

Model Explanation

We first had to clean our data to be able to use it to form a model.

This is done by removing irrelevant columns such as ID, Title, summary, and tagline. We then converted languages into one hot encoding, such that we can represent each language as a column of 0 and 1 of its own.

We then proceeded to one hot encoding of the genres too. As genres are saved as a string, we had to first convert it to an array and then change each genre type to be a different column of its own, with 0 being not of this genre, and 1 being of this genre.

Accounting for leap years, we converted the dates into a numerical value post a certain amount of time.

We then tested regression models, finding Boosted Decision Regression to perform the best. We ran the tests multiple times with different training/testing sets to try to avoid overfitting. We then tested combinations of hyperparameters, but found the defaults to be most accurate.

Recommendation

Title: The Uprising

Cast: Michelle Obama, Christopher Eisgruber, Tom Cruise

Genre: Thriller, Mystery, Horror

Runtime: 96.74 minutes

Budget: \$3,553,612

Release Date: 10/7/2024

- We came across this recommendation through testing of our model with generated inputs influenced by analysis of the dataset.