**Content Monetization Modeler**

**Project Overview**

In the rapidly evolving creator economy, understanding how different factors influence YouTube ad revenue is vital for content creators and media companies.  
This project focuses on building a **Linear Regression-based predictive model** that estimates **YouTube ad revenue** for individual videos based on various performance and contextual metrics.

The model will be integrated into a **Streamlit web application**, allowing users to interactively input video metrics and predict expected ad revenue

**Objectives**

* Predict **ad\_revenue\_usd** using regression techniques.
* Understand which factors most strongly influence revenue.
* Build a **Streamlit app** for interactive predictions and insights visualization.
* Provide data-driven recommendations for **content strategy optimization** and **revenue forecasting**.
* **Dataset Information**  
  **Format:** CSV  
  **Size:** ~122,000 rows  
  **Source:** Synthetic (for educational purposes)  
  **Target Variable:** ad\_revenue\_usd
* **Columns Overview**

| **Column** | **Description** |
| --- | --- |
| video\_id | Unique video identifier |
| date | Upload/reporting date |
| views | Total views |
| likes | Total likes |
| comments | Total comments |
| watch\_time\_minutes | Total watch time in minutes |
| video\_length\_minutes | Length of the video |
| subscribers | Channel’s subscriber count |
| category | Video category (e.g., Entertainment, Education) |
| device | Device type (e.g., Mobile, Desktop) |
| country | Viewer’s country |
| ad\_revenue\_usd | Ad revenue in USD (Target) |

**Exploratory Data Analysis (EDA)**

Perform a comprehensive EDA to uncover patterns and insights.

**Key Analyses**

* Distribution of target variable ad\_revenue\_usd.
* Correlation heatmap to identify relationships.
* Boxplots and scatterplots for detecting outliers.
* Category-wise revenue comparison (e.g., revenue by category, country, or device).

**Model Building**

**Models to Experiment With**

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

**Workflow**

1. Split data into **train/test** sets (e.g., 80/20).
2. Train each model and tune hyperparameters.
3. Evaluate performance using:
   * **R² Score**
   * **RMSE (Root Mean Squared Error)**

**Insights and Interpretation**

* **Views** and **Watch Time** are the strongest predictors of revenue.
* **Engagement Rate** positively correlates with higher ad revenue.
* Certain **categories (like Technology, Entertainment)** yield better monetization rates.
* **Country** and **device** types also affect revenue due to ad pricing differences

**Streamlit Web App**

**Features**

* Input fields for user metrics (views, likes, comments, etc.).
* Predicts expected **ad\_revenue\_usd** using trained model.
* Displays model insights and feature importance.
* Includes basic EDA visualizations.

