Phase-3 Submission

Student Name: Sandhiya M

Register Number: 712523104053

Institution: PPG Institute of Technology

Department: BE.Computer Science and Engineering

Date of Submission: 15.05.2025

Github Repository Link:
https://github.com/Sandhiyamrs/NM_Sandhiya_DS

---

1. Problem Statement

In today’s digital age, social media platforms like Twitter, Facebook, and Reddit have become major outlets for individuals to express thoughts, opinions, and emotions. These conversations often carry rich emotional content that reflects public sentiment on various topics, from political events and global crises to mental health and consumer experiences. However, extracting meaningful emotional insights from this unstructured data poses challenges due to informal language, slang, sarcasm, abbreviations, and multilingual usage. Traditional sentiment analysis methods often struggle to capture these nuances. This project aims to decode and classify emotions in social media posts using natural language processing (NLP) and machine learning techniques. The goal is to develop a system capable of identifying emotions such as joy, anger, sadness, fear, surprise, or disgust, enabling applications in mental health monitoring, public opinion analysis, customer experience enhancement, and social research.

The Challenges:

- Informal Language:
  Use of slang, abbreviations, and casual tone makes emotion detection harder for traditional NLP models.
- Sarcasm & Irony:
  Literal meanings often contradict true emotions, confusing the classifier.
- Multilingual & Mixed Language Posts:
  Users blend languages in single posts, increasing processing complexity.
- Ambiguous Emotions:
  Posts often express multiple or subtle emotions, making classification less clear-cut.

2. Abstract

□ Social media contains rich emotional data, but analyzing it is challenging due to slang, sarcasm, and multilingual content.
□ The project aims to classify emotions (joy, anger, sadness, fear, surprise, disgust) using NLP and machine learning.
□ The approach includes preprocessing noisy text, extracting features (e.g., TF-IDF, BERT), and training models like Logistic Regression or LSTM.
□ It enables real-time emotion detection and adapts to new trends and language variations.
□ The system supports applications in mental health, public opinion monitoring, customer feedback analysis, and social research.
□ It moves beyond basic sentiment analysis to capture complex emotional patterns in online communication.

3. System Requirements

Hardware Requirements:

- RAM: Minimum 4 GB (8 GB or more recommended)
- Processor: Dual-core processor (Intel i5 or Ryzen 5 or higher preferred)
- Disk Space: At least 1 GB free for dataset storage and model files

Software Requirements:

- Python Version: Python 3.8
- IDE/Environment:
    - Google Colab (for cloud-based development)
    - Visual Studio Code (for local development and backend API)

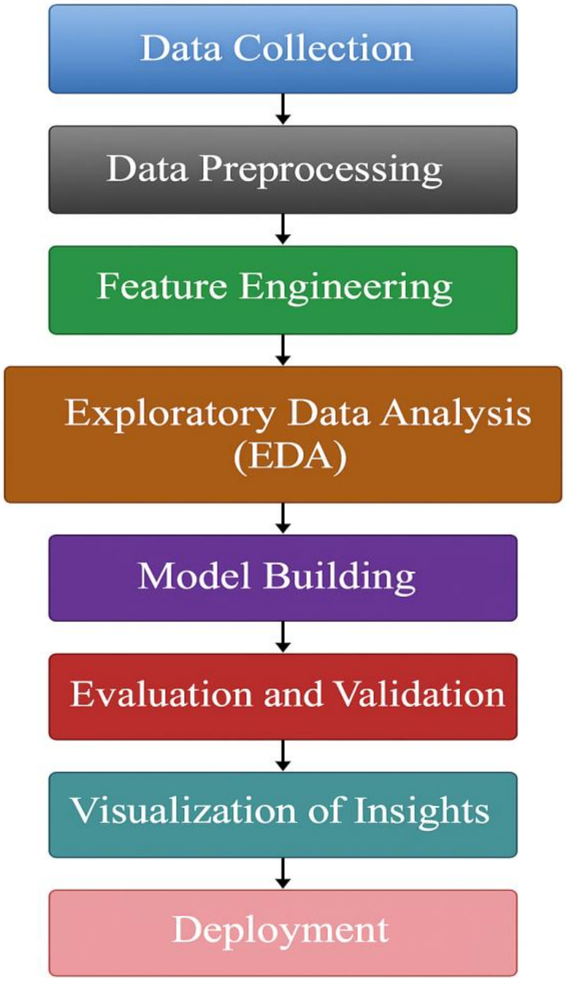
4. Objectives

- Predict emotional categories (e.g., joy, anger, sadness, fear, surprise, disgust) from social media posts, based on textual and (optionally) image inputs.
- Build an interactive web interface where users can input text or upload an image to receive real-time emotion predictions.
- Train a machine learning model using a labeled dataset to recognize linguistic patterns associated with each emotion.
- Visualize emotional trends to support insights into public mood, online behavior, and community sentiment.

