*Predictive Modelling of Churn Using Sentiment Analysis and Hypothesis Testing*



*Project work Submitted to*

The Department of Statistics, Pondicherry University

*By*

Ms. Annanya Singh

Regd. No.: 23375010

Mr. Hrithik Dineshan

Regd. No.: 20384308

As an Assignment for the Partial Fulfilment of the Course work

*Under the Guidance of*

Dr. Sudesh Pundir

Professor & Course Teacher

STAT-532 Testing of Statistical Hypotheses

*NOVEMBER – 2024*

**Contents**

1. **Abstract**
2. **Introduction**
3. **Literature Review**

* 3.1 Customer Churn Prediction Models
* 3.2 Sentiment Analysis in Churn Prediction
* 3.3 Satisfaction Scoring for Interpretive Insights
* 3.4 Hypothesis Testing in Model Validation

1. **Research Methodology**

* 4.1 Data Collection
* 4.2 Data Preparation
* 4.3 Sentiment Analysis
* 4.4 Churn Risk Clustering
* 4.5 Hypothesis Testing
* 4.6 Visualization and Interpretation of Results

1. **Modelling and Analysis**

* 5.1 Data Preprocessing
* 5.2 Sentiment Analysis
* 5.3 Cluster Analysis
* 5.4 Hypothesis Testing
* 5.5 Predictive Modelling

1. **Conclusion**
2. **Recommendations**
3. **References**
4. **Appendix**

* 9.1. Web Scrapping Code
* 9.2 Analysis Code

### Abstract

In the highly competitive streaming industry, understanding and mitigating customer churn is essential for sustained growth, as retaining subscribers is often more cost-effective than acquiring new ones. This project addresses this challenge by enhancing churn prediction models through the integration of sentiment analysis and satisfaction scoring derived from Netflix user reviews. Traditional churn models largely focus on behavioral data, such as user demographics and viewing habits, which, while informative, may overlook key experiential factors that influence user satisfaction and potential disengagement. By examining sentiment in customer feedback, we generate a Satisfaction Score that encapsulates both emotional and experiential indicators of satisfaction, providing a more comprehensive view of the customer experience and identifying factors not captured by behavioral metrics alone.

The Satisfaction Score serves as a vital feature within our predictive models, enabling us to identify high-risk users who may be likely to churn. To validate and refine our approach, we conduct hypothesis testing to explore key relationships within the data. Specifically, we assess correlations between sentiment and numerical ratings, the impact of sentiment on actual churn behavior, and variations in satisfaction and engagement across user segments and different product versions. This rigorous testing of hypotheses allows us to determine the reliability of sentiment-based predictions and gain insights into the influence of user sentiment on churn risk.

The results of our analysis provide a robust, data-driven framework that Netflix can use to identify and proactively address user dissatisfaction. By pinpointing potential churn risks early, Netflix can implement targeted retention strategies aimed at enhancing user satisfaction and loyalty. Ultimately, this project equips Netflix with actionable insights to foster long-term customer retention, allowing for timely interventions that strengthen overall customer satisfaction and reduce churn rates. By focusing on the emotional and experiential aspects of user feedback, this approach not only enhances predictive accuracy but also contributes to a deeper understanding of customer needs, paving the way for improved service delivery and content curation tailored to user preferences.

### Introduction

In recent years, the subscription streaming industry has become increasingly competitive, with platforms like Netflix striving to retain users in a crowded market. While traditional churn prediction methods leverage behavioral and demographic data, these approaches often miss important signals about a user's emotional and experiential satisfaction with the service. Customer reviews, such as those found on Netflix, offer valuable insights into these qualitative aspects of user experience, as they capture feedback that reflects both satisfaction and potential dissatisfaction. Analyzing this feedback to understand user sentiment can be key to identifying at-risk customers and taking proactive measures to improve retention.

Our project addresses this need by using sentiment analysis to evaluate Netflix user reviews, quantifying user sentiment into a Satisfaction Score that reflects overall satisfaction with the platform. By transforming textual feedback into a numerical score, we capture subjective opinions that are otherwise difficult to measure, thus providing a more comprehensive view of churn risk. This Satisfaction Score serves as a core component of our churn prediction model, augmenting traditional data inputs to create a predictive approach that goes beyond basic usage patterns.

