澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

榮譽學院
Colégio de Honra
Honours College

Machine Learning Approaches to Predict the Netflix APP Ratings

by

CHEN Xinyu, Hedy
(BC0-0504-7)

Advisor: PROF. LIAN Zhaotong

A report submitted in partial fulfillment of the requirements for the Honours College Certificate Programme in Honours College at the University of Macau

Year of Completion (2024)

SELF-DECLARATION

I declare that the report here submitted is original except for the source materials explicitly acknowledged and that this report, or parts of this report have not been previously submitted for the same course or for a different course.

I also acknowledge that I am aware of the Rules on Handling Student Academic Dishonesty and the Regulations of the Student Discipline of the University of Macau.

_CHEN XINYU______________	__4/30/2024____________________
Signature	Date
__CHEN Xinyu________________	_BC005047_____________________
Name	Student Number
HONR4001	
Course Code	

Abstract

Netflix, the unassailable streaming service, has experienced a significant surge in global subscribers since the pandemic outbreak. To gain insights into user reviews on the Google Play Store, this study utilized enhanced machine learning tools, Natural Language Processing techniques and data visualization applications. The cleaned dataset consists of 1,121,434 records spanning from 2017 to 2023, encompassing periods before, during, and after the pandemic. Four features were included: review text, review rating, likes, and the year of posting. A methodology was proposed to predict ratings based on user review comments, applying natural language processing and employing machine learning models, including random forest, XGBoost, support vector machines, and neural networks. Hyperparameters were fine-tuned, and accuracy was evaluated to determine the most effective model. By visualizing the data across different periods and using various plot formats, key patterns, trends, and prominent characteristics of user experiences are revealed. These insights provide valuable information to Netflix, enabling them to make informed decisions that focus on enhancing user satisfaction and deliver more personalized functions. The effectiveness of the approach was demonstrated through a comprehensive understanding of user sentiment through pandemic and the improvement of the overall user experience.

I. Introduction

The digital streaming entertainment landscape has experienced a significant shift in recent times, driven by the rapid expansion of streaming platforms. Among these platforms, Netflix stands out as an unassailable leader, commanding a substantial share of the global streaming market. In the end of 2017, Netflix has 110.64 million paid subscribers. As in the 4th quarter of 2023, Netflix has 260.28 million paid subscribers worldwide, realizing a more than double amount than 2017 (Netflix: Number of Subscribers Worldwide 2023 | Statista, 2024). The company's success has been propelled not only by its vast library of content but also by its commitment to delivering a personalized and immersive user experience.

In the first half of 2020, during the outbreak of COVID-19, Netflix achieved a net increase of over 26 million subscribers worldwide (ACCA - https://www.accaglobal.com, n.d.). The presence of COVID-19 is a double-edge sword chance for Netflix. Due to the lockdown policy, the cinema had been all shut down and people are mostly quarantined at home, leading to an increase in the subscribers of Netflix. On the other hand, the production of films and TV shows had been decreased sharply due to the limited movements and prevention from virus spread. Streaming services differentiate and outstand themselves from competitors relying on producing and delivering exclusive content to (Seetharaman, 2020). By offering unique shows, movies, and other forms of entertainment, streaming services aim to attract and retain subscribers. Although there is a decrease in the quantity of unique content, adults are devoting over one hour each day to streaming services. (Rahman & Arif, 2021). Netflix has emerged as the preferred platform for video streaming distribution in American households throughout the epidemic, capturing a significant 34% market share (Nielson, 2020). The longer time viewers spend on Netflix, the more experience, and comments they will have, and more advantages or disadvantages may be revealed. It is apparently important to understand and know customers’ experience during pandemic based on the personal description of viewing experience of Netflix. In Google Play, Netflix is a free-downloaded app for entertainment. People can freely give their comments and ratings to share opinions with others and give feedback to the company.

One research revealed that the subjective aspects of social media, such as user comments and ratings, exert a stronger influence on participants' decision-making compared to objective factors like experience or specialty. (Carbonell & Brand, 2017b). It stresses the importance of the importance in decision making and the contribution to the strategy improvement. However, other than the download platforms such as Google Play, App store, which enable the users to give both ratings and comments, some social platforms where people freely publish their thoughts would not require people to give ratings. Facebook and Twitter have remained popular social media platforms for a long time, attracting large user bases and offering various communication features. Additionally, TikTok and Instagram have gained significant traction, with their comments sections becoming increasingly important for engaging with content and connecting with others. A thorough approach is called for for examining user comments on social media. The goal is to provide a more comprehensive and in-depth method for analyzing user comments on social media in terms of both sentiment and content. (Alafwan et al., 2023)

Machine learning and Natural Language Processing (NLP) methods have become influential instruments for examining user comments on social media platforms. By leveraging the vast amounts of data generated by users, machine learning algorithms and NLP can realize real-time sentiment analysis of customers on social media platforms to understand public opinion and make appropriate responses (*Natural Language Processing in Data Science: Text Mining and Sentiment Analysis*, 2024). By employing approaches such as text categorization, sentiment analysis, and named entity recognition, machine learning and NLP enable us to extract valuable insights from user comments, uncover patterns, and gain a deeper understanding of user sentiments, opinions, and behaviors on social media.

