Developing a Text Classifier for Emotion Detection in Social Media

# 1. Introduction

Emotion detection in text is an important area of Natural Language Processing (NLP) that aims to identify the emotional tone behind written language. As online communication becomes increasingly common, the ability to automatically recognise emotions from short texts has become useful in many fields, including customer support, education, healthcare, and digital wellbeing. This project explores the development of a text classification system to detect emotions in short, user-generated messages.

The chosen dataset consists of social media posts, mainly tweets, each labelled with one of six basic emotions. The project applies a classical machine learning approach, using standard tools such as TF-IDF for text representation and logistic regression for classification. All work is performed locally without GPU acceleration, demonstrating that effective emotion classification is achievable using accessible, resource-efficient methods.

This report presents the problem area, defines the objectives of the project, and introduces the dataset used. It then describes the steps taken to process the data, build a baseline model, and evaluate its performance. The final sections reflect on the outcomes and discuss the potential applications and future improvements of the system.

# 2. Domain-Specific Problem Area

In today’s digital world, people use short messages to express themselves on platforms like social media, messaging apps, and online forums. These messages often contain emotional content, whether it’s happiness, anger, sadness, or surprise. Being able to automatically detect these emotions from text can be very useful. It can help build more human-like chatbots, support mental health tools, improve customer service, and make user experiences more personalised.

However, detecting emotions in text is not easy. People express feelings in many different ways, and emotions are often subtle or mixed. Unlike simple tasks like positive or negative sentiment analysis, emotion detection involves several possible categories and can be more challenging.

This project focuses on building a text classification model that can recognise emotions in short messages, such as tweets. The aim is to explore how well machine learning methods can understand emotions from written language, and how this could be applied in real-world systems that need to respond to users in a more thoughtful and emotionally aware way.

# 3. Objectives of the Project

The main goal of this project is to create and test a machine learning model that can recognise emotions from short texts. The model will be trined on a labelled dataset and should be able to predict the correct emotion in new, unseen messages.

To achieve this, the project will:

* Explore basic text pre-processing methods to prepare informal messages, such as tweets, for analysis.
* Use TF-IDF to turn the text into a numerical format suitable for machine learning.
* Build a simple classification model, such as logistic regression, to act as a baseline.
* Evaluate the model using common metrics like accuracy, precision, recall, and F1-score.

This project will help show how machine learning can be used to understand emotional language, and discuss the strengths and limitations of this approach. The solution will also be designed so that it could be reused or adapted for similar tasks in other areas.

# 4. Dataset Description

The dataset used in this project is the *Emotion* dataset by dair.ai, which is available through Hugging Face’s Datasets platform. It contains around 20,000 short text samples, mostly tweets, each labelled with one of six emotions: anger, fear, joy, love, sadness, and surprise.

The dataset is split into:

* 16,000 training samples
* 2,000 validation samples
* 2,000 test samples

Each entry includes a short text and a label in numeric form, which maps to a specific emotion. The labels are fairly balanced, though there are some small differences in how many examples each emotion has.

This dataset was chosen because it is clean, well-labelled, and focused on real-world, informal language — which makes it both challenging and relevant. It is a good size for training models locally and provides a realistic starting point for building an emotion classification system.