To ensure the reliability and effectiveness of this model, we incorporate hypothesis testing as a validation method. By systematically examining key hypotheses within the Netflix review data, we explore relationships between sentiment and churn-related variables. Specifically, we test hypotheses that assess (1) the correlation between sentiment scores and numerical ratings, (2) impact of review length on sentiment. This rigorous hypothesis testing approach helps to determine the robustness of sentiment-based predictions and clarifies the role of user sentiment as a churn indicator.

Through this combined approach, the project provides Netflix with a powerful tool for understanding and predicting customer churn. By integrating sentiment analysis and satisfaction scoring with a structured hypothesis testing framework, we offer actionable insights that can guide interventions to prevent churn, enhance customer satisfaction, and support long-term retention goals. This approach ultimately enables Netflix to make data-driven decisions that align with both user needs and business objectives, strengthening their competitive advantage in the streaming market.

### Literature Review

The use of predictive modeling for customer churn has evolved significantly, with traditional methods focusing on behavioral data such as customer usage patterns. However, sentiment analysis—particularly of user-generated feedback—provides a more nuanced approach to understanding customer dissatisfaction and predicting churn. This literature review discusses existing research on churn prediction, sentiment analysis, satisfaction scoring, and the use of A/B testing in predictive modelling, contextualized with Netflix reviews as a specific data source.

#### **3.1.Customer Churn Prediction Models**

Classic churn prediction models typically rely on quantitative user behavior and demographic data (Neslin et al., 2006; Verbeke et al., 2012). Techniques such as decision trees, logistic regression, and neural networks are often applied to identify users at risk of churning. However, these models may overlook the emotional indicators of churn that are often present in textual feedback, such as Netflix reviews, where users share their dissatisfaction or praise for content (Lemmens & Croux, 2006). Recent studies highlight the benefits of integrating sentiment-based variables into churn prediction models to provide a more holistic view of customer behavior (Coussement & Van den Poel, 2008).

#### **3.2.Sentiment Analysis in Churn Prediction**

Sentiment analysis has proven useful in churn prediction as it quantifies user emotions and opinions from textual feedback (Pang & Lee, 2008). In streaming services, where users regularly leave feedback on shows and platform experiences, sentiment analysis enables platforms like Netflix to gauge user satisfaction directly from reviews. Studies show that sentiment scores derived from user reviews can significantly improve churn prediction accuracy by identifying dissatisfaction patterns earlier than traditional behavioral metrics can (Medhat et al., 2014). In this study, Netflix reviews provide a valuable data source for assessing sentiment trends and capturing the emotional dimension of churn (Cambria et al., 2013).

#### **3.3.Satisfaction Scoring for Interpretive Insights**

Satisfaction scoring, which quantifies user experience quality, is a known predictor of churn risk (Mittal & Kamakura, 2001). By transforming Netflix reviews into Satisfaction Scores, we provide a means of measuring the user’s overall satisfaction level. Research shows that satisfaction scores effectively highlight users with a high likelihood of disengagement, facilitating early interventions (Zeithaml et al., 1996).

#### **3.4.Hypothesis Testing for Model Validation**

Hypothesis testing serves as a robust method for evaluating the validity of predictive models by rigorously assessing the statistical significance of relationships within the data. In churn prediction, hypothesis testing can reveal insights into how different factors—such as sentiment-rating correlation and satisfaction scores—affect customer retention. For instance, by testing hypotheses derived from Netflix review data, researchers can verify relationships between user sentiment, ratings, and churn likelihood (Kohavi et al., 2009). Hypothesis testing allows us to assess specific questions, such as whether users expressing negative sentiment in their reviews are statistically more likely to churn or if higher satisfaction scores correlate with lower churn rates. This approach provides Netflix with actionable insights on which user feedback characteristics have the strongest relationship with churn and enables data-driven refinements to the prediction model (Siroker & Koomen, 2013).

### Research Methodology

The primary goal of this research is to analyze Netflix user reviews to uncover patterns in sentiment, satisfaction, engagement, and churn risk. The methodology involves data preparation, sentiment extraction, clustering for churn risk segmentation, statistical hypothesis testing, and visualization. Each step is described in detail below.