II. Theoretical Background

a) Machine Learning:

Machine Learning is a specialized branch of Artificial Intelligence (AI) that concentrates on creating algorithms and models that can learn from data and draw conclusions or predictions without requiring explicit programming. Machine learning is the field that focuses on the application of statistical and computational methods (Sun & Wu, 2022, p. 17) to enable machines to autonomously learn patterns and relationships from data, which they may then use to make informed judgments or predictions. The three primary forms of this idea are reinforcement learning, unsupervised learning, and supervised learning. The categorization is based on the presence of labeled data and the specific learning purpose.

b) Natural Language Processing (NLP):

Natural Language Processing (NLP) is a field that aims to establish a connection between computers and human language. The primary objective is to facilitate machines in comprehending, analyzing, and producing human language (Donges, 2023) in a manner that is both significant and practical. Natural Language Processing (NLP) encompasses a range of tasks including text categorization, sentiment analysis, named entity identification, and question answering. It employs methodologies from linguistics, computer science, and machine learning to handle and examine textual data, extract significant insights, and empower machines to comprehend and produce human language.

c) Support Vector Machines (SVM):

A well-liked supervised learning technique for regression and classification problems is support vector machines. SVM looks for the best hyperplane with the largest margin in the feature space to divide various classes. Support Vector Machines (SVMs) may effectively handle data that is both linearly separable and non-linearly separable by utilizing several kernel functions, such as rbf, linear, and sigmoid, to transform the input data. Support Vector Machine have been extensively utilized in diverse fields such as text categorization, picture recognition, and bioinformatics (Mammone et al., 2009).

d) Random Forest:

Random Forest is an ensemble learning technique that utilizes many decision trees to generate predictions. It is a commonly employed method of supervised learning for tasks that involve categorization and prediction. The Random Forest algorithm constructs several decision trees using random selections of the training data and random subsets of the features. Throughout the prediction process, each tree in the forest independently creates a prediction, and the final prediction is determined by aggregating the individual guesses from all the trees. Random Forest is renowned for its capacity to effectively manage datasets with a large number of dimensions, effectively handle missing values, and offer reliable assessments of feature relevance (Genuer & Poggi, 2015).

e) XGBoost:

XGBoost, also known as Extreme Gradient Boosting, is an extensively enhanced implementation of the gradient boosting technique. Ensemble learning is a robust technique that integrates numerous weak prediction models, usually decision trees, to form a highly accurate predictive model. XGBoost constructs the model iteratively, with each iteration aiming to rectify the errors committed by the preceding iteration (*What Is XGBoost?*, n.d.). The primary objective is to enhance a particular loss function by the utilization of gradient descent. Additionally, it incorporates strategies like regularization to mitigate the risk of overfitting. XGBoost has been increasingly popular due to its high efficiency, scalability, and impressive performance in a wide range of machine learning contests and real-world scenarios.

f) Neural Network:

Neural Network, or Artificial Neural Networks (ANNs), is subset of machine learning which is designed based on the form and functionality of biological neural networks found in human brain. Neural networks consist of interconnected nodes, referred to as neurons, which are organized in layers. Each individual neuron receives input signals, applies an activation function to these inputs, and produces an output signal. Neural Networks undergo training by a technique known as backpropagation (Al-Masri, 2022), in which the model modifies the weights of its connections in order to reduce the discrepancy between expected and actual outputs. Artificial neural networks possess the ability to acquire knowledge about intricate non-linear connections within the data and reached notable achievements in diverse domains, ranging in image identification, natural language processing, and speech recognition (Alexandos, 2023).

III. Methodology

Organizations implement the SEMMA methodology in their machine learning and data mining endeavors to attain a competitive edge, enhance operational efficiency, and provide consumers with more valuable services (Hotz, 2023). It is designed by the SAS application company and widely used. It provides a structured approach for analyzing and transforming data to develop accurate and reliable predictive models. On the basis of the Google Play dataset, we use SEMMA methodology to provide a systematic framework for determining which of the four machine learning models listed above is the most accurate. After the data mining part, we use data visualization tools to make qualitative analysis on the low rating comments, to extract any insights for enhancing the understanding of user feedback and driving improvements in customer satisfaction and product development.

