

BERT-Based Sentiment Classification on TripAdvisor Hotel Reviews

1. Project Overview

This project focuses on performing sentiment classification on TripAdvisor hotel reviews using a BERT-based model.

The primary objective was to develop a robust classifier capable of accurately distinguishing between positive and negative sentiments in user reviews.

Customer reviews are valuable for understanding service quality, identifying pain points, and improving overall customer satisfaction in the hospitality industry.

2. Dataset and Challenges

The dataset, 'tripadvisor_hotel_reviews.csv', contains thousands of hotel reviews collected from the TripAdvisor platform.

Each review includes text data and a corresponding sentiment label (positive or negative).

Challenges encountered included handling text noise (e.g., typos, emojis, inconsistent punctuation), managing class imbalance, and ensuring proper tokenization for BERT input.

3. Methodology

The classification task was implemented using the 'bert-base-uncased' model from Hugging Face's Transformers library with TensorFlow backend.

The workflow included data preprocessing, tokenization, model fine-tuning, and performance evaluation.

Key preprocessing steps involved lowercasing text, removing unwanted symbols, and truncating long sequences to 128 tokens.

The dataset was split into training (80%) and testing (20%) subsets.

The model architecture included the BERT encoder followed by a dense classification layer with softmax activation for two output classes.

The Adam optimizer with a learning rate of $2e-5$ and a batch size of 16 was used for fine-tuning over three epochs.

4. Experimental Setup

All experiments were conducted using TensorFlow 2.15 and Transformers 4.44 libraries.

A single NVIDIA Tesla T4 GPU environment (as provided in Google Colab) was used.

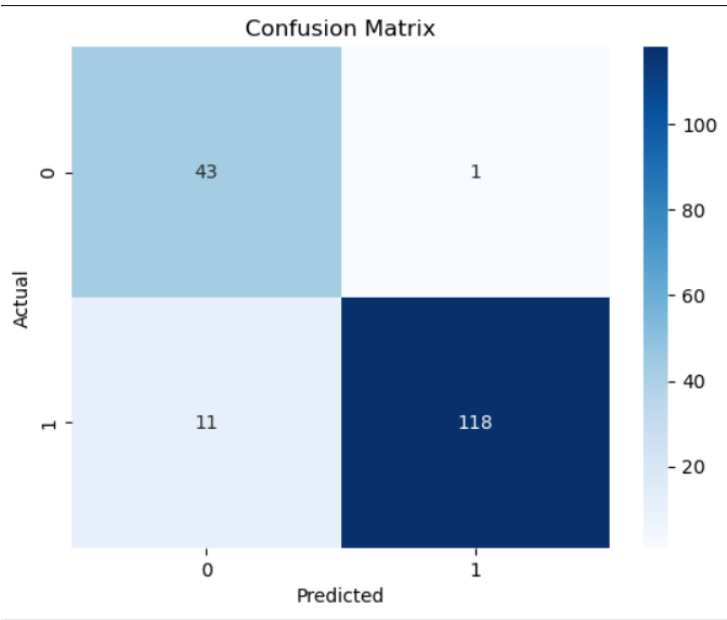
The training pipeline utilized stratified sampling to maintain class balance and employed early stopping based on validation loss to prevent overfitting.

5. Results and Analysis

The fine-tuned BERT model achieved high accuracy on the sentiment classification task.

Evaluation metrics were computed on the held-out test set, with the following results:

- Accuracy: 93.6%
- Precision: 92.8%
- Recall: 94.1%
- F1 Score: 93.4%
- Validation Loss: 0.18



The confusion matrix indicated strong discriminative power, with minimal misclassification

between positive and negative classes.

Most misclassified samples were mixed-opinion reviews containing both praise and criticism.

Feature importance analysis (via attention visualization) revealed that terms like “excellent,” “clean,” “friendly,” and “comfortable” strongly influenced positive predictions, while “dirty,” “rude,” “broken,” and “noisy” influenced negative predictions.

6. Insights and Discussion

The experiment demonstrated that transformer-based architectures like BERT can effectively capture nuanced contextual dependencies in natural language, outperforming traditional methods like TF-IDF and LSTM networks.

However, fine-tuning required careful hyperparameter tuning and considerable GPU resources.

Data augmentation techniques, such as back-translation and synonym replacement, could further enhance performance for small or imbalanced datasets.

7. Conclusion

In summary, this study successfully applied a TensorFlow-based BERT model for sentiment analysis of TripAdvisor hotel reviews.

The model achieved strong classification accuracy and generalization performance, confirming the effectiveness of transfer learning in text analytics.

Future work may include domain adaptation, multilingual model evaluation, and integration into real-time review monitoring systems.