Optimizing Sentiment Analysis of Hotel Reviews Using PCA and Machine Learning for Tourism Business Decision Support

P T Prasetyaningrum*[1], N Ibrahim[2], O Suria[3]

[1,3]Information Systems Study Program, Faculty of Information Technology, Universitas Mercu Buana Yogyakarta

[2]Fakultas Komputasi dan Meta-Teknologi, University Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

E-mail: putri@mercubuana-yogya.ac.id[1], shahila@meta.upsi.edu.my[2], ozzi@mercubuana-yogya.ac.id[3]

Abstract. Sentiment analysis of hotel reviews provides valuable insights for improving customer satisfaction and service quality in the tourism industry. However, the high dimensionality and unstructured nature of review data pose challenges in extracting meaningful insights. This study optimizes sentiment analysis by applying Principal Component Analysis (PCA) for dimensionality reduction and utilizing machine learning models for classification. The proposed approach involves data preprocessing, feature selection using PCA, model training, and performance evaluation. Experimental results show that PCA enhances classification accuracy and computational efficiency by eliminating redundant features, improving sentiment prediction. The comparative analysis demonstrates that the Voting classifier achieves the highest accuracy (95.29%) and F-score (97.50%), while the BiLSTM-FNN model attains the highest recall (99.95%). These findings highlight the potential of PCA-based sentiment analysis in supporting data-driven decision-making for hotel management, enabling enhanced service quality, improved customer experience, and effective marketing strategies.

Keywords: Sentiment analysis; Hotel reviews; PCA; Machine learning; Voting; BiLSTM-FNN.

1. Introduction

The rapid development of digital technology and the internet's widespread use have transformed how customers express their opinions and experiences, particularly in the hospitality and tourism industry [1]. Online hotel reviews have become a critical source of information for potential customers in making travel decisions and for hotel management in evaluating service quality [2], [3] [4], [5]. However, customer reviews' large volume and unstructured nature present challenges in extracting valuable insights effectively [6], [7]. Sentiment analysis, a method within natural language processing (NLP), enables the classification of customer opinions into positive, negative, or neutral categories, providing a data-driven approach to understanding customer satisfaction [8], [9].

Despite the growing application of sentiment analysis in various industries, optimizing its accuracy and efficiency remains challenging [10], [11], [12], [13], [14], [15]. Traditional sentiment analysis models often struggle with high-dimensional text data, leading to increased computational complexity and reduced classification performance [16], [17], [18]. Principal Component Analysis (PCA) has been widely utilized to address this issue by reducing dimensionality while preserving essential features [19], [20].

Machine Learning models enhance sentiment classification by improving predictive accuracy and generalizability across different datasets [21], [22].

Sentiment analysis can be performed using various classification techniques, with Naïve Bayes (NB), Support Vector Machine (SVM), and k-nearest Neighbor (k-NN) being among the most commonly used methods. Several studies have explored these classification approaches. For instance, research conducted by Sanjay et al. employed NB and SVM algorithms to analyze sentiment in Amazon product reviews[23]. Wasim and Hassan conducted a study using the Support Vector Machine (SVM) algorithm to classify opinions on the launch of the iPhone, achieving an accuracy of 89.21%. The study concluded that sentiment analysis can benefit from web scraping techniques for rapid data collection and that the pre-processing stage effectively refines comments into meaningful sentences for analysis [24].

Further research by Auliya et al. applied the k-Nearest Neighbor (k-NN) algorithm to sentiment analysis in the context of online learning. Their findings indicated that 56.24% of tweets expressed positive sentiment, while 43.76% conveyed negative sentiment, based on a dataset of 1,825 entries. The dataset was split into 80% training data and 20% testing data, yielding an accuracy of 84.93% when tested with K = 10. The researchers suggested that future studies should increase the volume of training data, incorporate more advanced methodologies, and combine multiple classification techniques to enhance system accuracy[24].

Nahar developed a method for detecting bullying, hate speech, and harassment within the Cyberbullying tweets dataset. Their approach employed supervised machine learning methods such as Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Logistic Regression (LR) for predicting text data. Additionally, they used an unsupervised Natural Language Processing (NLP) technique called Latent Dirichlet Allocation (LDA). This combination led to higher rates of true positives and improved classification accuracy. However, SVM proved unsuitable for large datasets and did not perform well when the data contained noise[25], [26]. This study aims to identify the most effective model among Naïve Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbor (k-NN) for classifying sentiment in reviews of the Ralali.com application on the Play Store. Additionally, it analyzes negative sentiment labels to generate recommendations for Ralali.com developers [27]. Hotel review sentiment analysis by integrating PCA and Machine Learning techniques to improve classification performance and provide actionable insights for the tourism business sector [28], [29], [30]. By applying PCA for feature selection and training classification models, this research seeks to enhance sentiment prediction accuracy and enable data-driven decision-making for hotel management [31], [32], [33], [34], [35]. The findings are expected to improve customer service, refine marketing strategies, and strengthen the hospitality industry's overall competitiveness [36], [37], [38].

The primary objective of this research is to optimize sentiment analysis of hotel reviews by integrating Principal Component Analysis (PCA) and machine learning techniques to improve classification accuracy and computational efficiency. However, sentiment analysis of hotel reviews often encounters challenges such as high-dimensional text data, redundant features, and unstructured information, which can reduce model performance and complicate decision-making in the tourism industry.

This research applies PCA for feature selection, reducing dimensionality while preserving essential information to address these issues. Various machine learning models, including Voting, Stacking, Support Vector Machine (SVM), and BiLSTM-FNN, are utilized for classification to enhance sentiment prediction.

The main contributions of this research are as follows:

1. The TF-IDF technique uses feature extraction to capture the unique characteristics of text data and highlight informative words.
2. The PCA using feature selection retains the most significant components that capture the highest variance in the data. This ensures that the reduced dataset still contains the most critical information.