IV. Data Mining

a) Sample – Data Collection

In this stage, we select a dataset and then partition a portion of the dataset with the required size for constructing the model. The objective of this preliminary phase is to ascertain the variables or characteristics that have an impact on the process. Subsequently, the gathered data is categorized into preparation and validation groups. Our dataset is downloaded from Kaggle which is grasped and extracted from Google Store. The dataset paints a good picture on what is the public's perception of the app over the years from 2011- 2023. It contains in total 1.53 million records, including the column review_id, pseudo_author_id, author_name, review_text, review_rating, review_likes, author_app_version, review_timestamp and is extracted from all world regions aims to capture sentiments and trends and use topic modeling to identify the pain points of the application.

To focus on the influence of ratings and reviews provided by users, certain variables such as review_id, pseudo_author_id, author_name, and author_app_version are removed from the dataset to create a cleaned version. Our primary objective is to investigate the recent years, specifically the periods before, during, and after the pandemic. Consequently, the dataset is filtered to include data only from 2017 to 2023, enabling a targeted analysis of the impact of ratings and reviews during this time frame.

To simplify the feature 'review_timestamp' and make it more suitable for our analysis, we have decided to remove the minute and hour information. Instead, we will create two new features: 'review_date' and 'review_year', which will be derived from the original timestamp. After data preprocessing, the dataset includes 1121433 rows of records and no duplicate row exists. Date ranges from 1/1/2017 to 11/15/2023 and features include review text, ratings and years.

b) Explore – Data Visualization

During this stage, both univariate and multivariate analysis are performed to examine the interconnections between data items and to discover any deficiencies in the data. An exhaustive analysis is conducted on all the aspects that may impact the conclusion of the study, with a strong emphasis on data visualization.

The dataset used for the analysis includes the variables ‘review_text’, ‘review_rating’ and ‘review_year’ (Appendix 1). The 'review_rating' variable has a mean of 3.93, indicating that the average rating given by users is slightly below 4. The standard deviation of 1.59 suggests a moderate level of variability of ratings. 'Review_year' variable covers from 2017 to 2023, with a mean of 2019.88, around 2020. These descriptive statistics provide an overview of the key characteristics of the variables and serve as a foundation for further analysis.

In order to examine the distribution of review information, we utilize a bar chart to present the count of reviews from various years (Appendix 2). Notably, the year 2020 stands out with the highest number of reviews, totaling approximately 300,000 records. Conversely, the lowest number of reviews is observed in 2023, with approximately 100,000 records.

The distribution of reviews across different ratings is not evenly distributed. Ratings of 5 receive the highest number of reviews, with approximately 700,000 records, indicating a strong positive sentiment. On the other hand, ratings of 2 receive the lowest number of reviews, suggesting a relatively lower level of satisfaction or negative feedback (Appendix 3).

To gain a broad and comprehensive understanding of the text, word cloud visualizations are generated to depict the frequency of individual words appearing in different commentsThe size of each word in the word cloud is proportional to its frequency, with greater font sizes representing higher occurrence. Positive words such as 'good', 'great', 'love', and 'awesome' are among the most frequently used, suggesting positive sentiments expressed by users. Conversely, words like 'update', 'money', 'subscription', and 'download' may indicate potential negative experiences encountered by users. Additionally, words like 'Netflix' and 'app' are prominent topic words but do not carry a specific sentiment or meaning.

c) Modify – Data Preprocessing

In modify stage, the conception gained during the exploration phase are derived by applying business logic to the data collected in the sample stage. The data is parsed and cleaned to ensure its quality and reliability. This cleaned data is then passed on to the modeling stage, where various techniques and algorithms are applied to extract insights and make predictions.

In the data preprocessing phase, several steps are taken to handle emojis and convert them into text-based words. This is done to ensure consistent and meaningful representation of emotions conveyed by users. Additionally, non-alphabetic words, such as 'Netflix' and 'app', which may not contribute significantly to sentiment analysis, are removed to reduce noise in the data. Furthermore, irrelevant characters like webpage links and special characters are eliminated to focus on the relevant text content. To enhance the speed and quality of text analysis, often occurring words with minimal semantic significance, known as stop words, are eliminated (Silva, 2003). Ultimately, the text undergoes tokenization, which involves breaking it down into discrete words or tokens, enabling more detailed analysis.