**Example Layout**

|---------------------------------------|

| YouTube Ad Revenue Predictor |

|---------------------------------------|

| [Enter Views] |

| [Enter Likes] |

| [Enter Comments] |

| [Enter Watch Time] |

| [Select Category / Device / Country] |

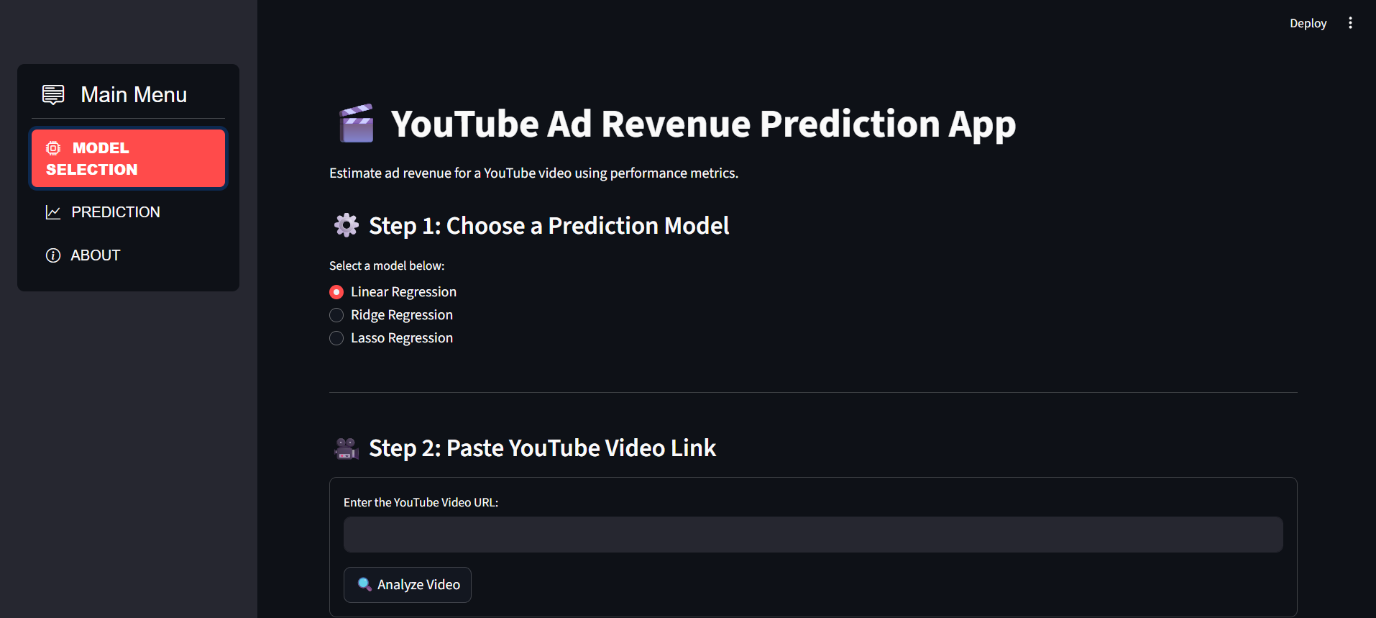
| [Predict Button] |

| Predicted Revenue: $XXX.XX |

|---------------------------------------|

| [Visualizations Section] |

|---------------------------------------|

**Project Deliverables**

1. **Jupyter Notebook / Python Script**
   * EDA
   * Preprocessing
   * Model Training
   * Evaluation
   * Insights
