CONTENT MONETIZATION MODELER

Introduction

The rapid growth of digital content platforms, particularly YouTube, has created new opportunities for creators and media companies to monetize their content. Accurately predicting ad revenue is critical for optimizing content strategy, planning uploads, and maximizing earnings. This project develops a **regression-based monetization model** that estimates YouTube revenue using video performance metrics and contextual features, and deploys the results in an **interactive Streamlit application**

Problem Statement

As video creators and media companies increasingly rely on platforms like YouTube for revenue, predicting potential ad earnings has become essential for business planning and content strategy. Without reliable forecasting, creators cannot accurately determine which content drives the highest returns, and media companies face uncertainty in revenue projections.

The objective of this project is to build a Linear Regression model capable of estimating YouTube ad revenue for individual videos based on performance metrics (views, likes, comments, watch time, subscribers) and contextual features (category, device, country). The results are deployed in a Streamlit web application, enabling users to interactively predict revenue and gain insights into the factors influencing monetization.

Objectives

- Estimate YouTube ad revenue accurately using regression models.
- Provide insights into the factors driving revenue.
- Develop a tool that supports creators, advertisers, and businesses.
- Deliver a user-friendly application for testing and predictions.

Dataset Description

- **Name**: YouTube Monetization Modeler
- **Format**: CSV (~122,000 rows)
- **Target Variable**: ad_revenue_usd

- **Features**: Video engagement metrics (views, likes, comments), watch time, video length, subscriber count, category, device, country.

Exploratory Data Analysis (EDA)

- Distribution of revenue, views, and engagement.
- Correlation analysis to identify strong predictors.
- Outlier detection in revenue and views.
- Category-wise and country-wise revenue trends.

Data Preprocessing & Feature Engineering

- Handled ~5% missing values in critical fields.
- Removed ~2% duplicates.
- Encoded categorical variables (category, device, country).
- Created new features such as **engagement rate = (likes + comments) / views**.
- Normalized features to improve model performance.

Model Building & Selection

Five regression models

1. **Linear Regression**
2. **Ridge Regression**
3. **Lasso Regression**
4. **Random Forest Regressor**
5. **XGBoost Regressor**

Model Evaluation Metrics

Models were compared using:

- **R^2 Score** – Goodness of fit.
- **Root Mean Squared Error (RMSE)** – Error magnitude.
- **Mean Absolute Error (MAE)** – Average prediction deviation.

The model with the best trade-off between accuracy and generalization was selected.

Results & Insights

- The final regression model achieved strong predictive accuracy.

- Key revenue drivers identified: **views, watch time, engagement rate, and subscribers**.
- Category and country also influenced monetization significantly.
- A cleaned dataset and trained model were produced for reuse.

Streamlit App Development

- Built an interactive web application for:
    - Uploading input data or entering video stats manually.
    - Predicting ad revenue dynamically. ◦ Visualizing revenue trends by category, device, or geography.

Business Use Cases

- **Content Strategy Optimization** – Identify highperforming content types.
- **Revenue Forecasting** – Support financial planning for creators & media firms.
- **Creator Support Tools** – Embed as part of analytics dashboards.

- **Ad Campaign Planning** – Estimate ROI before investing in campaigns.

Conclusion

The Content Monetization Modeler effectively predicts YouTube ad revenue using video performance and contextual features.

It provides actionable insights into key revenue drivers and supports creators, advertisers, and media companies in decision-making.

The interactive Streamlit app enables dynamic revenue predictions, making content strategy and monetization planning more data-driven.