Tokenization is an essential procedure in natural language processing (NLP) that entails dividing a text content into separate words or tokens. Tokenization allows the system to process and manipulate distinct components of the text (Pandey, 2023). Once the text has been tokenized, it can be converted into a numerical representation, typically as vectors, to enable computational analysis. This process is known as vectorization or word embedding.

	review_text
0	I'd love to use it with my new membership, but...
1	Right know I can not download vids or movies s...
2	How it has New moves😊😊😊😚😍
3	I like are should I say love it because you ca...
4	Yay

```
[[   0    0    0 ...    7   18 1055]
 [   0    0    0 ...    0    0 2928]
 [   0    0    0 ...  292    7   12]
 ...
 [   0    0    0 ...  634  158  364]
 [   0    0    0 ...  222  128   77]
 [   0    0    0 ...   57  612  659]]
```

Also, based on the insights we gained from the exploration stage, the imbalanced dataset is a potential factor to affect the performance of the result of the model. To mitigate the impact of imbalanced data and improve model performance, random sampling and resampling techniques are applied. Random sampling creates a balanced representation of different classes by selecting a subset of the original dataset (Horton, 2022). Additionally, resampling techniques generate synthetic examples or downsample the major class to build a more balanced dataset. In the case of the immense dataset with over one million records, random sampling and resampling ensure an equal number of records for each year from 2017 to 2023.

d) Model – Machine Learning Models

Once the variables have been improved and the data has been cleansed, the modeling step utilizes various data mining techniques to provide a projected model that demonstrates how this data leads to the desired conclusion of the process.

The overall goal of the research is to classify the comments into five ratings. Four appropriate models: SVM, Random Forest, XGBoost and Neural Network have been introduced for further use. By employing these diverse models, the research aims to explore and compare their performance in accurately classifying comments into their respective ratings. Through the modeling phase, we aim to identify the most suitable model that yields the highest accuracy and predictive performance for the classification task.

In evaluating the outcome among four built models on the original dataset of 16,822 records, several metrics were used, including accuracy, precision, recall, and F1-score.

To ensure the evaluation of the models is reliable, the dataset was split into a training set and a test set using a 70-30 ratio. 70% of the original dataset was allocated to the training set, while the remaining 30% was designated for the test set.

In our analysis, we observed that the input variable X exhibits a wide range, spanning from 1 to 12,000, the number of words in the vocabulary and label Y only ranges from 1-5. Considering this wide range of values, we made the decision to apply feature scaling to X before training the models. We compared the results of applying feature scaling versus not applying feature scaling for each of the four models. Feature scaling was undertaken to normalize the range of X, ensuring that all features contribute equally to the model training process. By scaling the features, we aimed to prevent any potential dominance of features with larger values, thus promoting a more balanced learning environment for the models.

During the model building process, we conducted hyperparameter tuning to optimize the performance of the models. For the SVM model, we experimented with various kernel functions, including the radial basis function (RBF), linear, sigmoid, and others. It determines the type of decision boundary that the model can learn to separate the classes in the input space. For the Random Forest and XGBoost models, we set the random state parameter to be 42. The random state is a seed value that is used to initialize the random number generator in the algorithms, which is an arbitrary value. By setting it to a specific value, such as 42, we ensure reproducibility of the results across different runs of the models. In the case of the Neural Network model, we explored different numbers of layers, ranging from 5 to 10. We trained and evaluated models with varying layer configurations to determine the optimal number of layers. After assessing the performance of each model, we found that the model with 7 layers demonstrated the best performance among the tested configurations.

e) Assess – Model Evaluation

During the last stage of SEMMA, the model is assessed to determine its utility and dependability in relation to the issue under study. Now, the data may be evaluated and utilized to measure the effectiveness of its performance. In the current study, we selected the best analysis approach by using the highest discriminant accuracy among the implemented analysis methods as a model-specific performance validation indicator. (Chen & Lee, 2021b) These metrics provide a comprehensive assessment of the model's performance in terms of classification accuracy, precision (true positive rate), recall (sensitivity), and the balance between precision and recall (F1-score). By considering these evaluation metrics, we were able to assess and validate the effectiveness and efficiency of our analysis approach.

Among the models, the Random Forest and Neural Network models demonstrated higher accuracy and F1-scores compared to SVM and XGBoost. The Random Forest model achieved a training accuracy of 95.3% and a testing accuracy of 67.9%, while the Neural Network model achieved a training accuracy of 89.0% and a testing accuracy of 72.4%.

Regarding precision, it quantifies the model's capacity to accurately recognize positive cases. Neural Network model outperformed the other models with a precision score of 0.661 on the testing set. The Random Forest and XGBoost models achieved precision scores of 0.568 and 0.591, respectively, while the SVM model had the lowest precision score of 0.408.

When evaluating recall, which quantifies the model's capacity to correctly recognize all positive instances, the Neural Network model again showed the highest performance with a recall score of 0.724 on the testing set. The XGBoost model achieved a recall score of 0.690, followed by the Random Forest model with a score of 0.663. The SVM model had the lowest recall score of 0.639.

