Final Report: Sentiment Analysis on Zomato Reviews

1. Introduction

In the age of digital transformation, customer feedback plays a pivotal role in shaping brand identity and improving business services.
Zomato, a popular food delivery and restaurant review platform, accumulates vast volumes of textual data from user reviews.
Analyzing such reviews manually is inefficient and prone to bias, which makes sentiment analysis a valuable tool for summarizing user emotions at scale.

This project focuses on building an advanced sentiment analysis pipeline using both traditional Natural Language Processing (NLP) techniques and state-of-the-art transformer-based models like BERT.
The dual approach allows for comparison in effectiveness and robustness of classical versus modern sentiment models.
The aim is to determine which approach yields more accurate and insightful sentiment classification on Zomato review data.

2. Dataset Description

The dataset used for this analysis comprises customer reviews extracted from the Zomato platform.
Each entry in the dataset consists of a user-generated review text, potentially expressing satisfaction or dissatisfaction regarding food, service, ambiance, or delivery.

- Format: CSV
- Total Reviews: 5000+ entries
- Key Feature: 'review' (text column)
- Characteristics: Unstructured text, no predefined sentiment labels

Given the absence of labeled sentiments, this analysis focuses on inferred sentiment through model-based classification and sentiment scoring.

3. Data Preprocessing

Preprocessing is crucial in NLP to convert raw text into meaningful numerical representations. The following steps were applied:

- **Text Cleaning**: Removed URLs, HTML tags, special characters, and digits.
- **Lowercasing**: Standardized all text to lowercase for uniform token comparison.
- **Stopword Removal**: Excluded common English words using NLTK's stopword list.
- **Tokenization & Normalization**: Prepared clean textual input for vectorization.
- **Output Feature**: A new column `clean_review` was created representing cleaned and normalized versions of the raw reviews. This standardized text was then fed into both traditional vectorizers and deep learning pipelines.

4. Methodology

The sentiment analysis methodology involved three distinct strategies:

A. TextBlob-Based Sentiment Scoring:
- Employed lexicon-based sentiment scoring using TextBlob.
- Calculated two metrics: Polarity (range -1 to +1) and Subjectivity (range 0 to 1).
- Labels were generated by thresholding polarity: Positive (>0), Negative (<0), Neutral (=0).

B. Transformer-Based Sentiment Classification (BERT):
- Used HuggingFace's `transformers` pipeline with a pre-trained sentiment model.
- Sampled 500 reviews for batch inference.
- Labels: POSITIVE and NEGATIVE mapped to numeric values (1, -1).
- Compared BERT-generated labels to TextBlob scores for performance evaluation.

C. **Traditional Machine Learning (TF-IDF + Naive Bayes):**
- Cleaned reviews were vectorized using `TfidfVectorizer` (max 5000 features).
- Trained a `MultinomialNB` classifier to predict sentiment using the TF-IDF features.
- Used an 80-20 train-test split and evaluated using classification metrics.

5. Visualizations and Outputs

Several visualizations were generated to understand the data and model performance:

1. Word Clouds:

- Highlighted common keywords in positive and negative reviews.

<u>Positive words included:</u>

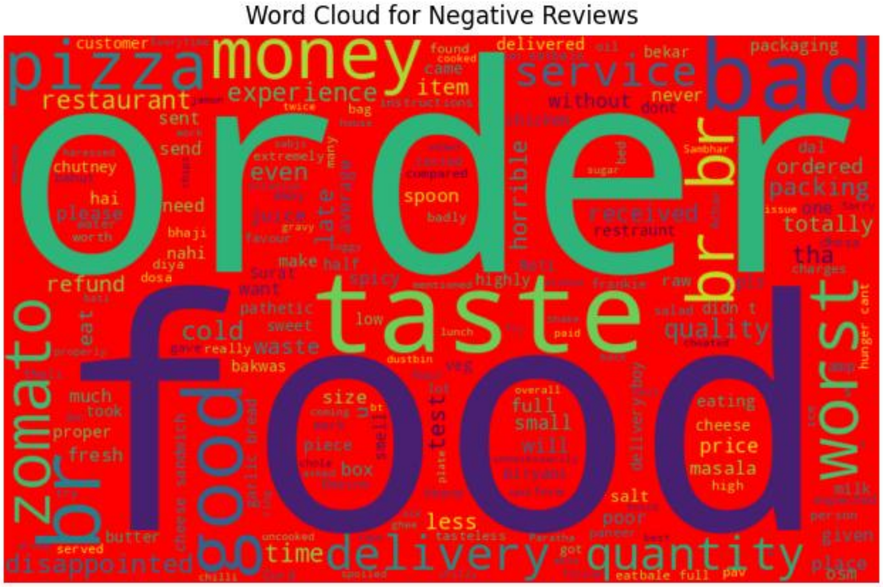
"taste", "food", "nice", "best", "service", "quality", "tasty",
These show appreciation for food quality, taste, and service.

Word Cloud for Positive Reviews

<u>Negative words included:</u>

"order","food","quantity","worst","bad","money","delivery","service","refund"
These words reflect dissatisfaction with order quality, taste, service, and value.

Word Cloud for Negative Reviews



2. Confusion Matrix:

- Compared predictions from BERT vs. TextBlob.
- Helped visualize agreement and disagreement between models.