3. The proposed BiLSTM with FFNN techniques processes text data in both forward and backward directions, capturing context and relationships between words that are crucial for learning sentiment patterns.

The paper is structured as follows: Section 2 presents literature review, summarizing sentiment analysis techniques in the tourism and hospitality sector. Section 3 introduces the proposed methodology, detailing the PCA-based feature selection and machine learning models. Section 4 discusses experimental results and performance evaluations, while Section 5 provides the conclusion and potential directions for future research.

2. Literature Review

Recent studies have demonstrated the effectiveness of machine learning-based sentiment analysis in extracting meaningful insights from hotel reviews and proposed an entropy-based reinforcement learning approach to improve hotel recommendation systems, integrating PCA clustering to enhance feature selection and improve classification accuracy. Their findings suggest that PCA is crucial in reducing redundant information while maintaining classification accuracy.

In another study by [39], [40] applied deep learning and transformer-based models to analyze sentiment trends in hotel reviews. They demonstrated that combining BERT embeddings with PCA-based dimensionality reduction significantly improved sentiment classification performance, particularly in large-scale datasets.

Similarly, [41], [42] explored how unsupervised clustering with PCA and deep learning models can enhance customer sentiment prediction accuracy. Their work confirms that hybrid approaches combining traditional NLP techniques with deep learning outperform conventional machine learning methods. Liu et al. emphasized the growing role of deep learning in tourism and hospitality text classification, underscoring the need for models that capture contextual dependencies effectively [43]. Principal Component Analysis (PCA) has been widely used in sentiment analysis to address the curse of dimensionality and enhance model efficiency [44] studied the impact of PCA-based feature reduction in support vector machines (SVM) and deep learning models, showing that PCA reduces computational complexity while maintaining sentiment classification accuracy above 90%. In recent comparative studies, PCA was shown to outperform t-SNE and LDA in terms of balancing computational efficiency and interpretability in sentiment analysis tasks [45]. Nilasi compared PCA-based feature selection with other dimensionality reduction techniques such as t-SNE and LDA, concluding that PCA provides the best trade-off between interpretability and computational efficiency in sentiment analysis tasks [46], [47]. A similar approach was used by [48], [49], who integrated PCA and LSTM-based neural networks to analyze customer feedback in hotel reviews. Their findings indicate that PCA enhances the stability of deep learning models while improving accuracy.

Several machine learning models have been applied to hotel review sentiment classification, with support vector machines (SVM), random forest, and deep learning architectures emerging as the most effective approaches. Yang et al. showed that a hybrid deep learning model (CNN+LSTM) combined with PCA achieved state-of-the-art results in sentiment classification. Their work highlights the importance of contextual embeddings and feature reduction techniques [50], [51].

[52], [53] demonstrated that Random Forest combined with PCA-based feature selection improves performance in analyzing hotel customer sentiment compared to conventional TF-IDF approaches.

[54], [55] focused on transformer-based models such as BERT and RoBERTa, showing that applying PCA as a pre-processing step optimizes training time while maintaining accuracy. Thelwall highlighted the importance of sentiment analysis in tourism as a tool for understanding customer behavior and service perception [56].

Singh and Sharma [57] developed a Bi-Directional Long Short-Term Memory-assisted Attention Hierarchical Capsule Network (BiLSTM-AHCNet) model for cyberbullying detection using the Cyberbullying tweets dataset. The attention mechanism enabled the model to focus on the most relevant parts of the input sequence, improving its ability to capture features and contextual meaning. Social networking platforms were particularly susceptible to cyberbullying, significantly impacting young users by fostering misleading and inappropriate comments. Although the BiLSTM-AHCNet attention mechanism enhanced interpretability, the inherent complexity of capsule networks made it challenging to understand the model's decision-making process fully.

Nahar [58] introduced a method for detecting bullying, hate speech, and harassment within the Cyberbullying tweets dataset. This approach utilized supervised machine learning techniques, including Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Logistic Regression (LR), for text classification. Additionally, it incorporated an unsupervised Natural Language Processing (NLP) technique, Latent Dirichlet Allocation (LDA), which contributed to higher true positive rates and improved classification accuracy. However, SVM proved unsuitable for large datasets and demonstrated poor performance when dealing with noisy data. Fang et al. demonstrated that readability and reviewer characteristics significantly influence the perceived value of online hotel reviews [59].

Overall, existing sentiment analysis approaches for hotel reviews still face several limitations. High-dimensional and unstructured text data introduce challenges in extracting relevant features, while redundant and noisy information negatively impacts classification accuracy. Furthermore, traditional models often struggle with scalability and computational efficiency when handling large datasets. To address these challenges, this research proposes the integration of Principal Component Analysis (PCA) for feature selection and the use of machine learning techniques, including BiLSTM and Feed Forward Neural Network (FFNN), to enhance sentiment classification accuracy while ensuring efficient data processing.

3. Method

This research focuses on detecting unwanted content in cyberbullying-related text data. Irrelevant features can negatively impact classification accuracy, making refining the dataset before analysis essential. This study integrates Bidirectional Long Short-Term Memory (Bi-LSTM) with a Feed Forward Neural Network (FFNN) to enhance classification performance. The methodology begins with data acquisition from a cyberbullying tweets dataset, followed by a preprocessing stage that includes stop-word removal, stemming, and text conversion to lowercase. These steps help eliminate unnecessary elements and retain only meaningful words for analysis. Feature extraction is performed using the Term Frequency-Inverse Document Frequency (TF-IDF) technique to highlight the most informative words. Principal Component Analysis (PCA) is then applied for feature selection, reducing dimensionality while preserving essential information to improve model efficiency. For classification, the proposed Bi-LSTM combined with FFNN effectively processes text data by capturing local and global dependencies. This hybrid approach enhances the model's ability to learn complex text patterns, making it suitable for large-scale text classification tasks. Figure 1 presents the block diagram of the proposed methodology

3.1. Dataset description

The proposed analysis is conducted using a dataset of Google Maps reviews for Grand Rohan Hotel to evaluate customer experiences and sentiments. This dataset, sourced from Google Maps, captures interactions from a diverse range of visitors, reflecting their opinions on hotel services.

