



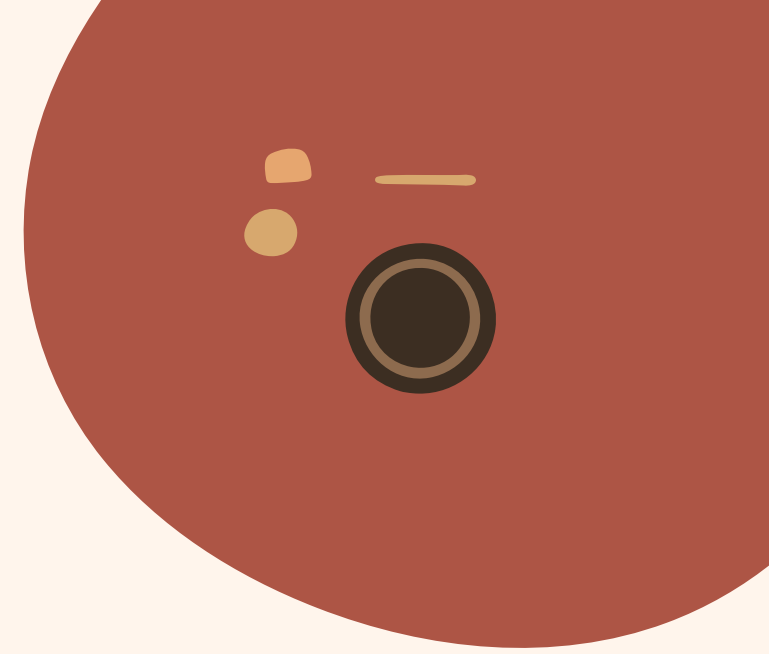
TERM EXTRACTION OF ASPECT BASED SENTIMENTAL ANALYSIS

FOUNDATION OF DATA ANALYTICS

Review-3

START



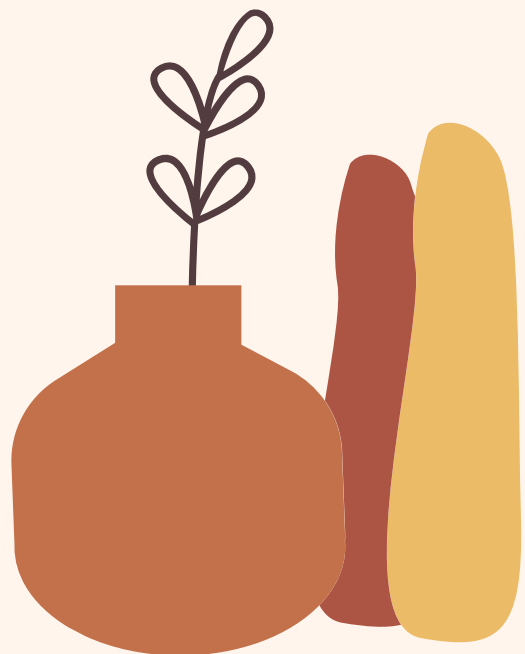


Our Team


Sakshi Chauhan
(20MIA1164)

Aditya Kumar Singh
(20MIA1088)


Sudeeksha Ganguli
(20BCE1549)

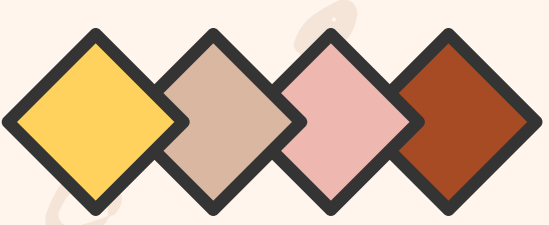


Topic description



Aspect-Based Sentiment Analysis (ABSA) is a sophisticated text analysis technique that categorises data by aspect and identifies the sentiment attributed to each one. It is particularly useful in analyzing customer feedback by associating specific sentiments with different aspects of a product or service. For instance, in the context of hotel reviews, aspects could include service, rooms, location, value, cleanliness, sleep quality, and business service.



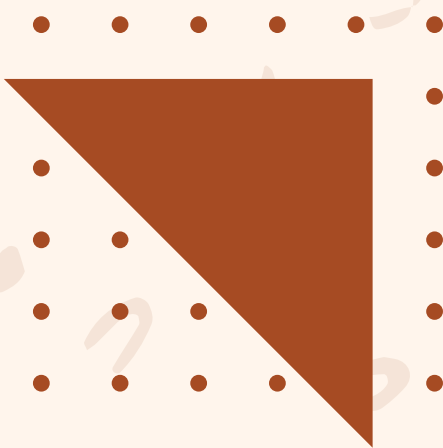


About the Dataset

The dataset provides a wealth of information about each review, including the name and geographical location of the hotel, the date of the review, the nationality of the reviewer, and the content of the review itself. The review content is further divided into positive and negative sections, allowing for a more nuanced understanding of the customer's experience. It has been used in various projects, such as creating sentiment analysis models, building recommendation systems, and performing exploratory data analysis.



