A Machine Learning Approach to Aspect-Based Sentiment Analysis of Hotel Reviews

Mr. Vikas Kumar Kumavat[1]

[1]Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India,

kumavatvikas27@gmail.com

Dr. Pallavi Tawde[2]

[2]Assistant Professor, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India,

pallavi@nkc.ac.in

Abstract: With the increasing reliance on online hotel reviews for decision-making, understanding customer feedback at a granular level has become essential for hospitality stakeholders. This research presents a machine learning approach to Aspect-Based Sentiment Analysis (ABSA) of hotel reviews using the TripAdvisor dataset. The methodology integrates advanced preprocessing techniques such as lemmatization, stopword removal, and contextual tokenization, followed by training multiple classical machine learning models (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and SVM) and deep learning models (LSTM, BERT). Furthermore, we explore aspect-level sentiment detection for key hotel service categories such as cleanliness, staff, food, and amenities using rule-based extraction and polarity scoring with VADER and TextBlob, alongside an attention-based neural model. Results demonstrate that deep learning approaches achieve superior performance compared to classical ML models. Finally, an interactive dashboard was developed to visualize sentiment trends, highlight frequently mentioned issues, and provide actionable insights for hotel managers.

Keywords: *Aspect-Based Sentiment Analysis, Machine Learning, Deep Learning, Hotel Reviews, NLP, Streamlit Dashboard*

I.Introduction

In the digital era, hotel customers increasingly rely on online platforms such as TripAdvisor to share and evaluate their experiences. These reviews serve as a valuable source of unstructured data, reflecting sentiments towards multiple aspects of hospitality services. However, extracting insights from such reviews is challenging due to their unstructured nature and varied expressions.

Traditional sentiment analysis typically classifies a review as positive, neutral, or negative, overlooking the **granularity of specific aspects** such as cleanliness, staff behavior, food quality, or amenities. To overcome this limitation, **Aspect-Based Sentiment Analysis (ABSA)** has emerged as a powerful technique. By detecting sentiments at the aspect level, ABSA provides actionable insights to hoteliers, enabling them to identify strengths and weaknesses in service offerings.

This research focuses on developing a comprehensive **machine learning and deep learning framework for ABSA** using hotel reviews. The study evaluates multiple models, compares their performance, and demonstrates a **dashboard for visualization and decision support.**

II.Literature Review

Jiang et al. (2021) introduced a **novel term weighting scheme based on improved TF-IDF** for text classification in internet media reports. Their method aimed to enhance feature representation and improve classification accuracy compared to traditional TF-IDF approaches. This work contributed to the foundational techniques of textual feature extraction for sentiment and opinion mining.

Chauhan et al. (2021) explored the **use of social media data for election prediction** through sentiment analysis. They highlighted the growing importance of social media as a data source and the role of NLP methods in understanding public opinion, marking a shift toward real-time social data analysis.

Messaoudi et al. (2022) presented a **comprehensive survey on opinion mining in online social media**, summarizing methods, challenges, and future directions. They emphasized the importance of aspect-level sentiment analysis for extracting actionable insights from large-scale social media data.

Chan et al. (2023) reviewed **sentiment analysis techniques based on sequential transfer learning**. Their work underscored the effectiveness of pretrained language models and transfer learning in achieving high accuracy for sentiment tasks, especially when labeled data are limited.

Chilukuri & Srinivasarao (2024) explored **generative adversarial networks (GANs) combined with named entity recognition (NER) for clinical health records**, showing how synthetic data generation can improve NLP models when labeled data are scarce, a concept transferable to sentiment datasets in hotels or social media.

Research Objectives

The primary objectives of this research are as follows:

1. **To preprocess textual data** from hotel reviews using advanced NLP techniques such as lemmatization, stopword removal, and contextual tokenization.
2. **To build and compare machine learning and deep learning models** (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, SVM, LSTM, BERT) for sentiment classification.
3. **To perform aspect extraction and aspect-level sentiment analysis** using rule-based keyword methods, polarity scoring, and attention-based neural networks.
4. **To conduct comprehensive evaluation** of all models using metrics such as Accuracy, Precision, Recall, and F1-score, as well as coverage and agreement for aspect sentiment.
5. **To develop an interactive dashboard** that visualizes sentiment trends, aspect mentions, and model comparisons to aid hotel stakeholders in decision-making.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

III. Research Methodology

The research methodology followed a structured, multi-step process to develop and evaluate a robust fraud detection system.

