International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Deep Opinion Representation Learning for Dynamic Hospitality Demand Estimation from Customer Review Ecosystems

G Raviraju[1*], Chanda Harshitha[1], Naram Durgabhavani[1], Lakkinichatty Akhil[1], Shaik Reshma[1], Uppala Meghana Naga Sai Pavani[1]
[1]Department of Electronics and Communication Engineering, Mother Teresa Institute of Science and Technology, Sanketika Nagar, Kothuru, Sathupally, Khammam, 507303, Telangana, India
*Correspondence: G Raviraju

ABSTRACT
Sentiment analysis has emerged as a crucial application in Natural Language Processing (NLP), enabling the identification and interpretation of opinions and emotions expressed in textual data. With the rapid expansion of social media and online review platforms, customers actively share their experiences regarding products and services, generating large volumes of opinion-rich data. Analyzing this information has become essential for businesses to understand customer preferences, enhance service quality, and remain competitive in a technology-driven environment. Various algorithms have been explored for sentiment classification, including both traditional Machine Learning (ML) techniques and advanced Deep Learning (DL) models. In this work, DL approaches such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) were applied to hotel review data, achieving accuracy rates of 86% and 84%, respectively, demonstrating their ability to capture contextual and sequential patterns effectively. Additionally, traditional classifiers such as Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) were evaluated, producing accuracies of 75%, 71%, 82%, and 71%. The comparative results indicated that DL-based models outperformed conventional methods in sentiment prediction tasks, providing more accurate and reliable insights from customer reviews. These findings highlight the effectiveness of DL techniques in improving business intelligence by enabling better understanding of customer feedback and supporting informed decision-making.

Key words: Sentiment Analysis, Natural Language Processing (NLP), Opinion Mining, Text Classification, Machine Learning (ML), Deep Learning (DL)

1. INTRODUCTION

The rapid expansion of online booking platforms has significantly increased the availability of hotel reviews, making them an essential resource for travellers when selecting accommodations. However, manually processing and interpreting the large volume of unstructured textual data is both time-consuming and inefficient, creating a strong need for automated solutions. Sentiment analysis, also referred to as opinion mining, addresses this challenge by automatically extracting and classifying subjective information from text based on the underlying emotional tone, typically categorizing it as positive, negative, or neutral.

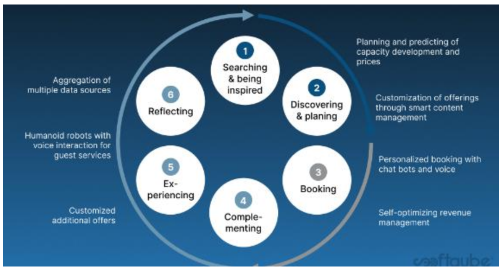

Fig. 1: The Flow of Hotel Review Prediction with Artificial Intelligence

This technique has gained widespread importance across multiple domains, including marketing, customer experience management, and product optimization, where understanding user feedback is critical for strategic decision-making. In this work, a ML-based framework was developed to perform sentiment classification on hotel review datasets, enabling the automated interpretation of customer opinions. The dataset consisted of labeled reviews, which were preprocessed through steps such as noise removal, stop word elimination, tokenization, and feature transformation techniques like vectorization to convert textual data into structured numerical representations. Multiple ML algorithms, including Naive Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest, were implemented and compared to determine the most effective model for sentiment prediction. The performance of each model was evaluated using standard metrics such as accuracy, precision, recall, and F1-score to ensure a comprehensive assessment.

The results demonstrated that automated sentiment analysis systems can efficiently process large-scale review data while maintaining reliable performance. Such systems offer significant practical value in the hospitality industry by enabling businesses to monitor customer satisfaction, identify service gaps, and enhance overall user experience. Furthermore, the proposed framework is adaptable and can be extended to other domains such as e-commerce, social media analytics, and healthcare applications, where large-scale opinion analysis plays a vital role in understanding user behavior and improving service quality.

2. LITERATURE REVIEW

Chengai Sun, et al. [1] proposed an advanced fake review detection approach that combined textual analysis with product-related contextual features to improve classification performance. Their study recognized that traditional text-based methods often fail to capture the relationship between reviews and associated products. To address this, they extracted additional features such as product attributes, reviewer-product consistency, and contextual dependencies across multiple reviews. By integrating these features with Machine Learning (ML) classifiers, they significantly enhanced the model's ability to distinguish between genuine and deceptive reviews. Their work demonstrated that incorporating domain-specific and relational features leads to more accurate and robust fake review detection systems.

N. Jindal, et al. [2] introduced the concept of opinion spam and conducted one of the earliest large-scale analyses of fake reviews in online platforms. Their research categorized spam into different types, including deceptive reviews, duplicate reviews, and irrelevant or promotional content. They applied data mining techniques to identify abnormal patterns in review behavior, such as repetitive postings and unusual rating distributions. Their study highlighted the widespread presence of spam in online review systems and established a foundational framework for detecting fraudulent content, which has influenced many subsequent research efforts in this domain.

