NETFLIX ORIGINAL RELEASES AND STOCK PERFORMANCE

A study into the effect of Netflix original release on its stock performance

Contents

Business Understanding ... 2

Business Problem ... 2

Dataset Choice - Why Original Releases? ... 2

Recent Implication of Original Releases ... 2

Data Mining Solutions ... 3

Data Understanding ... 3

Identify Necessary Dataset ... 3

Why did we choose NASDAQ 100? ... 3

Data Preparation ... 4

"Movie and TV series" dataset ... 4

"*Nasdaq Index and Netflix Price*" dataset ... 4

Feature Engineering ... 5

Modeling ... 6

Evaluation ... 8

Deployment ... 8

Deployment of Results ... 8

Potential Deployment Issues ... 9

Ethical Consideration ... 9

Risks and Mitigations ... 10

Appendix A – Team Contributions ... 10

Appendix B – Figures ... 11

Section A Team 2 – Data Science Final Report

Name: Minh Pham, Renqi (Kevin) Chen, Tina Chen

Cecily Cai, Yingdong Yang, Kunal Nadkarni

Business Understanding

Business Problem

The business problem we have chosen is what effect does a good Netflix original release has on Netflix's financial performance. Originally, we have defined financial performance as subscriber count, as that metric directly correlates with how well Netflix is doing. However, as subscriber count data is not publicly available, we have chosen NFLX stock price over the years as a proxy.

Dataset Choice - Why Original Releases?

Original releases are very important to Netflix because it allows Netflix to stop relying on studio partners for content on its site. As Netflix matures and needs more content to capture new subscribers, it transforms from just being an entertainment hub to entertainment providers, which can be beneficial to provide a better customized experience for its subscribers. However, original releases are incredibly expensive – currently, Netflix spends around $5.21B on producing original content, a whopping 40% of total content spending in 2021. Netflix is betting big on original releases, and we want to find out whether that has been a good decision throughout the past decade.

Recent Implication of Original Releases

Netflix was struggling in 2021, with subscriber counts decreasing quarter-over-quarter, and international expansion has not seen adoption as much as domestically. In Q2 2021, Netflix lost nearly 400,000 subscribers. However, at the end of Q2, just a few weeks ago, Squid Game happened. Squid Game became the most popular show on Netflix within 7 days of release and drove an international storm of supporters. Netflix has gained 1.54 million new subscribers since Q2. Squid Game was a successful original release for Netflix, but are all of them a success?

Data Mining Solutions

We believe that using the models learned in Data Science class, we can determine the causation of Netflix's original releases on its financial performance. Since our business problem revolves around explaining said causation, and the fact that original releases have been Netflix's core business strategy in the past decade, we believe that data mining solution will get us close to the answer we were looking for.

Data Understanding

Identify Necessary Dataset

Looking at the question at hand, we decided to look for datasets that contain the following information:

1. Netflix's TV show and movie releases, their individual IMDB score, and release date
2. Netflix's stock price in the relevant period
3. NASDAQ performance in the same period

These data will allow us to determine how "good" an original release was, and its impact to NLFX stock price in the weeks ensuing the release.

Why did we choose NASDAQ 100?

Our final goal is to learn about Netflix's performance over the years, related to its original releases. It makes sense to compare NLFX stock performance with the NASDAQ database, so then we can learn whether Netflix is doing well compared to the market. We chose the NASDAQ 100 instead of the NASDAQ composite because we think the 100 represents Netflix's competition better than the entire NASDAQ repository.

Data Preparation

"Movie and TV series" dataset

It was difficult to find Netflix's original TV series with both date and ratings, so we looked for those two datasets respectively and then merged them. For TV dataset, it contains information included genre, dates; for another TV datasets, it contains rating (IMDb score). We merge those two datasets in order to have all the three columns: TV show names, IMDb score. We then joined TV dataset date with TV ratings on column 'type'. We know that 'type' is a primary key for both table and can identify each unique row. Next, we add week for each data set based on the exact release date for each TV show and movie. Now, we have all the columns we need in dataset for TV (tv.csv). Then, we change the 'date' from char into date in both datasets. And find the min/max date in order to see whether it is logical and reasonable to include. Next, we add the 'week' column based on the dates in two separates dataset since we need to have a corresponding 'week' with Nasdaq stock price, which we will discuss later. Lastly, we merge the above two datasets into a final single dataset, by used 'rbind' to union TV series and movie in order to get our cleaned data (tv_movies.csv).

For the missing value in "Netflix Movie and TV series" dataset, there are only 8 missing values for over 1000 data. We have considered two ways to deal with them: 1. Using KNN to impute the missing values; 2. We drop the 8 missing values. After discussion, we don't think given the data available, KNN will be good enough to predict the values of those empty cells. Hence, we dropped the 8 rows.