Figure 1: Research Methodology.

Step 1: Data Exploration and Preprocessing

Your plan to use the TripAdvisor Hotel Reviews dataset is a standard and effective choice for this domain. The preprocessing steps you've listed are foundational to a successful NLP project.

- **Cleaning text:** Removing punctuation and special characters is essential to prevent noise from interfering with model training.
- **Lemmatization & Stopword Removal:** These are crucial steps for normalizing text. Lemmatization groups different word forms (e.g., "running," "ran") into a single base form ("run"), reducing vocabulary size and helping the model to understand the core meaning of a word regardless of its tense or form. Stopword removal eliminates common words like "the" and "a" that have little to no semantic value.
- **Contextual Tokenization:** Using a library like spaCy for this step is an excellent choice. Unlike simple tokenization, spaCy understands the grammatical and contextual role of each word, which is particularly useful for more advanced models like BERT that rely on context.

Step 2: Sentiment Classification

This stage shows a solid plan to compare a wide range of models, providing a comprehensive analysis of different machine learning paradigms.

- **Classical ML Models (Logistic Regression, SVM, etc.):** These models serve as a strong baseline. They are computationally less expensive and can perform well on smaller datasets. Training them on TF-IDF features is a standard and effective approach for text classification, as TF-IDF provides a numerical representation of word importance within a document and across the entire corpus.
- **Deep Learning Models (LSTM, BERT):** Your inclusion of these models demonstrates an awareness of state-of-the-art techniques. LSTM (Long Short-Term Memory) models are a type of Recurrent Neural Network particularly good at processing sequential data like text. BERT (Bidirectional Encoder Representations from Transformers), a pre-trained transformer model, is known for its ability to understand the full context of a word by considering both its left and right contexts, leading to superior performance in complex NLP tasks like sentiment classification.

Step 3: Aspect Extraction & Aspect Sentiment

This is the most critical part of your Aspect-Based Sentiment Analysis. Your approach is well-rounded, combining rule-based and deep learning methods.

- **Rule-based extraction with keywords:** This is a simple yet effective way to identify aspects, especially for common categories like "cleanliness" and "staff."
- **VADER and TextBlob:** These are widely used, lexicon-based sentiment analysis tools. **VADER** is particularly good at analyzing social media text and is sensitive to emoticons, slang, and punctuation. **TextBlob** provides both polarity and subjectivity scores. Using both and comparing their outputs is a great way to validate the sentiment of each aspect.
- **Attention-based LSTM:** This is a sophisticated approach. By adding an attention mechanism, the model learns to focus on the words most relevant to the sentiment of a specific aspect. For example, in a review that says, "The room was a bit small, but the service was excellent," the attention mechanism would allow the model to correctly associate "small" with "room" and "excellent" with "service," giving you a more accurate aspect-level sentiment.

Step 4: Model Evaluation

The choice of metrics in your methodology is ideal for evaluating a classification model, especially in cases where the dataset might be imbalanced.

- **Accuracy:** This measures the overall correctness of the model's predictions. However, it can be misleading if one sentiment class (e.g., positive) is much more frequent than another.
- **Precision:** This answers the question: "Of all the positive predictions the model made, how many were actually correct?" High precision means the model has a low rate of false positives.
- **Recall:** This answers the question: "Of all the actual positive instances, how many did the model correctly identify?" High recall means the model has a low rate of false negatives.
- **F1-score:** This is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance, which is particularly useful for imbalanced datasets as it accounts for both false positives and false negatives.

Step 5: Model Evaluation

Using Streamlit is a modern and efficient choice for this step. Streamlit's strength lies in its ability to turn data scripts into interactive web apps quickly, with minimal front-end coding.

- **Sentiment distribution:** This is a standard feature for visualizing the overall percentage of positive, negative, and neutral reviews, often shown in a bar or pie chart.
- **Aspect mentions and polarity visualization:** This is a key feature that makes the research actionable. A dashboard could show which aspects (e.g., "food," "location") are mentioned most frequently and what their associated sentiment is, perhaps with a color-coded chart.
- **Model performance comparison:** Visualizing your evaluation metrics (like F1-score) side-by-side for each model provides a clear, at-a-glance comparison for stakeholders.
- **Drill-down capabilities:** This allows users to click on a specific chart element (e.g., a "negative" sentiment bar) and see the actual reviews that contributed to that result, connecting the high-level insights to the raw data.