M. Ott, et al. [3] developed a supervised learning-based approach for detecting deceptive opinion spam by constructing a benchmark dataset containing both truthful and intentionally fabricated reviews. Their methodology involved extracting linguistic features such as n-grams, part-of-speech patterns, and psycholinguistic indicators to train classification models. They demonstrated that deceptive reviews, despite being written to appear authentic, exhibit subtle differences in language usage, emotional tone, and writing style. Their results showed that ML models can effectively identify these hidden patterns, making their approach a significant contribution to text-based deception detection.

M. Ott, et al. [4] extended their previous work by focusing specifically on negative deceptive opinion spam, which is often used to manipulate public perception and damage product reputation. Their study analyzed how deceptive negative reviews are crafted and identified distinctive linguistic patterns associated with such content. They emphasized that both positive and negative deceptive reviews must be considered in detection systems to ensure comprehensive coverage. Their findings contributed to a deeper understanding of sentiment-driven manipulation in online review platforms.

J. W. Pennebaker, et al. [5] introduced the Linguistic Inquiry and Word Count (LIWC) framework, a powerful tool designed to analyze the psychological and linguistic characteristics of text. Although originally developed for psychological studies, LIWC has been widely adopted in deception detection research. It enables the extraction of features related to emotional expression, cognitive processes, and writing style. Their work provided a strong analytical foundation for identifying subtle linguistic cues that differentiate truthful and deceptive reviews, making it a valuable resource for fake review detection systems.

S. Feng, et al. [6] proposed a syntactic stylometry-based method for detecting deceptive content by analyzing grammatical structures rather than relying solely on lexical features. Their approach focused on syntactic patterns such as sentence construction, parse trees, and structural consistency, which are more difficult for deceptive authors to manipulate intentionally. They demonstrated that syntactic features provide deeper insights into writing behavior and improve the robustness and generalization of detection models. Their study highlighted the importance of moving beyond surface-level text analysis to achieve more reliable deception detection.

J. Li, et al. [7] investigated the development of generalized rules for identifying deceptive opinion spam across multiple domains. Their study focused on analyzing linguistic, semantic, and structural patterns present in deceptive reviews collected from different datasets. They aimed to identify features that remain consistent regardless of domain-specific variations, such as product type or review context. By extracting these invariant characteristics, they developed models that demonstrated strong cross-domain adaptability. Their results showed that leveraging generalized patterns significantly improves the robustness and scalability of fake review detection systems, making them more effective in real-world applications where data diversity is high.

E. P. Lim, et al. [8] proposed a behavior-based approach for detecting fraudulent reviewers by analyzing user activity patterns instead of relying solely on textual content. Their study examined features such as rating deviations, frequency of reviews, temporal posting patterns, and reviewer consistency across different products. They demonstrated that abnormal behavioral patterns are strong indicators of spam activity and can effectively identify malicious users. Their work highlighted that combining behavioral analysis with text-based methods provides a more comprehensive fraud detection framework, improving accuracy and reducing false positives in identifying deceptive reviews.

3. PROPOSED SYSTEM:

The proposed system for hotel review analysis was designed as a comprehensive deep learning pipeline that transforms unstructured customer feedback into meaningful business intelligence through a hybrid CNN–LSTM architecture. The framework begins with a data acquisition module where large-scale hotel review datasets are collected from online platforms. These reviews undergo an extensive preprocessing stage that includes text cleaning, removal of noise and stop words, tokenization, lemmatization, and normalization to ensure consistency and quality of input data. Following preprocessing, the textual data is converted into numerical representations using word embeddings, which capture semantic and contextual relationships between words, allowing the model to understand similarities and contextual meanings within the reviews.

The embedded sequences are then passed into the CNN component, which performs convolution operations using multiple filters to extract local patterns such as key phrases, sentiment-rich expressions, and important contextual cues. Pooling layers are applied to reduce dimensionality while preserving the most significant features, thereby improving computational efficiency and reducing overfitting. The feature maps generated by the CNN are subsequently forwarded to the LSTM network, which processes the sequential nature of the data. The LSTM component, equipped with memory cells and gating mechanisms, captures long-term dependencies and contextual flow across sentences, enabling the system to understand complex linguistic structures and multi-aspect sentiments present in customer reviews. To further enhance performance, the system can incorporate attention mechanisms that allow the model to focus on the most relevant parts of the text when making predictions. The final representation is passed through fully connected layers and an output layer to perform tasks such as sentiment classification and business outcome prediction. The system is trained using optimized loss functions and evaluation metrics to ensure high accuracy and generalization capability. Beyond sentiment classification, the proposed framework is capable of extracting actionable insights such as customer satisfaction levels, service quality indicators, and potential areas for improvement. A deployment layer enables real-time analysis of incoming reviews, while a visualization dashboard presents results in an interpretable format for hotel management. Compared to traditional ML approaches, this DL-based system offers superior scalability, adaptability to large datasets, and the ability to capture complex hidden relationships within textual data. The proposed system provides a robust, intelligent, and data-driven solution that supports strategic decision-making and enhances customer experience in the hospitality industry.