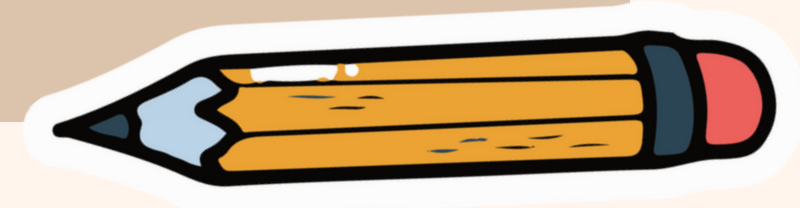
Dataset link: <https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe>





Project Novelty



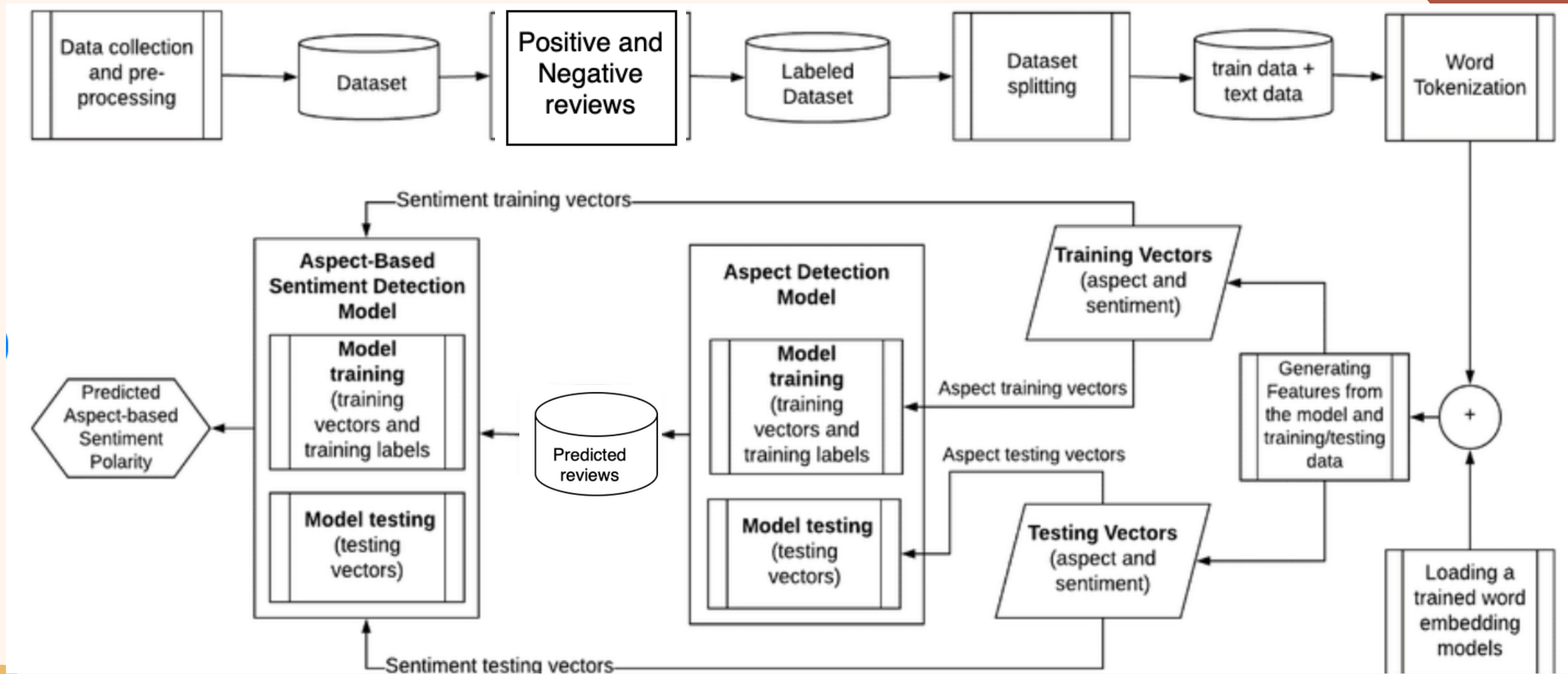
1. **Large-scale Application:** The dataset contains over half a million reviews, making it one of the largest datasets used for aspect-based sentiment analysis in the hospitality industry. This large-scale application lead us to a more robust and generalisable model usage.
2. **Geographical Diversity:** The dataset covers 1,493 luxury hotels across Europe, providing a wide geographical range. This diversity allowed our project to uncover regional differences in customer sentiment and preferences, which has not been extensively explored in previous studies.
3. **Integration of Positive and Negative Reviews:** The dataset separates review content into positive and negative sections. This unique feature enabled the project to perform a more nuanced analysis of customer sentiment, providing valuable insights into what customers like and dislike about their hotel experiences.



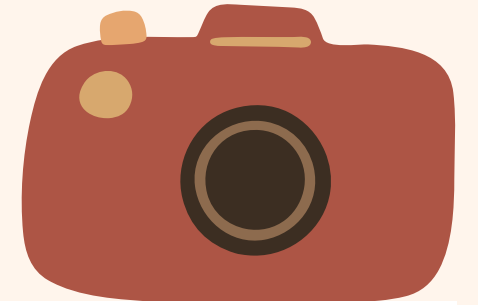
Challenges

- 
1. **Noisy data:** Hotel reviews can be noisy, with users providing star ratings that may not accurately reflect their sentiments. This made it difficult to establish a reliable ground truth for training and evaluating the machine learning models which we used initially.
 2. **Text preprocessing:** Hotel reviews can contain various linguistic nuances, such as slang, abbreviations, and misspellings, which often complicated text preprocessing and feature extraction.
 3. **Aspect identification:** Identifying relevant aspects in the reviews was challenging, as they were expressed both implicitly and explicitly, and they were also intertwined with other aspects. This made it difficult to accurately associate sentiments with specific aspects.
 4. **Sentiment classification:** Sentiment classification was challenging due to the context-dependent nature of sentiments. For example, a positive sentiment in one context may be negative in another, making it difficult to accurately classify sentiments for each aspect.
- 