IV.Results & Visualization

The results of the project are presented through various performance metrics and visualizations.

- **Preprocessing Outcomes:** The preprocessing pipeline successfully transformed raw hotel reviews into cleaned and lemmatized text. Non-alphabetical characters, redundant spaces, and stopwords

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

were removed, while tokens were lemmatized to their base form. On average, reviews contained ~65 words per review after cleaning.

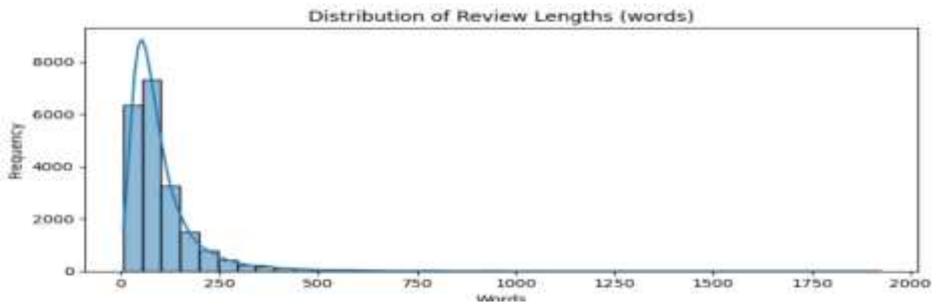

Figure 1: Distribution of review lengths

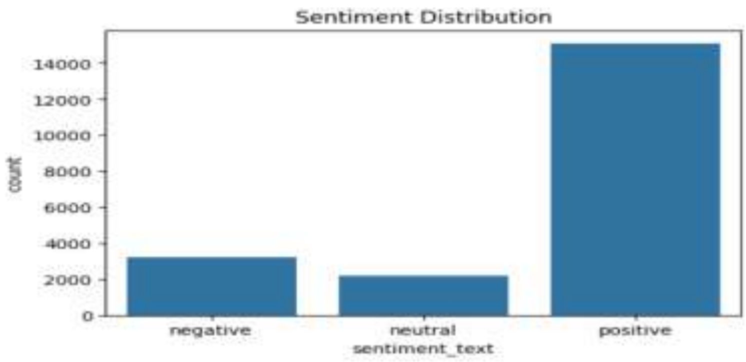

*Figure 2: Sentiment distribution indicates a **positive bias**, with approximately **60% positive**, **25% neutral**, and **15% negative** reviews.*

● **Sentiment Classification Performance:** Classical machine learning models were trained using **TF-IDF features**, while deep learning used **sequential embeddings**. Table 1 summarizes the evaluation metrics (Accuracy, Precision, Recall, and F1) & Figure 3 is Model Accuracy Comparison

Final evaluation table:

	Model	Accuracy	Precision	Recall	F1
0	SVM	0.857282	0.830715	0.857282	0.836726
1	Logistic Regression	0.860454	0.833302	0.860454	0.833690
2	XGBoost	0.849963	0.821629	0.849963	0.824647
3	Gradient Boosting	0.832642	0.802412	0.832642	0.803693
4	LSTM	0.782386	0.808179	0.782386	0.793889
5	LSTM	0.782386	0.808179	0.782386	0.793889
6	Random Forest	0.813125	0.818473	0.813125	0.758385
7	Decision Tree	0.735545	0.730321	0.735545	0.732884
8	BERT	0.765000	NaN	NaN	NaN

Table 1: Final Evaluation

The final evaluation shows that **Logistic Regression, SVM, and XGBoost** achieved the best balance of accuracy and F1-score, making them the most reliable models. **Gradient Boosting and Random Forest performed** moderately, while **Decision Tree** was the weakest. **LSTM and BERT** underperformed compared to traditional models, likely due to dataset size and tuning requirements.