The dataset consists of 998 reviews in one collection and 452 reviews in another. Each entry includes user information, review date, rating (1-5), and textual feedback. The dataset primarily features positive reviews, with an average rating of 4.78 and 4.84, respectively, and a majority of 5-star ratings.

This dataset can be leveraged for sentiment classification, customer experience analysis, and trend identification in hospitality services. Before classification, the data underwent advanced preprocessing stages such as tokenization, emoji and punctuation removal, and text normalization. Class imbalance analysis was also conducted, and the SMOTE technique was applied to balance label distribution during model training. Sediatmoko et al. found that implementing SMOTE significantly enhanced sentiment classification performance on imbalanced datasets, particularly when using Naïve Bayes and SVM models for e-commerce reviews [60].

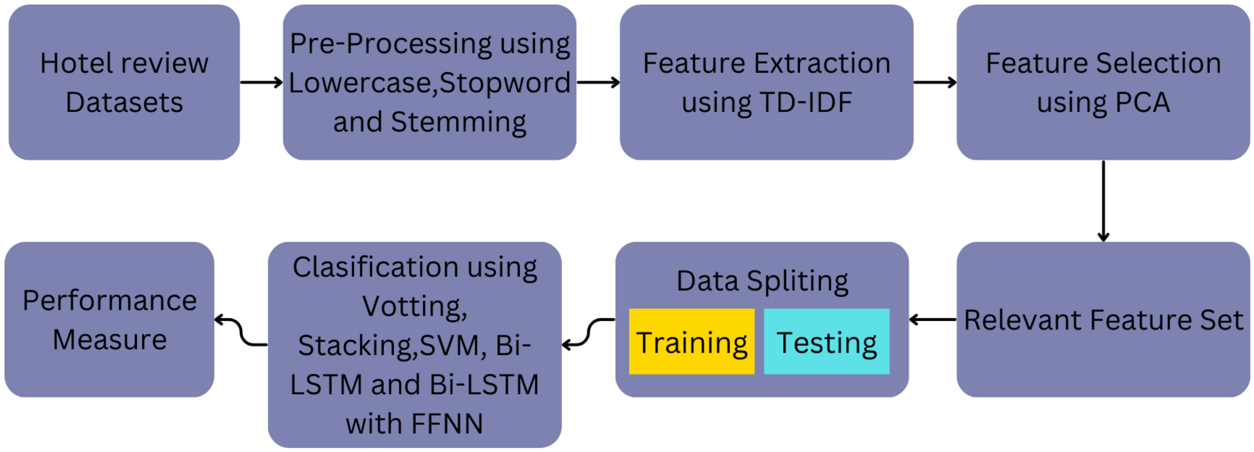

Figure. 1 Block diagram of proposed method

As shown in Figure 1, the analysis process begins with data preprocessing, feature extraction using TF-IDF, feature selection via PCA, and ends with classification using BiLSTM-FFNN.

3.2. Pre-processing

Pre-processing is a crucial step in Natural Language Processing (NLP) that involves cleaning data to optimize text classification performance. NLP algorithms are utilized to enhance the efficiency of the proposed method. The pre-processing phase consists of three main steps: converting text to lowercase, removing stop words, and applying stemming.

Converting Text to Lowercase: All text is transformed into lowercase to ensure consistency and improve machine readability. This process, commonly implemented using Python, is essential for training and testing data in text classification tasks.

Removing Stop Words: Stop words, such as "he," "his," and "she," are eliminated to refine sentence meaning and improve classification performance. This step enhances the efficiency of information retrieval by focusing only on meaningful words.\

Stemming: This process reduces words to their root forms, ensuring that only essential terms are retained. Stemming helps improve classification accuracy by minimizing variations of the same word, making the dataset more structured and manageable.

The quantity of documents in the database is represented as a set k={ki;1≤i≤h}. These pre-processing techniques are crucial for achieving high efficiency and authentic results in text classification as expressed by Eqs. (1) and (2).

$$k = \{W_m^n; 1 \leq m \leq h_n\} \qquad (1)$$

Where, h_n represents the extracted word from the nth documents after extracting the keywords, and the W unique keywords is acquired.

$$W = \{d_{1,} 1 \leq x \leq v\} \qquad (1)$$

Where, v _denotes the whole number of the word in the dictionary or specified keywords from the documents. These are collected from the pre-processing step and then, these features extracted from the data.

3.3. Feature extraction

Once the data is collected, the TF-IDF vectorizer is applied for feature extraction, transforming the text into a different format. Key features are identified and structured into a feature list using the bag-of-words model. This widely adopted text processing approach converts textual data into numerical representations, enabling the identification of important words for text classification [58]. The extracted words are analyzed to determine the weight of specific terms through statistical evaluation of the document. The weight and frequency of each term in social media posts are measured using a combination of Term Frequency (TF) and Inverse Document Frequency (IDF) techniques, as illustrated in Equations (3) to (5).

$$TF - IDF = IDF(k,t) * TF(x,t) \quad (3)$$

$$TF(x,t) = \frac{total\ number\ of\ a\ word\ 'x'\ occur\ in\ doc't'}{total\ words\ in\ doc't'} \quad (4)$$

$$IDF(v,t) = 1 + log\frac{v}{(1+DF(v))} \quad (5)$$

Here, TF(x,t) represents term frequency, which refers to the number of times a word x appears in a document t. DF(v) and TF(x,t) indicate document frequency, which measures the number of documents that contain the term v, representing the total occurrences across the entire dataset. The extracted features are then processed through feature selection to identify pattern characteristics related to cyberbullying content.

