# **News Video Sentiment & Topic Analyzer**

A Python application designed for analyzing news videos using advanced **NLP techniques**. The project extracts speech from videos, transcribes it, translates to English (if needed), detects topics, and performs **sentiment analysis** using both rule-based and transformer-based methods.

## **🚀 Features**

* **Video-to-Audio Conversion**: Converts news videos into audio files using FFmpeg.
* **Speech-to-Text (Transcription)**: Transcribes speech using the **Faster Whisper** model.
* **Language Detection & Translation**: Automatically detects the spoken language and translates to English if necessary.
* **Topic Modeling**: Identifies main topics from transcripts using **LDA (Latent Dirichlet Allocation)**.
* **Sentiment Analysis**:  
  + **VADER** for rule-based sentiment scoring.
  + **Transformer model (nlptown/bert-base-multilingual-uncased-sentiment)** for deep learning sentiment prediction.
* **Overall Summary**: Aggregates sentiment across multiple videos to determine the general tone of the day’s news.

## **📂 Project Structure**

* app.py → Main script for processing news videos.
* transcriptions/ → Stores generated transcripts.
* outputs/ → Stores analysis results (topics, sentiment reports).
* requirements.txt → Dependencies list.

## **⚙️ Installation**

Clone the repository:

bash

CopyEdit

git clone https://github.com/Muhammad-Kamran-Hussain/NLP-Project.git

cd nlp-news-analysis

(Optional) Create a virtual environment:

bash

CopyEdit

python3 -m venv venv

source venv/bin/activate # On Windows: venv\Scripts\activate

Install the dependencies:

bash

CopyEdit

pip install -r requirements.txt

## **▶️ Usage**

1. Place your news videos (.mp4) in the project directory.

Run the application:  
  
 bash  
CopyEdit  
python app.py

1. Enter video paths when prompted.
2. The system will:  
   * Convert video → audio
   * Transcribe speech
   * Translate (if non-English)
   * Extract topics
   * Perform sentiment analysis
   * Generate an **overall sentiment summary**

## **📦 Dependencies**

* **FFmpeg** → Video-to-audio conversion
* **Faster Whisper** → Speech-to-text transcription
* **Transformers (HuggingFace)** → Sentiment analysis (BERT model)
* **NLTK (VADER)** → Rule-based sentiment scoring
* **SentenceTransformers & Sklearn** → Topic modeling & clustering
* **Googletrans** → Translation
* **Matplotlib** → Visualization (optional)

## **📊 Example Output**

* **Detected Language**: Urdu
* **Translated Transcript**: "The government announced new policies today…"
* **Topics**: government, policy, economy
* **Sentiment (VADER)**: {pos: 0.23, neu: 0.60, neg: 0.17}
* **Sentiment (Transformer)**: positive
* **Overall Daily Sentiment**: Neutral

✅ This project showcases how **NLP + Speech Recognition** can be applied to automate **news analysis**, making it easier to understand the **tone, topics, and sentiment** of multilingual broadcasts.