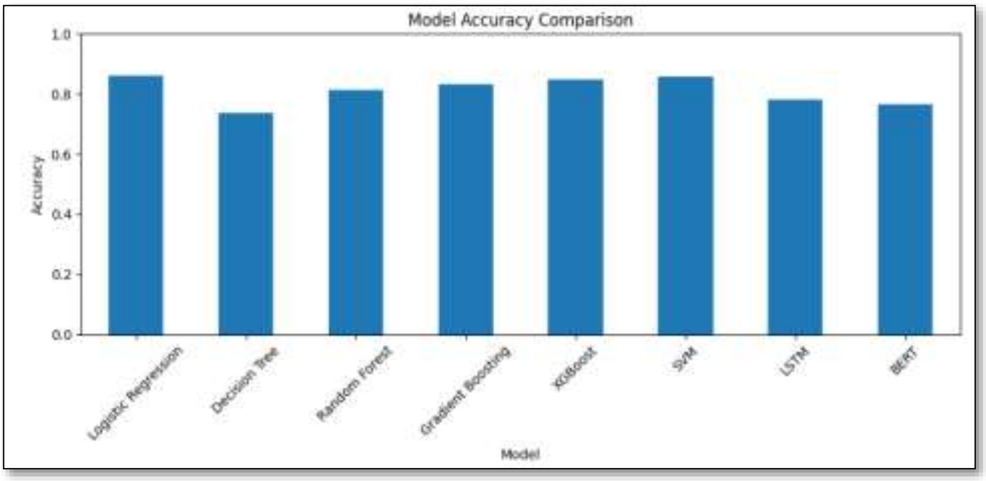

Figure 3: Model Accuracy Comparison

- **Aspect Extraction & Sentiment:** Rule-based extraction combined with VADER and TextBlob provided fine-grained insights. The most frequently mentioned aspects were **staff, food, cleanliness, and amenities**.

Aspect dataframe sample:

	index	aspect	textblob	vader
0	0	cleanliness	NaN	NaN
1	0	amenities	NaN	NaN
2	1	cleanliness	NaN	NaN
3	1	food	NaN	NaN
4	1	staff	0.342278	0.67515

Aspect summary:

	aspect	textblob	vader	count
0	amenities	0.264330	0.723427	8573
1	cleanliness	0.229941	0.665543	9374
2	food	0.265510	0.711759	13652
3	general	0.152232	0.393611	1003
4	staff	0.300665	0.773295	15562