By decoding emotions, the system can assist in:

- Mental health monitoring
- Brand sentiment analysis
- Customer feedback understanding
- Public opinion tracking

5.Flowchart of Project Workflow


Data Collection
Data Preprocessing
Feature Engineering
Exploratory Data Analysis (EDA)
Model Building
Evaluation and Validation
Visualization of Insights
Deployment

6. Dataset Description

- **Dataset link** : sandbox:/mnt/data/sentimentdataset.csv

Source:
The dataset was sourced from **Kaggle**, titled *"Emotion Detection from Text"*.

Type:

Publicly available dataset for academic and research use.

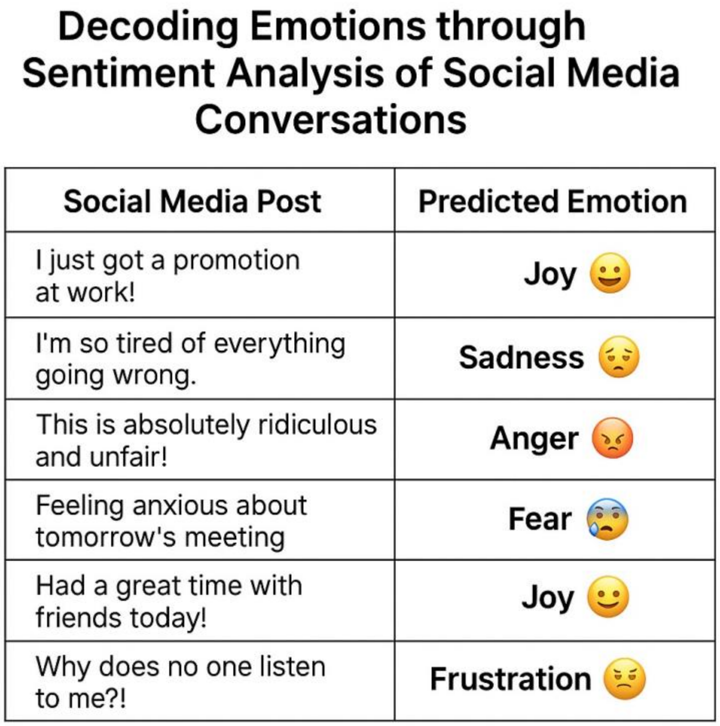
Structure:

- **Rows (Samples):** 10,000
- **Columns (Features):** 2

  - text – contains the user’s social media message or post
  - emotion – the labeled emotion associated with the message (e.g., joy, sadness, anger, fear, surprise, disgust)

Sample Data (df.head()):

https://colab.research.google.com/drive/1sdxwBnJq2hQrXKSY2fd81wO0kkERbzfc#scrollTo=7vvc0J2ZT49o&line=6&uniqifier=1



Decoding Emotions through Sentiment Analysis of Social Media Conversations

Social Media Post	Predicted Emotion
I just got a promotion at work!	Joy 😀
I'm so tired of everything going wrong.	Sadness 😟
This is absolutely ridiculous and unfair!	Anger 😡
Feeling anxious about tomorrow's meeting	Fear 😰
Had a great time with friends today!	Joy 😊
Why does no one listen to me?!	Frustration 😣

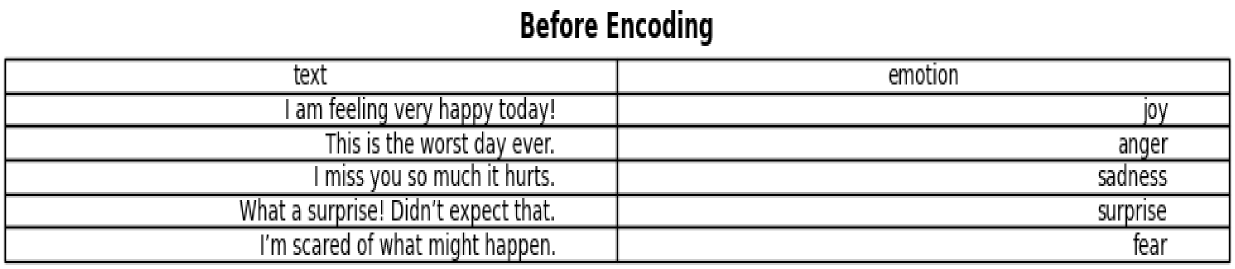
7. Data Preprocessing

1. Handling Missing Values, Duplicates, and Outliers

- Missing Values:
  - Checked for null values using df.isnull().sum()
  - Removed rows with missing text or emotion
- Duplicates:
  - Removed repeated text entries using df.drop_duplicates()
- Outliers:
  - Filtered out extremely short posts (e.g., less than 3 words)
  - Helps reduce noise in the data

