Project ⬇️ Disaster Tweet Classification using NLP

A comprehensive machine learning project for real-time disaster detection on social media platforms

By Deepa Pathak | NeXT Hikes IT Solutions |

GitHub link: DataWithDeepa (Deepa Pathak)

DataWithDeepa

The Critical Need for Disaster Detection

Real-Time Alerts

Social media platforms like Twitter serve as crucial channels for immediate disaster communication and emergency response coordination.

Human Challenge

Identifying genuine disaster tweets accurately remains challenging even for human moderators due to context and language complexity.

Classification Challenges

Ambiguous Language

Words like "fire" can refer to actual disasters or metaphorical expressions, creating classification confusion.

Slang & Variations

Internet slang, abbreviations, and regional language variations complicate automated text analysis.

Context Dependency

Distinguishing between real disaster reports and casual mentions requires deep contextual understanding.

Project Objectives

Build Classification Model

Develop a robust machine learning model capable of accurately distinguishing disaster tweets from regular content.

High Precision & Accuracy

Achieve optimal performance metrics to minimize false positives and ensure reliable disaster detection.

Real-Time Scalability

Create a solution that can process tweets in real-time for immediate emergency response applications.

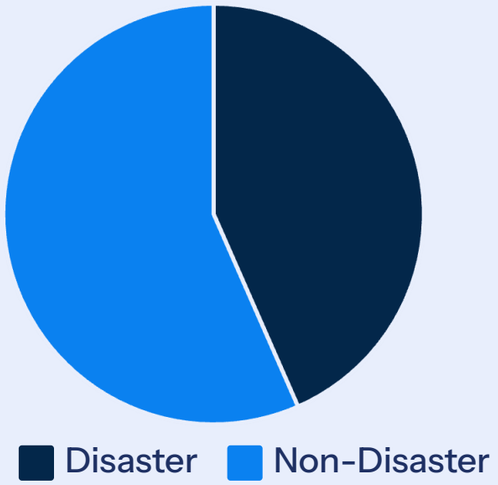
Data Exploration Insights

Dataset Overview

Analyzed 10,000 labeled tweets with binary classification (disaster/non-disaster).
Dataset includes diverse tweet types, hashtags, and emergency-related content.

Key Findings

- Balanced distribution of disaster vs non-disaster tweets
- High variation in tweet length and structure
- Common disaster keywords identified