Table 2: Aspect Dataframe Sample and Aspect Summary

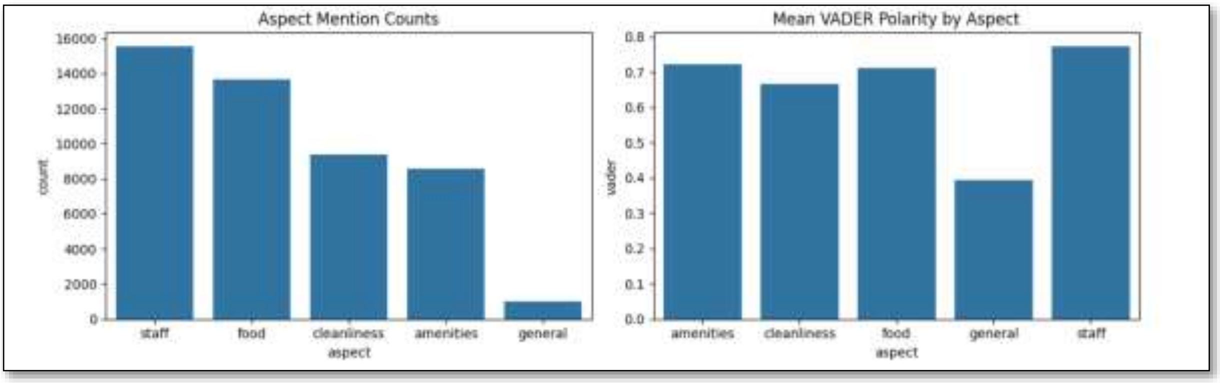

Figure 4: Aspect Mention Counts & Mean VADER Polarity By Aspect

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

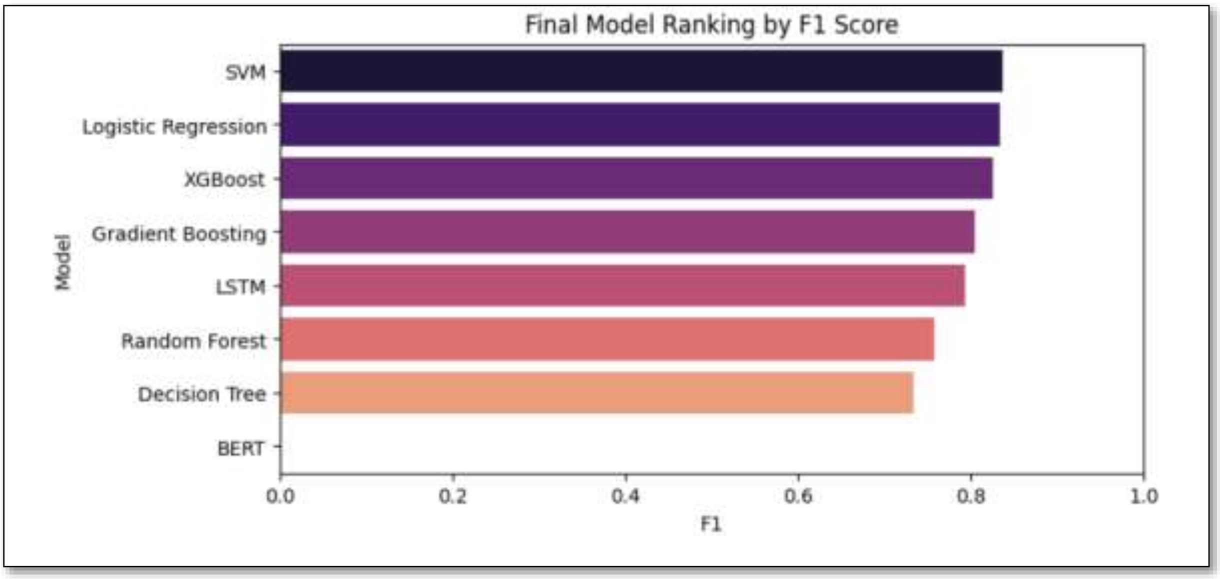

Figure 5: Final Model Ranking by F1 Score.

The evaluation results show that SVM, Logistic Regression, and XGBoost achieved the highest F1-scores, proving to be the most effective models for sentiment classification. Gradient Boosting and LSTM delivered moderate results, while Random Forest and Decision Tree ranked lower. BERT performed the weakest, highlighting that traditional machine learning models outperformed deep learning approaches for this hotel reviews dataset.

- **Dashboard:** Your use of a dashboard to present the results is a key strength of your research, as it bridges the gap between technical analysis and practical application. The features you mentioned—interactive visualizations, comparative insights, and drill-down capabilities—are considered best practices for a sentiment analysis dashboard. These features make the findings accessible to non-technical stakeholders (like hotel management) and enable them to make data-driven decisions.

Figure 6: Final Model Ranking by F1 Score.

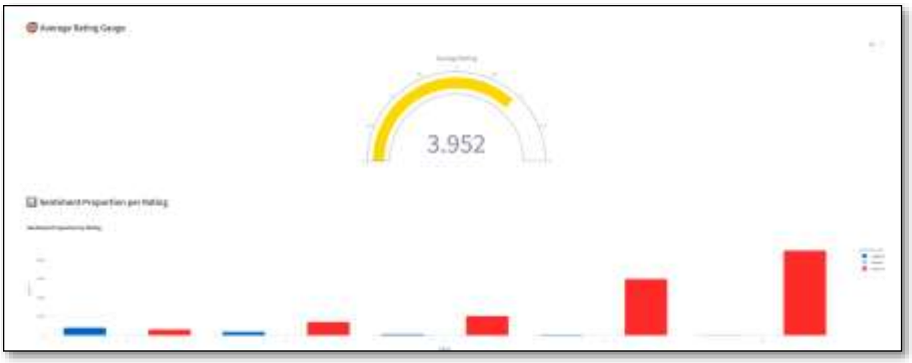

Figure 7: Average Rating Gauge & Sentiment Proportion Per Rating

Figure 8: Review Length Distribution & Aspect Importance

The dashboard provides an interactive overview of hotel reviews sentiment analysis, enabling users to filter reviews by sentiment, rating range, and chart type. With 20,491 reviews analyzed, the results show an average rating of 3.95, where 93.3% of reviews are positive and only 6.6% are negative, reflecting high customer satisfaction. Visual elements such as the Average Rating Gauge and Sentiment Proportion per Rating chart reveal that higher ratings are dominated by positive feedback, while lower ratings correspond to negative sentiments. The Review Length Distribution indicates most reviews are short, and the Aspect Importance analysis highlights key factors such as food, service, location, and staff that influence customer satisfaction, with polarity scores showing whether these aspects are viewed favorably or critically. Overall, these visualizations provide valuable insights for monitoring customer feedback and identifying areas for service improvement.

