ML MINI PROJECT

Aspect Based Sentiment Analysis for Hotel Reviews

Authors: Nikhilesh Venkat PES1UG23AM192, Prem M Thakur PES1UG23AM214,

Rohan Suresh PES1UG23AM240

1. Introduction

This project focuses on analyzing hotel reviews to determine sentiments associated with specific aspects such as room quality, service, food, and cleanliness. Aspect-Based Sentiment Analysis (ABSA) provides fine-grained insights by identifying opinions tied to distinct attributes rather than general sentiments.

2. Objective

The primary objective is to perform aspect-level sentiment classification for hotel reviews. This involves extracting aspects from text and determining whether the sentiment expressed toward each aspect is positive, negative, or neutral.

3. Dataset & Preprocessing

Dataset: Kaggle
Total Samples: 20,000
Features: Review text, aspect labels, sentiment labels.
Preprocessing: Text cleaning, tokenization, stopword removal, lemmatization, and aspect annotation.

4. Methodology / Model

The approach consists of two main stages:
1. Aspect Extraction: Identify specific aspects (e.g., room, service, food) using dependency parsing or attention-based models.
2. Sentiment Classification: Use machine learning or transformer-based models (SVM) to determine sentiment polarity for each aspect.

Libraries Used: Python, Pandas, NumPy, Scikit-learn, Transformers, NLTK.

5. Experiments & Evaluation

Baseline Models: Logistic Regression, Naive Bayes, LSTM
Metrics: Accuracy, Precision, Recall, F1-Score.
Data Split: 80% training, 20% testing.

6. Results

Train Accuracy: 100%
Test Accuracy: 85.90%
F1-Score: 83.99%
Observations: The transformer-based model achieved superior accuracy compared to traditional ML baselines.

7. Error Analysis & Discussion

Common misclassifications occurred in reviews expressing mixed sentiments about multiple aspects. Further fine-tuning on aspect-specific subdatasets could improve precision.

8. Conclusion & Future Work

The project demonstrates that Aspect-Based Sentiment Analysis provides deeper insights into customer opinions compared to overall sentiment classification. Future work includes expanding the dataset, integrating multilingual models, and developing a web-based dashboard for visualization.

9. References

1. Pontiki et al., 'SemEval-2014 Task 4: Aspect Based Sentiment Analysis'
2. Devlin et al., 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'