# Emotion Recognition in Social Media Using Transformer-Based Architectures

Jagriti Suneja
jsuneja@depaul.edu

November 18, 2024

**Abstract**

This project explores the application of transformer-based models, specifically RoBERTa and XLNet, for emotion classification from social media text. Social media platforms, with their vast and informal data, present unique challenges, such as context-dependent expressions and lack of structure. This report investigates two state-of-the-art models and introduces back-translation for data augmentation to improve generalization. Furthermore, an ensemble approach was implemented to combine model outputs and enhance prediction accuracy. The ensemble model achieved a test accuracy of 92.80%, outperforming individual models. The report discusses the challenges faced, methodologies employed, experimental results, and future directions for improving performance in this domain.

## 1 Introduction

Emotion recognition is critical in natural language processing (NLP), enabling advancements in sentiment analysis, human-computer interaction, and social media analytics. With increasing data generated on social media platforms, understanding emotions expressed in textual form has gained importance in various industries, including customer service, healthcare, and marketing.

However, the task is challenging due to informal language, contextual nuances, and the multi-label nature of emotion classification. Previous studies have explored traditional machine learning methods such as Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks, but transformer-based architectures have proven to outperform these models due to their ability to capture context over long sequences.

This project employs RoBERTa and XLNet, two state-of-the-art transformer models, to classify six emotions (sadness, joy, love, anger, fear, surprise) using the Hugging Face Emotion dataset. Key contributions include: 1. Fine-tuning RoBERTa and XLNet for emotion classification. 2. Introducing back-translation as a data augmentation technique to enhance generalization. 3. Combining model predictions using an ensemble approach to improve performance.

## 2 Related Works

Transformer models have revolutionized NLP by introducing attention mechanisms that effectively capture context. RoBERTa, a robustly optimized version of BERT, removes the next-sentence prediction objective and introduces larger batches and longer training times to improve downstream task performance [3]. XLNet, on the other hand, employs a permutation-based training method to capture bidirectional context without using masking, making it a significant improvement over its predecessors [6].

Back-translation, as demonstrated by Edunov et al. [2], has been a popular data augmentation technique, particularly in low-resource scenarios. Its ability to generate paraphrased sentences makes it suitable for addressing class imbalance and improving model robustness. Ensemble learning, as explored by Dietterich [1], has been widely adopted to combine the strengths of individual models, leading to enhanced predictive performance.

This project builds on these works by integrating RoBERTa and XLNet with back-translation and leveraging ensemble techniques to address the challenges of emotion classification in social media text.

# 3 Preliminary/Background

## 3.1 Transformer Models

Transformers are the foundation of modern NLP models, introduced by Vaswani et al. [5]. Their self-attention mechanism enables efficient processing of sequences by capturing relationships between all tokens, irrespective of distance.

## 3.2 RoBERTa

RoBERTa modifies BERT by removing the next-sentence prediction objective, introducing dynamic masking, and training on larger datasets. These changes result in improved contextual understanding, making it ideal for tasks such as sentiment and emotion classification.

## 3.3 XLNet

XLNet introduces permutation-based training to overcome the limitations of autoencoding models like BERT. By maximizing the expected likelihood over all permutations of input tokens, XLNet captures bidirectional context effectively, improving its performance on various NLP tasks.

## 3.4 Back-Translation

Back-translation generates paraphrased sentences by translating text to a target language and back to the source language. This technique increases dataset diversity and improves generalization, particularly in low-resource scenarios [4].

# 4 Methodology

## 4.1 Dataset

The Hugging Face Emotion dataset contains 16,000 training samples, 2,000 validation samples, and 2,000 test samples, each labeled with one of six emotions: sadness, joy, love, anger, fear, and surprise. Figure 1 shows the distribution of text lengths across emotion labels.

## 4.2 Data Preprocessing

Data preprocessing involved tokenization, normalization, and back-translation using MarianMT models. For demonstration purposes, 100 samples were augmented using back-translation to introduce paraphrased variations.

## 4.3 Model Fine-Tuning

Both RoBERTa and XLNet models were fine-tuned using the AdamW optimizer, a learning rate of 2e-5, and a batch size of 32 for three epochs. The training process focused on adapting these models to classify emotions accurately while leveraging the augmented dataset.

