

# Project Documentation: Social Media Trend Analysis Tool

## Table of Contents

1. Introduction
2. System Overview
3. Technology Stack
4. Usage
5. Features
6. Data Flow
7. Scraping Process
8. Topic Extraction
9. Filtering and Prioritization
10. Frontend
11. Conclusion

## 1. Introduction

The Social Media Trend Analysis Tool is a software project designed to identify trending topics related to a specified keyword across various social media platforms. This document provides an overview of the project, its features, technology stack, and usage instructions.

## 2. System Overview

The system follows a systematic approach:

1. **Keyword Input:** Users provide a specific keyword (e.g., "Young Innovator").
2. **Scraping:** The system automates web scraping using Selenium and BeautifulSoup to gather information from social media platforms (e.g., Pinterest, Facebook, Reddit).
3. **Data Storage:** Scraped data is stored in a MongoDB database for in-depth analysis.
4. **Topic Extraction:** The system utilizes TF-IDF Vectorization and Non-Negative Matrix Factorization (NMF) for extracting relevant topics from the collected data.

5. **Post Relevance:** BERT-based tokenization and cosine similarity are employed to determine the relevance of posts to the specified keyword.
6. **Filtering:** Users can apply filters (region-based, country-based, age-based) to customize trending topics.
7. **Display:** The prioritized and filtered topics are displayed in a Streamlit frontend.

### 3. Technology Stack

- Web Scraping: Selenium, BeautifulSoup
- Database: MongoDB
- Topic Extraction: TfidfVectorizer, NMF
- Post Relevance: PyTorch, BertTokenizer, BertModel, cosine\_similarity
- Frontend: Streamlit

### 4. Usage

1. Access the Streamlit app through a web browser.
2. Enter the desired keyword in the input field.
3. Adjust filters (optional).
4. Click the "Submit" button to initiate the analysis.
5. View the prioritized and filtered trending topics.

### 5. Features

- Automated web scraping from multiple social media platforms.
- MongoDB storage for efficient data management.
- Topic extraction for identifying key subjects related to the specified keyword.
- Post relevance analysis using BERT-based models.
- Customizable filters for enhanced user experience.

### 6. Data Flow

1. User inputs a keyword in the Streamlit frontend.
2. Selenium and BeautifulSoup scrape data from social media platforms.
3. Scraped data is stored in MongoDB.
4. TF-IDF Vectorization and NMF extract topics from the stored data.
5. BERT-based models assess post relevance.

6. Prioritized and filtered topics are displayed in the Streamlit frontend.

## **7. Scraping Process**

- The scraping process involves navigating through the social media platforms, identifying relevant posts, and extracting textual information.

## **8. Topic Extraction**

1. TF-IDF Vectorization and NMF are used to identify key topics within the scraped data.
2. The process involves transforming text data into numerical vectors and applying matrix factorization for topic extraction.

## **9. Filtering and Prioritization**

- Users can apply filters such as region, country, and age to customize the displayed trending topics.
- The system prioritizes topics based on relevance and engagement metrics.

## **10. Frontend**

- Streamlit is used for creating an interactive and user-friendly frontend.
- The interface allows users to input keywords, apply filters, and view the results in a structured format.

## **11. Conclusion**

The Social Media Trend Analysis Tool provides a comprehensive solution for identifying and customizing trending topics across various social media platforms. Its modular architecture, efficient scraping mechanisms, and advanced analysis techniques make it a valuable tool for content creators, marketers, and social media enthusiasts.