The F1-score, which combines both precision and recall into a single metric, also favored the Neural Network model, which achieved a score of 0.681 on the testing set. The Random Forest, XGBoost, and SVM models achieved F1-scores of 0.603, 0.630, and 0.498, respectively.

The Neural Network model demonstrated the best accuracy, precision, recall, and F1-score, showing its outstanding overall performance in this classification task.

Evaluation of Models (Original Dataset)

Model	SVM	Random Forest	XG Boost	Neural Network
Train Accuracy	0.643	0.953	0.864	0.890
Test Accuracy	0.647	0.679	0.687	0.724
Test Precision	0.408	0.568	0.591	0.661
Test Recall	0.639	0.663	0.690	0.724
Test F1-score	0.498	0.603	0.630	0.681

For the resample dataset, it shows quite similar results with the original dataset. Among them, Neural Network achieved the highest Test Accuracy of 0.714, followed by XGBoost with 0.676 and Random Forest with 0.663. In terms of precision, Neural Network had the highest value of 0.645, while Random Forest achieved the highest recall of 0.663. The Neural Network model also had the highest F1-score of 0.673. Overall, Neural Network consistently performed well across multiple metrics, followed by XGBoost and Random Forest. SVM had relatively lower performance compared to the other models.

Evaluation of Models (Resample Dataset)

Model	SVM	Random Forest	XG Boost	Neural Network
Train Accuracy	0.647	0.952	0.865	0.868
Test Accuracy	0.626	0.663	0.676	0.714
Test Precision	0.481	0.564	0.571	0.645
Test Recall	0.626	0.663	0.682	0.714
Test F1-score	0.525	0.600	0.611	0.673

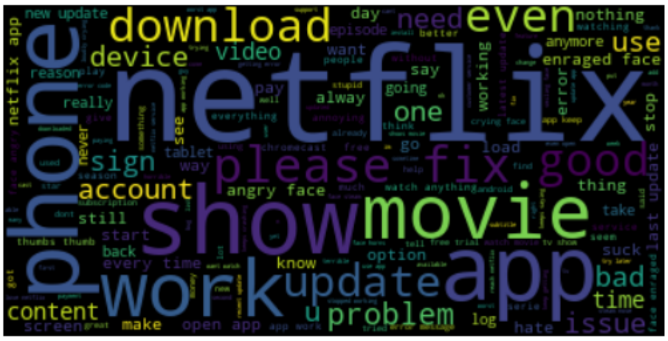
The neural network model demonstrated the highest accuracy among the tested models for classifying ratings based on the social media comments, without employing any resampling techniques. It outperformed other models such as SVM, Random Forest, and XGBoost in accurately categorizing the ratings.

V. Data Visualization

After training and obtaining the most accurate model to predict ratings based on the comments from the social media, the model can be further used in different social media which do not require ratings explicitly when they comment. The comments can be extracted from social media platforms using techniques such as web scraping, and subsequently processed through natural language processing (NLP) for further analysis (Jamin Rahman Jim et al., 2024) and prediction of ratings.

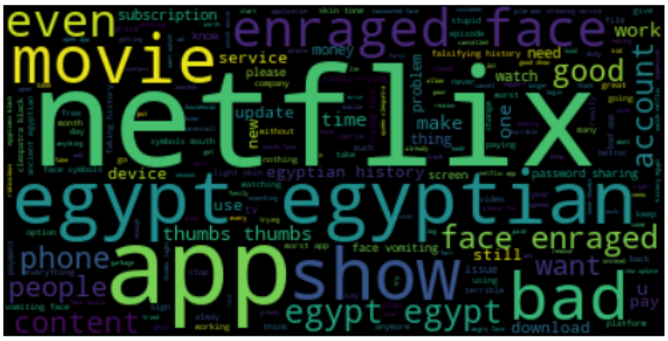
Once the ratings are predicted, they can be further analyzed and interpreted using data visualization tools. These tools can help uncover valuable business insights from the predicted ratings. In order to uncover the problems, our focus should primarily be on analyzing the comments associated with low ratings. We generate a word cloud specifically using low rating (rating from 1 to 2) comments from different year groups, including before, during, and after the pandemic, to gain insights into prevalent issues.

Word Cloud (2017-2019, Rating 1-2)



Word Cloud (2020-2022, Rating 1-2)

Word Cloud (2023, Rating 1-2)



The analysis of different year periods reveals distinct high-frequency keywords. In the 2017-2019 period, keywords such as 'download,' 'sign,' 'account,' 'update,' and 'content' highlight potential issues. For 2020-2022, keywords like 'download,' 'device,' 'subscription,' and 'option' draw attention. In 2023, keywords such as 'Egypt,' 'Egyptian,' 'history,' and 'content' provide valuable insights for further investigation.