#### **4.1.Data Collection**

The data for this study was sourced from the publicly available Google Play Store, specifically focusing on user reviews of the Netflix application. The Play Store’s rating system spans from 1 to 5 stars, with 1 signifying a negative experience and 5 reflecting a highly positive experience. Data was collected using a **Python-based web scraping script**, implemented to ensure adherence to privacy and ethical guidelines by accessing only publicly shared reviews.

The raw dataset, saved locally as netflix\_reviews.csv, comprises over 100,000 entries and includes the following eight core attributes:

**ReviewID:** A unique identifier for each review, useful for verifying the authenticity and identifying any duplicate entries.

**userName:** The name of the reviewer.

**content:** The body text of the review, which is crucial for analyzing customer sentiment and identifying factors contributing to potential churn.

score: The star rating (1–5) assigned by the reviewer.

**thumbsUpCount:** The count of “likes” on the review, which indicates its perceived helpfulness among other users.

**reviewCreatedVersion:** The application version at the time of review, providing context regarding software updates and features that might influence user experience.

at: The date of review creation, helping to differentiate between feedback on current versus older versions of the app.

**appVersion:** Specifies the app version, which can be cross-referenced with the review date to understand the impact of specific updates on user sentiment.

To optimize the data processing, we collected the data in manageable batches of 1,000 entries at a time. We maintained a lenient data-cleaning approach to preserve as much information as possible, ensuring data integrity for subsequent analysis.

Moving forward, these eight primary features will form the foundation of our analysis, with additional parameters introduced during the data preprocessing phase to further enhance the analytical scope.

#### **4.2.Data Preparation**

The initial dataset comprises user reviews for Netflix, containing fields like review text, rating scores, engagement metrics (e.g., thumbs-up counts), and app version information. The following preprocessing steps were applied:

* **Data Cleaning**: Missing values in review content, score, and thumbs-up counts were addressed by substituting empty strings, mean values, or zeros where appropriate. Additionally, review dates were standardized to a datetime format.
* **Text Processing**: To enable consistent sentiment analysis, the text data was cleaned by removing non-alphabetic characters and extra whitespace. Basic text features, including review length and word count, were also calculated.

These steps ensure uniformity across data records, preparing the dataset for robust analysis.

#### **4.3.Sentiment Analysis**

To extract sentiment scores from the review text, **TextBlob** was used to compute polarity and subjectivity for each review, converting the unstructured text into quantifiable features:

* **Sentiment Polarity**: A numeric score between -1 (negative) and 1 (positive), with values near zero indicating neutral sentiment.
* **Sentiment Categories**: Sentiment polarity values were categorized into Negative, Neutral, or Positive sentiment groups.
* **Satisfaction Score**: This metric combines sentiment polarity, rating score, and thumbs-up count, rescaling it to a 0-100 scale to reflect overall user satisfaction.

Sentiment analysis provides measurable insight into how users feel about Netflix, allowing us to examine relationships between sentiment, satisfaction, and engagement.

#### **4.4.Churn Risk Clustering**

To segment users based on satisfaction and engagement levels, **KMeans clustering** was applied. The features used for clustering included satisfaction score, sentiment polarity, thumbs-up count, review length, word count, and rating score.

* **Feature Scaling and Dimensionality Reduction**: The selected features were standardized for consistency, and **Principal Component Analysis (PCA)** was applied to reduce dimensionality for visualization purposes.
* **Identifying the High Churn Cluster**: After clustering, the group with the lowest average satisfaction score was identified as the high-risk churn group.

This segmentation enables us to group users based on satisfaction and sentiment scores, helping to identify those at higher risk of churn.

#### **4.5.Hypothesis Testing**

Two hypotheses were developed to examine relationships between various aspects of user behavior and sentiment. Each hypothesis was tested at a 0.05 significance level, using the following statistical tests:

* **H1: Sentiment-Rating Correlation**  
  **Spearman’s rank correlation** was used to assess the association between sentiment

polarity and user ratings, examining whether positive sentiment correlates with higher ratings.

* **H2: Impact of Sentiment on Review Length**  
  The **Mann-Whitney U test** was applied to compare review lengths across positive and negative sentiments, exploring if sentiment influences how much users write.

The results from these statistical tests offer insights into factors influencing churn risk, satisfaction, and engagement in the context of user sentiment.