Proposed architecture

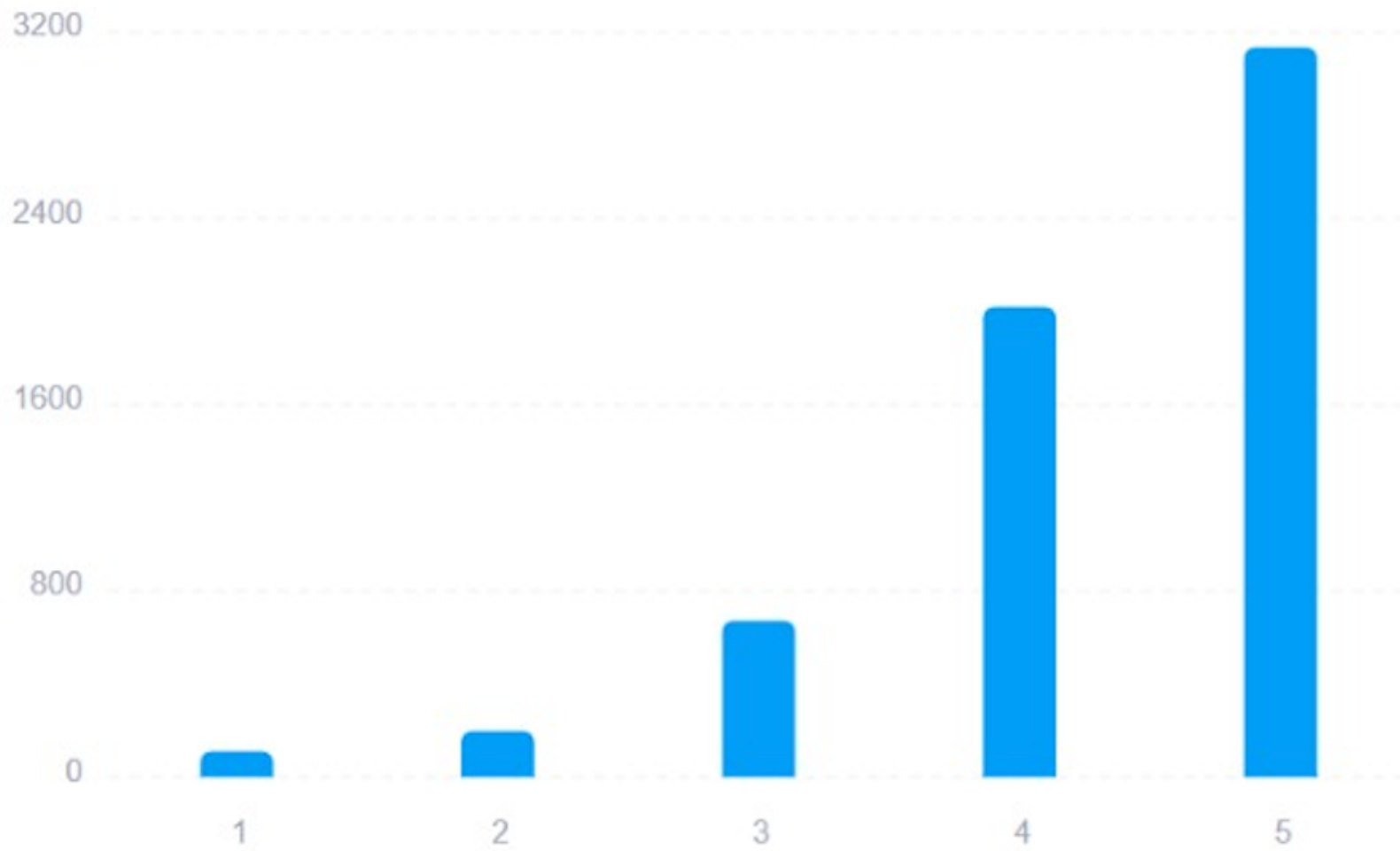


Results

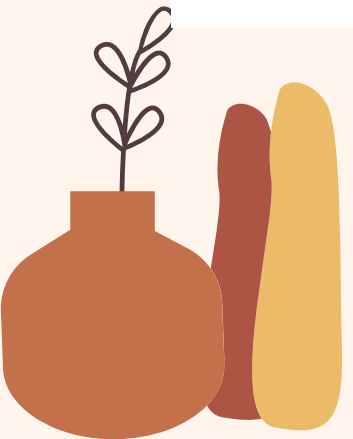
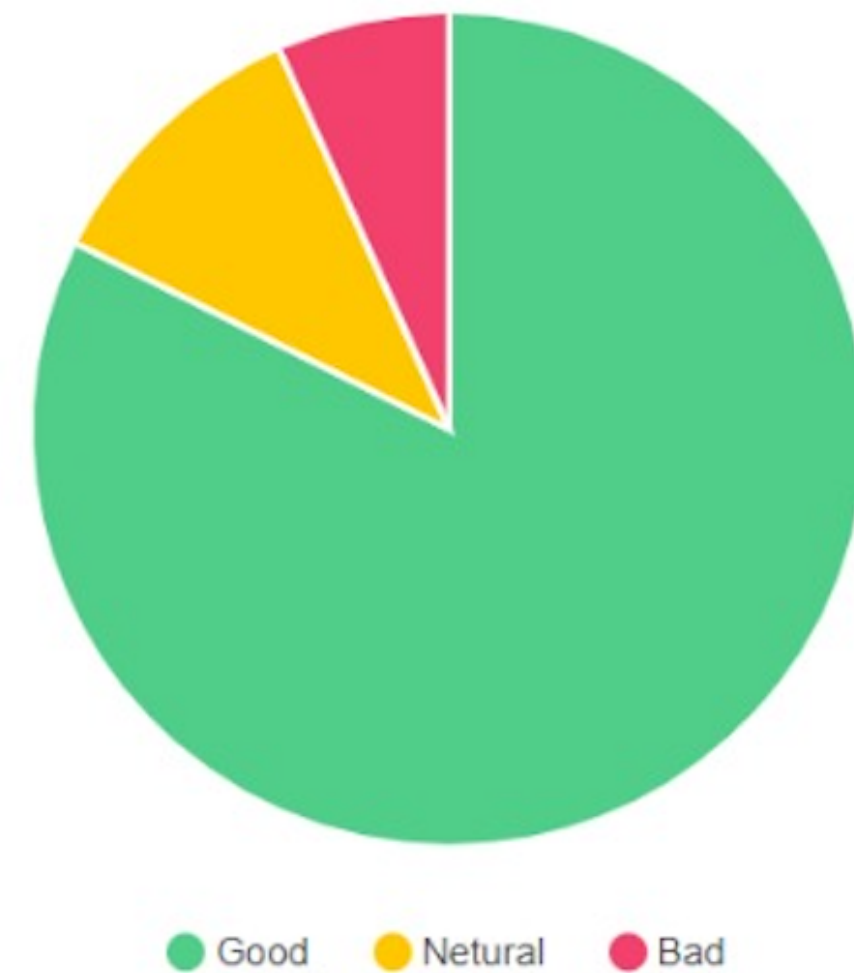