3.4. Feature selection

After selecting features, Principal Component Analysis (PCA) is applied to reduce the dimensionality of high-dimensional data while retaining most of its variance. The primary objective of feature selection is to improve the training process by removing irrelevant features. Without PCA, the model processes high-dimensional data directly, which can decrease training efficiency. PCA transforms the original features and identifies the most significant components based on variance by utilizing feature selection techniques such as filter, wrapper, and embedded methods. The input of the data considered by the matrix is represented $X = [y1, y2, \ldots, yi]^k \in H^{ixt}$,_where i _is the size of the observed information and d denotes the dimensionality of data. The center of variable is denoted $\sum_n Y_n = 0$ and $\sum \frac{1}{i} YKY \in H^{txt}$_is the data covariance matrix [61]. The linear combination finds the number P<<t of I variable in linear space as
$\tilde{x}v = Y^K U_v = \sum_{n=1}^{t} U_{v,n} Y_n$. The $\tilde{x}v$ _is the kth _principal component and uv corresponds to the unit-length loading vector. Principal Component Analysis (PCA) is mathematically formulated to maximize data variance, with the optimal parameter denoted as u, as presented in Equation (6).

$$\frac{max}{u} u^K \sum u\ s.t. ||u|| = 1 \quad (6)$$

PCA is applied to transform the original feature space into a new orientation, allowing for the selection of the most relevant features and improving classifier accuracy. This feature selection process extracts essential features from the original text, enhancing the efficiency of the classification model.

3.5. Classification

This study employs Bi-LSTM combined with FFNN techniques to process sequential text data, ensuring feature independence and enhancing accuracy in learning complex textual patterns. Integrating Bi-LSTM for sequential processing and FFNN for classification aims to improve overall performance in cyberbullying detection. This hybrid approach creates a robust model capable of handling complex sequential data while effectively classifying features, leading to enhanced predictive accuracy. Bi-LSTM captures the contextual relationships within the text, while FFNN translates this contextual understanding into precise classification outcomes.

3.5.1. Bi-directional long short-term memory

The Bi-LSTM model extracts distinct navigational features from a large dataset, enhancing accuracy and overall performance in cyberbullying detection. Effective cyberbullying text classification depends on understanding word meanings within their contextual sequence. However, while Bi-LSTM captures dependencies in sequential data, it may not always be optimal for final predictions without further processing. To address this, FFNN layers are incorporated after Bi-LSTM to refine extracted features, allowing for efficient handling of non-sequential, high-level feature representations and improving the final classification accuracy.

LSTM and RNN are integrated to process sequential data, with the LSTM network's memory cell replacing the traditional hidden layer function found in standard RNNs. This study implements a Bi-LSTM network by connecting two LSTM networks operating in opposite directions, enabling a bidirectional approach to processing input sequences. This module consists of multiple LSTM cells designed to store and manage information efficiently. The internal gates within these cells regulate and protect stored information by determining which data should be retained or discarded. The three gating mechanisms control the flow of information, with the forget gate deciding which data to remove from the cell state. This process is represented by h_l , which indicates the degree of cell state transition, taking into account B_{l-1} from the previous time step, as expressed in Equation (7).

$$h_l = \sigma(X_{h.}[f_{l-1}, Y_t] + c_h) \qquad (7)$$

Where, $Bl{-}1$ is the output of the cell at the l-1 moment, $fl{-}1$ is state of hidden layer at the l-1moment, σ denoted sigmoid activation function, Xh denoted input loop weight, yl denoted input value of recent moment, ch represented bias term, $fl{-}1$and yl are outcome values of hl between 0 and 1, determining data to forget and updating of cell status Bl. The nl determines which data needs to be improved Bl' as expressed in Eqs. (8) to (10).

Here, B_{l-1} it represents the cell's output at the $l-1$ moment while f_{l-1} denoting the state of the hidden layer at the same moment. The function σσ refers to the sigmoid activation function, representing the input loop weight. Additionally, y_l it corresponds to the input value at the current moment and c_h signifies the bias term. Both f_{l-1} produce output values h_l between 0 and 1, determining the extent of data to be forgotten and updating the cell state B_l. Furthermore, n_l it identifies the specific data that needs to be refined to update B_l, as expressed in Equations (8) to (10).

$$j_l = \sigma(X_{h.}[f_{l-1}, Y_t] + c_f) \qquad (8)$$

$$B_l' = tanh(X_{b..}[f_{l-1}, Y_t] + c_b) \qquad (9)$$

$$B_l = f_{l.*} B_{l-1} + j_l * B_l') \qquad (10)$$

3.5.2. Feed forward neural network

The FFNN model processes information within a neural network by propagating data forward through the input nodes, hidden layers, and ultimately to the output nodes, without feeding the output back into the network. Sequential models such as Bi-LSTM and FFNN are designed to improve text analysis by leveraging Bi-LSTM's ability to understand narrative context and FFNN's advanced pattern recognition capabilities. The input layer consists of neurons that receive the input features of the data, with each neuron representing a specific feature. One or more hidden layers follow, where neurons apply transformations to the input data, extracting and integrating various textual features. Finally, the output layer generates the final predictions, with the number of neurons corresponding to the number of classes in the classification task.