To gain a comprehensive understanding of what these words are specifically addressing, we utilize the pandas library in Python. We extract a whole page of detailed information related to these keywords, providing a more comprehensive context and insights into the topics being discussed. The ‘Words’ column indicates if any of the word shown in the comment, they will be count. And the ‘Frequency’ column represents the times the words appear in the comments. ‘Example’ are representative comments extracted manually that demonstrate the potential value of these insights for driving business improvements.

One potential problem raised most frequently is the possibility of the updated app version crashing or being incompatible with certain devices. Additionally, there may be inefficiencies in content updates, failing to adequately meet customer needs. Concerns about account privacy and information leakage are also raised. Age policy inaccuracies and discrepancies in content restrictions between countries are other areas of contention. Subscription options and pricing are highlighted as well. Notably, the mention of 'Egypt' uncovers an event where Netflix allegedly published a series that falsified Egyptian history, leading to significant awareness and opposition from the Egyptian community (Yee, 2023).

Table 9

Low Rating Comments Detail

Words	Frequency	Example
'update', 'updating'	26278	•I tried this on two android devices and it carshz after it finish loading Update: doesnt work on WiFi •Still great to watch good movies, but updating the movies section, tv shows, and such, needs to be updated asap because it gets boring after awhile.
'account', 'sign'	18696	•I love the app, but recently I deleted it and downloaded it a again and my profile on the account was missing. •I have no serious problem with the app I just wish that there could be another account category option, at the moment there's 18+ and 7+ and I just thought that if there was one in the middle of the age groups like 15+ that would be much better other than being stuck with watching 7+ shows.
'content'	14401	•Good but need more content in uae region •im so disappointed that every person pay the same money and cannot watch the same content as others im from pakistan and alot of movies like avengers and dont mess with zohan i cannot watch them... america have all the content and we have very verylimited content and both pays the same money. So wrong 🙁
'subscription', 'subscribe'	10848	•your monthly rent is high plz make yearly plan like amazon prime video with lowest rent so all can easily subscribe netflix.plz make yearly plan wih low cost for mobile user.
'HBO', 'Hulu', 'Amazon', 'Disney'	7985	•The same issue that Amazon has: lack of foreign subtitles. •Worst tv app ever!! Its so costly, amazon prime subscription was just rs 499 for an year!! And netflix charges 800 a month!!! And cant even pay for that!!! •Not top of the line service LIKE Disney Got all the Movies in 4k HDR Atmos unlike NETFLIX DISNEY CHARGE AROUND $9 MONTH AND NETFLIX $16 4 screen I want for 1 screen only HD MOVIES ON NETFLIX ALWAYS HAS LIKE DOTS And Amazon Prime is awesome lots of titles
'egypt', 'egyptian'	5014	•Forging Egypt's civilized history? This will not work again and I think you are messing with us, and now it is time for the real war and we are ready Well done Netflix, bye bye •Netflix is falsifying Egyptian history and heritage and supports blackwashing ancient egyptian history.

VI. Limitation

Several limitations should be considered in the analysis presented. Firstly, the study relied on limited data sources, primarily extracting comments from Google Play Store. In order to improve the reliability of the analysis, it would be advantageous to incorporate data from supplementary sources such as Apple Store reviews and official download URLs.

Another limitation is the incomplete coverage of internet slang and its nuanced meanings. Internet slang is constantly evolving, and certain unique expressions or terms may not have been adequately captured in the analysis. This could potentially impact the accuracy and comprehensiveness of the findings.

Furthermore, personal preferences for repetitive or specially-meaningful emojis may introduce subjectivity into the analysis. Emojis can convey a range of emotions and interpretations, and individual users may attribute different meanings to specific emojis. Accounting for these subjective interpretations can be challenging and may affect the overall sentiment analysis.

The significance of BOLD characters in conveying emotions may not have been fully captured (Bernazzani, 2017). Textual emphasis, such as the use of bold characters, can play a role in expressing sentiment or intensity. However, the analysis primarily focused on textual content, potentially overlooking the emotional impact of formatting choices.

Additionally, computational limitations may have resulted in a reduced dataset. Due to resource constraints, it was not feasible to analyze the entire corpus of comments, potentially leading to a smaller sample size. This limitation could impact the representativeness and generalizability of the findings (Gobo, 2004).

VII. Conclusion

In conclusion, the neural network, as a deep learning technique, achieved the highest accuracy with 89% on the training set and 73% on the test set. This model can be effectively utilized to predict comments on Netflix from other social platforms. However, it is noteworthy to address that although the random forest model had the highest training accuracy, it exhibited overfitting issues that cannot be disregarded.

