VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science Engineering and Information Systems

Winter Semester 2024-2025

SWE2020 – Software Metrics - EPJ

Faculty: Prof. Rahamathunnisa U

Slot: B1

REVIEW 1

MOVIE BOX OFFICE PREDICTION USING SOFTWARE METRICS

Team Members:

Name	Reg No.
Thanigaivel V	22MIS0040
Affan Rahmathullah K	22MIS0064
Mohammed Affan M	22MIS0164

Problem Definition

Predicting a movie's box office revenue is a challenging task due to the large number of influencing factors, including budget, cast, genre, production house, and audience sentiment. Filmmakers, producers, and distributors can benefit from accurate predictions for better decision-making and risk management. This project aims to develop a predictive model that estimates a movie's box office revenue based on these factors, ensuring accuracy and performance through the application of various software metrics.

Introduction

The entertainment industry relies heavily on data-driven decisions to maximize the success of movies. Box office revenue is a key indicator of a film's performance and is influenced by numerous factors such as production budget, cast popularity, release date, and critical reviews. Accurately predicting box office revenue can help stakeholders make informed decisions regarding marketing strategies, resource allocation, and financial planning. In this project, we aim to build a machine learning-based predictive model to estimate box office revenue. The model development process will be guided by software metrics to ensure high performance, maintainability, and efficiency.

Tool Description:

Orange is an open-source data analysis and visualization tool designed for both beginners and advanced users. It offers a user-friendly, drag-and-drop interface for building machine learning workflows without needing to write code. Orange includes a variety of machine learning algorithms for classification, regression, clustering, and more, as well as tools for data preprocessing and visualization. It is highly extensible, allowing users to integrate Python scripting for more advanced tasks. Ideal for exploring data, building predictive models, and visualizing results, Orange is widely used in fields like business analytics, research, and education.

Literature Survey:

S. No	Title	Author	Advantages	Disadvantages	Tool/Method	Result
1	Box Office Sales and Social Media: A Cross-Platform Comparison of Predictive Ability and Mechanisms	Matthias Bogaert, Michel Ballings, Dirk Van den Poel, Asil Ozteki	- Social media data improves box office prediction accuracy. - Facebook data outperforms Twitter data.	- Twitter data has lower predictive power due to lower source credibility.	- Machine Learning algorithms: Random Forest, Gradient Boosting, Neural Networks, etc.	Facebook-based models improve prediction accuracy by 12-47%.
2	Predicting Movie Box-Office Revenues Using Deep Neural Networks	Yao Zhou, Lei Zhang, Zhang Yi	- Uses a multimodal deep neural network (DNN) incorporating movie poster features for revenue prediction.	- Model complexity increases due to multimodal learning. - High dependence on quality and availability of movie posters.	- Deep learning techniques (CNN, DNN) - IMDB dataset - Feature extraction from movie posters	Multimodal DNN outperforms traditional models. - Poster-based features significantly improve prediction accuracy.
3	Pre-Production Box-Office Success Quotient Forecasting	Usman Ahmed, Humaira Waqas, Muhammad Tanvir Afzal	- Forecasts box-office success at the pre-production stage.	- Accuracy depends on the completeness and quality of historical data. - The model does not incorporate real-time social media trends.	Machine Learning classifiers: Support Vector Machine (SVM), Random Forest, Gradient Boosting, Extreme Gradient Boosting.	- Achieves 85% accuracy, outperforming previous models.
4	A Movie Box Office Revenue Prediction Model Based on Deep Multimodal Features	Canaan Tinotenda Madongo, Tang Zhongjun	Uses DMFCNN, ResNet50, and metadata for better accuracy.	High computational cost, dataset limitations.	ResNet50, DNN, Transfer Learning	59.3% accuracy, 93.2% within one category

5	Movie Box-Office Revenue Prediction Using Deep Features from Trailers	Kumar P et al.	Uses RNNs to extract deep features from trailers.	Requires high-quality video datasets, complex preprocessing.	LSTM, CNN, Feature Extraction	Improves revenue prediction with trailer-based features
6	Predictive Model Performance: Offline and Online Evaluations	Zhang Y, Liu M	Compares offline and online evaluation methods for revenue models.	Online evaluations depend on user interactions, bias issues.	Regression models, Neural Networks, A/B testing	Demonstrates model effectiveness in real-world applications
7	Predictive Metrics: Transforming Engineering Productivity and Software Quality	Saumen Biswas	Uses predictive metrics like cycle time, LTFC, and defect density for software quality improvement.	High dependency on data accuracy and team adoption challenges.	CI/CD pipelines, automated reporting, machine learning	Enhances software productivity, aligns metrics with business goals
8	Software Reliability: Development of Software Defect Prediction Models Using Advanced Techniques	Mayur Jagtap, Praveen Katragadda, Pooja Satelkar	Uses Neural Networks and Fuzzy Logic for defect prediction, improving reliability estimation.	Computationally intensive, dataset-sensitive performance.	ANN, Fuzzy Logic, MATLAB for modeling	85%+ prediction accuracy, better risk management
9	Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models	Vagelis Plevris, German Solorzano, Nikolaos Bakas, Mohamed El Amine Ben Seghier	Compares 14 regression metrics, evaluates error measures for ML models.	Some metrics show inconsistencies in real-world scenarios.	Monte Carlo simulations, regression models, error metrics	Identifies most effective performance metrics for ML models

10	Classification Model Evaluation Metrics	Željko Đ. Vujović	Analyzes classification metrics and models (BayesNet, NaiveBayes, MLP, J48).	Models performed poorly; data preprocessing needed.	WEKA, evaluation metrics (Accuracy, Kappa, MAE, RMSE, etc.).	Models failed; preprocessing required for better results.

Conclusion

Predicting box office revenue is a valuable yet challenging task due to the numerous variables influencing a movie's success. This project aims to develop an accurate and efficient predictive model while emphasizing software metrics such as performance, maintainability, and complexity. By using tools like SonarQube to monitor code quality and libraries such as Matplotlib for performance visualization, we ensure a structured development process.

The integration of software metrics will help us evaluate and optimize the model at each stage, resulting in a solution that is not only accurate but also scalable and maintainable. Ultimately, this approach will demonstrate how predictive modelling and software metrics can complement each other to deliver a high-quality solution for real-world business challenges in the entertainment industry.