2. **Streamlit App**
   * Interactive prediction interface
   * Visualization of key insights
3. **README.md**
   * Overview
   * Setup instructions
   * How to run the notebook and app

**Tech Stack**

* **Programming Language:** Python
* **Libraries:** Pandas, NumPy, Scikit-learn, Seaborn, Streamlit,pickle, googleapiclient.discovery ,request,isodate, urllib.parse,PIL.

**Evaluation Metrics**

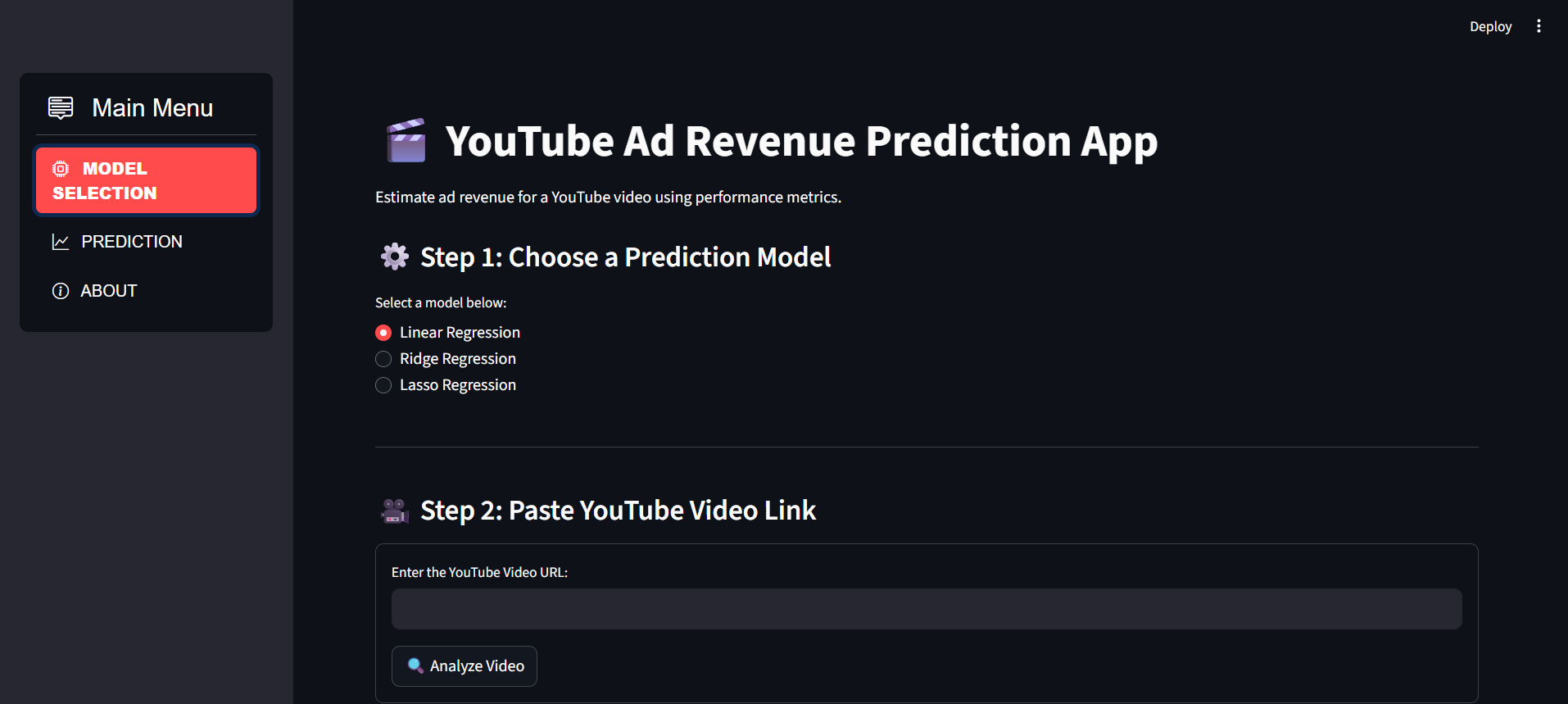
* **R² Score**
* **Root Mean Squared Error (RMSE)**
* **Mean Absolute Error (MAE)**
* **Code Quality & Documentation**
* **EDA & Insight Quality**
* **App Functionality and Usability**