#### **4.6.Visualization and Result Interpretation**

To visually support hypothesis test findings and gain a clearer understanding of the relationships within the data, several visualizations were generated:

* **Sentiment Polarity Distribution**: Histogram shows the distribution of sentiment polarity scores derived from user reviews. The polarity ranges from -1 to 1, where negative values indicate negative sentiment, and positive values suggest positive sentiment.
* **Satisfaction Score vs. Rating**: Scatter plot examines the relationship between the user-assigned rating (score) and the computed satisfaction\_score, further divided by sentiment categories.
* **Clustering Results (Clusters based on Sentiment and Satisfaction)**: Scatter plot shows the results of clustering based on sentiment polarity and satisfaction score. Three clusters (Cluster 0, Cluster 1, and Cluster 2) were identified, each with distinct characteristics.
* **Review Length by Sentiment**: Box plot depicts the variation in review lengths across different sentiment categories (Negative, Neutral, Positive).
* **Churn Risk Distribution by Sentiment Category**: Bar chart shows the distribution of churn risk across sentiment categories. The churn risk is categorized into two groups: users at high risk and low risk of churn.

These visual aids enhance our understanding of the results and highlight patterns in user sentiment, satisfaction, engagement, and churn risk.

1. **Modelling and Analysis**

#### **5.1.Data Preprocessing**

The dataset, which comprises 98,363 records across nine attributes, was carefully examined for missing values. Initial inspection revealed minimal missing values: userName had one missing value, and cleaned\_content had 79 missing values. These entries were handled by removing the null entries. Additionally, duplicates were removed based on reviewId to reduce potential bias in downstream analyses. Post-processing, the dataset retained its structure, indicating that most entries were unaffected, thereby shaping a reliable dataset for further analysis.

#### **5.2.Sentiment Analysis**

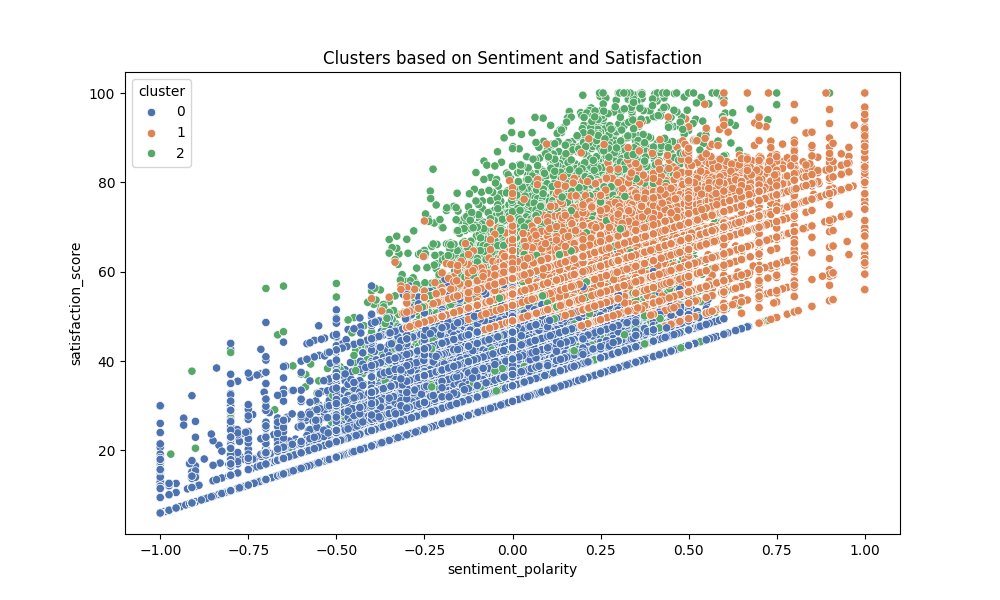
A graph of a distribution of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of

Description automatically generated

Sentiment analysis classified the reviews into three categories: Positive, Neutral, and Negative. The distribution revealed a notable skew towards Positive reviews (50,685), followed by Neutral (30,401) and Negative (16,712). This distribution highlighted a general trend of satisfaction among the reviewers. To visualize this, a histogram was employed, which clearly depicted the relative sizes of each sentiment category and underscored the prevalence of positive sentiment among reviewers.