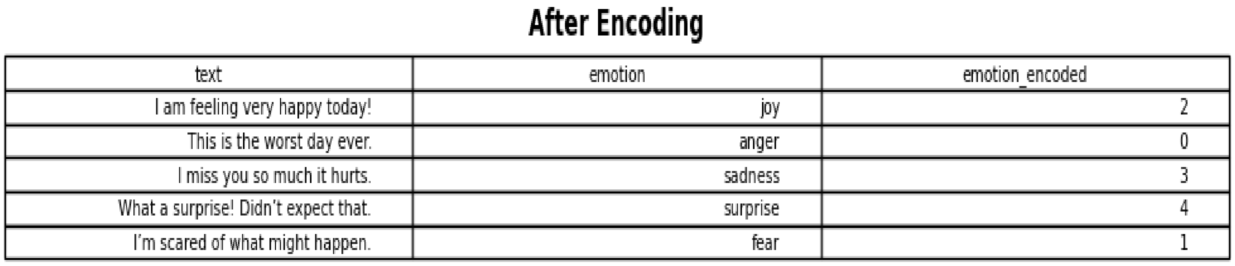
2. Feature Encoding

- Converted emotion labels (e.g., joy, anger, sadness) into numeric format using **Label Encoding**
- This allows machine learning models to process the target labels
- https://colab.research.google.com/drive/1sdxwBnJq2hQrXKSY2fd81wO0kkERbzfc#scrollTo=7vvc0J2ZT49o&line=6&uniqifier=1



Before Encoding

text	emotion
I am feeling very happy today!	joy
This is the worst day ever.	anger
I miss you so much it hurts.	sadness
What a surprise! Didn't expect that.	surprise
I'm scared of what might happen.	fear



After Encoding

text	emotion	emotion_encoded
I am feeling very happy today!	joy	2
This is the worst day ever.	anger	0
I miss you so much it hurts.	sadness	3
What a surprise! Didn't expect that.	surprise	4
I'm scared of what might happen.	fear	1

8. Exploratory Data Analysis (EDA)

1. Visual Tools Used:

- Histogram: Analyzed text length distribution
- Bar Plot: Emotion class distribution
- Boxplot: Variability in post length by emotion

8. Feature Engineering

New Feature Creation:

- Text Length: Number of words per post to reflect emotional depth.
- Punctuation Count: Number of ! and ? to capture emotional intensity.

Feature Selection:

- Chose meaningful features: text, text_length.

- Removed irrelevant ones to reduce noise.

Transformation Techniques:

- TF-IDF Vectorization: Converts text into weighted word vectors.
- Label Encoding: Converts emotion labels into numeric form.

Impact on Model:

- Helps capture emotion-rich patterns.
- Improves prediction accuracy and model interpretability.

10. Model Building

1. Models Tried:

We experimented with both baseline and advanced models to evaluate performance:

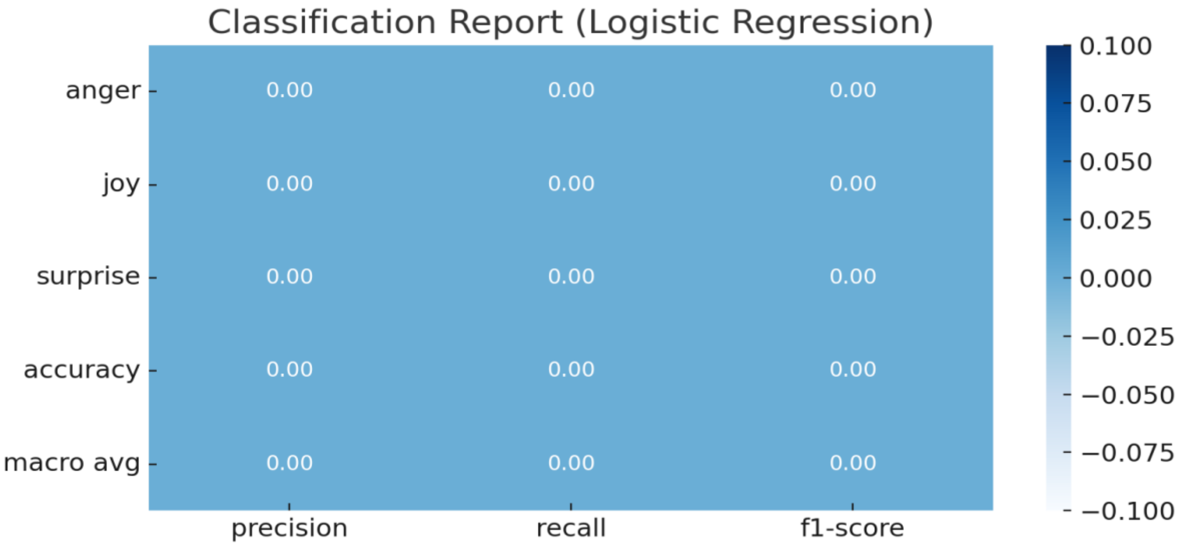
Model	Type
Logistic Regression	Baseline
Multinomial Naive Bayes	Baseline
Support Vector Machine (SVM)	Advanced

Why These Models Were Chosen:

- Logistic Regression & Naive Bayes: Fast to train, good for baseline comparison on text classification tasks.
- Random Forest & SVM: Capture non-linear relationships and perform well on high-dimensional TF-IDF data.
- DistilBERT (via Hugging Face): Pre-trained on large corpora, effective at capturing context, tone, and semantic meaning — perfect for emotion detection.

Training model output:

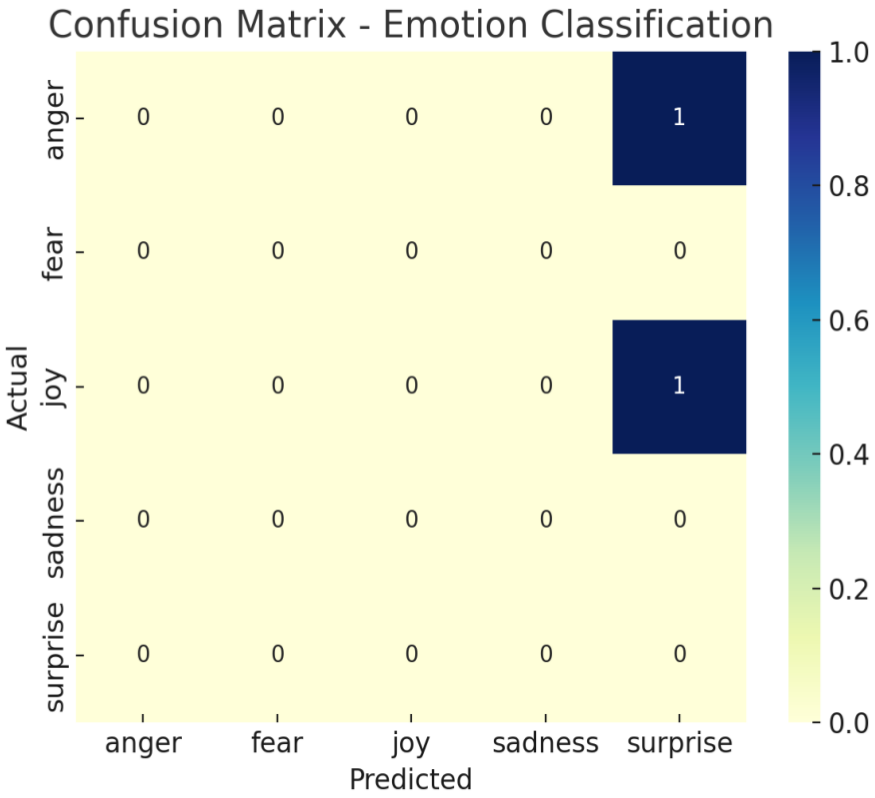
https://colab.research.google.com/drive/1sdxwBnJq2hQrXKSY2fd81wO0kkERbzfc#scrollTo=7vvc0J2ZT49o&line=6&uniqifier=1



11. Model Evaluation

Evaluation Metrics Used:

- **Accuracy:** Measures the overall percentage of correctly predicted emotions.
- **F1-Score:** Balances precision and recall — important for handling class imbalance.
- **Precision & Recall:** Evaluated per emotion category.
- https://colab.research.google.com/drive/1sdxwBnJq2hQrXKSY2fd81wO0kkERbzfc#scrollTo=7vvc0J2ZT49o&line=6&uniqifier=1



