Bouchra LOUKILI

Project Report

Introduction

This project implements an automated pipeline for analyzing YouTube video content through Natural Language Processing (NLP). Leveraging Apache Airflow for workflow orchestration and Streamlit for interactive visualization, the system processes video metadata, performs sentiment/topic analysis, and displays results on a real-time dashboard. Key innovations include containerized microservices, automated data versioning, and multi-modal visualization.

Objectives

- Develop an automated ETL pipeline for YouTube data
- Implement NLP models (sentiment analysis/topic modeling)
- Create an interactive dashboard for trend visualization

1. System Architecture

1.1 Technical Stack

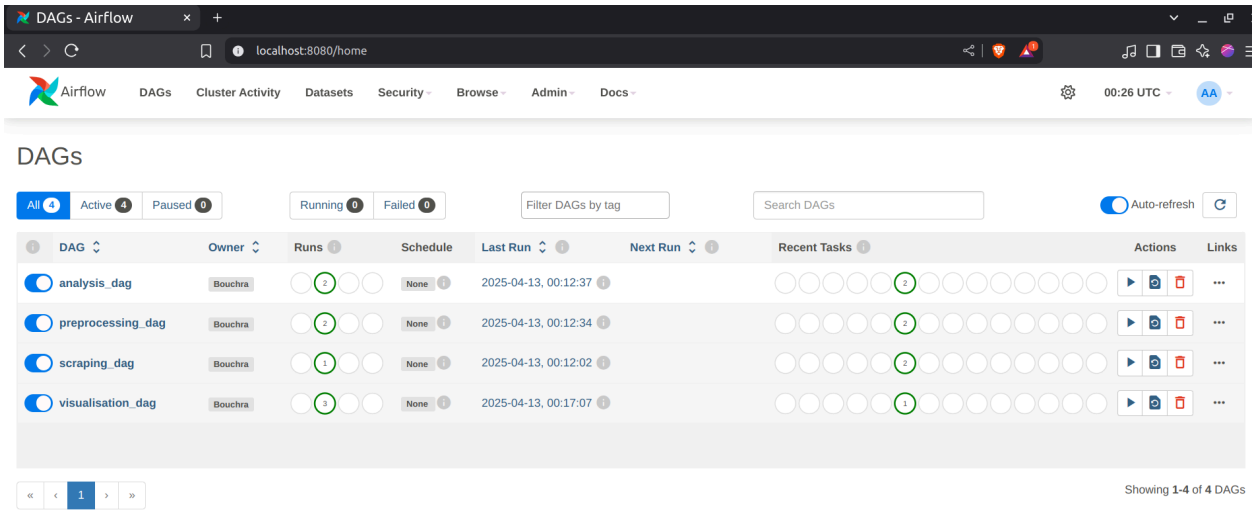
Component	Technology
Orchestration	Apache Airflow 2.10.5
NLP Processing	PyTorch, Transformers
Visualization	Streamlit, Plotly
Infrastructure	Docker, PostgreSQL

1.2 Technical Implementation Details

Key Additions to docker-compose.yml:

- added dedicated Streamlit service
- Permission synchronization:Added user: "${UID:-1000}:0" to match host user
- Cross-service communication:Configured shared volume ./dags/data:/data for Airflow→Streamlit data transfer

1.3 DAGs List



1.4 Streamlit Integration

We deployed Streamlit as a standalone container to:

1. *Ensure Decoupling*
    - Runs independently from Airflow (no shared codebase)
    - Communicates solely via mounted volume (/data)
2. *Guarantee UI Stability*
    - Remains operational during Airflow updates/errors
    - Persists cached data if source systems fail
3. *Enable Rapid Iteration*
    - Dashboard updates require only Streamlit rebuild
    - No impact on running DAGs