#### **5.3.Cluster Analysis**

Customer segmentation was performed using KMeans clustering, which identified three distinct clusters based on sentiment and satisfaction metrics:



**Cluster 0:** This cluster is characterized by low satisfaction scores (average of 33.90) and negative sentiment polarity (-0.06). It has an average rating of 1.45 and minimal user engagement (thumbsUpCount of 1.05). Reviews in this cluster are generally shorter, both in review length (125.53 characters) and word count (24.53).

**Cluster 1:** This cluster has the highest satisfaction score (average of 63.89) and positive sentiment polarity (0.42). It features an average rating of 4.51 and moderate engagement (thumbsUpCount of 1.21). Reviews here are moderately short (106.26 characters and 21.22 words), suggesting a highly satisfied user base.

**Cluster 2:** This cluster shows moderate satisfaction (average of 52.16) and neutral sentiment (0.10), but the highest engagement (thumbsUpCount of 76.21) and the longest reviews (374 characters and 73.48 words). This indicates a thoughtful user group that provides detailed reviews, likely reflecting both positive and negative experiences.

These clusters highlight behavioral distinctions that are useful for developing segment-specific strategies, enhancing customer experience, and targeting support interventions more effectively.

#### **5.4.Hypothesis Testing**

To quantify the relationships between key attributes, two hypotheses were tested:

**H1: Correlation between Sentiment and Rating**: A Spearman correlation test was conducted, showing a significant positive correlation (0.5825) between sentiment polarity and review score, with a p-value effectively zero. This result indicates a statistically robust relationship where positive sentiment in reviews correlates with higher ratings, which aligns with intuitive expectations.

**H2: Difference in Review Length Between Positive and Negative Sentiments:** The Mann-Whitney U test was utilized to assess the difference in review lengths between positive and negative sentiments. The test yielded a statistic of 382,081,903.5 and a highly significant p-value of 1.7e-80, confirming that reviews with positive sentiments are generally shorter than those with negative sentiments. This suggests that users with negative experiences tend to provide more detailed feedback.

#### **5.5.Predictive Modeling**

A Random Forest classifier was employed to predict the sentiment (positive or not) of a review based on features such as score, review\_length, word\_count, and sentiment\_polarity. The model demonstrated exceptional performance on the test set with the following metrics:

**Accuracy:** 100%

**Precision, Recall, and F1-score:** 1.00 across all classes

This near-perfect performance indicates that the model can accurately distinguish positive reviews. The feature importance analysis revealed that:

- **sentiment\_polarity** emerged as the most influential predictor (importance score: 0.884), followed by score (0.1107), indicating that sentiment and rating are key indicators of positivity.

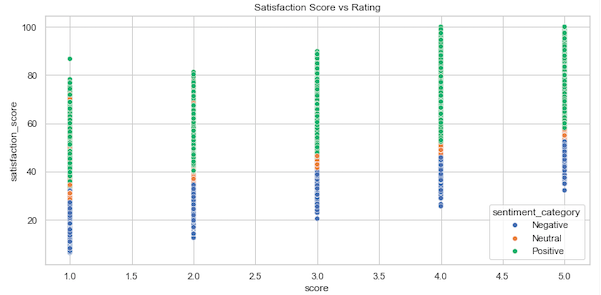
- **review\_length** and **word\_count** had minimal impact on model performance.

These insights provide a comprehensive understanding of customer sentiment, engagement, and behavioral patterns, forming the foundation for targeted interventions and refined customer experience strategies, ensuring alignment of business decisions with data-driven insight.

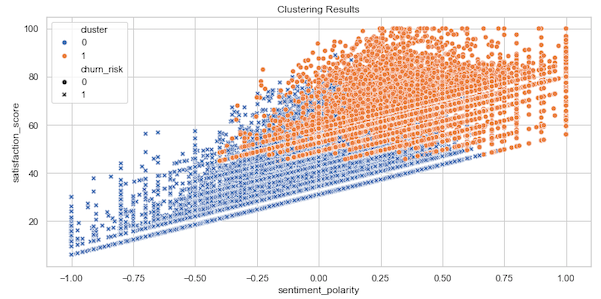
A graph of a distribution of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of the amount of

