Emotion Analysis Report

Project Title: Emotion Analysis from Tweets
Author: Ayman Mostafa Negm
Date: April 2025

Table of Contents

---

1. Executive Summary

This project focuses on performing an emotion analysis on a dataset of tweets. Through data loading, cleaning, and exploratory analysis, we map the tweets to specific sentiments and visualize their distributions. Finally, a Machine Learning model is built to predict the emotion associated with unseen tweets.

2. Installation of Required Libraries

To ensure all dependencies are met, the following libraries were installed:

- pandas
- numpy
- seaborn
- matplotlib
- scikit-learn
- wordcloud

Installation commands were executed via pip to facilitate a smooth setup.

3. Importing Necessary Libraries

The key libraries were imported for:

- Data manipulation (pandas, numpy)
- Data visualization (matplotlib, seaborn, wordcloud)
- Machine Learning (scikit-learn)

4. Loading the Dataset

A CSV file containing tweets and their corresponding emotions was loaded into a pandas DataFrame. Initial exploration confirmed the dataset's structure and integrity.

5. Data Exploration

Exploratory steps included:

- Checking for null values
- Summarizing class distribution across emotions
- Viewing sample entries to understand content

6. Data Cleaning

Basic cleaning operations included:

- Removing missing values
- Stripping unwanted characters
- Ensuring uniform text formatting

This step ensured that only meaningful and properly formatted tweets were used for modeling.

7. Sentiment Mapping

Sentiment categories were mapped to numerical labels for Machine Learning purposes. This transformation allowed classification models to process and predict emotions effectively.

8. Visualization of Sentiment Distribution

Using bar plots and pie charts, the distribution of tweets across various sentiment classes was visualized. Word clouds were generated to highlight commonly used terms associated with each emotion.

9. Machine Learning Model

A Machine Learning pipeline was developed which included:

- Text vectorization using TF-IDF
- Model training with Logistic Regression
- Model evaluation via accuracy, precision, recall, and F1 score

Results demonstrated a promising ability to predict emotions from raw tweet text.

10. Conclusion

This project successfully demonstrated the workflow of emotion analysis on text data, from loading and cleaning to modeling and evaluation. Future improvements could include:

- Testing advanced NLP models (e.g., BERT)
- Expanding to multi-label classification
- Incorporating deeper text preprocessing (lemmatization, stopwords removal)

End of Report