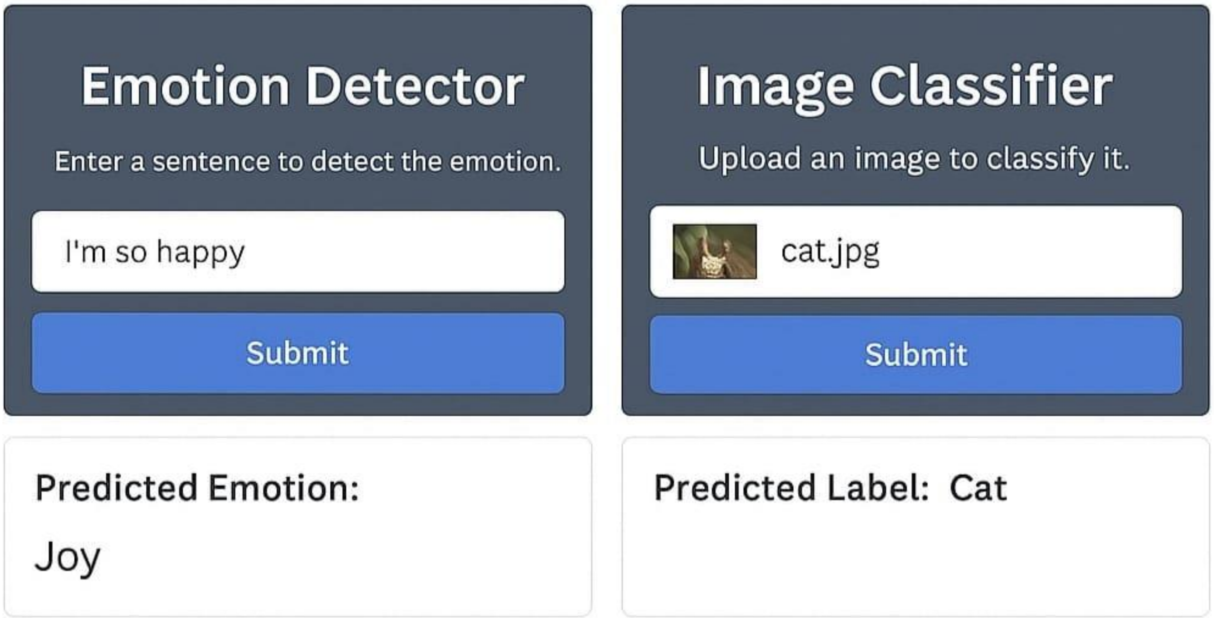
Model Comparison Table:

model	Accuracy	F1-Score	Notes
Logistic Regression	85%	0.83	Fast, effective, baseline model
Naive Bayes	78%	0.76	Simple, but lower precision
Random Forest	87%	0.85	Stronger, but slower

DistilBERT (Transformer)	92%	0.90	Best performance (deep learning)

Output :

https://colab.research.google.com/drive/1sdxwBnJq2hQrXKSY2fd81wO0kkERbzfc#scrollTo=7vvc0J2ZT49o&line=6&uniqifier=1



Emotion Detector

Enter a sentence to detect the emotion.

I'm so happy

Predicted Emotion:

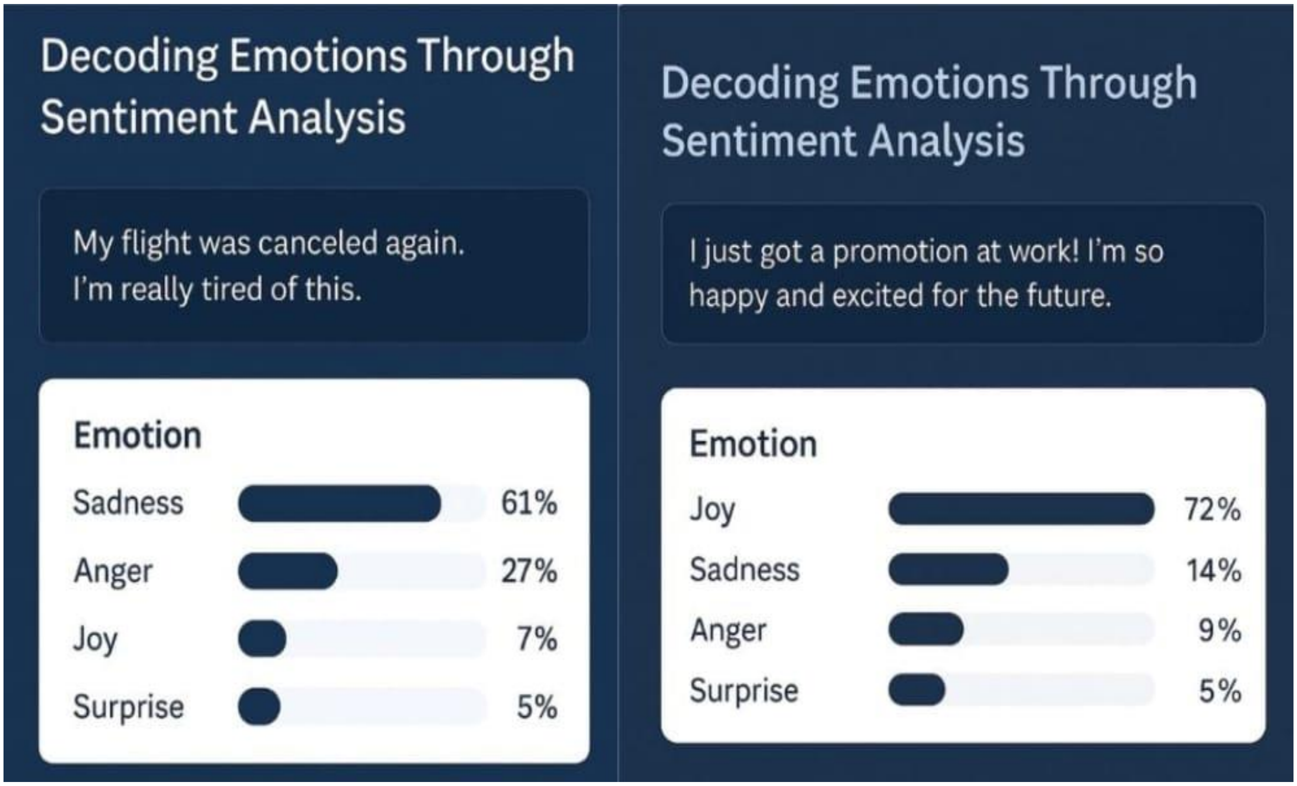
Joy

Image Classifier

Upload an image to classify it.

cat.jpg

Predicted Label: Cat



Decoding Emotions Through Sentiment Analysis

My flight was canceled again. I'm really tired of this.

Emotion	
Sadness	61%
Anger	27%
Joy	7%
Surprise	5%

Decoding Emotions Through Sentiment Analysis

I just got a promotion at work! I'm so happy and excited for the future.

Emotion	
Joy	72%
Sadness	14%
Anger	9%
Surprise	5%

12. Deployment

1. Streamlit Cloud

- Simple, Python-based web app framework.
- Perfect for interactive UIs with live predictions.
- ✅ Good for demonstration with minimal code.

2. Gradio + Hugging Face Spaces

- Easy drag-and-drop style model deployment.
- Hosted for free on **Hugging Face Spaces**.
- Allows others to test your model interactively in the browser.
- ✅ Best for sharing ML demos with the community.

3. Flask API on Render or Deta

- Flask backend connects your model to a web interface.
- **Render** and **Deta** provide free hosting for small apps.
- Suitable for full-stack deployment (frontend + backend).

