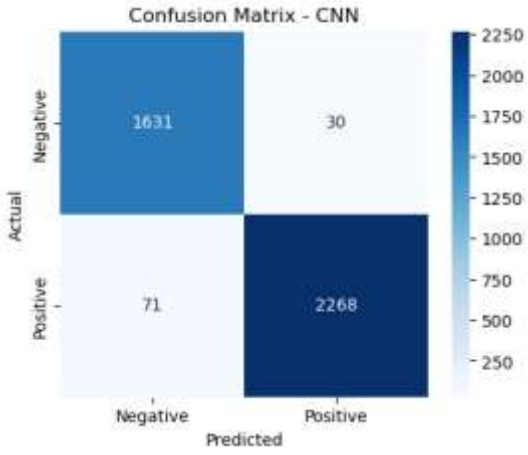

Figure 3. Confusion matrix for CNN

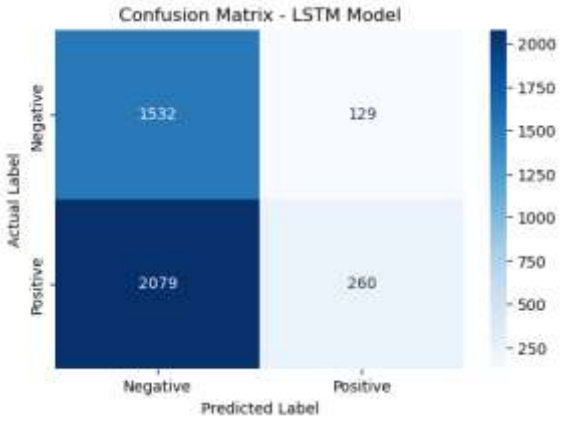

Figure 4. Confusion matrix for LSTM

The Convolutional Neural Network (CNN) had the highest accuracy at 97%, misclassifying just 3% of the test reviews. This shows that CNNs work well for natural language processing because their filters can pick up on important word combinations and phrases that reveal sentiment. This likely helped the CNN generalize better and make fewer mistakes than the other models.

The Feedforward Neural Network (FNN) also did well, reaching 93% accuracy and misclassifying 7% of reviews. While FNNs do not capture the order of words, they can still work well if given good feature representations, like word embeddings or tokenized inputs. In this study, the FNN learned useful overall patterns from the text, which led to high accuracy. However, since FNNs look at each input feature separately, they might miss some important context in the text. Even with this limitation, the FNN performed strongly and its results were close to those of the CNN.

The Long Short-Term Memory (LSTM) model had the lowest accuracy at 45%, much lower than the CNN and FNN. LSTMs are meant to capture the order of words and long-term context in text, so they should work well for sentiment analysis. However, in this experiment, the LSTM did not seem to learn useful patterns from the data.

The CNN model was effective in classifying text reviews sentiment. CNN have been shown to be efficient for NLP tasks with excellent results at semantic parsing (Yih, He, & Meek, 2014). The results in this research are similar to those obtained by (Rani & Kumar, 2019) where they performed sentiment analysis on movie reviews written in Hindi. The model achieved an accuracy of 95% which was better than traditional machine learning algorithms. The findings suggest that both FNN and CNN were highly effective in capturing sentiment-related patterns in the review dataset, despite their architectural differences. Particular, CNN's ability to extract discriminative local features such as key phrases and sentiment-bearing n-grams likely contributed to its high performance.

The FFN, although has a simple architecture leveraged on the word embedding and feature extraction layer to learn patterns. This result may not be obtainable with a larger dataset. Traditional neural networks struggle with training large datasets because of the model's inability to save information at a long term, therefore a loss in efficiency occurs (Abbas & George, 2020) notwithstanding FNN can be a robust model that can handle noisy data (Despotovic & Tanikic, 2017).

Contrary to expectations and much of the existing literature, the LSTM model performed poorly under the training configuration applied in this study. The LSTM misclassified more than half of the test data. Several factors may explain the relatively poor performance of the LSTM model. First, LSTM networks often require larger training datasets to fully exploit their ability to learn long-range dependencies. If the dataset is not sufficiently large or the training process is not optimized, the model may struggle to capture meaningful patterns. Additionally, issues related to data preprocessing, such as tokenization, padding, or vocabulary size, may negatively affect model performance.