3.1 LSTM

Long Short-Term Memory (LSTM) networks are a special kind of recurrent neural network (RNN) capable of learning long-term dependencies. They excel at sequence prediction problems where the order of data is crucial, outperforming traditional RNNs which suffer from vanishing gradients. This report covers LSTM's architecture, workings, applications, advantages, and limitations.

1. Recurrent Neural Networks (RNNs) and the Vanishing Gradient Problem:

RNNs process sequential data by maintaining a hidden state that carries information from previous time steps. However, traditional RNNs struggle with long sequences because the gradient, used to update the network's weights during training, can vanish as it's backpropagated through many time steps. This vanishing gradient problem prevents the network from learning long-range dependencies, making it difficult to remember information from earlier time steps.

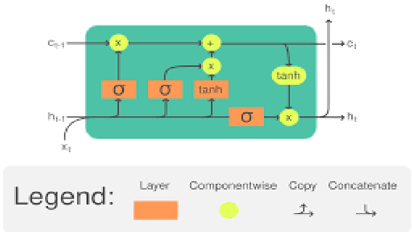

Fig. 2: LSTM Internal workflow

2. The LSTM Solution:

LSTMs address the vanishing gradient problem by introducing a special "memory cell" that can store information for extended periods. Instead of simply updating the hidden state, LSTMs use gates to control the flow of information into and out of the memory cell.

3. LSTM Architecture:

An LSTM unit as shown in Fig. 1 consists of:

- **Cell State (C_t):** The memory cell, which stores information over time.
- **Hidden State (h_t):** A filtered version of the cell state, used for making predictions.
- **Input Gate (i_t):** Controls which information from the current input is added to the cell state.

- **Forget Gate (f_t):** Controls which information from the previous cell state is discarded.
- **Output Gate (o_t):** Controls which information from the cell state is used to update the hidden state.

4. LSTM Working Mechanism (Step-by-Step):

1. **Forget Gate:** Decides what information to throw away from the cell state based on the previous hidden state and the current input. It outputs a value between 0 and 1 for each number in the cell state (0 means completely forget, 1 means keep).
2. **Input Gate:** Decides what new information to store in the cell state. It uses a sigmoid function to decide which values to update and a tanh function to create a vector of new candidate values.
3. **Cell State Update:** The forget gate multiplies the previous cell state, discarding the information deemed irrelevant. The input gate then adds the new candidate values, updating the cell state with new information.
4. **Output Gate:** Decides what information to output based on the cell state. It uses a sigmoid function to filter the cell state and a tanh function to create a new hidden state.

The LSTM model is then trained to understand the sequential nature of the reviews and capture long-term dependencies in the text, which helps in sentiment analysis or rating prediction. The model is evaluated using metrics to measure its performance. LSTM's ability to remember contextual information from previous words in a review makes it particularly effective for understanding the sentiment or feedback in hotel reviews, offering valuable insights into customer experiences.

4. CONCLUSION

In today's digital landscape, advancements in technology have significantly reshaped how people interact with services, especially in the travel and hospitality sector. The increasing use of internet-based platforms has enabled users to conveniently access and utilize online reservation systems without physical effort. Hotel booking platforms, in particular, have simplified the process of selecting and reserving accommodations by allowing users to compare options, view details, and confirm bookings remotely. This convenience has encouraged more individuals to travel and explore various destinations with better planning and reduced uncertainty. Additionally, such systems provide valuable information about pricing, facilities, and locations, improving overall user experience. The growing dependence on these platforms continues to transform travel behavior and service accessibility. In the future, incorporating advanced features such as intelligent recommendations and real-time updates can further enhance efficiency and user satisfaction.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, 2016.

[2] Jindal, N., Liu, B.: Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 219–230. ACM, New York, NY, USA (2008)

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.

[4] M. Ott, C. Cardie, and J. T. Hancock, “Negative deceptive opinion spam,” in Proc. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol., 2013, pp. 497–501.

[5] Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.

[6] J. W. Pennebaker, M. E. Francis, and R. J. Booth, Linguistic Inquiry and Word Count: Liwc 2001, vol. 71, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2001.

[7] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers— Volume 2, pp. 171–175, Association for Computational Linguistics, 2012.

[8] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL ’14), pp. 1566–1576, June 2014.

[9] Kalae, U. K. (2020). Developing scalable Power BI dashboards for enhanced data analysis and strategic business decision-making. International Journal of Enhanced Research in Science, Technology & Engineering, 9(3), 8–15.

[10] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detecting product review spammers using rating behaviors,” in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010