

Yelp Review Sentiment Analysis using DistilRoBERTa

Anas Khalil
Mourad Mahmoudi

February 9, 2026

Abstract

This project focuses on multi-class sentiment analysis of Yelp restaurant reviews using a transformer-based deep learning model. The objective is to predict user sentiment across five rating levels while maintaining an efficient training time. We fine-tune a DistilRoBERTa model and evaluate its performance on the Yelp Review Full dataset.

1 Introduction

Sentiment analysis is a key Natural Language Processing (NLP) task widely used in opinion mining, recommendation systems, and customer feedback analysis. In this project, we address a five-class sentiment classification problem using Yelp reviews, where each review is associated with a star rating from 1 to 5.

Our objective is to build an efficient yet accurate model capable of capturing semantic nuances in user reviews while maintaining reasonable computational costs.

2 Dataset

We use the **Yelp Review Full** dataset available on Hugging Face. It contains restaurant reviews labeled from 1 to 5 stars.

- Training set: used for training and validation
- Test set: strictly reserved for final evaluation

3 Data Preprocessing

The following preprocessing steps were applied:

- Removal of URLs and special characters
- Normalization of whitespace and line breaks
- Truncation to a maximum sequence length of 256 tokens

These steps help reduce noise while preserving the semantic structure of the reviews.

4 Model Architecture

We fine-tuned a **DistilRoBERTa-base** model for sequence classification with five output classes. DistilRoBERTa was chosen as a compromise between:

- High performance

- Reduced model size
- Faster training compared to larger transformer models

5 Training Configuration

The model was trained under the following settings:

- GPU: NVIDIA V100
- Batch size: 16
- Learning rate: 2×10^{-5}
- Epochs: 1
- Mixed precision training (FP16)

6 Project Structure

The project follows a modular and reproducible structure as shown below:

```
yelp_sa/
docs/
figures/
models/
  distilroberta_yelp_run1/
    config/
    tokenizer/
    tokenizer_config/
    model.safetensors
    training_args.bin
notebooks/
outputs/
  run1/
    checkpoint-500/
    checkpoint-1000/
```

7 Training Time

The total training time was measured as follows:

- Training time: 592 seconds
- Training time: 9.87 minutes

This demonstrates the efficiency of the DistilRoBERTa architecture for large-scale NLP tasks.

8 Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, and F1-score.

Class	Precision	Recall	F1-score
0	0.784	0.762	0.773
1	0.596	0.634	0.614
2	0.629	0.562	0.594
3	0.584	0.614	0.598
4	0.742	0.757	0.749
Weighted Avg	0.667	0.666	0.666

Table 1: Classification performance on the test set

Overall accuracy achieved was **66.59%**.

9 Discussion

Most classification errors occur between neighboring sentiment classes (e.g., 3 vs 4 stars), which is expected due to subjective interpretation of ratings. Despite training for only one epoch, the model achieves strong performance, highlighting the effectiveness of transformer-based representations.

10 Conclusion and Future Work

This project demonstrates that DistilRoBERTa is an effective and efficient solution for multi-class sentiment analysis. Future improvements could include:

- Training for additional epochs
- Using larger transformer models
- Applying class-weighting or focal loss
- Data augmentation techniques