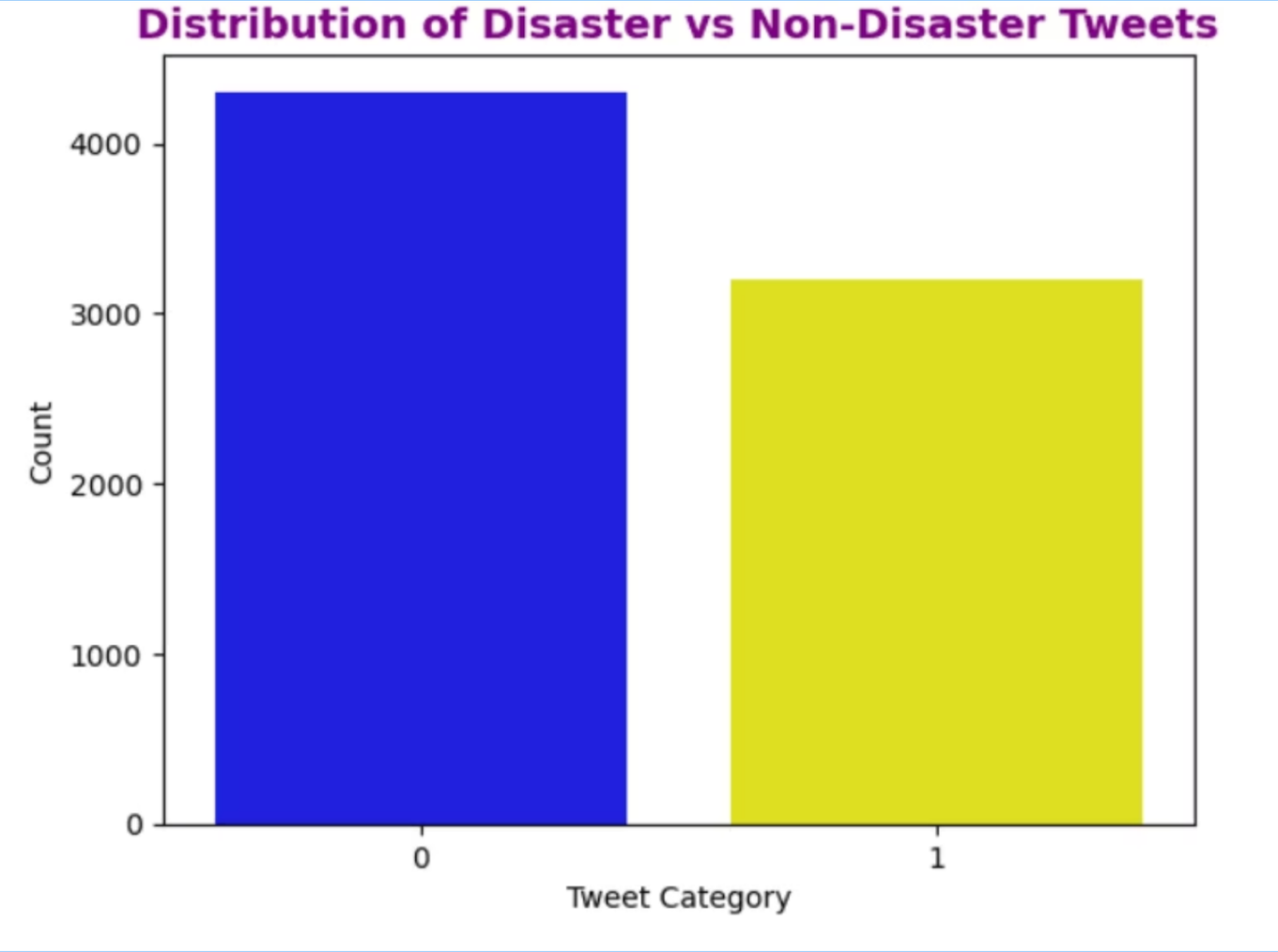


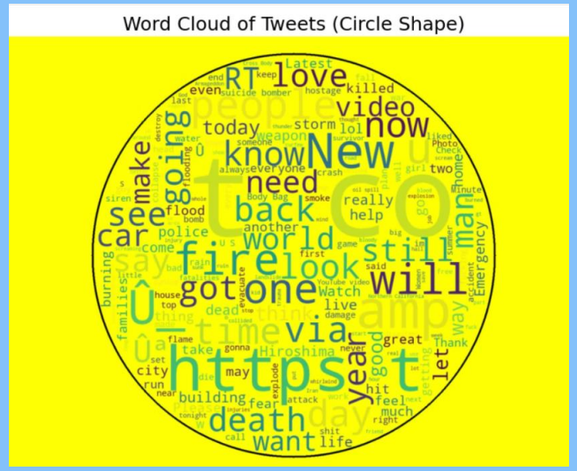
✔ Insight:

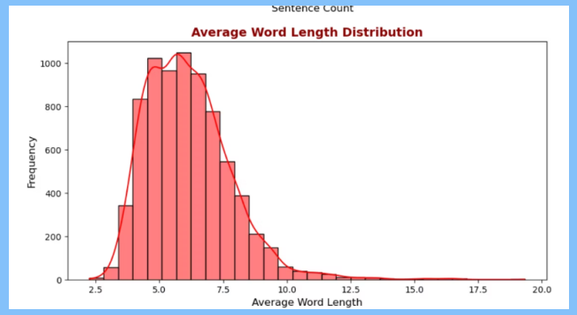
The dataset contains **tweets** with a target column:

- `1` → Disaster Tweet
- `0` → Non-Disaster Tweet

Data Exploration Insights: Distribution of Disaster vs non-disaster tweets & Word cloud

