Capstone Project 2 - Proposal

Problem Statement - Predict the average user rating of a movie based on its budget, genre, cast popularity, and release year. Using a dataset of 10000+ movies from TMDB. The goal is to build a model with at least 70% accuracy to support strategic content investment decisions, by the end of the third quarter.

Context - Streaming platforms and production studios are constantly seeking to understand what makes a movie successful. By analyzing features from large public dataset like TMDB, we can explore which factors are related to audience ratings.

Criteria For Success

- A clean, wrangled dataset ready for modeling.
- Identification of strong predictors of high ratings.

Scope of Solution Space

- Focus on exploratory data analysis, preprocessing, and feature engineering.
- Target variable : vote_average

Constraints

- No data about user behavior (example: Their personal views, or their sentiment)
- Potential imbalance in rating distribution

Stakeholders

Streaming platform data science team
Marketing Team & Studio Content Acquisitions team

Data Sources

- TMDB 10000+ movie_ratings (https://www.kaggle.com/datasets/kunalduttads/tmdb-top-10000-10k-movies-dataset)