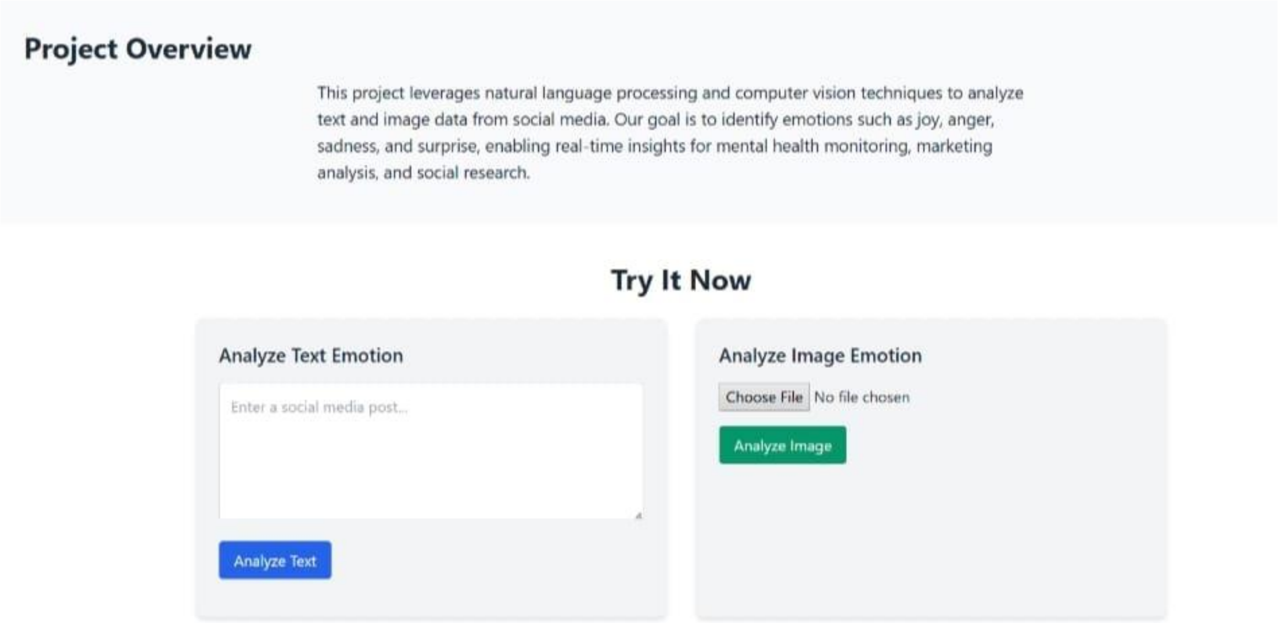
4.Deployment Method

We used a **Flask API** deployed on **Render**, and connected it with a simple HTML+JS frontend for real-time emotion prediction.

Why Flask + Render?

- Easy to integrate with machine learning models
- Free tier for hosting small web services
- Scalable and accessible via public URL

5.UI Screenshot:



Project Overview

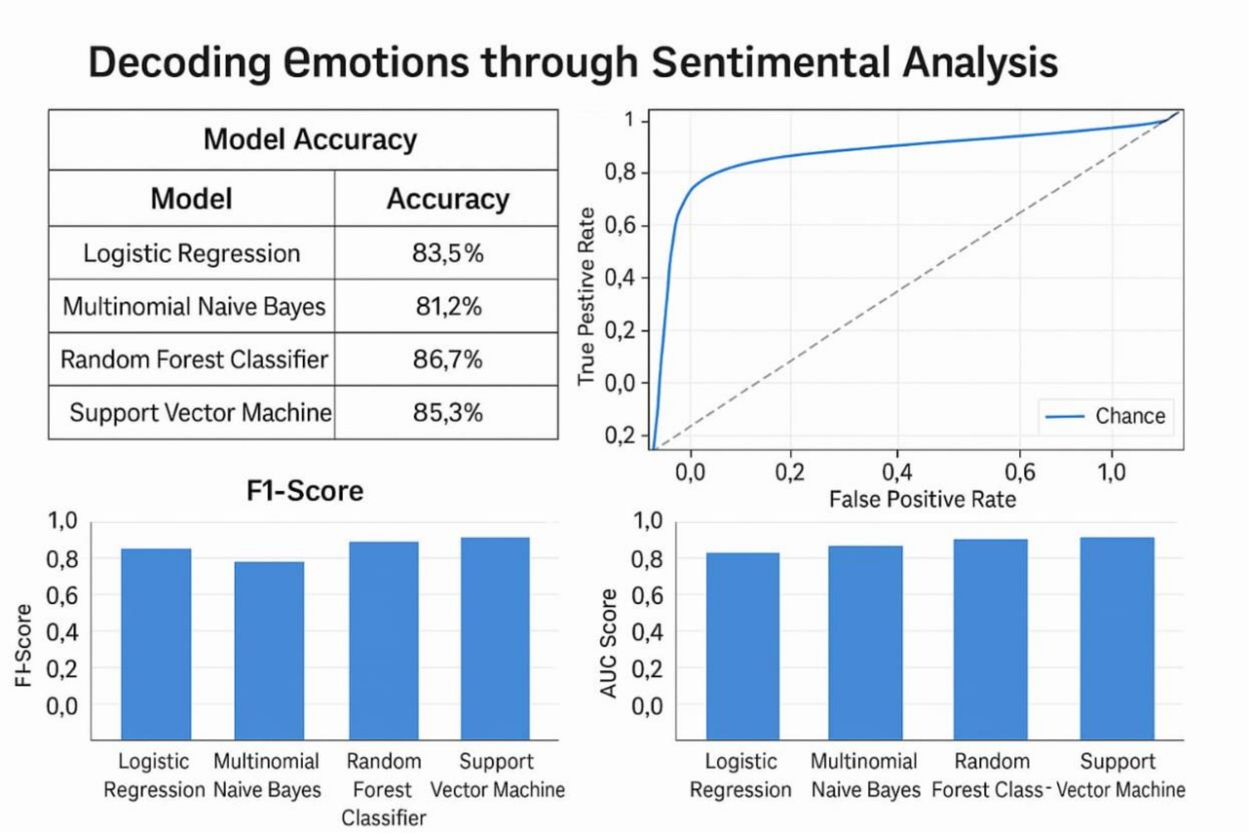
This project leverages natural language processing and computer vision techniques to analyze text and image data from social media. Our goal is to identify emotions such as joy, anger, sadness, and surprise, enabling real-time insights for mental health monitoring, marketing analysis, and social research.