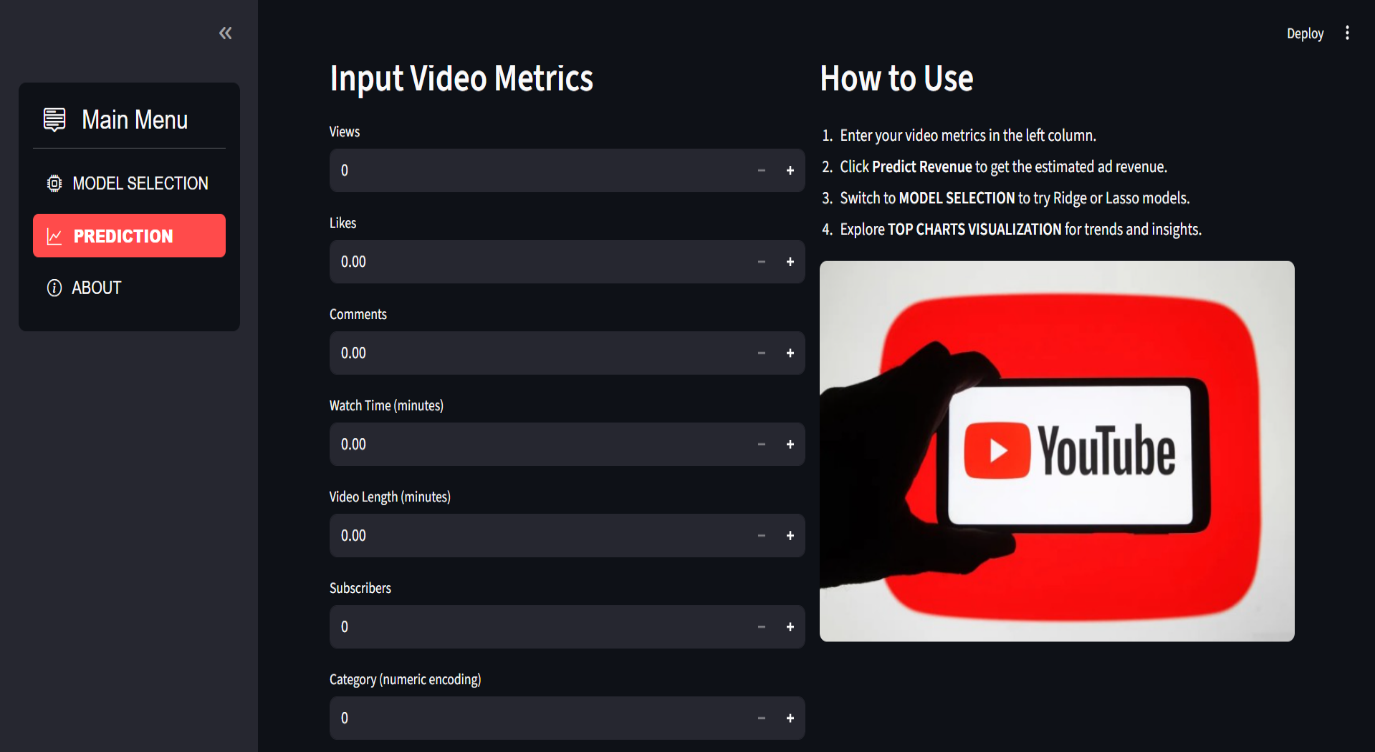
**Project Flow Summary**

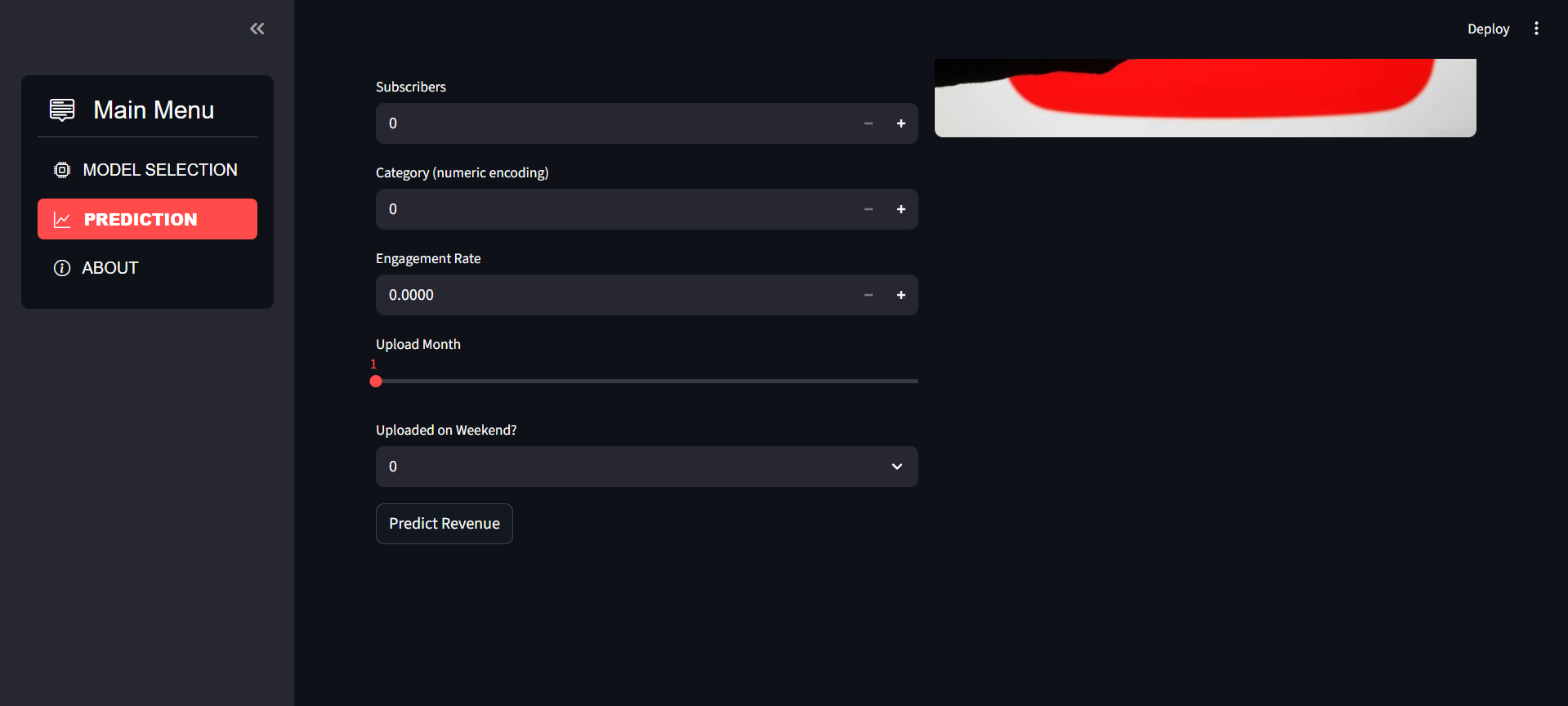
Data Loading → EDA → Preprocessing → Feature Engineering → Model Training → Evaluation → Insights → Streamlit App Deployment