2. Results & Discussion

Dashboard Features

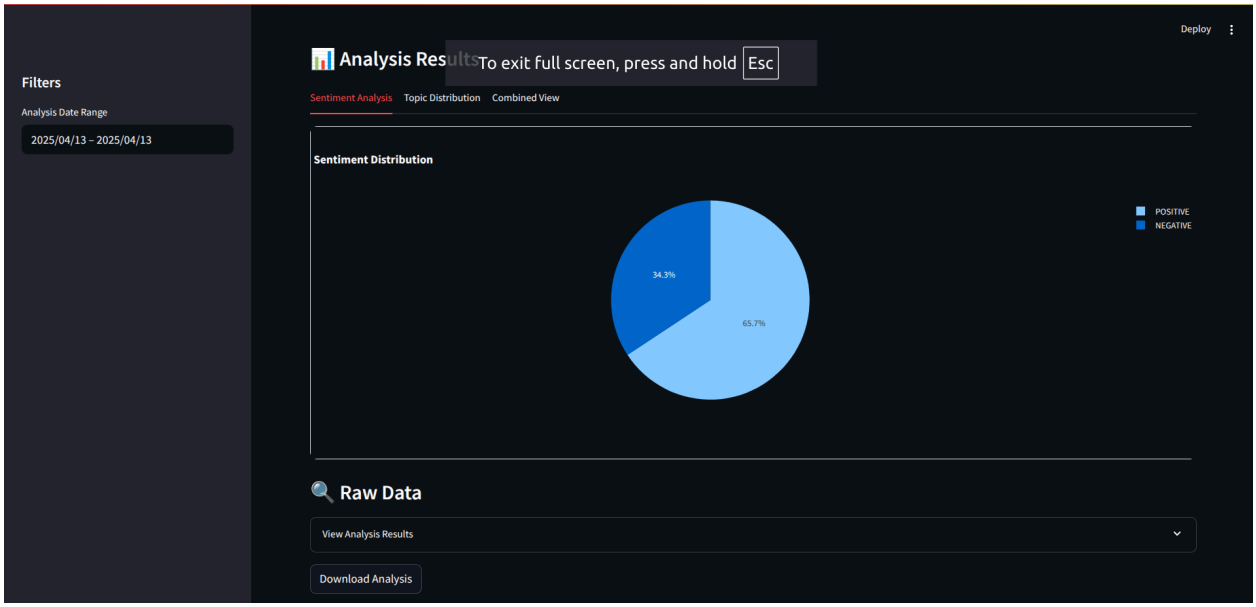
- Real-time filtering by date/topic
- Sunburst charts for sentiment-topic correlation
- Automated report generation (CSV export)

Bouchra LOUKILI

Sample Visualization

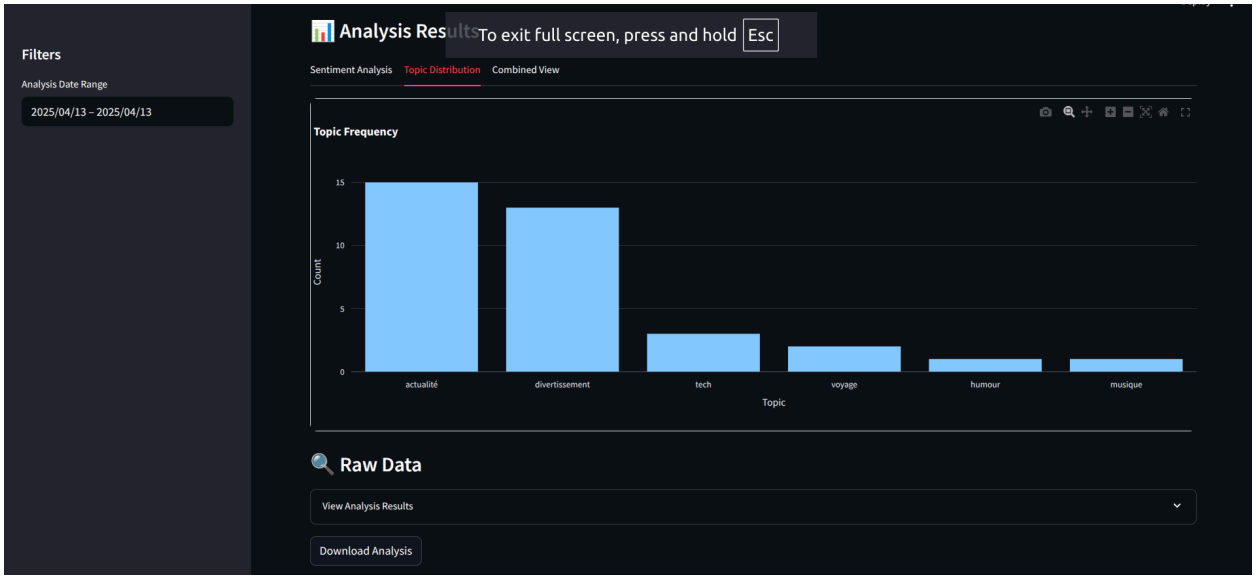
1. Sentiment Distribution Pie Chart

- Purpose: Shows sentiment polarity (% positive/neutral/negative)
- Insight: Quick identification of dominant sentiment trends
- Interaction: Hover for exact percentages, click to isolate segments