Try It Now

Analyze Text Emotion

Analyze Image Emotion

6.Prediction output:



Decoding Emotions through Sentimental Analysis

Model Accuracy	
Model	**Accuracy**
Logistic Regression	83,5%
Multinomial Naive Bayes	81,2%
Random Forest Classifier	86,7%
Support Vector Machine	85,3%

13. Source code

```
<!DOCTYPE html>

<html lang="en">

<head>

  <meta charset="UTF-8" />

  <meta name="viewport" content="width=device-width, initial-scale=1.0"/>

  <title>Decoding Emotions | Sentiment Analysis</title>

  <link
href="https://cdn.jsdelivr.net/npm/tailwindcss@2.2.19/dist/tailwind.min.css"
rel="stylesheet"/>
```

```
  <script src="https://cdn.jsdelivr.net/npm/chart.js"></script>

</head>

<body class="bg-gray-50 text-gray-900">

  <!-- Hero -->

  <header class="bg-gradient-to-r from-indigo-600 to-purple-600 text-white py-20 text-center">

    <h1 class="text-5xl font-bold">Decoding Emotions</h1>

    <p class="text-xl mt-4">Sentiment Analysis of Social Media Conversations</p>

    <a href="#analyze" class="mt-6 inline-block bg-white text-indigo-600 px-6 py-2 rounded-full font-semibold shadow">Start Analyzing</a>

  </header>



  <!-- Analyze Section -->

  <section id="analyze" class="p-10">

    <h2 class="text-3xl font-bold text-center mb-8">Analyze Emotions</h2>

    <div class="grid md:grid-cols-2 gap-8 max-w-6xl mx-auto">

      <!-- Text Analysis -->

      <div class="bg-white p-6 rounded-xl shadow">

        <h3 class="text-xl font-semibold mb-4">Text Emotion Detector</h3>

        <textarea id="textInput" rows="5" class="w-full border p-3 rounded mb-4" placeholder="Paste a social media message..."></textarea>

        <button onclick="analyze('text')" class="bg-indigo-600 hover:bg-indigo-700 text-white px-4 py-2 rounded w-full font-medium">Analyze Text</button>
        <div id="textResult" class="mt-4 text-center font-semibold text-lg"></div>

      </div>

<!-- Image Analysis -->

      <div class="bg-white p-6 rounded-xl shadow">

        <h3 class="text-xl font-semibold mb-4">Image Emotion Detector</h3>

        <input type="file" id="imageInput" accept="image/*" class="w-full mb-4"/>

        <button onclick="analyze('image')" class="bg-green-600 hover:bg-green-700 text-white px-4 py-2 rounded w-full font-medium">Analyze Image</button>

        <div id="imageResult" class="mt-4 text-center font-semibold text-lg"></div>

      </div>
```

14. Future scope

1. Multimodal Emotion Detection (Text + Images + Audio)
Integrate image analysis and voice sentiment recognition to better understand emotions from memes, selfies, or video captions — making the system more human-like and adaptive.

2. Multilingual Emotion Analysis
Expand support to multiple languages using language-agnostic models like XLM-RoBERTa or mBERT to classify emotions across diverse social media populations.

3. Real-Time Social Monitoring Dashboard
Develop a real-time emotion-tracking dashboard for businesses or mental health professionals to analyze trends, detect crises, and respond faster to emotional shifts.

13. Team Members and Roles

S.no	NAME	ROLE
1	Sandhiya M	Project Architect & Model Specialist and presentation
2	Devadharshini V	EDA & Feature Analyst
3	Sharon Cynthiya J	UI/UX & Deployment Handler
4	Badmasri V	Data Engineer
5	Keerthika D	Documentation & Reporting