**Final Project Deliverable 3**

**Angela Wilburn**

**Information Technology, Walsh College**

**IT 445 VI Spring 2024**

**Chris Heiden**

**May 20, 2024**

**Machine Learning Model**

The data analytics team at Wanted Brothers Studios developed a supervised machine learning model using logistic regression to predict whether movie reviews are positive or negative. The model was trained on a dataset of 25,000 movie reviews and sentiment labels from Mathew Walke's IMDB Movie Reviews for Binary Sentiment Analysis. It has an accuracy score of 0.8912, a precision score of 0.8805, a recall score of 0.9032, and an F1 Score of 0.9032. With this strong performance, the team moved on to the next phase.

In the next phase, the team used Aditya Pal, Abhilash Barigidad, and Abhijit Mustafi’s (2020) IMDb Movie Reviews Dataset, which contains 1,259,472 movie reviews. The model predicted 894,048 positive sentiment reviews and 365,424 negative reviews. These predictions will help the team analyze the audience's sentiments. The final step for the data analytics team was to review which movie characters tend to have higher and lower sentiments.

**Recommendations**

**Genre recommendation**

Biography, animation, music, drama, and sports are the movie genres with the highest user sentiment, and horror, mystery, sci-fi, fantasy, and thriller are the lowest. We recommend that Wanted Brothers Studios' marketing only invests in the top five genres: Biography, animation, music, drama, and sports.

**Country recommendation**

Mexico, Sweden, Italy, Australia, and Japan are the top countries movies were released in with the highest user sentiment, Mexico with a 0.909 sentiment rating (0 being negative and 1 being positive). Movies released in the Netherlands, Ireland, USA, South Korea, and France have the lowest user sentiment. We recommend that Wanted Brothers Studios' marketing expands to include movies produced and released in more of the top 5 countries with the highest sentiment, especially Mexico.

**Further Data Analysis and Investigation Recommendations**

The data team is proposing a project aimed at mining data from movie scripts, as well as collecting information about actors and directors from the top 20 and bottom movie sentiment ratings. The objective is to uncover potential correlations between movie themes, actors, and audience sentiment. Additionally, the data team has observed a steady decline in user sentiment over the years, dropping from 0.81 in 2000 to 0.71. This suggests that people have lower sentiment overall for movies now than they did in the early 2000s. Further investigation into this trend is recommended.

Further analysis of these user sentiment findings will ultimately assist us in producing movies that deeply resonate with our audience and maximize profitability. Furthermore, we advise exploring the years with high and low sentiment for further insights.

**Conclusion**

In conclusion, the data analytics team has provided several key recommendations based on our machine learning model and movie review sentiment data analysis. By implementing these data-driven recommendations, Wanted Brothers Studios can strategically focus its marketing resources on the genres, regions, and creative elements that data shows will maximize audience enthusiasm and viewership. The team's work provides an objective, quantified framework to guide decisions for maximum revenue potential.