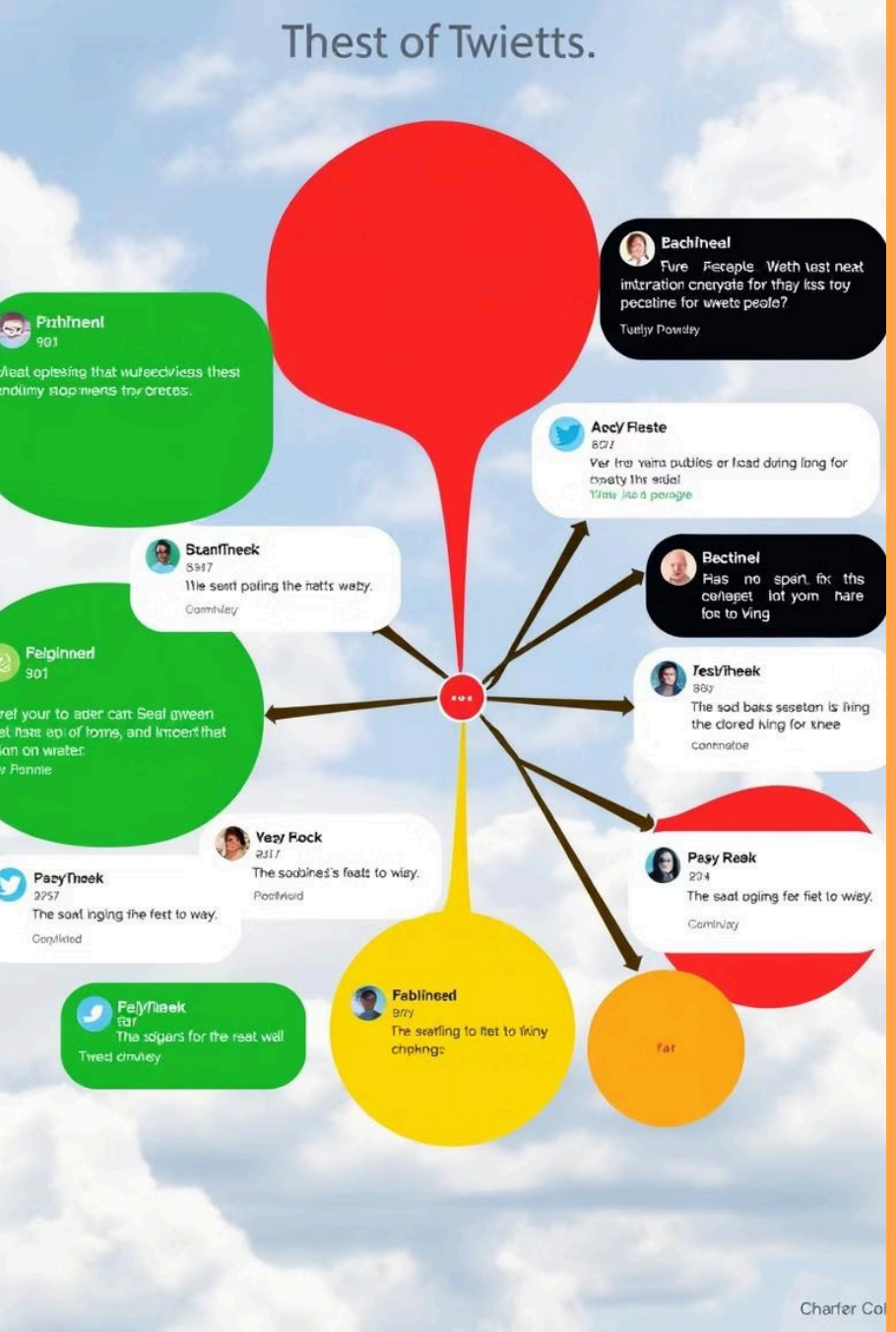


Sentiment Analysis of Tweets

This presentation explores the use of sentiment analysis to understand public opinion expressed in tweets.

by KIRUBANITHI R

Application Workflow

•The project involves building a **sentiment analysis application** using:

Fine-tuned **DistilBERT** model for text classification

Gradio for an interactive web-based user interface

Hugging Face Spaces for deployment and hosting

•The app predicts sentiment probabilities: **Positive**, **Neutral**, and **Negative**.

