Emotion Detection Using Machine Learning

Arnab Dana

M.Tech, CSE
NIT Uttarakhand

March 27, 2025

Introduction

- **Emotion Detection**: Classifies user comments into emotions.
- **Dataset**: Kaggle dataset with two columns:
  - **Text**: Normal user comments.
  - **Label**: Values from 0 to 5 representing emotions (Sad, Happy, Angry, etc.).
- **Objective**: Build a model to classify text into emotional categories.

Data Preprocessing

Preprocessing Steps:

- Removing stop words (e.g., *is, the, and*).
- Lowercasing text.
- Removing special characters and punctuation.

Why Preprocessing?

- Improves model accuracy.
- Reduces noise in data.

Feature Extraction (Vectorization)

Techniques Used:

- **TF-IDF (Term Frequency - Inverse Document Frequency)**
  - Assigns importance to words based on their frequency.
- **Bag of Words (BoW)**
  - Converts text into a matrix of word occurrences.

Note: Any one technique can be used for model training.

Classification Models

Algorithms Used:

- **Logistic Regression**: Predicts the emotion score based on text features.
- **Random Forest**: Ensemble learning technique for better accuracy.

Model Performance:

- Achieves **90% accuracy** on the test dataset.

Conclusion & Future Scope

Key Takeaways:

- Emotion detection helps in analyzing user sentiments.
- Preprocessing and feature extraction improve classification results.
- The model achieves high accuracy with simple ML techniques.

Future Scope:

- Testing with deep learning models like LSTMs.
- Using a larger, real-world dataset for better generalization.