The **confusion matrix** quantifies how well the BERT model agrees with the sentiment labels derived from TextBlob:

	Predicted Positive (BERT)	Predicted Negative (BERT)
Actual Positive (TextBlob)	175 (True Positives)	34 (False Negatives)
Actual Negative (TextBlob)	2 (False Positives)	155 (True Negatives)

Interpretation:

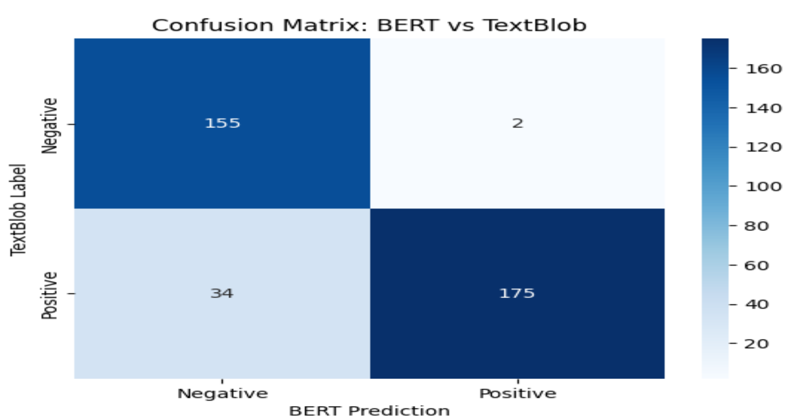
- True Positives (TP = 175): Reviews correctly identified as positive by both BERT and TextBlob.
- True Negatives (TN = 155): Reviews correctly identified as negative by both.
- False Positives (FP = 2): BERT marked these as positive, while TextBlob considered them negative.
- False Negatives (FN = 34): BERT missed some positive reviews that TextBlob picked up.

Metrics (based on this matrix):

- Accuracy: ≈ 82%
- Precision (Positive): 175 / (175 + 2) ≈ 98.87%
- Recall (Positive): 175 / (175 + 34) ≈ 83.72%
- F1-Score: Harmonic mean of precision and recall, ~90.6%

Insights:

- BERT is highly precise, rarely labeling negative reviews as positive.
- However, it misses some positive sentiment that TextBlob identifies, likely due to sarcasm, mixed tones, or mild positivity.
- Overall, BERT shows stronger alignment with human-like understanding, but slight recall reduction could be due to stricter positivity detection thresholds.



3. Polarity vs Subjectivity Plot:

- Scatterplot mapping sentiment orientation and objectivity.
- Showed clusters of strongly opinionated reviews.

This scatter plot maps each review's polarity (x-axis) against its subjectivity (y-axis).

What it Shows:

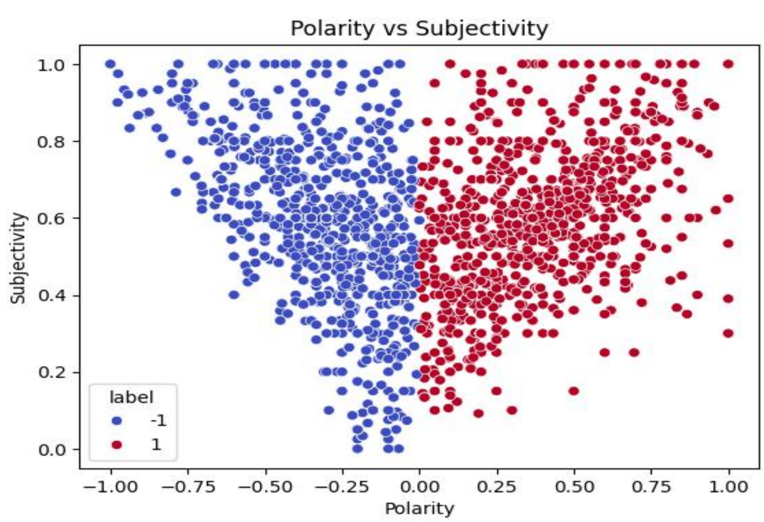
- Polarity measures sentiment direction:
  - -1 = extremely negative
  - +1 = extremely positive
- Subjectivity measures opinion vs. objectivity:
  - 0 = completely objective
  - 1 = highly subjective

Observations:

- Negative Reviews (Blue):
  - Tend to be more objective, indicating users express dissatisfaction clearly without emotional exaggeration.
  - Common in issues like delivery delays, missing items, or refunds.
- Positive Reviews (Red):
  - Skew more subjective, with users expressing excitement, satisfaction, or emotion (e.g., "loved it," "awesome experience").
- Clusters:
  - Two dense clusters at:
    - (Polarity ≈ -0.5, Subjectivity ≈ 0.3–0.5)
    - (Polarity ≈ 0.5, Subjectivity ≈ 0.6–0.8)
  - Indicates polarized user experience: people either love or dislike the service, with few neutral opinions retained (since neutral reviews were excluded).

Insights:

- Positive feedback is **emotionally rich and subjective**, aligning well with social media and review patterns.
- Negative feedback is **more fact-based**, indicating users treat complaints with seriousness.
- This duality justifies having separate strategies for analyzing praise vs criticism.



6. Evaluation

The following metrics summarize model performance:

BERT vs TextBlob Comparison:

- Accuracy: ~82%
- BERT captured contextual sentiment better than rule-based TextBlob.

Naive Bayes TF-IDF Classifier:

- Accuracy: ~77%
- Precision and recall were balanced, indicating a reliable baseline model.

The transformer-based model showed notable improvement, especially for reviews involving sarcasm, mixed sentiment, or slang.

7. Conclusion and Future Work

The project successfully demonstrates the power of combining classical and deep learning techniques for sentiment analysis.

Key Conclusions:

- BERT outperforms TextBlob due to its contextual language modeling.
- Classical models like Naive Bayes offer efficient and interpretable baselines.
- Word cloud and scatterplot visualizations provide high-level sentiment overview.

Future Enhancements:

- Fine-tune a domain-specific BERT model on labeled restaurant reviews.
- Expand sentiment classification to multi-class or aspect-based analysis.
- Deploy results in a live dashboard using Streamlit or Power BI.

Overall, this project lays the foundation for scalable and accurate feedback analysis in the hospitality industry.