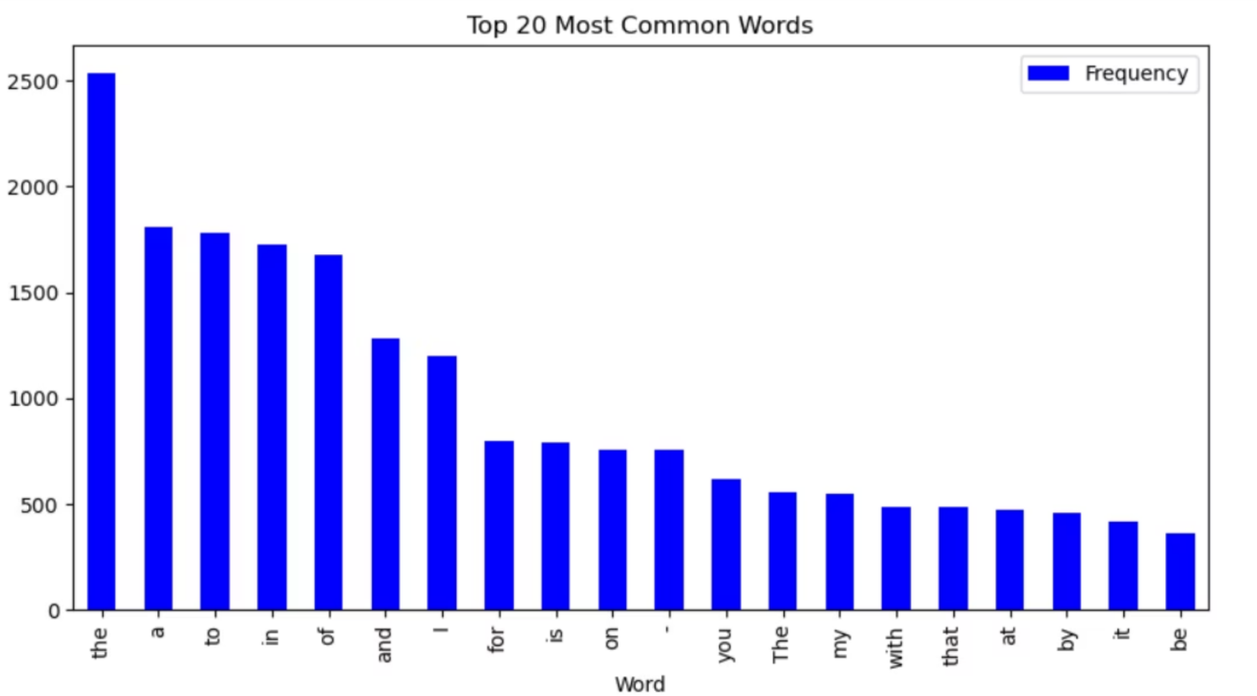


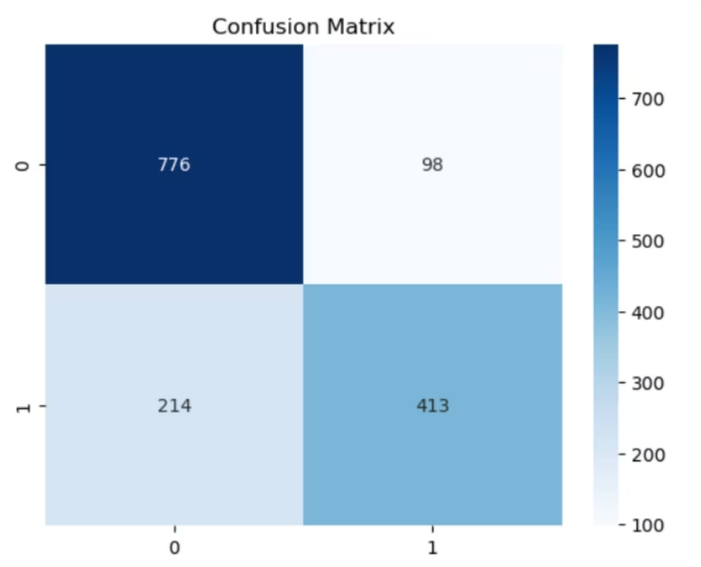




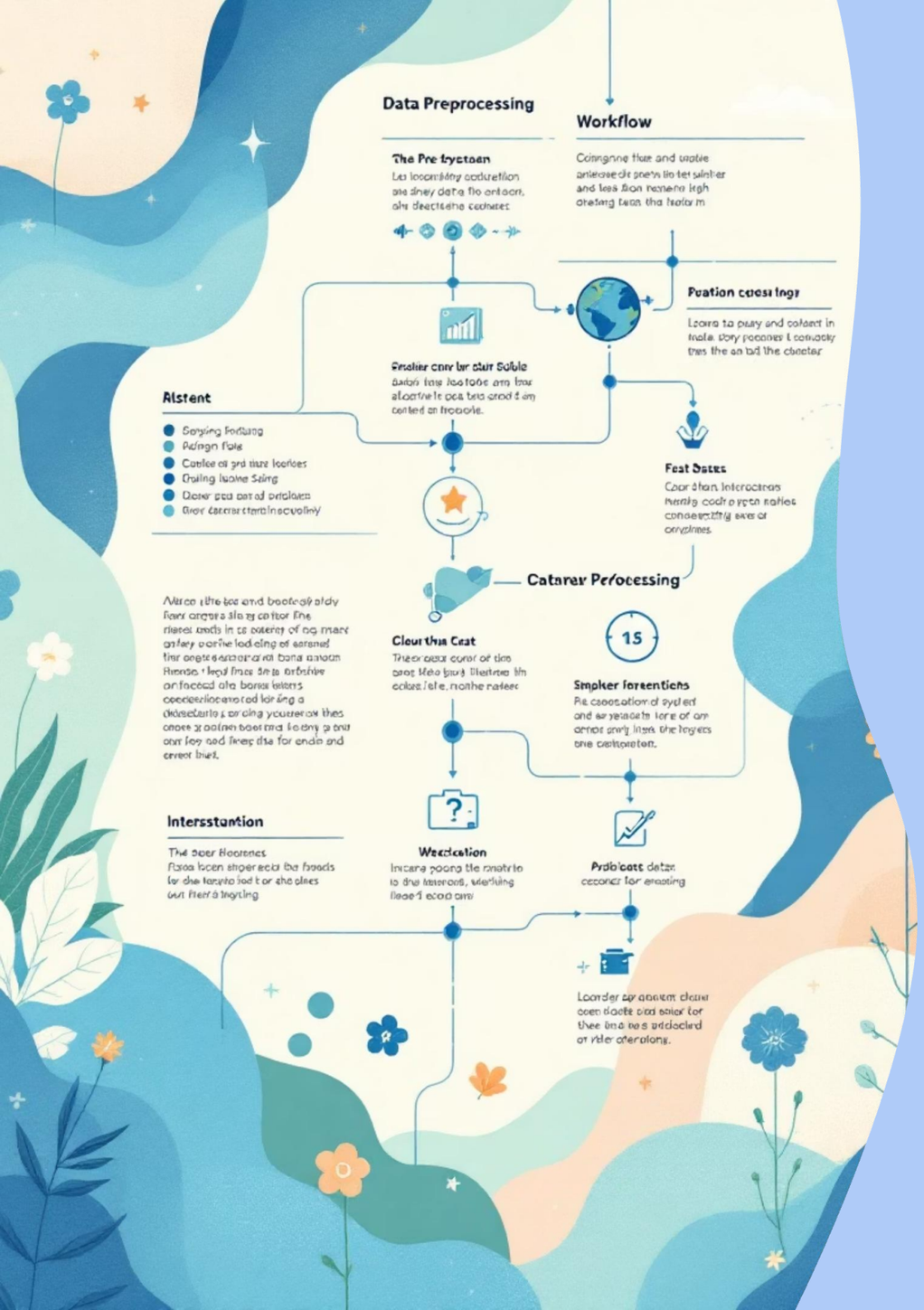
DataWithDeepa

Data Exploration Insights: top 20 most common words & Confusion Matrix





DataWithDeepa



Data Preparation Pipeline

1

Raw Tweets

Original tweet text with URLs, mentions, hashtags, and special characters.

2

Text Cleaning

Remove URLs, special characters, punctuation, and normalize text format.

3

Tokenization

Split cleaned text into individual words and remove stopwords.

4

Train/Test Split

Divide processed data into training and testing sets for model validation.

Feature Engineering & Model Selection

TF-IDF Vectorization

Convert text to numerical features using Term Frequency-Inverse Document Frequency weighting.

Word Embeddings

Implement Word2Vec and GloVe embeddings for semantic text representation.

Additional Features

Extract tweet length, hashtag count, mention frequency, and URL presence as supplementary features.

Model	Accuracy	Precision	Recall
Logistic Regression	82.1%	80.5%	83.2%
Random Forest	79.8%	78.9%	81.1%
Neural Network	85.3%	84.7%	86.0%

Model Performance & Validation

79%

Accuracy

Overall model accuracy on test dataset

81%

Precision

Precision score for disaster tweet detection

89.0%

Recall

Recall rate for identifying true disasters

83%

F1-Score

Balanced performance metric

Neural Network model selected based on superior performance across all evaluation metrics. Confusion matrix and ROC curve analysis confirmed robust classification capability.