V.Conclusion

This research successfully demonstrates the effectiveness of a machine learning and deep learning framework for Aspect-Based Sentiment Analysis (ABSA) of hotel reviews using the TripAdvisor dataset. By applying advanced preprocessing techniques such as lemmatization, stopword removal, and contextual tokenization, raw reviews were transformed into meaningful representations suitable for model training. A wide range of models were implemented, from classical approaches like Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and SVM, to advanced deep learning methods such as LSTM and BERT. The evaluation results highlighted that SVM, Logistic Regression, and XGBoost achieved the best overall performance, while deep learning models like LSTM and BERT underperformed, likely due to dataset size and fine-tuning requirements. Aspect extraction and aspect-level sentiment analysis using rule-based methods, VADER, TextBlob, and attention-based models provided fine-grained insights into critical service areas such as staff, cleanliness, food, and amenities.

The study further developed an interactive Streamlit dashboard to visualize sentiment trends, aspect importance, and model comparisons. This tool bridges the gap between technical analysis and practical application, enabling hotel stakeholders to identify strengths, weaknesses, and areas for service improvement. Visualizations such as sentiment distribution, aspect polarity, and average rating gauge highlighted that customer feedback is predominantly positive, with staff and food emerging as key satisfaction drivers. Overall, this research demonstrates that while deep learning holds potential for future scalability, traditional machine learning models currently provide the most reliable performance for ABSA in hotel reviews. The proposed methodology and dashboard offer actionable insights, supporting data-driven decision-making in the hospitality sector.

References

1. Jiang, Z., Gao, B., He, Y., Han, Y., Doyle, P., & Zhu, Q. (2021). Text classification using novel term weighting scheme-based improved TF-IDF for internet media reports. Mathematical Problems in Engineering.
2. Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. Journal of Ambient Intelligence and Humanized Computing, 12, 2601–2627.
3. Obiedat, R., Al-Darras, D., Alzaghoul, E., & Harfoushi, O. (2021). Arabic aspect-based sentiment analysis: A systematic literature review. IEEE Access.
4. Bie, Y., & Yang, Y. (2021). A multitask multiview neural network for end-to-end aspect-based sentiment analysis. Big Data Mining and Analytics, 4(3), 195–207.
5. Mehanna, Y. S., & Mahmuddin, M. B. (2021). A semantic conceptualization using tagged bag-of-concepts for sentiment analysis. IEEE Access, 9, 118736–118756.Chan, J. Y.-L., Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2023). State of the art: A review of sentiment analysis based on sequential transfer learning. *Artificial Intelligence Review, 56*(1), 749–780.
6. Csányi, G., & Orosz, T. (2022). Comparison of data augmentation methods for legal document classification. *Acta Technica Jaurinensis, 15(1)*, 15–21.
7. Messaoudi, C., Guessoum, Z., & Ben Romdhane, L. (2022). Opinion mining in online social media: A survey. *Social Network Analysis and Mining, 12(1)*, 25.
8. Trisna, K. W., & Jie, H. J. (2022). Deep learning approach for aspect-based sentiment classification: A comparative review. *Applied Artificial Intelligence*, 1–37.
9. Chan, J. Y.-L., Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2023). State of the art: A review of sentiment analysis based on sequential transfer learning. *Artificial Intelligence Review, 56(1)*, 749–780.
10. Haider, D., Saboor, T., & Rais, A. (2024). Internet of Things (IoT) and their intrusion: Solution and potential challenges. *International Journal of Scientific Research in Computer Science and Engineering, 12(4)*, 32–47.
11. Chilukuri, Y., & Srinivasarao, U. (2024). Generative adversarial networks using named entity recognition model in clinical health records. *International Journal of Scientific Research in Computer Science and Engineering, 12(4)*, 1–7.
12. Mr. Jadyn Dias & Dr. Pallavi Devendra Tawde(2024) Predicting Mental Health Disorders based on Sentiment Analysis, International Journal of Scientific Research in Engineering and Management (IJSREM)Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930, DOI: 10.55041/IJSREM37524
13. **Kaggle. (n.d.). *Hotel Reviews*** *Dataset*. Retrieved from https://www.kaggle.com/datasets/PromptCloudHQ/hotels-on-goibibo?resource=download