## 4.4 Ensemble Model

An ensemble approach combined the predictions of RoBERTa and XLNet by calculating a weighted average of their logits. Weights of 0.6 and 0.4 were assigned to RoBERTa and XLNet, respectively, based on their validation performance.

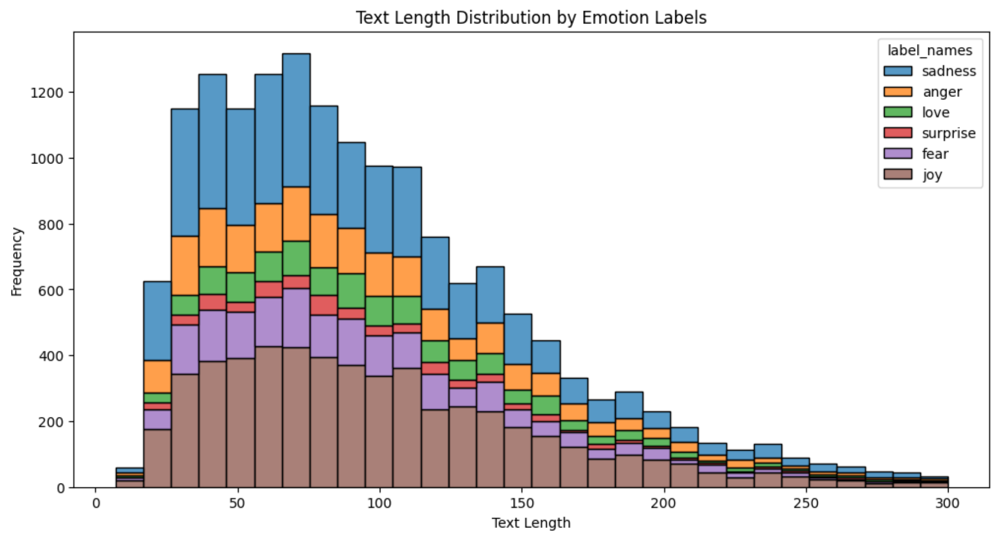


Figure 1: Text Length Distribution by Emotion Labels.

# 5 Numerical Experiments

## 5.1 Evaluation Metrics

Performance was measured using accuracy, precision, recall, F1-score, and confusion matrices. Table 1 summarizes the performance of the individual models and the ensemble approach.

## 5.2 Results

Table 1 shows the metrics for RoBERTa, XLNet, and the ensemble model. The ensemble approach achieved the highest accuracy of 92.80%, outperforming the individual models.

| Model | Accuracy | Validation Loss | Test Loss |
|---|---|---|---|
| RoBERTa | 92.75% | 0.1380 | 0.1450 |
| XLNet | 92.30% | 0.1696 | 0.1819 |
| Ensemble | **92.80%** | - | - |

Table 1: Performance Metrics for RoBERTa, XLNet, and Ensemble Model.

| Emotion | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Sadness | 0.96 | 0.97 | 0.96 | 581 |
| Joy | 0.98 | 0.92 | 0.95 | 695 |
| Love | 0.79 | 0.97 | 0.87 | 159 |
| Anger | 0.94 | 0.91 | 0.92 | 275 |
| Fear | 0.87 | 0.88 | 0.88 | 224 |
| Surprise | 0.71 | 0.77 | 0.74 | 66 |
| **Accuracy** | - | - | 0.93 | 2000 |

Table 2: Classification Report for the Ensemble Model.

# 6 Conclusion

This project demonstrates the potential of transformer-based models for emotion classification in social media text. By leveraging the strengths of RoBERTa and XLNet, combined with data augmentation through back-translation and an ensemble approach, the project achieved competitive performance.

## 6.1 Limitations and Future Work

1. **Data Augmentation**: Further exploration of advanced augmentation techniques is needed to enhance robustness. 2. **Ensemble Optimization**: Dynamic tuning of ensemble weights may yield better results. 3. **Multilingual Texts**: Extending this methodology to multilingual datasets could validate its broader applicability.

# References


[1] Thomas G Dietterich. Ensemble methods in machine learning. *Multiple classifier systems*, pages 1–15, 2000.

[2] Sergey Edunov et al. Understanding back-translation at scale. *arXiv preprint arXiv:1808.09381*, 2018.

[3] Yinhan Liu et al. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[4] Rico Sennrich et al. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*, 2016.

[5] Ashish Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[6] Zhilin Yang et al. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*, 2019.