PREDICTING OSCAR NOMINEES – PROGRESS UPDATE

Hui, Macarious
Wu, Victor
Zhang, Tianyi

Project Setup and Literature Review

- Define project scope and objectives
- Conduct literature review and review/gather datasets from existing sources
- Created shared folder, GitHub Repository, Teams group chats and shared Notebooks

Data Preparation

• Performed EDA on the Golden Globes dataset. • Cleaned datasets by removing records related to television and non-Oscar nominated awards. • Converted Golden Globes categories into features. • Analyzed the relevancy of Golden Globes categories in predicting Oscar nominations.	• Merged Golden Globe datasets and Oscar rewards dataset by implementing fuzzy matching based on film titles; verified manually. • Resolved the no common key problem due to movie title mismatch • Differentiated remake versions of the film by comparing metadata.	• . Analyzed the IMDb dataset and The Movies dataset from Kaggle. • Filter out non-relevant data from datasets. • Created a master dataset to be used for training containing features and labels. • Add various new features based on the movie metadata.

Model Development

• Prepare movie names to be matched with our Master dataset.	• Clean the SAG dataset similar to the Golden Globe Dataset. • Join datasets with our Master Dataset.	• Analyze repeated production company names and reorganize data accordingly. • Add actor/actresses/director related features to dataset.

Model Testing and Optimization

• Train, test, and optimize machine learning models for K-nearest neighbors and logistic regression.	• Create multi-label decision trees and random forests.	• Train, test, and optimize a neural network for multi-label classification using Sigmoid activation function in the output layer.

Documentation and Reporting

- Documentation and reporting.