In the research done by (Mohammed, Taysir, Taloba, & Farghaly, 2022) LSTM performed poorly because it deals with targets on a par with other context words. In the view of (Murthy, Allu, Andhavarapu, Bagadi, & Belusonti , 2020) LSTM perform better with more training data. While LSTM architectures are theoretically designed to capture long-term dependencies in sequential data, this advantage was not reflected in the results obtained. Although in the work of (Venkata Srinivas, Ch.Satyanarayana, Ch.Divakar, & Sirisha, 2021) LSTM out-performed CNN and FNN with an accuracy score of 87% while CNN had 82%.

5. CONCLUSION

The findings show that the CNN is the most effective architecture for the sentiment analysis task in this study, with the FNN model coming next, whereas the LSTM model fared considerably worse under the present experimental setup. This underscores the need to choose an appropriate neural network architecture according to the features of the dataset and the type of classification problem.

Future research might focus on enhancing the LSTM model by means of hyperparameter optimization, using more complex architectures, or by employing hybrid approaches such as CNN-LSTM models. Moreover, more advanced transformer-based architectures could be looked into in order to further improve performance in the task of sentiment classification when using Nigerian customer review data.

In general, the results of this study show that more complex architectures do not necessarily ensure better performance in sentiment analysis. When the dataset includes strong sentiment cues or when the training conditions are not suitable for sequential learning, simpler models such as the FNN and architectures that focus on features such as CNN can end up performing better than recurrent models.

Yet this study provides valuable insights by showing that FNN and CNN can act as reliable and efficient alternatives for the tasks of sentiment analysis and review classification.

REFERENCES

Abbas, S. K., & George, L. E. (2020). The Performance Differences between Using Recurrent Neural Networks. *Iraqi Journal of Science, 61*(6), 1512-1524. doi:DOI: 10.24996/ijs.2020.61.6.31

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks*, 157-166.

Deriu, J., & Cieliebak, M. (2016). Sentiment Analysis using Convolutional Neural Networks with Multi-Task Training and Distant Supervision on Italian Tweets. *EVALITA. Evaluation of NLP and Speech Tools for Italians.* Naples.

Despotovic, V., & Tanikic, D. (2017). Sentiment Analysis of Microblogs Using Multilayer Feedforward Artificial Neural Network. *Computing and Informatics, 36*, 1127-1142. doi:doi:10.4149/cai 2017 5 1127

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Computation and, V2*. Retrieved from arXiv:1810.04805

Hochreiter, S., & Schmidhuber, J. (1997, November 01). Long Short-Term Memory. *Neural Computation, 9*(8), pp. 1735 - 1780. doi:https://doi.org/10.1162/neco.1997.9.8.1735

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies.* doi:http://dx.doi.org/10.2200/S00416ED1V01Y201204HLT016

Mohammed, A. M., Taysir, S. H., Taloba, A., & Farghaly, M. F. (2022). Arabic aspect based sentiment analysis using bidirectional GRU based models. *Journal of king saud university computer and information sciences*, Pages 6652-6662.

Murthy, G. N., Allu, S. R., Andhavarapu, B., Bagadi, M., & Belusonti , M. (2020, May). Text based Sentiment Analysis using LSTM. *International Journal of Engineering Research & Technology, 9*(5).

Rani, S., & Kumar, P. (2019). Deep Learning Based Sentiment Analysis Using Convolution Neural Network. *Arabian Journal for Science and Engineering, 44*, 3305-3314.

Sanchez, O. (n.d.). Feedforward Artificial Neuronal Network for Analysis of Feelings Based on the Unitec Twitter Social Network. *Academia.edu*.

Venkata Srinivas, k. C., Ch.Satyanarayana, Ch.Divakar, & Sirisha, K. P. (2021). Sentiment Analysis using Neural Network and LSTM. *IOP* (p. 1074). IOP.

Yih, W.-t., He, X., & Meek, C. (2014). *Semantic Parsing for Single Relation Question Answering.* Redmond, WA 98052, USA: Microsoft Research.