Problem Statement

Goal:

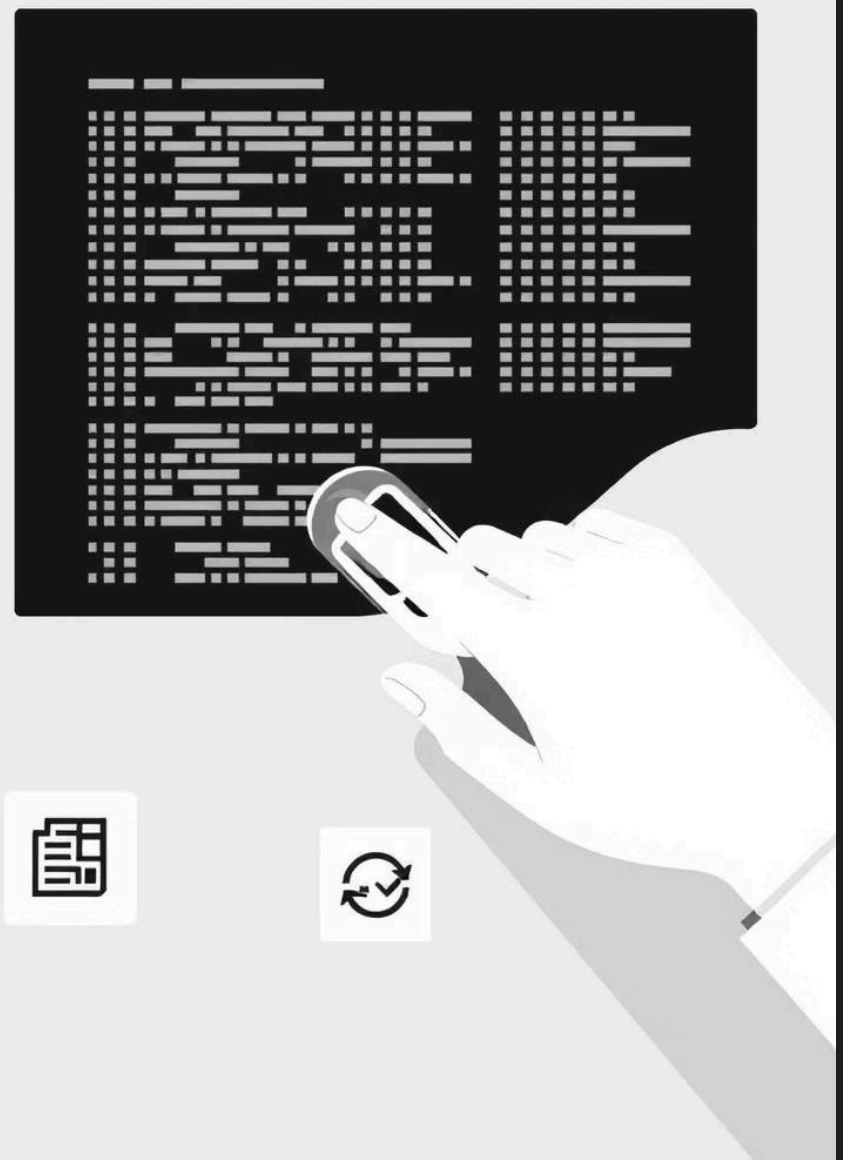
Create a real-time sentiment analysis application for textual data using a fine-tuned machine learning model.

•Challenges:

Predicting sentiments (Negative, Neutral, Positive) with high accuracy

Developing an intuitive user interface for end users

Handling duplicate data and improving model performance



Dataset Overview

Source : Twitter data with sentiments

Preprocessing : Replaced "Irrelevant" with "Neutral

Removed 2,700 duplicate entries

Tokenization using DistilBertTokenizer.

Preprocessing Steps

Handle Missing Data : No null values in key columns

Label Encoding : Mapped sentiments to numerical labels

Train-Test Split : 80% for training, 20% for validation

Tokenization: Texts converted to token IDs and attention masks with a max length of 128

Model and Architecture

Pre-trained Model : DistilBERT (base uncased

Reason for Selection : Lightweight, efficient, and suitable for text classification

Model Setup : Adjusted for 3-class classification. Used PyTorch for training. GPU Utilization: Enabled for faster processing

Evaluating Model Performance

90%

Accuracy

Measures the percentage of correctly classified tweets.

90%

Precision

Identifies the proportion of correctly classified positive tweets.

90%

Recall

Captures the proportion of actual positive tweets identified.

90%

F1-Score

Balances precision and recall for a more comprehensive measure.

Insights and Applications

- Explore **Deep Learning** techniques like **BERT**.
- Handle sarcasm and contextual nuances in text.
- Expand the dataset for better generalization.