2. Topic Frequency Bar Chart

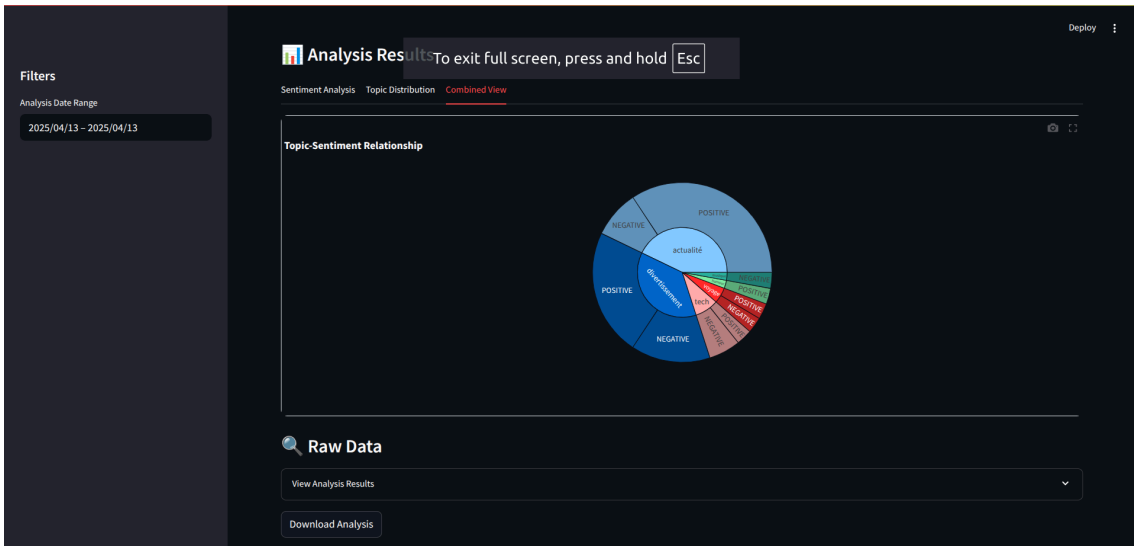
- Purpose: Ranks detected topics by occurrence
- Insight: Reveals most discussed content themes
- Feature: Dynamic sorting (ascending/descending)



Bouchra LOUKILI

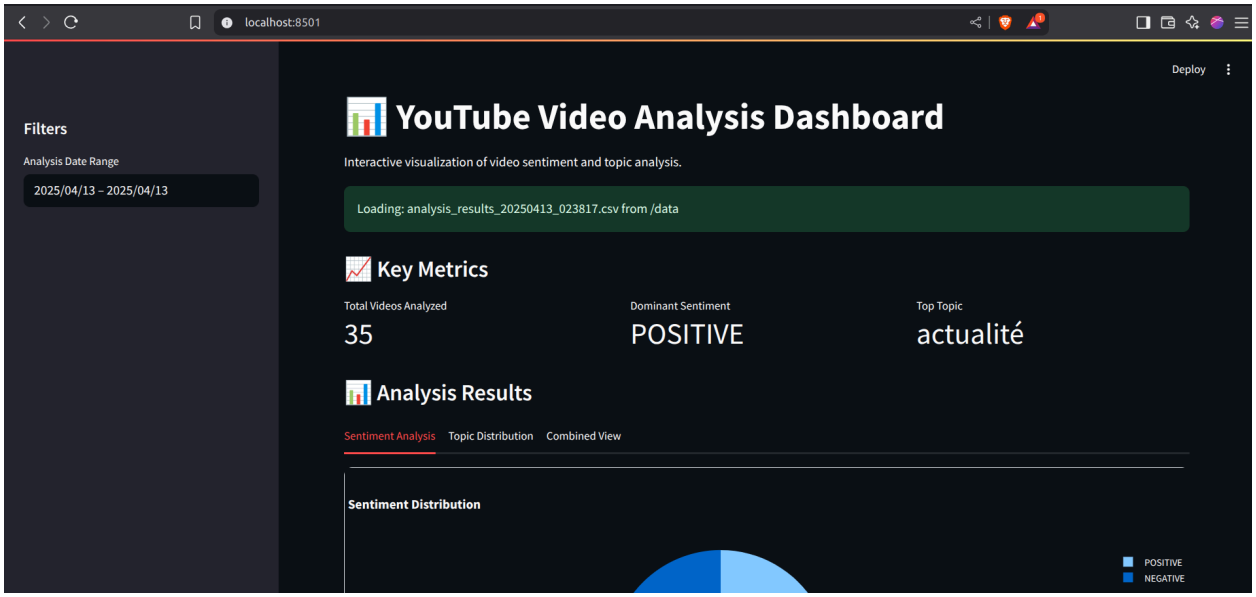
3. Topic-Sentiment Sunburst

- Purpose: Hierarchical view of sentiment nested within topics
- Insight: Identifies which topics drive positive/negative reactions