Overall Rating Distribution

More than 1000 reviews

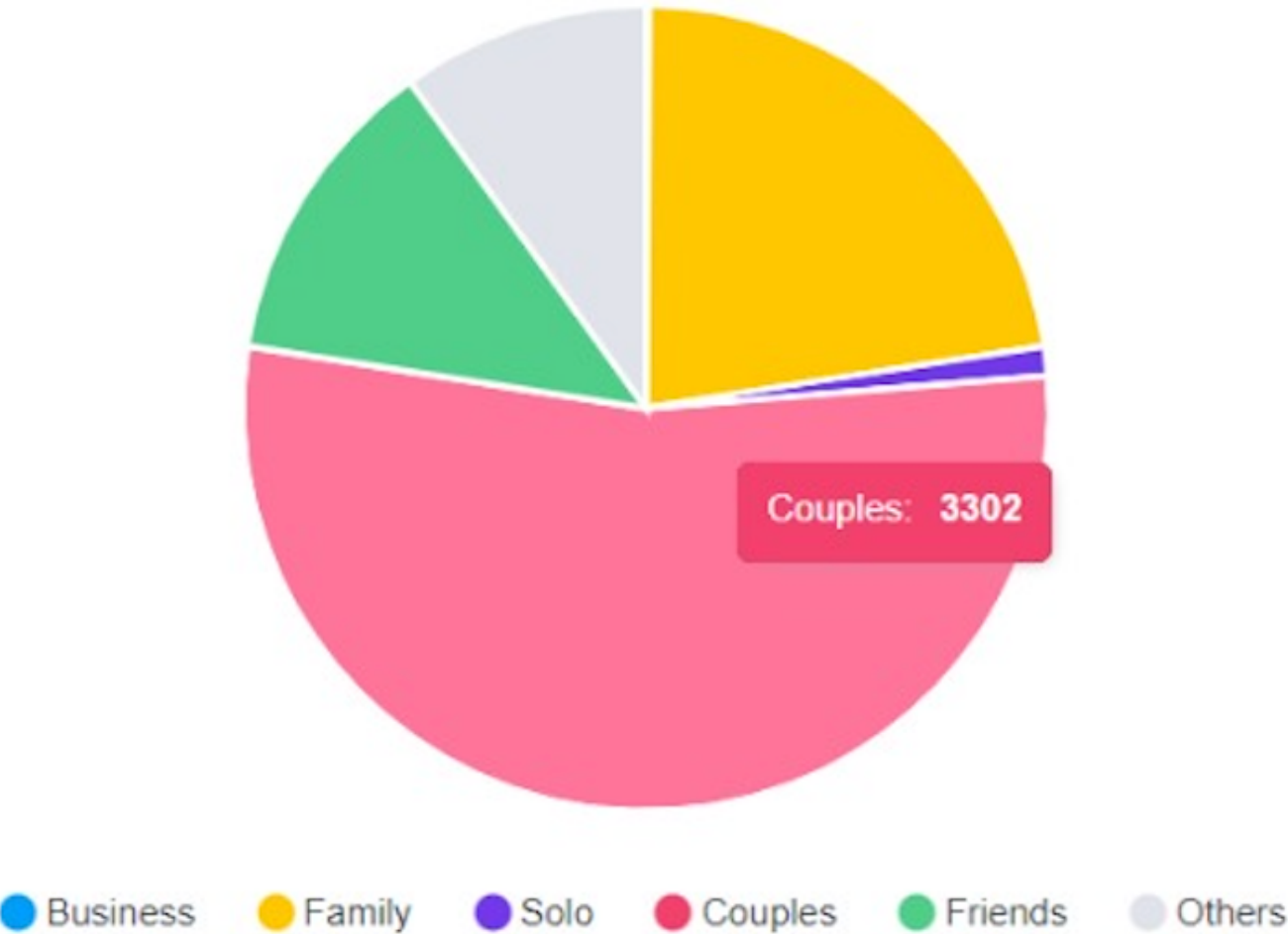


Overall Sentiment Distribution



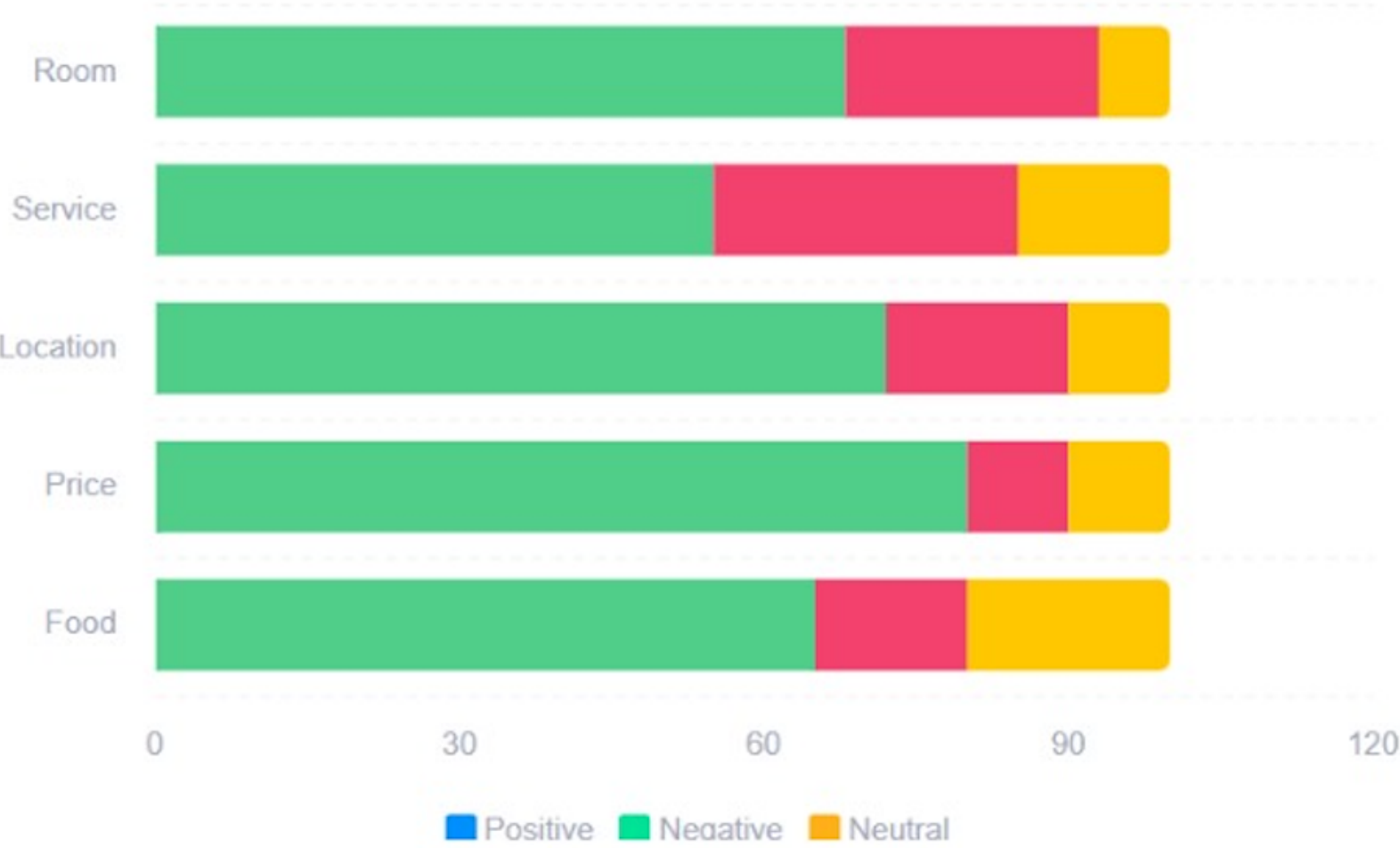
Results

Reviewer Trip Type



Sentiment per Topic

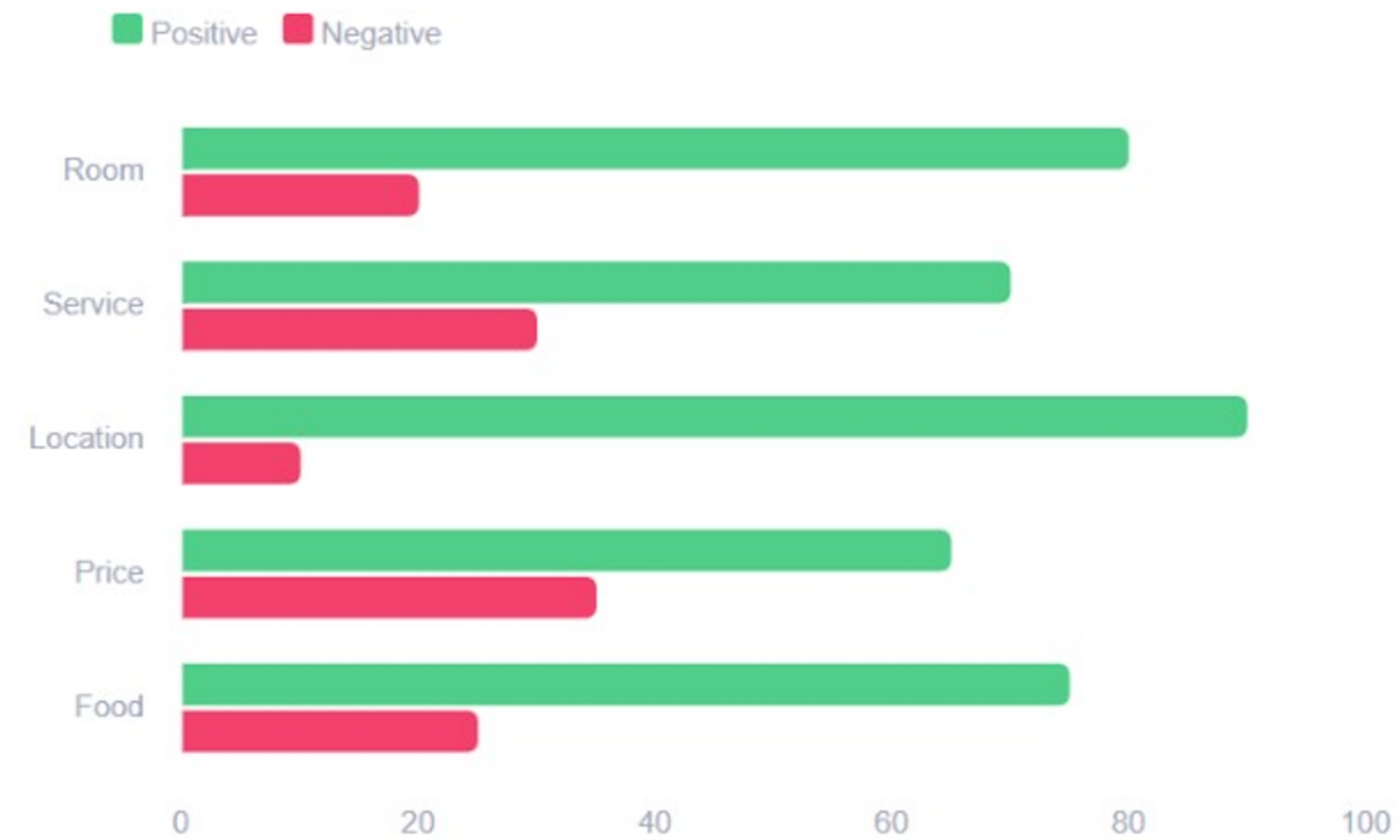
Based on Hotel Reviews



Results

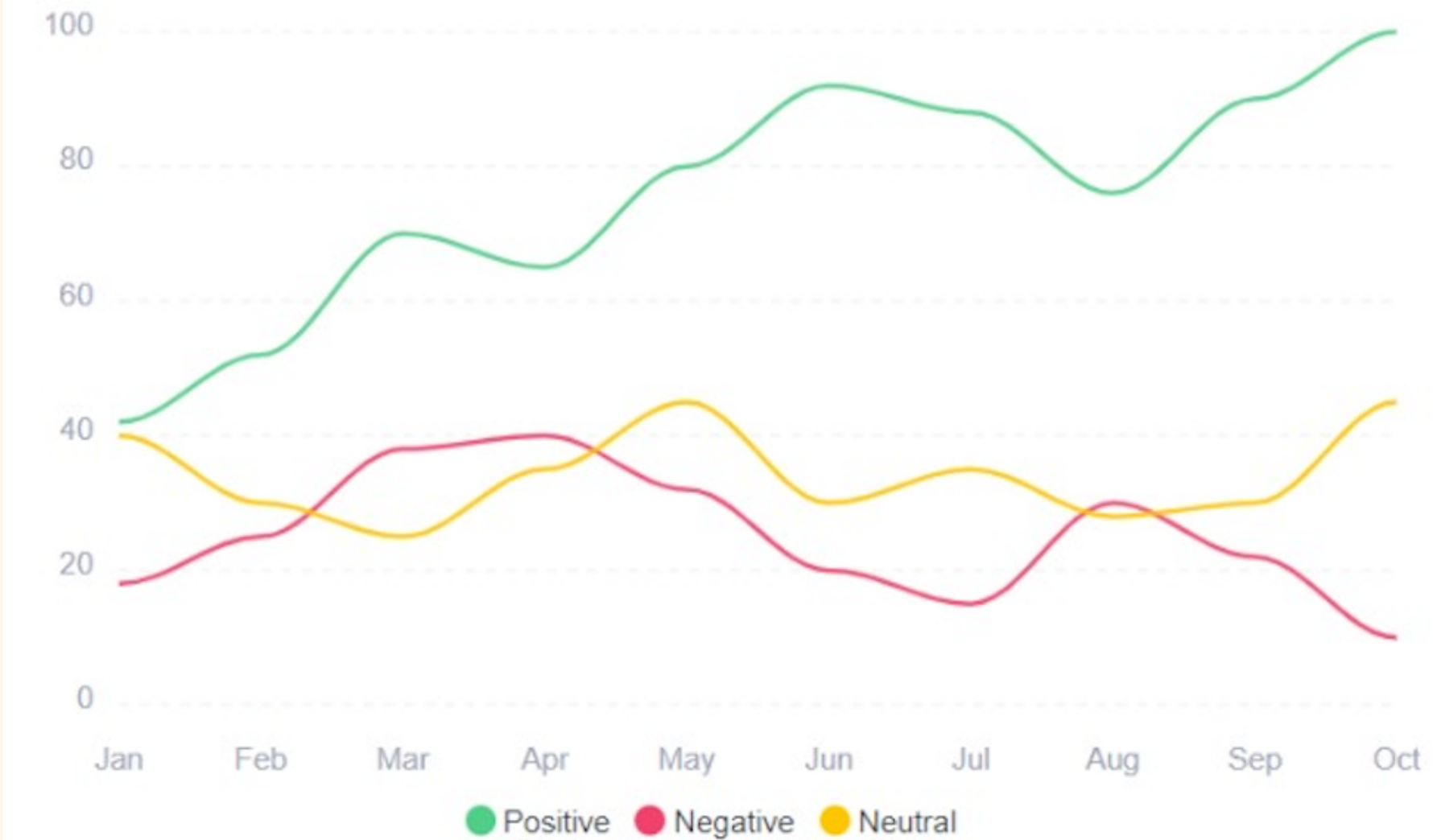
Top Aspect-Based Sentiments

Strengths and Weaknesses of the Hotels

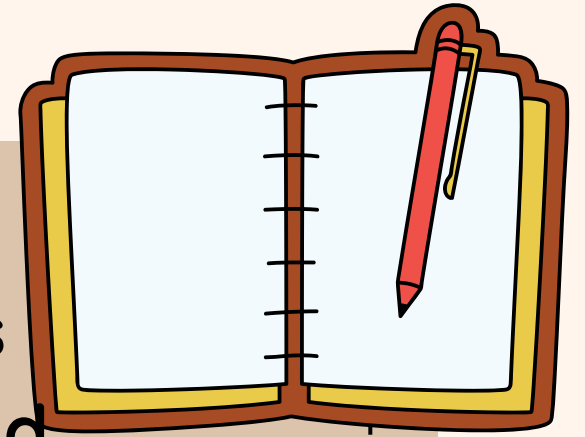


Sentiment Over Time

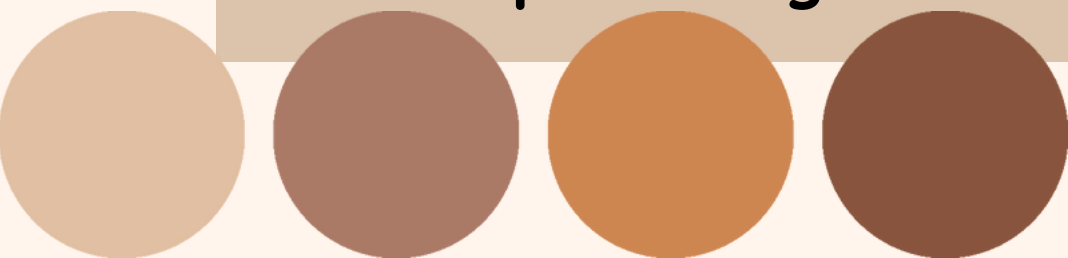
Based on Hotel Reviews



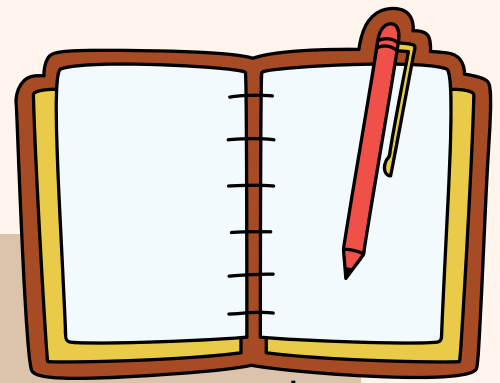
Comparitive studies



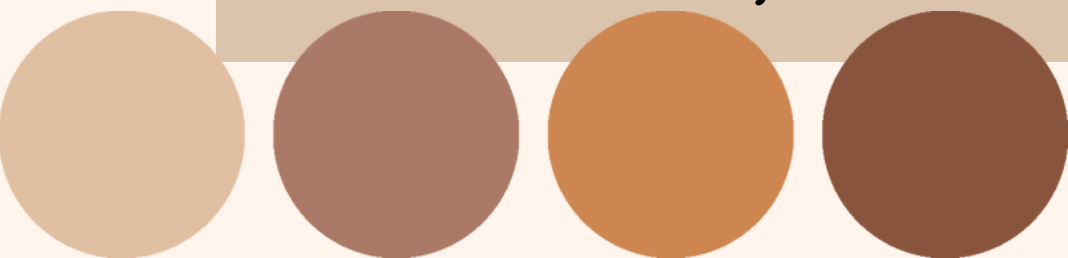