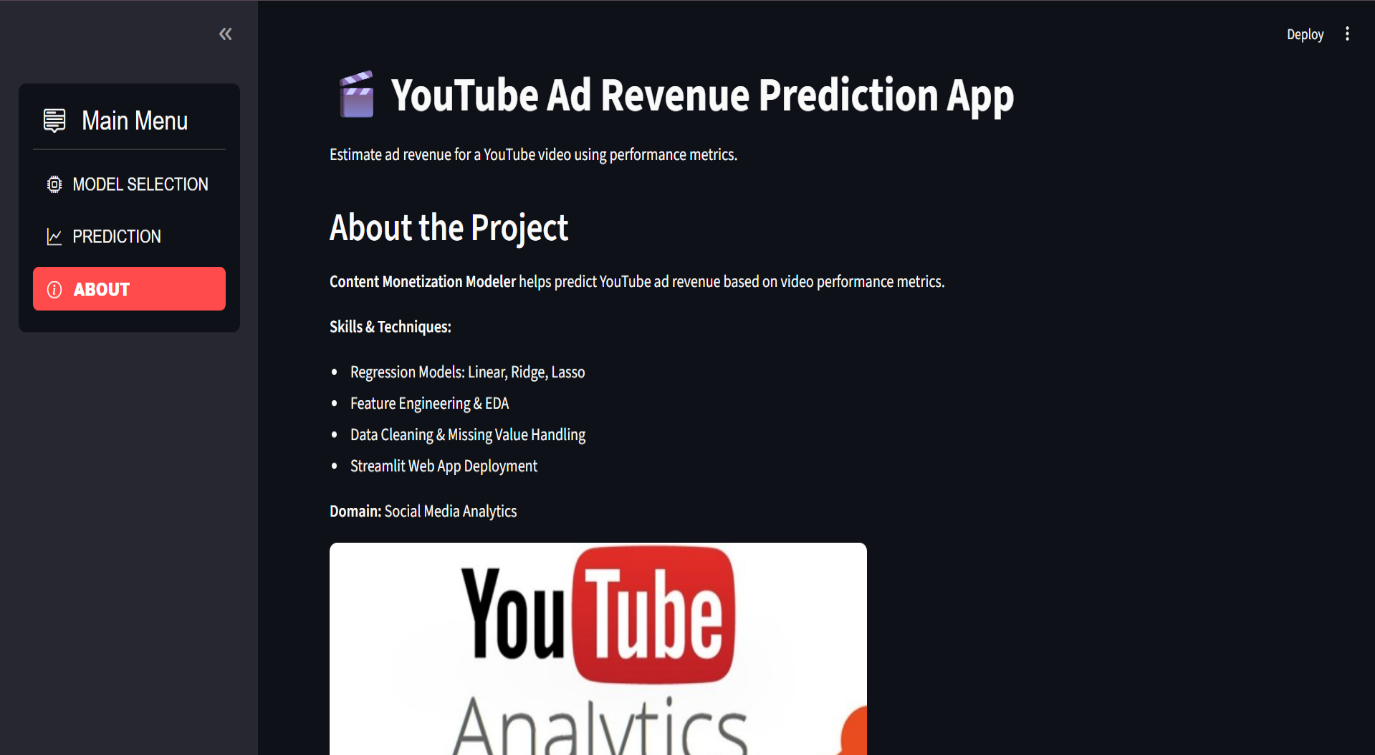
**Future Improvements**

* Include **time-based trends** (upload frequency, seasonal revenue).
* Integrate **YouTube API** for real-world testing.
* Deploy app on **Streamlit Cloud / Hugging Face Spaces**.
* Use **Deep Learning** (e.g., DNN Regressor) for complex nonlinear patterns.

**ScreenShot**







**Conclusion**

This project successfully developed a **Linear Regression model** capable of predicting YouTube ad revenue based on video performance and contextual features such as views, likes, comments, engagement rate, watch time, and video length. By analyzing these metrics, the model helps creators and media companies estimate potential earnings and make data-driven decisions for optimizing content strategies.

The model demonstrated that engagement-related factors — particularly **views**, **watch time**, and **engagement rate** — have the strongest influence on ad revenue. While Linear Regression provides a clear and interpretable baseline for prediction, future improvements could involve testing more advanced machine learning algorithms (e.g., Random Forest, Gradient Boosting, or Neural Networks) to capture nonlinear relationships between features and revenue.