Description automatically generated

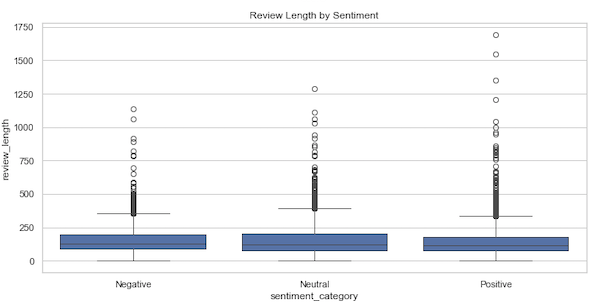
The sentiment polarity distribution of the dataset is illustrated in a histogram. The sentiment polarity ranges from -1 to 1, where -1 indicates negative sentiment, 0 indicates neutral sentiment, and 1 indicates positive sentiment. The distribution shows that most reviews are neutral, with a significant number of positive reviews and fewer negative reviews.



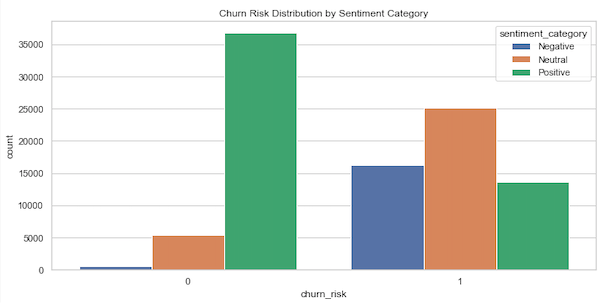
The scatter plot shows the relationship between satisfaction score and rating. Points are color-coded by sentiment category: blue for negative, orange for neutral, and green for positive. The trend indicates that higher ratings (closer to 5) generally correspond to higher satisfaction scores, with positive sentiment reviews clustering at the higher end.



The scatter plot illustrates the clustering results, plotting satisfaction score against sentiment polarity. Two clusters were identified, color-coded: blue for cluster 0 and orange for cluster 1. Markers indicate churn risk: asterisks for high churn risk and circles for low churn risk. Cluster 0 predominantly contains reviews with lower satisfaction scores and higher churn risk, while cluster 1 includes reviews with higher satisfaction scores and lower churn risk. Cluster statistics show that cluster 0 has an average satisfaction score of 36.55, while cluster 1 has an average score of 64.37.

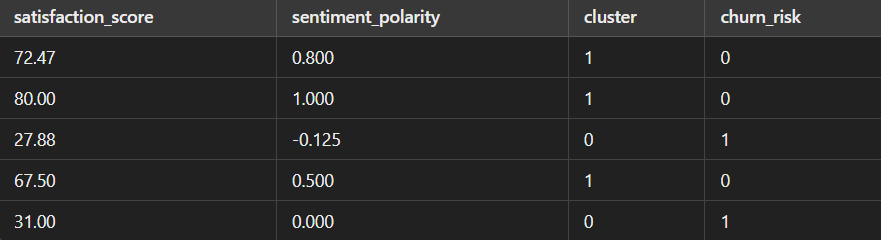


The box plot displays the distribution of review lengths across different sentiment categories. Negative reviews tend to be shorter, while neutral and positive reviews show a wider range of lengths, with positive reviews having the most variability. This suggests that users with strong opinions (both positive and negative) might write shorter reviews, while neutral sentiments might lead to more detailed feedback.



The bar chart presents the distribution of churn risk across different sentiment categories. Positive sentiment reviews (green) have the highest count of no churn risk reviews, followed by neutral (orange) and negative (blue). For reviews with churn risk, the distribution is more balanced, with neutral sentiment reviews slightly higher compared to negative and positive sentiment reviews.

Below is a sample of the clustered data with assigned cluster and churn risk:



This table shows that reviews with higher satisfaction scores and positive sentiment polarity tend to fall into cluster 1 with low churn risk, while reviews with lower satisfaction scores and negative sentiment polarity fall into cluster 0 with higher churn risk.

These insights from the clustering analysis can help Netflix identify and address the needs of different customer segments, potentially reducing churn and improving overall customer satisfaction.