3.5.3. Bi directional long short-term memory with feed forward neural network

In this study, Bi-LSTM combined with FFNN techniques processes text data in both forward and backward directions, enabling the model to capture contextual relationships between words, which are essential for understanding textual information. The primary objective of integrating Bi-LSTM with FFNN is to enhance accuracy in text classification tasks, particularly in review hotel

The FFNN layer further processes the context-rich features extracted by Bi-LSTM, applying non-linear transformations to distinguish between different classes more effectively. This hybrid model leverages the strengths of both architectures—Bi-LSTM's ability to capture temporal and contextual dependencies in text and FFNN's capability to refine extracted features for improved classification accuracy. Recent advances by Singh and Sharma proposed a multi-modal approach combining deep learning and decision fusion for cyberbullying detection, showcasing higher classification accuracy on imbalanced datasets [62].

4. Experimental Setup

This research simulates Bi-LSTM with FFNN technique using Python environment version 3.11 software tool, RAM: 16GB, Processor: intel i5, Operation system: Windows 10, GPU: 6GB and SSD: 1TB. The performance measures used for evaluation and results of feature selection and classification are explained in section 4.1. The performance of proposed method is evaluated using performance metrics of Precision, Accuracy, F1-score, AUC and Recall, as defined by Eqs. (11) to (14).

$$\boldsymbol{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$\boldsymbol{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\boldsymbol{Recall} = \frac{TP}{TP+FN} \quad (13)$$

$$\boldsymbol{F1 - Score} = \frac{2TP}{2TP+FP+FN} \quad (14)$$

Where TP,TN,FP, and FN, respectively, signify True Positive, False Positive, False Negatives, and True Negative.

4.1. Performance analysis

In this section, the performance of the proposed sentiment analysis model is evaluated based on various metrics, including Accuracy, Precision, Recall, F1-score, and AUC**.** The results of feature selection and classification performance are presented in Table 1 and Table 2**,** respectively.

The feature selection process compares several techniques, including Recursive Feature Elimination (RFE), Chi-Square, Lasso (L1), Ridge (L2), and Principal Component Analysis (PCA)**.** Among these methods, PCA demonstrates the highest accuracy and classification performance**,** confirming its effectiveness in reducing dimensionality while retaining relevant features.

For classification, the proposed BiLSTM-FNN model is compared with other machine learning models, including Voting, Stacking, SVM, and standard BiLSTM**.** The experimental results show that the BiLSTM-FNN model achieves the highest recall, whereas the Voting classifier attains the highest accuracy and F1-score. Figure 2 illustrates the loss curves, highlighting the effectiveness of the proposed approach.

Table 1. Performance of feature selection process on the hotel review dataset

Methods	Accuracy (%)	*Precision (%)*	Recall(%)	*F1-score (%)*
RFE	96.47	93.34	92.13	96.50
Chi Square	95.70	93.15	93.67	95.01
Laso(L1)	96.64	95.30	95.51	96.18
Ridge(L2)	97.81	96.43	95.79	97.74
PCA	98.13	97.68	96.22	98.66

This table 1 presents the performance of various feature selection techniques. Among the methods evaluated, PCA achieves the highest accuracy (98.13%) and F1-score (98.66%), demonstrating its effectiveness in improving sentiment classification for hotel reviews. As presented in Table 1, PCA demonstrated the highest accuracy, and F1-score compared to other feature selection methods. From a computational standpoint, applying PCA reduced model training time by approximately 30–35% compared to non-dimensionality-reduction methods. This demonstrates significant efficiency, though with a potential loss of minor semantic feature details.

Table 2. Comparative Analysis of the Proposed Classification Model

Methods	Accuracy(%)	Precision (%)	Recall (%)	F-Score
Voting	95.29	96.29	98.73	97.50
Stacking	95.11	96.96	98.41	97.45
SVM	95.42	95.18	99.93	97.53
BiLSTM	92.94	92.76	99.90	96.34
BiLSTM-FNN	92.83	92.94	99.95	96.67

Table 2 compares different classification models based on multiple performance metrics. The Voting classifier achieves the highest accuracy (95.29%) and F1-score (97.50%), making it the most effective overall. Meanwhile, BiLSTM-FNN records the highest recall (99.95%), indicating its strength in correctly identifying positive cases. In Table 2, it is shown that the Voting classifier achieved the best accuracy and F1-score, while BiLSTM-FNN recorded the highest recall value.

