INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration With

UNIVERSITY OF WESTMINSTER

Cine Cast: Box Office Forecasting Using Hybrid ML Model & XAI.

A Thesis By

Mr. Amjad Umerthambi

Supervised By

Mr. Saadh Jawwadh

Submitted in partial fulfilment of the requirements for the BEng Software Engineering degree at the University of Westminster.

April 2025

ABSTRACT

Cinema industry is a form of entertainment that makes the audiences dive into the movie creators world. However, behind the scenes there are a lot of stakeholders which is the reason why it's a risky business yet a multi billion dollar industry. While a movie can earn from selling their movies to different online streaming services the main source of its revenue comes from its box office collection.. In order to make sure that everyone (audiences, producers, movie exhibitors etc) gain something an effective box office prediction is vital.

A stacked ML model will be developed where initially three models which are decision tree, XGBoost and random forest are trained. Then it will then be used to combine the predictions of each model which will result in a better accurate result using a meta model random forest regressor. Moreover, for the transparency of the model XAI method SHAP will be integrated and provide SHAP values and visualizations.

The chosen model evaluation metrics were R squared, MAE and MSE. According to the model testing conducted, out of the three base models trained and developed random forest has been the best performing base model by achieving a result of 0.7235 for R squared, 1.0431 for MSE and 0.7311 for MAE. The lowest performing base model was decision tree by achieving a result of 0.5994 for R squared, 1.5114 for MSE and 0.8802 for MAE. Whereas the meta-model also achieved a better result of 0.6745 for R squared, 1.2278 for MSE and 0.7702 for MAE.

Keywords: Machine learning, Forecasting system, Explainable artificial intelligence, Meta-model

Subject Descriptors

- Computing methodologies→ Machine learning → Machine learning algorithm
- Computing methodologies→ Machine learning → Cross-validation
- Computing methodologies→ Machine learning → Machine learning approaches → Feature selection