

NLP for Data Science Group 2 Final Project

PROJECT PROPOSAL

Luke Wu, Xiao Qi

Emotions are an integral part of human life, and the ability to identify and express them effectively is crucial for maintaining mental health and well-being. In recent years, the growing availability of social media platforms has resulted in an explosion of data containing textual expressions of emotions. Multiclass emotion classification is the task of automatically identifying emotions expressed in text and assigning them to predefined categories.

In this project, we are trying to explore the effectiveness of machine learning models for multiclass emotion classification on a textual dataset. The dataset comes from Kaggle (<https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp>), which contains labeled text samples belonging to six distinct emotions (sadness, fear, angry, joy, love, surprise), and the goal is to develop a model that accurately classifies these emotions.

The project involves various stages such as data preprocessing, model selection, hyperparameter tuning and model evaluation. We will build the LSTM model and the fine-tuned transform model at the same time and compare the results of different models. After that, we will do some model interpretation work, in order to understand the models better.

The results of this project can have practical applications in various fields, including mental health, marketing, and social media analysis.