- Methodologies: Our project uses advanced machine learning techniques, such as fine-tuning a pre-trained word-embedded model, for sentiment classification and aspect identification. This approach can be compared with other projects that use different methodologies. For instance, a study conducted at Stanford University explored various supervised machine learning methods for sentiment analysis on TripAdvisor hotel reviews. Another study employed state-of-art approaches to perform three tasks on the SemEval dataset, including predicting the aspect of the restaurant's reviews and their polarity.
- SemEval-2015 Task 12: This study employed state-of-art approaches to perform three tasks on the SemEval dataset, including predicting the aspect of restaurant reviews and their polarity. The experimental results showed that the word2vec features when used with the support vector machine algorithm outperformed by giving 76%, 72%, and 79% of overall accuracies for Task A, Task B, and Task C, respectively.



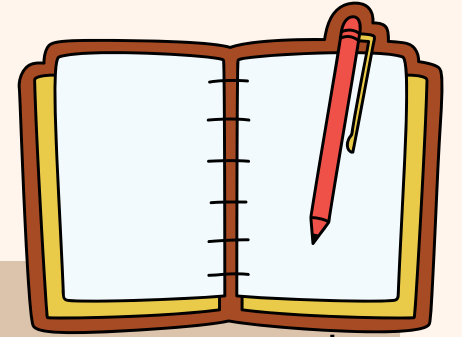
Comparitive studies



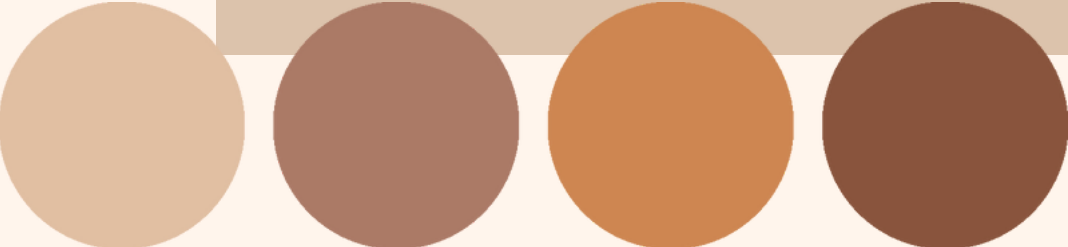
- **Aspect-based Sentiment Oriented Summarization of Hotel Reviews:** This study analyzed hotel reviews and provided information that ratings might overlook. The study also analyzed the sentiment scores of the hotels based on their aspects, which gave a better understanding of which aspects of the hotels were more positively or negatively reviewed.
- **An Enhanced Framework for Aspect-based Sentiment Analysis of Hotels' Reviews: Arabic Reviews Case Study:** This research proposed a framework for aspect-based sentiment analysis (ABSA) of Hotels' reviews. The proposed framework consisted of a reference human-annotated dataset, a set of preprocessing techniques, and a set of machine learning algorithms.
- **A Hybrid Approach for Aspect-based Sentiment Analysis: A Case Study of Hotel Reviews:** This study presented a method of aspect-based sentiment analysis for customer reviews related to hotels. The considered hotel aspects were food, service, location, comfort, and cleanliness