Based on the insights gained from the analysis and through data visualization, several areas for business improvement can be suggested. These include focusing on enhancing user experience, addressing version update concerns, refining subscription policies, and improving content creation and supervision. These issues were found to be significant during and after the pandemic, indicating their importance in driving positive customer experiences and satisfaction.

Furthermore, specific insights for Netflix can guide future actions. It is recommended to conduct thorough testing of updated versions before release to ensure stability and minimize crashes. Implementing robust measures to prevent misrepresentation of religion, country, history, or race is crucial in maintaining a responsible and inclusive content platform. Regular evaluation and refinement of age policies should be undertaken to ensure compliance with legal requirements. Additionally, customizing content offerings based on cultural sensitivities and preferences of different regions can enhance user engagement. Lastly, the subscription during pandemic has greatly increased and decreased after pandemic, indicating that it not a pricy payment if the viewing experience is not top priority (Paramasivan & Ying, 2023). Therefore, analyzing usage patterns and customer feedback related to different subscription choices can inform the provision of more flexible plans tailored to individual preferences.

Netflix can strategically utilize social media data, employ sophisticated natural language processing (NLP) techniques, visualize valuable insights, and implement improvements to efficiently address customer needs, maintain a competitive advantage, and foster long-term success in the ever-changing streaming industry. Other applications can also utilize this technology to improve the services they offer. Through the analysis of social media data, the extraction of significant insights, and the utilization of NLP approaches, firms can acquire a more profound comprehension of customer preferences, concerns, and suggestions. These observations can subsequently be utilized to create well-informed judgments, enhance consumer satisfaction, fine-tune product offers, and maintain a competitive advantage. Companies can cultivate enduring success and establish robust client connections by actively and promptly addressing customer requirements. Thus, these approaches can be applied not just by Netflix but also by other firms to improve their services and stimulate expansion.

VIII. Reference

Alafwan, B., Siallagan, M., & Putro, U. S. (2023). *Comments analysis on social media: A review*. ICST Transactions on Scalable Information Systems. https://doi.org/10.4108/eetsis.3843

Al-Masri, A. (2022, October 31). *Backpropagation in a Neural Network: Explained | Built In*. Builtin.com. https://builtin.com/machine-learning/backpropagation-neural-network

ACCA - https://www.accaglobal.com. (n.d.). *The unassailable rise of Netflix | ACCA Global*. https://www.accaglobal.com/gb/en/student/sa/features/netflix.html

Bernazzani, S. (2017, January 2). *Fonts & Feelings: Does Typography Connote Emotions?* Blog.hubspot.com. https://blog.hubspot.com/marketing/typography-emotions

Chen, Q., & Lee, S. (2021b). *A machine learning approach to predict customer usage of a home workout platform*. *Applied Sciences*, *11*(21), 9927. https://doi.org/10.3390/app11219927

Carbonell, G., & Brand, M. (2017b). *Choosing a Physician on Social Media: Comments and Ratings of Users are More Important than the Qualification of a Physician*. International Journal of Human-computer Interaction, 34(2), 117–128. https://doi.org/10.1080/10447318.2017.1330803

Donges, N. (2023, July 31). *Introduction to NLP*. Built In. https://builtin.com/data-science/introduction-nlp

Genuer, Robin & Poggi, Jean-Michel & Tuleau-Malot, Christine & Villa-Vialaneix, Nathalie. (2015). *Random Forests for Big Data. Big Data Research*. 9. 10.1016/j.bdr.2017.07.003.

Gobo, G. (2004). *Sampling, representativeness and generalizability. Qualitative research practice*, 405, 426.

Horton, M. (2022, March 28). *Simple Random Sample: Advantages and Disadvantages*. Investopedia. https://www.investopedia.com/ask/answers/042815/what-are-disadvantages-using-simple-random-sample-approximate-larger-population.asp

Hotz, N. (2023, August 12). *What is SEMMA? - Data Science Process Alliance*. Data Science Process Alliance. https://www.datascience-pm.com/semma/

Jamin Rahman Jim, Apon, M., Partha Malakar, Md Mohsin Kabir, Nur, K., & M.F. Mridha. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *100059*, 100059–100059. https://doi.org/10.1016/j.nlp.2024.100059

Mammone, Alessia & Turchi, Marco & Cristianini, Nello. (2009*). Support vector machines. Wiley Interdisciplinary Reviews: Computational Statistics*. 1. 283 - 289. 10.1002/wics.49.