Now, we have the cleaned data set with four variables: Title, week, premiere (Date), IMDb; 1126 observations.

"*Nasdaq Index and Netflix Price*" dataset

We used the Netflix stock price, Nasdaq composite (stock market index that includes around 3000 companies), Nasdaq 100 (stock market index of 100 largest non-financial companies) historical data from Yahoo finance. The final dataset should include three columns: Date of every Monday starting 2012/12/31, cumulative week number (from 2012/12/31 to 2020/09/20), percentage changes of stock/index weekly.

When cleaning the data, we created two columns to show the percentage changes of the stock prices:

1. Percentage changes from week open (Monday) to week close (Friday).
2. Percentage changes from the previous week close (Friday) to week close (Friday).

Feature Engineering

Why did we choose Friday to Friday?

We choose the timeline of each week from Friday to Friday because that will account for all the after-hours trade and volatility in the market. Plus, it will be easier for us to clean the data since Yahoo Finance provides a pretty

Why and how did we use alpha (excess return)?

Alpha refers to the excess return earned on investment (in this case, the stock of Netflix: NASDAQ: NFLX). We chose to use this value as it takes away the movement of the general market (NASDAQ index), and thus helps us control variables such as general economic trend, industry specific changes, or regulation changes to some degree. Note that this does have some limitations as it is not a good controller for some other omitted variables.

We calculated alpha from the available data, using the difference between closing prices of each week, and then subtracting that from the Netflix price change. We further separated alpha into 6 categories, each with an increment of 2.5 as shown in the following table:

Alpha	<= -5	<= -2.5	<= 0	<= 2.5	<= 5	>5
Encoding	1	2	3	4	5	6

For example, as of right now, Netflix has an approximated alpha of 0.5 and would belong to category 4.

We chose the above scale by looking at the distribution of alpha values, as shown in Figure 1.

Why did we cluster IMDb score?

The next question we want to tackle is how to define whether a release is good or not? Using k-mean, we have clustered the Netflix original releases IMDBA scores into 4 categories: Excellent, Good, Normal, and Disappointed. In doing so, we can easily explain the variance of each of these clusters, and clustering will also make it easier for us to run our models. As

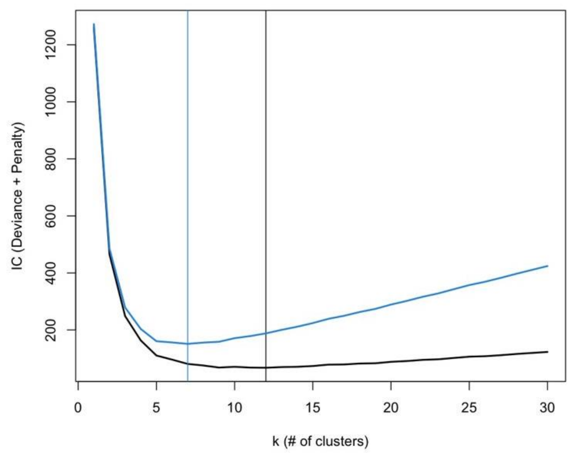

Figure 2 shows – IC starts to descend from the beginning and the decremental derivative is not significant from K=4 (the slope is becoming flat). Although the optimized K would be 7, it is much harder to interpret with only a few benefits. Therefore, balancing the technical and business-related factors, 4 clusters are determined to be used.

Modeling

The final training dataset includes 13 columns: movie/TV show count under each IMDb rating criteria (Excellent, Good, Normal, Disappointed) of recent 3 weeks and alpha classes of Netflix stock price (class 1-6). Our goal is to predict which alpha class Netflix Stock will fall under based on the rating scores of movies/TV shows Netflix released within the recent 3 weeks. To train the dataset, we decided to use three supervised algorithms: Random Forest, Decision Tree, and Multinomial Logistic Regression.

Random Forest

The random forest presents quick identification of significant information from the dataset and considers results of more than one algorithm of the same/different kinds of classifications. For our dataset, the pro of using random forest is reduced error. The random forest model takes input from all the trees, ensuring that individual errors of trees are minimized, and overall variance is reduced. For example, random forest can eliminate the outliers of our training samples (overrated/underrated movies/TV shows). The con of using random forest is that we should have some predictive power for the model to work. We need to consider what other proxies (ex. Released of competitors' original movies/TV shows) could affect the outcome (alpha class of Netflix stock price) to gain a more accurate prediction.

Decision Tree

The decision tree model provides a way to present algorithms with conditional control statements working as decision-making steps that lead to a favorable result. Using the decision tree algorithm requires fewer data cleaning and outliers have less significance. Moreover, it is easy to interpret without requiring statistical knowledge. It will be useful when presenting the information directly to clients. One of the disadvantages of using the decision tree model in our project is its unstable nature. For example, a minor change in our data (ex. Adding new proxies) can result in a major change in the structure of the decision tree which also leads to a different result.

