YouTube Ad Revenue Prediction

by Kanish Midhun K

Problem Statement

Creators and marketers struggle to estimate YouTube ad revenue accurately. Ad earnings depend on multiple factors — views, engagement rate, watch time, and demographics. This project uses machine learning to predict revenue efficiently.

Dataset Overview

- Data collected from YouTube analytics and video statistics API.
- Features include: Views, Likes, Comments, Watch Time, Engagement Rate, Video Length, and Subscribers.
- Target variable: Ad Revenue (USD).
- Cleaned missing values using median and normalized numerical features.

Exploratory Data Analysis

- Near-zero skewness → symmetric distribution
- Negative kurtosis → platykurtic (flatter than normal)
- Indicates no heavy tails or extreme outliers

Model Performance Comparison

Model	MAE	RMSE	R^2	Time(s)
Lasso	3.10	13.47	0.9526	0.20
Ridge	3.11	13.48	0.9525	0.11
Linear Regression	3.11	13.48	0.9525	0.20
Gradient Boosting	4.06	13.59	0.9518	12.94
Random Forest	3.65	13.94	0.9492	10.93
Decision Tree	5.43	19.90	0.8966	1.39

Streamlit Web App

- ✅ User can input values manually or paste a YouTube link.
- ✅ Fetches video details using YouTube Data API.
- ✅ Predicts ad revenue instantly using trained ML model.
- ✅ Allows manual override for country, category, and device type.

Results & Insights

- Lasso Regression was chosen as the final model.
- $R^2 = 0.95$ → strong model accuracy.
- RMSE ≈ 13 → prediction error around ±$13.
- Random Forest and Boosting added interpretability for feature importance.

Conclusion

✅ Built an end-to-end ML app for ad revenue prediction.
✅ Used regression techniques and feature engineering.
✅ Deployed model via Streamlit with YouTube API integration.