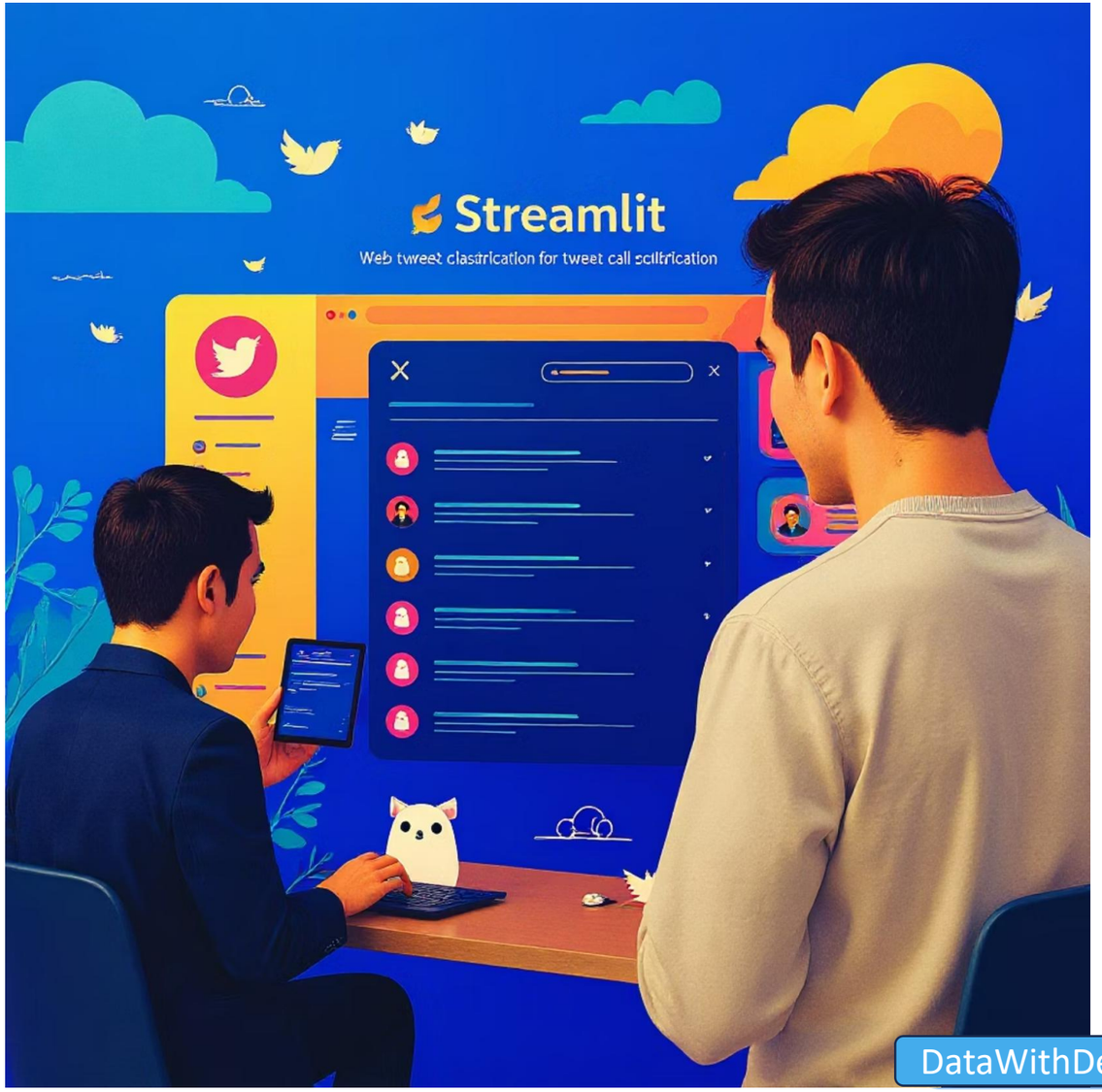
Web Application Deployment: link: Streamlit

Streamlit Web Interface

- Interactive web app for real-time tweet classification
- Multi-language support (English/Hindi)
- Instant prediction with confidence scores
- Emoji-based result visualization

