Netflix Success Predictor Project Proposal

TO: Reed Hastings, Co-CEO, Netflix
FROM: Adelaida Dominguez (ad667), Andrea Siby (as3246), Kayla Runkel (kmr227)
CLASS CODE: 4741

March 17, 2024

1 Background

In today's entertainment landscape, the success of movies and TV shows on streaming platforms like Netflix depend on various factors ranging from genre, cast, budget, release date, and marketing buzz. Understanding these key drivers for success is crucial for producers, directors, and investors to make informed decisions. In this project, we plan to use machine learning algorithms to predict the success of Netflix movies and TV shows using views as a proxy for success.

Television and movies are central to modern entertainment, offering immersive storytelling and cultural reflection. We watch them for relaxation, education, and social connection. With a plethora of options available across many platforms, choosing what to watch is influenced by personal preferences, recommendations, and trends. As viewers, we navigate this diverse landscape to find narratives that resonate with us and reflect our own experiences, values, and connections. For directors, producers, and actors, achieving success entails a combination of factors such as compelling storytelling, engaging performances, effective marketing, and resonating with audience preferences.

Ultimately, the question is what contributes to the success of movies and television shows. Moreover, can we develop an accurate predictive model using movie/TV show features and viewing data to forecast the success of Netflix content offerings?

2 Importance

Movie success prediction can play a highly important role when deciding to proceed with the creation of the show. As mentioned by Lash and Lao [5], movie investment is a risky business where investors make a large investment upfront in the initial stages of production. Netflix's competitive strategy comes from the offering of exclusive content [6]. However, with the increase in the number of streaming platforms, Netflix has started venturing into investing billions of dollars to create its own shows. Exclusive original content distinguishes Netflix from rivals. This means the investment risk lies heavily on Netflix, wanting to create low-risk, successful shows. This increases the need for a good predictor of which show types perform well generally.

Our model would therefore give Netflix better ability to predict the success of a movie or show. In addition, we also recognise the presence of numerous small scale production companies and directors who are trying to break into the movie industry. Doing so will necessitate them bearing large amounts of financial burdens, considering how they are new to the industry and may not have adequate financial ability to produce a show on their own. Many such producers often take hefty loans while facing extensive risk of not producing a successful show. The presence of our model may provide better insights into which movies are likely successful, reducing their investment risk.

Additionally, accurate success prediction enables Netflix to prioritize potentially popular shows, optimize marketing spend, and reduce costly gambles on riskier projects. Ultimately, delivering more hits that resonate with audiences directly impacts subscriber growth and retention - the key to Netflix's revenue. Our model's predictive capabilities allow data-driven decisions around multi-million dollar productions, better catering to demand and maximizing returns.

3 Data

To build a robust success prediction model, we will leverage several datasets providing valuable information on Netflix's content offerings and performance. Our primary dataset comes directly from Netflix's "Global List" of popular titles, updated weekly since January 2021 [1]. As of the week of March 11, the dataset contains nearly 6,000 entries providing details such as the title of show or film, weekly hours viewed in that week, runtime, weekly views and whether the release of the show is staggered or not.

To complement this dataset, we aim to incorporate additional sources, including the Netflix Engagement Report [2] and two Kaggle datasets titled "Netflix TV Shows and Movies" [3] and "Netflix Shows" [4]. These supplementary datasets offer insights into factors like global availability, release dates, genres and audience ratings. Given our approach of combining various datasets, we anticipate the need for data aggregation and feature engineering techniques to address missing values and ensure consistent data alignment.

By merging these diverse datasets, we can gain a comprehensive understanding of the factors influencing the success of Netflix's content offerings. The combination of quantitative metrics, release details and audience engagement signals will enable us to identify key drivers of success and develop predictive models to inform content strategy.

Our approach involves applying regression methods to identify relationships between these features and the success metric. We plan to validate models using future time-series data releases, leveraging the weekly availability of the new Global List data. The large sample size and variety of features present in our compiled datasets make them well-suited for training accurate predictive models using regression and machine learning techniques. With a robust dataset spanning multiple years of Netflix content releases and performance metrics, we can effectively capture the evolving trends and audience preferences, ensuring our model's predictions remain relevant and actionable.

While the available data sets offer valuable insights, we acknowledge potential limitations and challenges. The data is constrained to releases within the past 5 years, though the rapid popularity of streaming platforms in the past decade still make this span highly valuable. Additionally, defining views as our success metric embeds some bias, as it assumes that viewership equates to success for a show or movie. It will be important to scrutinize this assumption while exploring the data to ensure it holds. Incorporating financial data like budgets and revenue figures could further add to this analysis but finding this comprehensive data has been a challenge.

Through this project, we aim to develop an accurate predictive model that can guide Netflix's multi-billion-dollar content investment decisions. With the insights from this project, Netflix can optimize their content strategy, better cater to audience demand and maximize returns. Our project has the potential to create significant value for the company and its stakeholders in this competitive environment. With an accurate success prediction tool, Netflix can better cater to audience demand, reduce costly gambles, and deliver more hit content that drives subscriber growth and retention – the key to sustainable revenue generation.

4 References

1. https://www.netflix.com/tudum/top10/most-popular?week=2023-12-03
2. https://about.netflix.com/en/news/what-we-watched-a-netflix-engagement-report
3. https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies?select=titles.csv
4. https://www.kaggle.com/datasets/shivamb/netflix-shows
5. https://www.researchgate.net/publication/278733924_Early_Predictions_of_Movie_Success_The_Who_What_and_When_of_Profitability
6. https://www.investopedia.com/articles/investing/060815/how-netflix-changing-tv-industry.asp

Link to our repository: https://github.com/adelaidad/ORIE4741FinalProject.git