Natural Language Processing in Data Science: Text Mining and Sentiment Analysis. (2024, January 15). https://moldstud.com/articles/p-natural-language-processing-in-data-science-text-mining-and-sentiment-analysis

Netflix - apps on Google Play. (n.d.). https://play.google.com/store/apps/details?id=com.netflix.mediaclient&hl=en_US&pli=1

Netflix: number of subscribers worldwide 2023 | Statista. (2024, January 31). Statista. https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/

Paramasivan, Kehxheni & Ying, Julia & Syazana, Iffah & Wenji, Jin & KG, Akanksh. (2023). *Subscription and Customer Loyalty A Study of Netflix Before and After Covid-19 Pandemic.* Asian Pacific Journal of Management and Education. 6. 117-128. 10.32535/apjme.v6i3.2674.

Pandey, A. (2023, October 15). *NLP: HOW MACHINE UNDERSTANDS HUMAN LANGUAGE*. Medium. https://medium.com/@annie_/nlp-how-machine-understands-human-language-009ec4000a66

Rahman, K. T., & Arif, M. Z. U. (2021). Impacts of Binge-Watching on Netflix during the COVID-19 pandemic. *South Asian Journal of Marketing*, *2*(1), 97–112. https://doi.org/10.1108/sajm-05-2021-0070

Sun, Z., & Wu, Z. (2022). *Handbook of Research on Foundations and Applications of Intelligent Business Analytics* (p. 17). IGI Global.

Silva, C.. (2003). *The importance of stop word removal on recall values in text categorization*. 3. 1661 - 1666 vol.3. 10.1109/IJCNN.2003.1223656.

Vasileiadis, Alexandos. (2024). *Artificial Neural Network and Its Applications*.

What is XGBoost? (n.d.). NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/xgboost/

Yee, V. (2023, May 10). *Whose Queen?* Netflix and Egypt Spar Over an African Cleopatra. *The New York Times*. https://www.nytimes.com/2023/05/10/world/middleeast/cleopatra-netflix-race-egypt.html

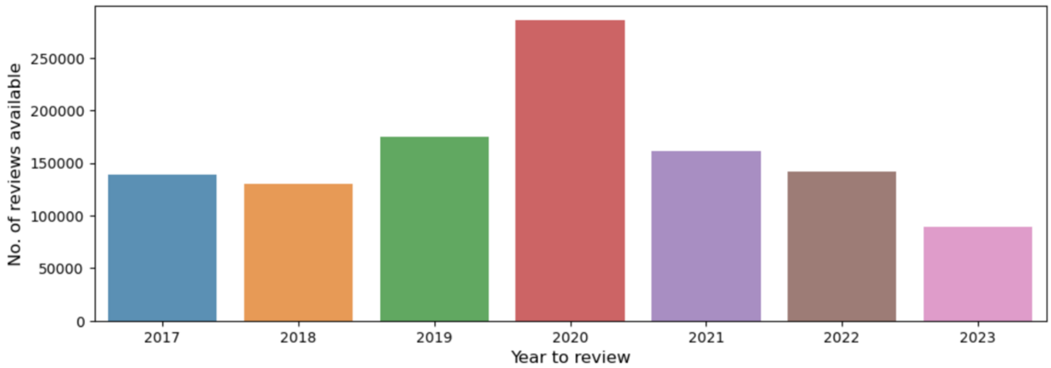
IX. Appendix

Appendix 1: Descriptive Statistics of variables

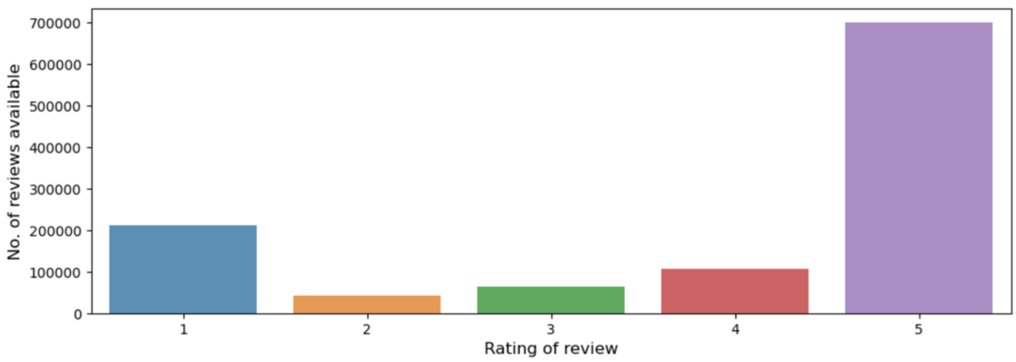
Descriptive Statistics

	review_rating	review_year
count	1121434	1121434
mean	3.931	2019.878
std	1.586	1.757
min	1	2017
25%	3	2019
50%	5	2020
75%	5	2021
max	5	2023

Appendix 2: Bar chart for distribution of review by year



Appendix 3: Bar chart for distribution of review by year



Appendix4: Programming code print (see attachment)