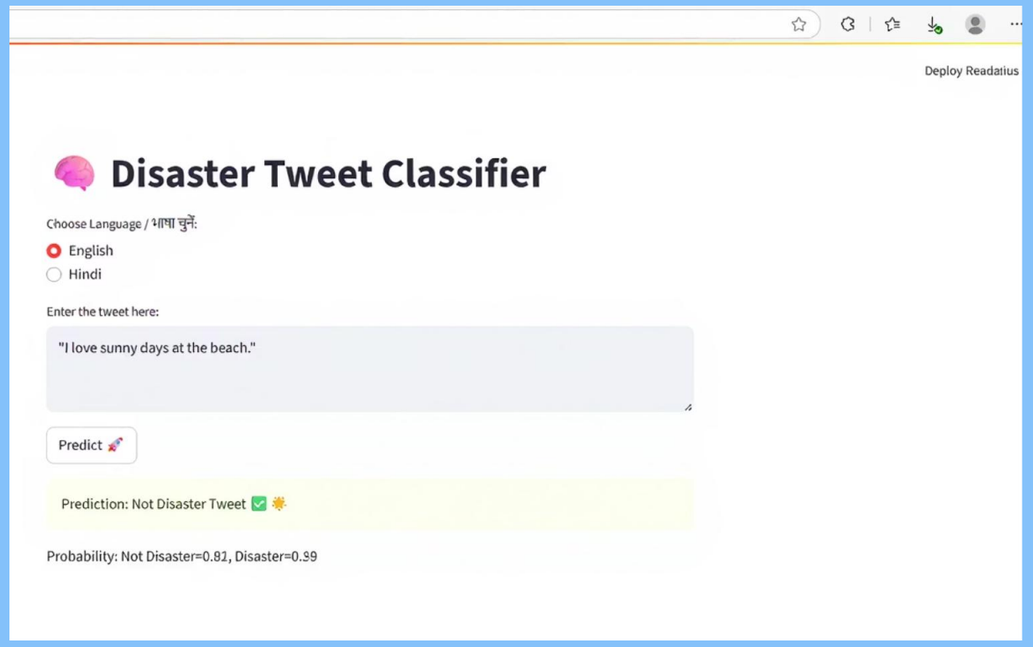
Deployment Features

- Model serialized as .pkl file
- Hosted on Streamlit Community Cloud
- User-friendly interface design

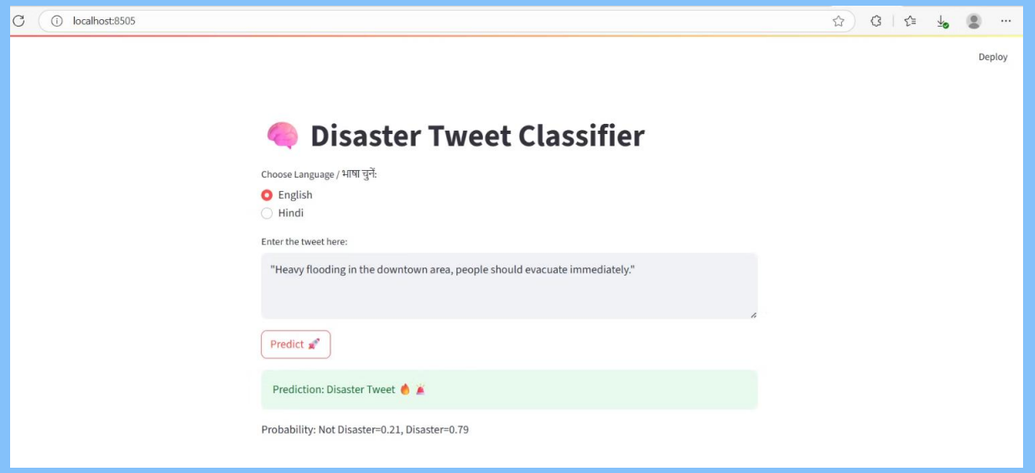


Web Application Deployment: Streamlit Web Interface with English Language tweet.

Non-Disaster Tweet Prediction



Disaster Tweet Prediction



DataWithDeepa

Streamlit Web Interface with Hindi Language tweet.Streamlit App link : Streamlit

Impact & Future Scope

Project Impact

- **Enables faster disaster response through automated tweet classification, supporting emergency services with real-time social media monitoring capabilities.**
- **Future Enhancements**
- **Multi-language expansion, sentiment-based alert systems, integration with emergency services, and advanced deep learning models for improved accuracy.**

Thank You

Advancing disaster response through intelligent NLP solutions