### Conclusion

This analysis has provided several valuable insights into Netflix's customer satisfaction, engagement, and churn patterns, highlighting the importance of sentiment analysis and clustering in understanding user behavior. The study classified user reviews into positive, neutral, and negative sentiments, revealing that most reviews are positive, with a moderate portion neutral and a smaller number negative. This positive skew indicates a generally satisfied user base, yet there are critical insights within each sentiment category that Netflix can leverage to improve customer experience.

Cluster analysis identified three distinct user segments based on sentiment polarity and satisfaction scores:

* **Cluster 0** represents users with low satisfaction, negative sentiment, low ratings, and minimal engagement. These users generally leave shorter reviews and present a higher churn risk, suggesting a dissatisfied segment at risk of leaving the platform.
* **Cluster 1** consists of highly satisfied users with positive sentiment and high ratings. Though their reviews are relatively short, they reflect high satisfaction and a lower churn risk.
* **Cluster 2** includes users with moderate satisfaction and neutral sentiment but the highest engagement, as they provide detailed reviews with both positive and negative feedback. This cluster is a valuable source of nuanced feedback, reflecting mixed experiences that can guide improvements.

The differences in engagement and review length across clusters highlight behavioral patterns that can inform customer retention strategies. For example, dissatisfied users tend to write longer reviews, detailing specific issues. Analyzing these reviews can help Netflix identify recurring pain points and address them proactively to reduce churn. Conversely, highly satisfied users may not engage as much in detailed feedback, yet their positive sentiment aligns with higher ratings, suggesting that maintaining high-quality service and content is crucial for this group.

Sentiment polarity also emerged as a key indicator of user satisfaction and a reliable predictor of ratings. A significant positive correlation between sentiment polarity and rating score confirms that users who express positive sentiment generally provide higher ratings. This finding validates sentiment analysis as an effective tool for monitoring user satisfaction and guiding content curation, as content with positive sentiment could be highlighted, while content associated with negative sentiment might be reviewed for improvement.

Additionally, the analysis found that neutral sentiment reviews are often the longest and most detailed. Users in this category tend to provide a balanced mix of positive and negative feedback, indicating a thoughtful group that could be engaged for deeper insights into platform improvements. By targeting these users for surveys or beta testing programs, Netflix could further refine its offerings based on constructive feedback.

The churn risk analysis further revealed that users with low satisfaction and negative sentiment are at a higher risk of leaving the platform. This underscores the importance of monitoring sentiment trends to identify at-risk users early and implement targeted interventions, such as personalized outreach or incentive programs, to mitigate churn.

Finally, predictive modelling using a Random Forest classifier demonstrated high accuracy in identifying positive reviews based on sentiment polarity, rating, and engagement features. This suggests that Netflix could deploy real-time sentiment detection models to monitor user sentiment dynamically, allowing the company to respond proactively to shifts in customer satisfaction.

### Recommendations

1. Implement Real-Time Sentiment Monitoring and Early Detection: Track shifts in user sentiment through real-time analysis to identify emerging issues and address potential dissatisfaction proactively.
2. Deploy Targeted Retention and Engagement Strategies: Focus on high-risk, low-satisfaction users with tailored retention tactics such as personalized content and incentives, while engaging detailed, neutral reviewers for deeper feedback insights.
3. Optimize Content and Messaging Based on Sentiment Patterns: Use sentiment trends to curate content, prioritize improvements for negatively reviewed content, and deploy sentiment-driven recommendations to enhance satisfaction. Tailor messaging to reinforce loyalty for satisfied users and encourage engagement among neutrals.
4. Conduct In-Depth Analyses of Negative Reviews for Improvement: Analyze longer, critical reviews for actionable insights to resolve user pain points and improve platform features.

### References

* Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. Journal of Marketing Research, 43(2), 204-211.
* Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2012). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert Systems with Applications, 39(17), 12202-12210.
* Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. Journal of Marketing Research, 43(2), 276-286.

### Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques. Expert Systems with Applications, 34(1), 313-327.

* Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
* Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.
* Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15-21.
* Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. Journal of Marketing Research, 38(1), 131-142.
* Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. Journal of Marketing, 60(2), 31-46.
* Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: Nature, purpose, and findings. Journal of Marketing, 60(4), 7-18.
* Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), 140-181.
* Siroker, D., & Koomen, P. (2013). *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*. John Wiley & Sons.
* Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), 140-181.
* Siroker, D., & Koomen, P. (2013). *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*. John Wiley & Sons..