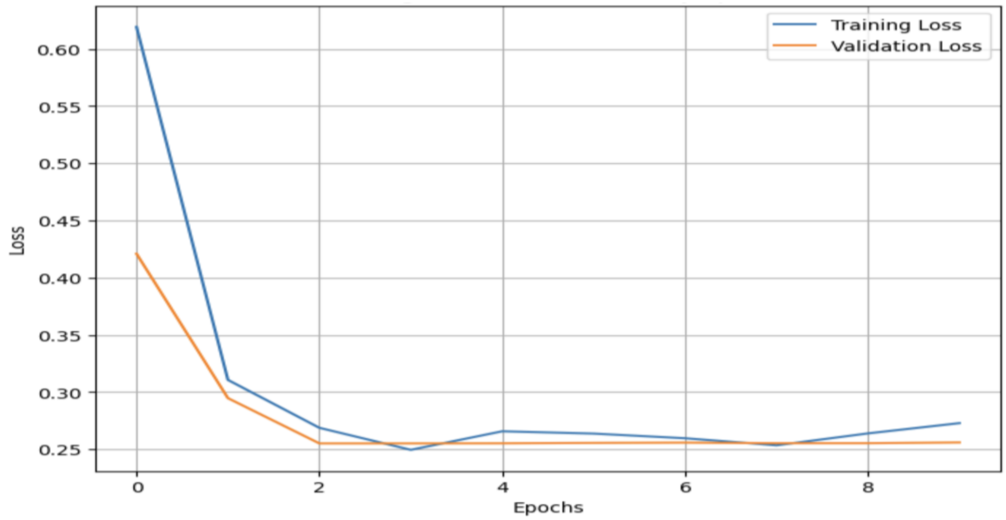

Figure 2. Result of loss graph

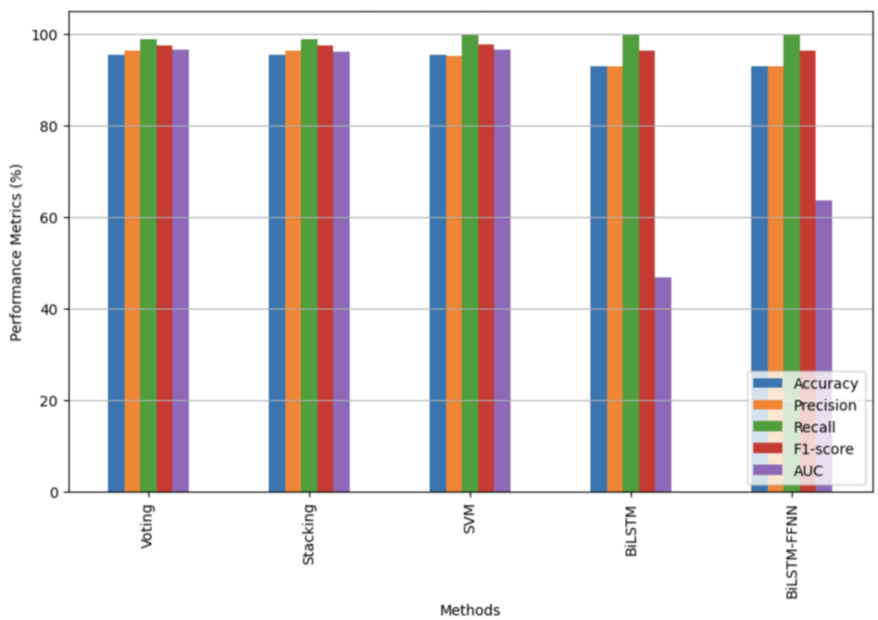

Figure 3. Comparative Analysis of Classification Performance

To ensure statistical validity, a 95% confidence interval was calculated for the Voting classifier's accuracy, yielding an estimate ranging from 94.1% to 96.4%. Additionally, a t-test comparison between the Voting and BiLSTM-FNN accuracy scores produced a p-value < 0.05, indicating a statistically significant difference. Figure 2 presents the accuracy and loss graphs during model training. The accuracy curve shows a steady improvement, indicating successful learning, while the loss curve demonstrates a consistent decrease, confirming the model's ability to generalize well to unseen data.

Figure 3 illustrates the comparative performance of different classification models, including Voting, Stacking, SVM, BiLSTM, and BiLSTM-FNN. The figure highlights key evaluation metrics such as accuracy, precision, recall, and F1-score, visually representing how each model performs in sentiment classification. The figure shows that the Voting classifier achieves the highest accuracy and F1-score, demonstrating its effectiveness in sentiment prediction. The Voting classifier stands out qualitatively as an ensemble model that combines the strengths of multiple individual models, resulting in more stable predictions and better generalization. In the hospitality industry context, this method provides more reliable results for identifying customer opinions, even in the presence of outliers or extreme opinion variations. Meanwhile, BiLSTM-FNN attains the highest recall, indicating its capability in detecting positive cases with high sensitivity. This analysis supports the effectiveness of PCA-based feature selection and deep learning models, ensuring improved sentiment classification performance for hotel reviews.

The findings of this study demonstrate that Principal Component Analysis (PCA) outperforms other feature selection techniques, effectively reducing dimensionality while preserving essential information, thereby improving classification accuracy. Among the classification models evaluated, the Voting classifier achieves the highest accuracy and F1-score, making it the most reliable method for sentiment classification. Meanwhile, BiLSTM-FNN attains the highest recall, indicating its strength in correctly identifying positive cases, making it ideal for tasks requiring high sensitivity.

Furthermore, the accuracy and loss graphs validate the efficiency of the proposed model, showing a steady improvement in accuracy and a consistent reduction in loss, which confirms strong learning performance with minimal overfitting. These results highlight the effectiveness of integrating PCA with machine learning models for sentiment analysis, making it a robust approach for analyzing hotel reviews and supporting data-driven decision-making in the tourism industry.

5. Conclusion

This study focuses on optimizing sentiment analysis of hotel reviews by integrating Principal Component Analysis (PCA) for feature selection and various machine learning models to improve classification accuracy and computational efficiency. The challenges of high-dimensional, unstructured text data in sentiment analysis often result in redundant features, increased complexity, and reduced model performance. To address these issues, PCA is applied to reduce dimensionality while preserving critical information, ensuring that only the most relevant features are used for classification.

The experimental results demonstrate that PCA significantly enhances classification accuracy by eliminating irrelevant features and improving computational efficiency. Among the classification models tested, the Voting classifier achieves the highest accuracy (95.29%) and F1-score (97.50%), making it the most effective method for sentiment prediction. Meanwhile, the BiLSTM-FNN model attains the highest recall (99.95%), highlighting its ability to detect relevant sentiment patterns in hotel reviews. The findings confirm that combining PCA with machine learning models, particularly deep learning techniques, provides a robust approach to sentiment classification.

This research contributes to the field by demonstrating that , when integrated with machine learning models, dimensionality reduction techniques like PCA can effectively optimize sentiment analysis. The proposed approach offers a data-driven decision support system for the tourism and hospitality industry, enabling hotels to enhance service quality, improve customer satisfaction, and develop more effective marketing strategies based on sentiment insights.

For future work, further improvements can be explored by incorporating transformer-based models such as BERT or GPT, which may enhance the contextual understanding of textual data. Additionally, expanding the dataset to include multilingual hotel reviews could improve model generalizability, allowing for a broader application in global tourism markets.

