Project Title:

Predicting Success of Netflix Shows using Supervised Learning Techniques

---

1. Background & Business Objective

Netflix invests heavily in original shows. However, not all shows achieve expected engagement or popularity. The Content Strategy team wants to build a predictive model that forecasts whether a new show will be successful based on historical data—before it's even released.

A “successful show” is defined as a show that ranks in the Top 10 globally within the first 30 days of release.

2. Business Problem

Netflix has asked for a data-driven solution to:

- Predict the potential success of upcoming shows.
- Identify the key features that contribute to a show’s success (e.g., genre, cast, language, release timing).
- Use the insights to support content investment and recommendation strategy.

3. Goal

Build a classification model to predict show success (Success = 1, Not Success = 0). Supplement this with:

- Statistical insights from the dataset (EDA).
- A Power BI dashboard for the business team.
- Model evaluation metrics.
- Ethical considerations and error risk assessment.

4. Scope of Work

In Scope:

- Perform exploratory data analysis (EDA) and feature engineering.
- Apply statistical analysis (correlation, distribution, hypothesis testing).
- Build a supervised classification model (Naive Bayes or Decision Tree).
- Evaluate model using accuracy, precision, recall, F1 score, confusion matrix.
- Deploy insights using a Power BI dashboard.
- Document assumptions, limitations and ethical risks.

5. Tools & Technologies

- Python (Pandas, NumPy, SciPy, Seaborn, Matplotlib, Sklearn)
- Jupyter Notebook
- Statistical principles

6. Deliverables

- Cleaned and documented dataset.
- EDA and insights notebook (Jupyter).
- Final ML model with explanation and metrics.
- Power BI Dashboard.
- GitHub repository with README, Project Proposal, and PACE document.
- Executive Summary for stakeholders.

Data Analyst: Shijin Ramesh