

Social Media Trend Prediction using NLP and Machine Learning

Mantripragada Venkata Karthikeya

Project Overview

- ▶ **Problem Statement:** Predict trending hashtags on social media platforms like Twitter and Reddit.
- ▶ **Solution:** A system that scrapes real-time social media data, preprocesses it using advanced NLP techniques, and predicts trends using machine learning models.
- ▶ **Technologies Used:**
 - ▶ Web scraping (Twitter, Reddit)
 - ▶ NLP (Word2Vec, Sentiment Polarity)
 - ▶ Clustering (K-Means)
 - ▶ Deployment with **Streamlit** for user interface.

Data Collection & Preprocessing

- ▶ **Automated Scraping:**

- ▶ Scrapes trending hashtags from **Twitter** and **Reddit**.
- ▶ Regularly fetches updated hashtags to stay relevant.

- ▶ **Text Preprocessing:**

- ▶ Cleans scraped text by removing noise (punctuation, stop words).
- ▶ Applies tokenization, stemming, and lemmatization for better NLP analysis.
- ▶ Prepares data for training by transforming raw text into **Word2Vec** embeddings.

Predictive Modeling

- ▶ **Word2Vec (Gensim):**
 - ▶ **Word Embeddings:** Converts words into vector representations that capture semantic relationships.
 - ▶ Helps in identifying similarity and context of hashtags.
- ▶ **Sentiment Polarity Analysis:**
 - ▶ Uses deep learning models to assess sentiment of posts and hashtags (positive, negative, neutral).
 - ▶ Provides insights into hashtag popularity.
- ▶ **K-Means Clustering:**
 - ▶ Categorizes hashtags based on their popularity.
 - ▶ Helps group related hashtags together for trend prediction.

User Interface & Deployment

- ▶ **Streamlit UI:**

- ▶ Provides an interactive interface for users to view predictions.
- ▶ Easy to use for end-users with no technical expertise.

- ▶ **Deployment:**

- ▶ Deployed as a web application accessible via `http://localhost:8501`.
- ▶ Allows users to input and view trending hashtag predictions in real time.