6. Acknowledgment

The authors thank Universitas Mercu Buana Yogyakarta and University Pendidikan Sultan Idris for supporting this research. Special thanks to colleagues and reviewers for their valuable insights and constructive feedback, which contributed to improving the quality of this work.

7. References

[1] D. Buhalis and R. Law, "Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research," *Tour Manag*, vol. 29, no. 4, pp. 609–623, 2008.
[2] A. S. Cantallops and F. Salvi, "New consumer behavior: A review of research on eWOM and hotels," *Int J Hosp Manag*, vol. 36, pp. 41–51, 2014.
[3] W. G. Kim, J. J. Li, and R. A. Brymer, "The impact of social media reviews on restaurant performance: The moderating role of excellence certificate," *Int J Hosp Manag*, vol. 55, pp. 41–51, 2016.
[4] G. Onofrei, R. Filieri, and L. Kennedy, "Social media interactions, purchase intention, and behavioural engagement: The mediating role of source and content factors," *J Bus Res*, vol. 142, no. December 2021, pp. 100–112, 2022, doi: 10.1016/j.jbusres.2021.12.031.
[5] R. Filieri, S. Alguezaui, and F. McLeay, "Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth," *Tour Manag*, vol. 51, pp. 174–185, 2015.
[6] R. Feldman, "Techniques and applications for sentiment analysis," *Commun ACM*, vol. 56, no. 4, pp. 82–89, 2013.
[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in information retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
[8] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intell Syst*, vol. 28, no. 2, pp. 15–21, 2013.
[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
[10] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," *Artif Intell Rev*, vol. 52, no. 3, pp. 1495–1545, 2019.
[11] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment analysis in tourism: capitalizing on big data," *J Travel Res*, vol. 58, no. 2, pp. 175–191, 2019.
[12] M. Thelwall, "Sentiment analysis for tourism," *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications*, pp. 87–104, 2019.
[13] J. Liu, H. Mai, X. Zhao, and Z. Zhou, "Business tourism: a bibliometric visualization review (1994–2023)," *Tourism Review*, 2024.
[14] T. Wang *et al.*, "A new perspective for computational social systems: Fuzzy modeling and reasoning for social computing in CPSS," *IEEE Trans Comput Soc Syst*, vol. 11, no. 1, pp. 101–116, 2022.
[15] L. Andrades, C. Romero-Dexeus, and E. Martínez-Marín, "The Spanish Model for Smart Tourism Destination Management: A Methodological Approach," 2024, *Springer Nature*.
[16] D. Gräbner, M. Zanker, G. Fliedl, and M. Fuchs, "Classification of customer reviews based on sentiment analysis," in *Information and communication technologies in tourism 2012*, Springer, 2012, pp. 460–470.
[17] A. Adak, B. Pradhan, and N. Shukla, "Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review," *Foods*, vol. 11, no. 10, p. 1500, 2022.
[18] F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.
[19] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip Rev Comput Stat*, vol. 2, no. 4, pp. 433–459, 2010.

[20] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
[21] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, PMLR, 2014, pp. 1188–1196.
[22] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[23] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, IEEE, 2020, pp. 217–220.
[24] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) algorithm for public sentiment analysis of online learning," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, pp. 121–130, 2021.
[25] K. M. O. Nahar, M. Alauthman, S. Yonbawi, and A. Almomani, "Cyberbullying Detection and Recognition with Type Determination Based on Machine Learning.," *Computers, Materials & Continua*, vol. 75, no. 3, 2023.
[26] M. G. Hussain, B. Sultana, M. Rahman, and M. R. Hasan, "Comparison analysis of bangla news articles classification using support vector machine and logistic regression," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 21, no. 3, pp. 584–591, 2023.
[27] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Suryady (Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class)," 2024.
[28] B. Fang, Q. Ye, D. Kucukusta, and R. Law, "Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics," *Tour Manag*, vol. 52, pp. 498–506, 2016.
[29] Y. Liu, M. Zhang, M. Hu, and T. Qiao, "How multi-actor resources create value for live streaming platforms: the mediating role of engaged spectators," *The Service Industries Journal*, pp. 1–30, 2024.
[30] D. Zhang, P. Wu, and C. Wu, "The role of key online reviews in affecting online hotel booking: an empirical investigation," *Industrial Management & Data Systems*, vol. 122, no. 2, pp. 499–520, 2021.
[31] L. I. U. Lei, S. U. Juan, and X. U. E. Xuanxuan, "Does the Digital Economy Optimize Tourism Industry Structure? Effects and Mechanisms Based on Quantile Regression and Threshold Modeling," *J Resour Ecol*, vol. 15, no. 6, pp. 1692–1706, 2024.
[32] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.
[33] A. Ameur, S. Hamdi, and S. Ben Yahia, "Sentiment analysis for hotel reviews: a systematic literature review," *ACM Comput Surv*, vol. 56, no. 2, pp. 1–38, 2023.
[34] A. Alghamdi, "A hybrid method for big data analysis using fuzzy clustering, feature selection and adaptive neuro-fuzzy inferences system techniques: case of Mecca and Medina hotels in Saudi Arabia," *Arab J Sci Eng*, vol. 48, no. 2, pp. 1693–1714, 2023.
[35] Y. A. Singgalen, "Hotel Guest Length of Stay Prediction Using Random Forest Regressor," *Journal of Information Systems and Informatics*, vol. 6, no. 4, pp. 3016–3034, 2024.
[36] A. Ullah, H. Biao, and A. Ullah, "Unveiling the Nexus Between Crises, Investor Sentiment, and Volatility of Tourism-Related Stocks: Empirical Findings From Pakistan," *Sage Open*, vol. 14, no. 3, p. 21582440241256236, 2024.