Multinomial Logistic Regression

Multinomial logistic regression is a classification method and a natural probabilistic view of class predictions. It generalizes logistic regression to multiclass problems. One advantage of using multinomial logistic regression is that it is less prone to over-fitting in a low dimensional dataset while the dataset has enough training samples. The con of using multinomial logistic regression is that on high dimensional datasets, it will lead to the model being over-fit on the training set. For instance, if we want to observe the impact of additional proxies on the alpha class of Netflix stock price, multinomial logistic regression may lead to overfitting.

We aimed at finding the relationship between IMDb ratings of Netflix's original movies/TV shows and Netflix stock prices. By training our sample set using random forest, decision tree, and multinomial logistic regression algorithms, we can compare the accuracy of each model's predictions and gain more insights into the relationships of the dataset's proxies. We used 10-fold cross-validation to test our models. Random forest gave better predictions on the test dataset than both decision tree and multinomial logistic regression (as shown in Figure 3). If alternative supervised machine learning algorithms are needed for the project, we can use the Support Vector Machine (SVM) and discriminant function analysis.

Evaluation

Our first evaluation approach is based on the performance of our predictive model. The red dot in the scatter plot below shows false prediction and the black dots show the true prediction. By using the best model mentioned above (random forest) and including the training dataset, the best accuracy of our model is around 32.2% (as shown in Figure 4), which we may conclude this is a bad model to use for predicting the Netflix Stock Price.

Our second approach to evaluate our model is to see the best features to predict the stock price. Based on Figure 5, we can draw two conclusions: 1) bad releases are less important. 2) lagged effect of good release. In more detail, compared to good Netflix releases, bad releases are less important in affecting the change in Netflix stock price return alpha.

Deployment

Deployment of Results

Using this result, we can see that Netflix's successful original releases have a big positive impact on its financial performance, while bad or disappointing releases have no implication – meaning it does not negatively affect NFLX stock performance. Thus, through our analysis, there are 2 recommendations our team can put forth for Netflix:

1) Netflix should invest more in original content, as it is the key indication of stock market performance. A good release has positive effects on the company's financial performance. An example at the beginning of the report is Netflix's Squid Game.
2) Netflix should not be afraid to green light more niche/unusual scripts. As indicated through our analysis above, a bad release has little to no effect on Netflix's financial performance. However, many of these niche/unusual scripts would be enjoyed by many subscribers worldwide and allow Netflix to bring in more subscribers in its global strategy. An example is the widely successful movie franchise in Vietnam called Camellia Sisters, produced by Netflix. Each of these 5 movies has very low IMDB scores, ranging from 4-5, which would be clustered in as a bad release. However, these movies are enjoyed by millions of Vietnamese, and is Netflix's breakthrough releases in Vietnam. After Camellia Sisters, Netflix gained millions of subscribers.

Generally, we found that it takes two weeks before the effect of a good release is seen on the NFLX stock performance.

Potential Deployment Issues

IMDB scores, although one of the most important indicators, might not be the only indication of Netflix's financial performance. There are other non-quantifiable indicators such as market sentiment, consumers' entertainment behavior, competitors' original releases, indirect entertainment competitors (gaming, VR, social media...).

Omitted variable bias is another issue for this project. We aren't sure what is the effect of related, but omitted, variables that go into the original release variable. For example, a release can be successful solely to the fact that it has famous actors/actresses, and Netflix spent a lot of resources on marketing that release. Regardless of its IMDB score, Netflix will see positive financial performance.

Ethical Consideration

There are 2 ethical considerations that Netflix should consider with these recommendations.

1. Investors should not use NFLX new releases as the basis of investment decision, as there are multiple factors involved in stock's performance, original releases being one of them.
2. "Bad" releases sometimes still have positive implication for Netflix, as the example of Camellia Sisters. Thus, labelling these releases as "bad" might not produce the intended results for our research.

Risks and Mitigations

The biggest risk in our recommendations is the cost of original release movie productions. Even though good productions produce positive effects for stock performance, it also costs Netflix a lot more to produce than licensing. To mitigate this, we need to do additional research on the right ratio, and causation on what factors go into releases to make them successful, so Netflix can produce more releases that are well-loved.

