AML 2203 2, FINAL PROJECT PROPOSAL

PROJECT TITLE	Machine Learning Model to Predict Movies's Popularity
SUBMITTED BY	Mary Gomez
DATE OF SUBMISSION	11/22/2023

First Name	Last Name	Student Number
Mary	Gomez	C0891136
Rehan	Khokhar	C0896278
Andres	Correal	C0872634
Viki	Patel	C0906295
Bhavya	Vadher	C0894977
Prashant	Bhattarai	C0898956

MOTIVATION	The size of the global movies and entertainment business was around USD 94.45 Billions in 2022 and it is expected to grow 80% by 2030. Therefore, there is a huge opportunity for those initiatives that attend this industry needs. We believe that having a tool able to predict movies' popularity would attend a genuine business need allowing industry stakeholders a better decision-making regarding investing or not into a particular movie. Therefore, the users of these tools could be studios during content creation along with streaming platforms to negotiate licensing fees for those movies with higher predicted popularity.
METHOD	• Most of the data will be taken from TMDB API (https://developer.themoviedb.org). However, we will explore the option of combining data from various sources • We will use primarily nltk, scikit-learn and Hugging Face to conduct our experiments. • We aim to train a supervised machine learning model able to predict the popularity of a movie.
INTENDED EXPERIMENTS	• Use a feature selection technique to determine the best predictors • Feature Engineer new columns • From the textual data, use the top words, bigrams, trigrams as features of the model. • We want to compare vectorization using TF-IDF vs using NLP (Natural Language Processing) embeddings from Hugging Face. • To test and compare several machine learning models • To optimize the model performance by doing hyperparameter tunning using GridSearch technique.
PLANNING AND MILESTONES	• Get the best predictors - Viki • Feature Engineer new columns – Andres • Choose bigrams and trigrams as features of the model – Prashant • Compare different vectorization methods - Mary • Compare machine learning models- Rehan • Hyperparameter tunning - Bhavya • Reporting - Mary

References:

Zion Market Research (April 2023), Movies and Entertainment Market. Retrieved from: https://www.zionmarketresearch.com/report/global-movies-entertainment-market#:~:text=The%20global%20movies%20and%20entertainment,7.21%25%20between%202023%20and%202030.