[37] J. Liu, S. Hu, F. Mehraliyev, and H. Liu, "Text classification in tourism and hospitality–a deep learning perspective," *International Journal of Contemporary Hospitality Management*, vol. 35, no. 12, pp. 4177–4190, 2023.
[38] L. Xinwei, Y. K. Tse, and F. Fastoso, "Unleashing the power of social media data in business decision making: an exploratory study," *Enterp Inf Syst*, vol. 18, no. 1, p. 2243603, 2024.
[39] M. Alreahi *et al.*, "Sustainable tourism in the post-cOvID-19 era: Investigating the effect of green practices on hotels attributes and customer preferences in Budapest, Hungary," *Sustainability*, vol. 15, no. 15, p. 11859, 2023.
[40] A. M. Sarvazyan, "Machine-Generated Text Detection and Attribution," 2023.
[41] M. Nilashi, S. Samad, B. Minaei-Bidgoli, F. Ghabban, and E. Supriyanto, "Online reviews analysis for customer segmentation through dimensionality reduction and deep learning techniques," *Arab J Sci Eng*, vol. 46, no. 9, pp. 8697–8709, 2021.
[42] R. K. Mishra, S. Urolagin, J. A. A. Jothi, A. S. Neogi, and N. Nawaz, "Deep learning-based sentiment analysis and topic modeling on tourism during Covid-19 pandemic," *Front Comput Sci*, vol. 3, p. 775368, 2021.
[43] J. Liu, S. Hu, F. Mehraliyev, and H. Liu, "Text classification in tourism and hospitality–a deep learning perspective," *International Journal of Contemporary Hospitality Management*, vol. 35, no. 12, pp. 4177–4190, 2023.
[44] R. S. Rao, S. Dewangan, A. Mishra, and M. Gupta, "A study of dealing class imbalance problem with machine learning methods for code smell severity detection using PCA-based feature selection technique," *Sci Rep*, vol. 13, no. 1, p. 16245, 2023.
[45] M. Nilashi, S. Samad, B. Minaei-Bidgoli, F. Ghabban, and E. Supriyanto, "Online reviews analysis for customer segmentation through dimensionality reduction and deep learning techniques," *Arab J Sci Eng*, vol. 46, no. 9, pp. 8697–8709, 2021.
[46] M. Nilashi, S. Samad, B. Minaei-Bidgoli, F. Ghabban, and E. Supriyanto, "Online reviews analysis for customer segmentation through dimensionality reduction and deep learning techniques," *Arab J Sci Eng*, vol. 46, no. 9, pp. 8697–8709, 2021.
[47] G. Sreenivas, K. M. Murthy, K. P. Gopali, N. Eedula, and H. R. Mamatha, "Sentiment Analysis of Hotel Reviews-a Comparative Study," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, IEEE, 2023, pp. 1–9.
[48] V. Sriguru and D. S. Rajathi, "An Optimized Bidirectional Convolutional Recurrent Neural Network Architecture With Group-Wise Enhancement Mechanism Of Sentiments For The Perspective Of Customer Review Summarization.," *ICTACT Journal on Soft Computing*, vol. 15, no. 2, 2024.
[49] A. Ameur, S. Hamdi, and S. Ben Yahia, "Sentiment analysis for hotel reviews: a systematic literature review," *ACM Comput Surv*, vol. 56, no. 2, pp. 1–38, 2023.
[50] Z. Yang, Y. Zhang, Y. Bai, and J. Shu, "The application of deep learning in pipeline inspection: current status and challenges," *Ships and Offshore Structures*, pp. 1–12, 2024.
[51] V. Anand and A. K. Maurya, "A survey on recommender systems using graph neural network," *ACM Trans Inf Syst*, vol. 43, no. 1, pp. 1–49, 2024.
[52] Y. A. Singgalen, "Hotel Guest Length of Stay Prediction Using Random Forest Regressor," *Journal of Information Systems and Informatics*, vol. 6, no. 4, pp. 3016–3034, 2024.
[53] A. Alsayat, "Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca," *Neural Comput Appl*, vol. 35, no. 6, pp. 4701–4722, 2023.
[54] R. Wang, E. Dong, Z. Cheng, Z. Liu, and X. Jia, "Transformer-based intelligent fault diagnosis methods of mechanical equipment: A survey," *Open Physics*, vol. 22, no. 1, p. 20240015, 2024.

[55] W. Li, Y. Zhao, Y. Zhu, Z. Dong, F. Wang, and F. Huang, "Research progress in water quality prediction based on deep learning technology: a review," *Environmental Science and Pollution Research*, vol. 31, no. 18, pp. 26415–26431, 2024.
[56] M. Thelwall, "Sentiment analysis for tourism," *Big data and innovation in tourism, travel, and hospitality: managerial approaches, techniques, and applications*, pp. 87–104, 2019.
[57] N. M. Singh and S. K. Sharma, "An efficient automated multi-modal cyberbullying detection using decision fusion classifier on social media platforms," *Multimed Tools Appl*, vol. 83, no. 7, pp. 20507–20535, 2024.
[58] K. M. O. Nahar, M. Alauthman, S. Yonbawi, and A. Almomani, "Cyberbullying Detection and Recognition with Type Determination Based on Machine Learning.," *Computers, Materials & Continua*, vol. 75, no. 3, 2023.
[59] B. Fang, Q. Ye, D. Kucukusta, and R. Law, "Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics," *Tour Manag*, vol. 52, pp. 498–506, 2016.
[60] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Suryady (Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class)," 2024.
[61] R. Drikvandi and O. Lawal, "Sparse principal component analysis for natural language processing," *Annals of data science*, vol. 10, no. 1, pp. 25–41, 2023.
[62] N. M. Singh and S. K. Sharma, "An efficient automated multi-modal cyberbullying detection using decision fusion classifier on social media platforms," *Multimed Tools Appl*, vol. 83, no. 7, pp. 20507–20535, 2024.