Appendix A – Team Contributions

Cecily Cai	Data Preparation, Modeling, Report Write up
Renqi (Kevin) Chen	Data Understanding, Data Preparation, Modeling, Presentation, Report Write up
Tina Chen	Data Preparation, Modeling, Report Write up
Kunal Nadkarni	Presentation Framework, Business Problem Research
Minh Pham	Data Preparation, Report Write up
Yingdong Yang	Data Understanding, Data Preparation, Modeling, Presentation, Report Write up

Appendix B – Figures

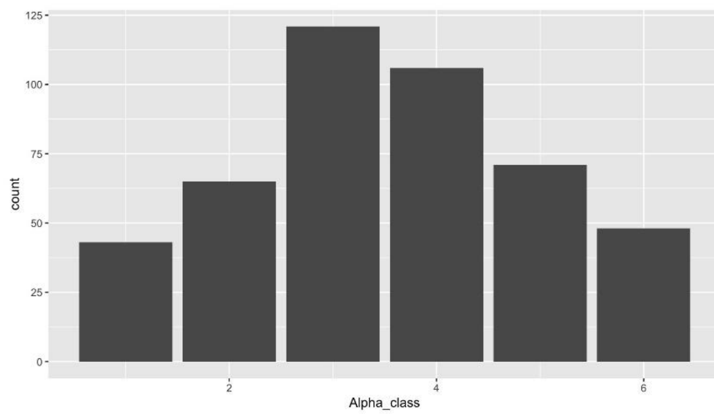

Figure 1. Count of Different Alpha Classes

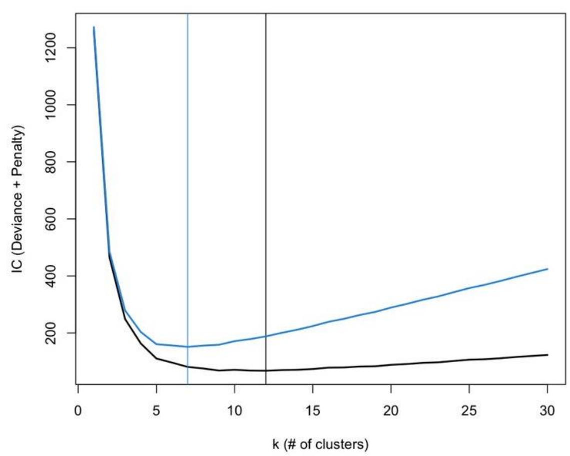

Figure 2 AIC/BIC vs. k (# of clusters)

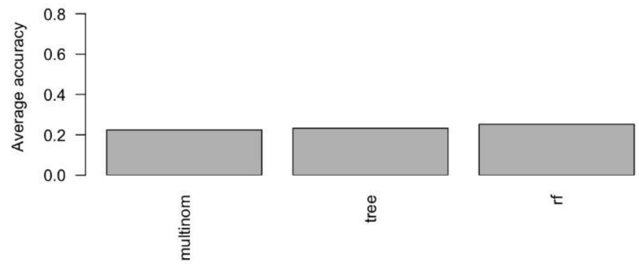

Figure 3 Average accuracy of each model: multinomial regression, decision tree, random forest

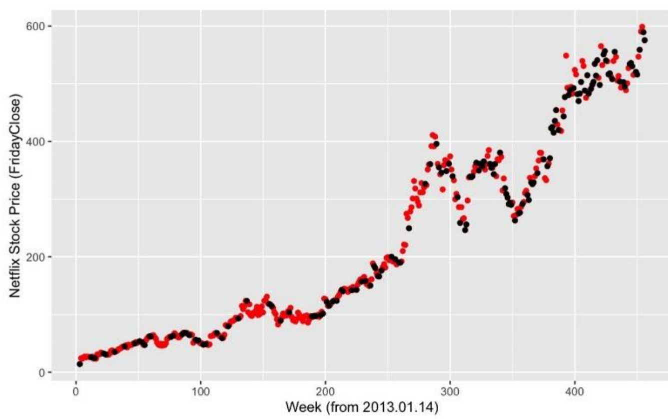

Figure 4 Performance of our model on predicting Netflix stock price

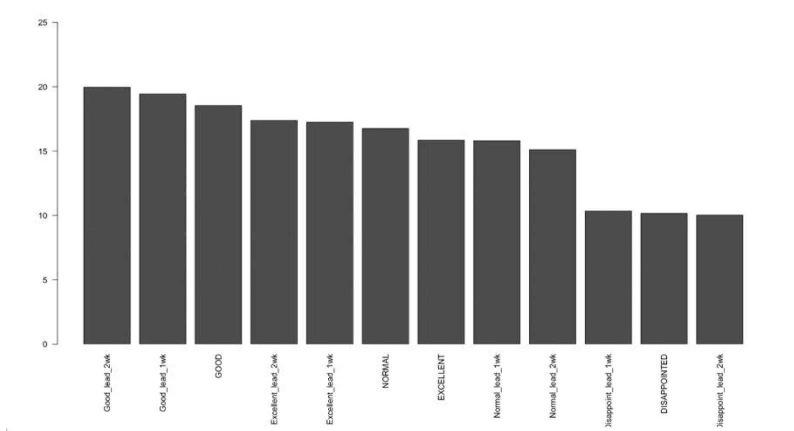

Figure 5 Feature importance plot