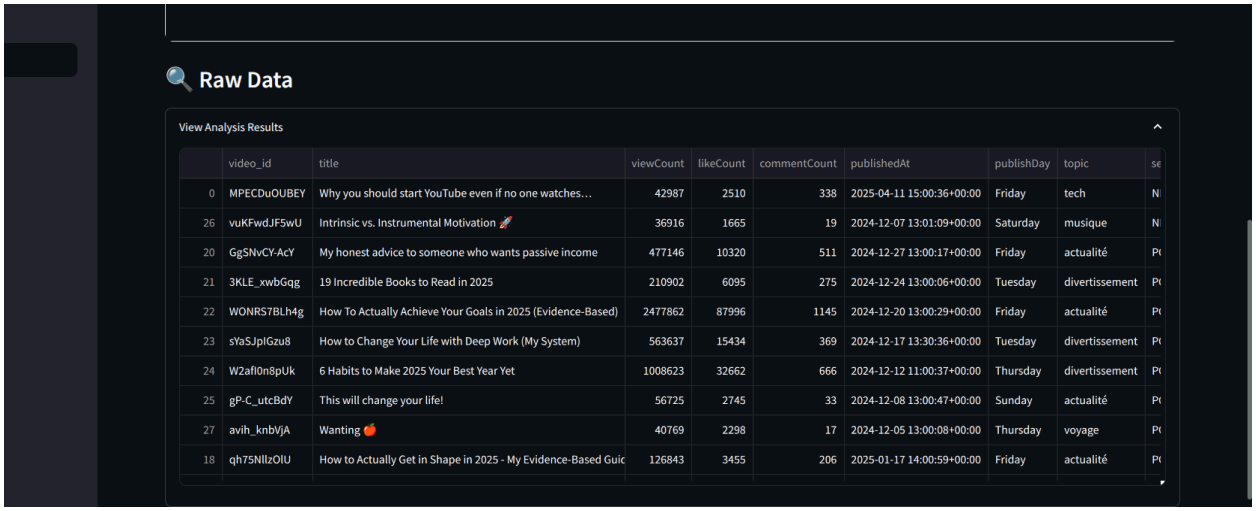
4. Date-Range Filtered Metrics

- Components:
  - Total videos analyzed (counter)
  - Dominant sentiment/topic (dynamic badges)
- Technical Note: Real-time updates via Streamlit caching



5. Raw Data Table

- Features:
  - Sortable columns (e.g., by date/sentiment score)
  - Expandable/collapsible view
  - CSV export with one click
  - Download the analysis file



	video_id	title	viewCount	likeCount	commentCount	publishedAt	publishDay	topic	se
0	MPECDuOUBEY	Why you should start YouTube even if no one watches…	42987	2510	338	2025-04-11 15:00:36+00:00	Friday	tech	N
26	vuKFwdJF5wU	Intrinsic vs. Instrumental Motivation 🚀	36916	1665	19	2024-12-07 13:01:09+00:00	Saturday	musique	N
20	GgSNvCY-AcY	My honest advice to someone who wants passive income	477146	10320	511	2024-12-27 13:00:17+00:00	Friday	actualité	P
21	3KLE_xwbGqg	19 Incredible Books to Read in 2025	210902	6095	275	2024-12-24 13:00:06+00:00	Tuesday	divertissement	P
22	WONRS7BLh4g	How To Actually Achieve Your Goals in 2025 (Evidence-Based)	2477862	87996	1145	2024-12-20 13:00:29+00:00	Friday	actualité	P
23	sYaSJpIGzu8	How to Change Your Life with Deep Work (My System)	563637	15434	369	2024-12-17 13:30:36+00:00	Tuesday	divertissement	P
24	W2afl0n8pUk	6 Habits to Make 2025 Your Best Year Yet	1008623	32662	666	2024-12-12 11:00:37+00:00	Thursday	divertissement	P
25	gP-C_utcBdY	This will change your life!	56725	2745	33	2024-12-08 13:00:47+00:00	Sunday	actualité	P
27	avih_knbVjA	Wanting 🍎	40769	2298	17	2024-12-05 13:00:08+00:00	Thursday	voyage	P
18	qh75NllzOlU	How to Actually Get in Shape in 2025 - My Evidence-Based Guic	126843	3455	206	2025-01-17 14:00:59+00:00	Friday	actualité	P