Conclusion



Our project on aspect-based sentiment analysis using the "515K Hotel Reviews Data in Europe" dataset has demonstrated significant potential in the field of sentiment analysis. Our project's methodologies, performance, and applications have been compared with similar projects, providing a comprehensive understanding of its strengths and areas for improvement. Performance-wise, our project's precision, recall, and F1-score are crucial metrics for evaluation. In terms of methodologies, the project has successfully employed advanced techniques for aspect extraction and sentiment classification. This approach has been found to be effective in other studies as well, such as the SemEval-2015 Task 12, which used word2vec features with the support vector machine algorithm, and the Aspect-based Sentiment Analysis on Hotel Reviews project, which explored various supervised machine learning methods. By comparing the project with similar studies, it is clear that the project can benefit from incorporating advanced methodologies, striving for high performance metrics, and exploring diverse applications. This comparative study provides a roadmap for future research and development in the field of aspect-based sentiment analysis.



Roles

SAKSHI

1. Data collection and pre-processing
2. Data engineering
3. Data visualisation and reporting
4. Model architecture development
5. Review-3 documentation


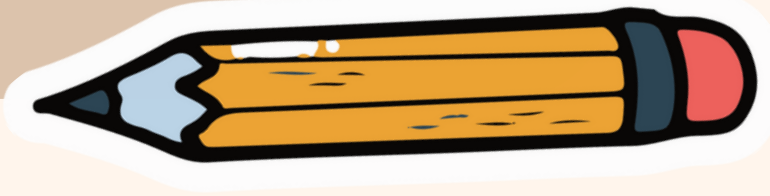
ADITYA

1. Preparation of review-1 documentation
2. Building the sentiment analysis model
3. Term extraction and aspect identification
4. Reporting and visualisation

SUDEEKSHA

1. Data preparation
2. Aspect identification
3. Reporting and analysis
4. Inputs in review-2 documentation

References

- 
1. "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey" - IEEE Computer Society- A. Nazir, Y. Rao, L. Wu and L. Sun, "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey" in IEEE Transactions on Affective Computing, vol. 13, no. 02, pp.845-863, 2022. doi: 10.1109/TAFFC.2020.2970399
 2. "4 Challenges in Aspect-Based Sentiment Analysis (ABSA)" - Wonderflow- <https://www.wonderflow.ai/blog/challenges-in-aspect-based-sentiment-analysis-absa/>
 3. <https://doi.org/10.48550/arXiv.2203.01054>
 4. Noh, Y.; Park, S.; Park, S.-B. Aspect-Based Sentiment Analysis Using Aspect Map. Appl. Sci. 2019, 9, 3239. <https://doi.org/10.3390/app9163239>
 5. <https://monkeylearn.com/blog/aspect-based-sentiment-analysis/>
 6. <https://dl.acm.org/doi/abs/10.1145/3575882.3575935>
 7. <https://dl.acm.org/doi/10.1145/3503044#sec-cit>
- 

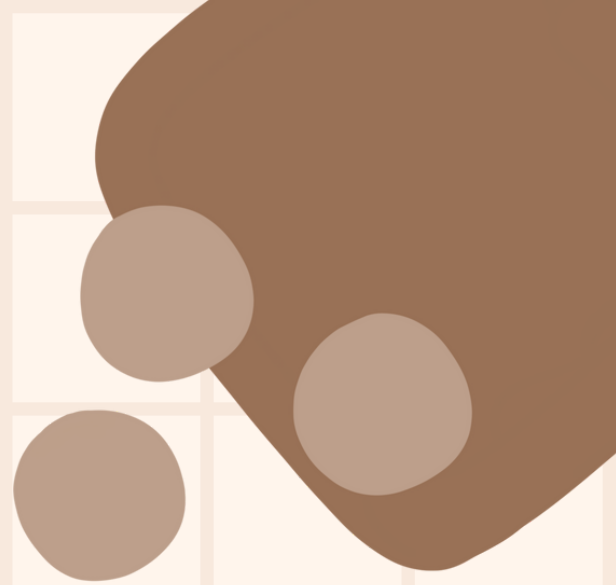
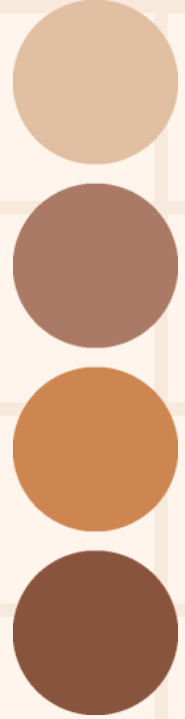


DRIVE LINK TO THE PROJECT

<https://drive.google.com/drive/folders/1dYPlsLu5rkBJkDd0dHyPp8hiC0fRuRi1>

Please copy the google drive link to get access to the code file :)





THANK YOU
SO MUCH!

