PYTHON PROGRAMMING PROJECT

YouTube Multiview Video Views Predictor

INTRODUCTION

With the increasing consumption of online gaming content on platforms like YouTube, predicting the potential popularity of a gaming video has become a useful tool for content creators. The YouTube Multiview Video Views Predictor is a machine learning-based application designed to estimate the number of views a gaming video might receive based on various features such as likes, dislikes, comments, duration, category, and multiview type.

This project simulates a realistic dataset inspired by YouTube video metadata and uses a regression model to make accurate predictions. It also includes an interactive web application built with Streamlit that allows users to input video attributes and instantly see predicted views.

2. Objective

The main goal of this project is to:

- Build a machine learning model that predicts video view counts.
- Create a synthetic dataset resembling real-world YouTube gaming data.
- Train and evaluate a regression model.
- Develop a user-friendly web app interface for real-time predictions.

3. Dataset Overview

As there was no direct access to the actual YouTube Multiview Video Games dataset, a synthetic dataset was created using Python libraries such as NumPy and Pandas. The dataset consists of 500 records with the following features:

- likes: Number of likes on the video
- dislikes: Number of dislikes
- comments: Total comments
- duration: Length of the video in minutes
- category: Game genre (e.g., Action, Puzzle, Sports, Strategy)
- multiview_type: The type of visual presentation (Single View, Split View, Third-person)
- views: Target variable representing the number of views

Views were calculated using a synthetic formula combining the other features and random noise to simulate realistic behavior.

4. Data Preprocessing

Before feeding the data into the model, the following steps were carried out:

- One-hot encoding of categorical features (category, multiview_type).
- Train-test split with 80% data used for training and 20% for testing.
- No missing values existed in the synthetic dataset, ensuring smooth preprocessing.

5. Model Building

A Random Forest Regressor from the scikit-learn library was chosen due to its robustness and ability to handle non-linear relationships.

Model Evaluation:

- Root Mean Square Error (RMSE) and R-squared score (R^2) were used as evaluation metrics.
- The model achieved high accuracy on the test set, showing its effectiveness at predicting view counts.

The trained model was saved using joblib for later integration with the web app.

6. Streamlit Web App

A simple yet functional web app was developed using Streamlit. The app allows users to input various video features using sliders and dropdowns. On clicking the "Predict Views" button, the app loads the trained model and displays the estimated view count.
Features of the Web App:

- Clean and intuitive UI
- Real-time predictions
- Sliders for numeric inputs (likes, dislikes, comments, duration)
- Dropdowns for category and multiview type
- Instant output display using st.success()

7. Project Structure

```
youtube-views-predictor/
| | -- app.py
| | -- yt_multiview_model.pkl
| | -- requirements.txt
| | -- README.md
```

8. Future Improvements

While this project achieves its current objectives, the following improvements are possible:

- Incorporating real YouTube data via API integration.
- Adding more features such as upload day/time, channel subscribers, etc.
- Using deep learning models for enhanced performance.
- Adding visualizations inside the app.

9. Conclusion

This project successfully demonstrates how machine learning can be applied to predict the performance of YouTube gaming videos. By simulating a realistic dataset and building a Streamlit web interface, this project presents a full pipeline from data creation to model deployment. The predictor is not only an academic exercise but also a potential prototype for gaming creators seeking insight into content performance.