

Classification for youtube video category using bag-of-word technique for Machine Learning

Introduction

In this project, we aimed to classify YouTube video content using machine learning techniques based on transcribed and translated audio data. The workflow involved multiple steps including data collection, audio processing, speech recognition, translation, and text-based classification. By combining tools such as yt-dlp, Whisper, and Deep Translator, the project demonstrates a practical approach to text classification using real-world multimedia data.

Dataset Overview

The key features of the dataset include from over ten thousand data :

- **Video_id (for download audio)**
- **Category**

Data processing

1. **Data Acquisition:** YouTube video audio files were downloaded using the yt-dlp library, based on the video IDs provided in the Kaggle dataset.
2. **Audio Transcription:** The downloaded audio files were transcribed into text using OpenAI's Whisper model, which converts speech into text with high accuracy.
3. **Text Translation:** The transcribed text was then translated into English using the deep-translator library to maintain consistency in language across all samples.
4. **Text Vectorization:** The translated text was converted into numerical form using the Bag of Words technique, preparing it for input into the classification model.
5. **Model Training:** The processed text data was used to train a classification model to predict the video category based on linguistic patterns.

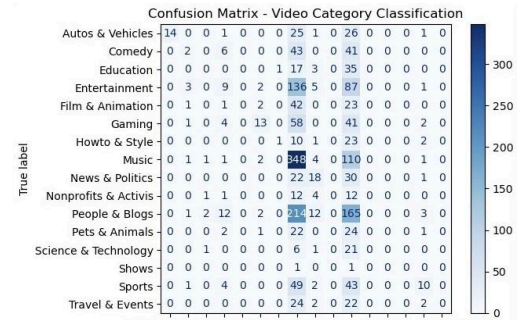
Graphs Results and Discussion

The accuracy of the model was evaluated by comparing the predicted categories to the actual categories of the YouTube videos. The comparison between the model's predicted output and the true labels demonstrates the model's classification effectiveness.

Normal Dataframe

Accuracy: 30%

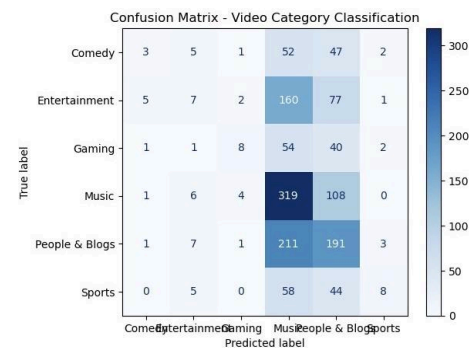
- Due to the unbalanced sample, it leads to an imbalance training, so the overall model have a low accuracy.



Filtered Dataframe with categories greater than 500

Accuracy: 37%

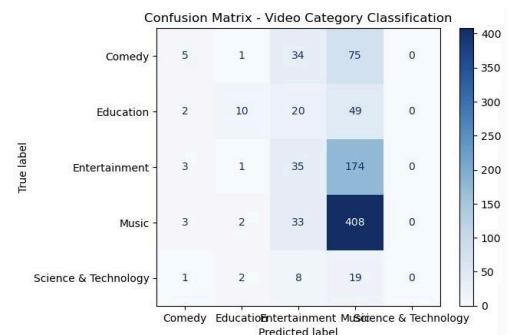
- In this model, we filtered only the categories that have more than 500 samples. This helps improve the accuracy.



Filtered Dataframe with some selected categories

Accuracy: 51%

- In this model, we've tried and selected the categories that give us the most accurate results. However, it leads into less class than the original datasets



Conclusion

This project used machine learning to classify YouTube videos based on transcribed and translated audio. Tools like yt-dlp, Whisper for transcription, and Deep Translator for translation were combined to process the data. The bag-of-words technique was applied to convert text into numerical data for classification. The model showed that linguistic patterns in the text were strong indicators of video categories, demonstrating the potential of multimedia data for classification tasks.