# PREDICTING CUSTOMER CHURN IN STREAMING SERVICES

DECEMBER 10, 2024

AHMAD HAMOOD
WSU – DSA 6000 Professor Liu

**Abstract**

This report analyzes customer churn for two streaming platforms, Netflix and a Hulu-like streaming service. Utilizing logistic regression, I identified critical predictors of churn and evaluated the model's performance. The Netflix model achieved perfect accuracy, precision, and recall, with tenure and monthly fees emerging as the most significant features. For the Hulu-like dataset, class imbalance reduced model effectiveness, but tenure remained a driver of churn among more significant features. These findings emphasize the importance of long-term customer retention strategies and pricing considerations in mitigating churn.

**Background**

The introduction of streaming services during the 2000s has flipped the entertainment industry on its head. Since Netflix launched its streaming platform during the 2000s, the concept of watching TV and movies at home was no longer limited to cable or satellite TV providers. Instead of watching episodes live as they premiered each week, Netflix and incoming streaming services provided viewers with an alternative option of having viewers' favorite programs fully completed and available to stream. Subscribers were able to access this content on devices other than their television at any time without interruption. Furthermore, Netflix was cost-friendly; while the average TV plan costs around $80 a month, Netflix had a monthly subscription plan as little as $8 a month.

Netflix's new model of watching TV and movies changed daytime and nighttime viewing entirely. Since 2014, roughly 20 million subscribers have left cable and satellite TV altogether and the number is expected to grow to 80 million by the year 2026.

Meanwhile, since Netflix's inception, many streaming services have been introduced and gained popularity. In recent years, Hulu, Disney+, Amazon Prime Video, Max (previously known as HBO Max) have put out content to match Netflix to entertain their millions of subscribers. However, with greater competition, rising prices, and greater demand to fulfill, customer churn and retaining customers has become the rising issue for these streaming services.

Many streaming platforms like Hulu, Disney+, Amazon Prime Video, and even Spotify keep much of their customer data private. Netflix, however, has publicly available datasets and because of this has been historically helpful in the academic and data science community. Other platforms choose not to provide similar data because of privacy concerns and competitive advantage. Using the Telco dataset gives us the opportunity to compare against Netflix and find any similarities or differences among factors of churn. For the sake of clarity this report refers to the Telco dataset as a "Hulu-like streaming service" because it is highly representative of a subscription-based service like Hulu by including factors like monthly charges, contract types, payment methods etc.

Customer churn is a critical challenge for subscription-based industries, particularly in the competitive landscape of streaming services. Churn occurs when customers discontinue their subscriptions, resulting in lost revenue and increased costs associated with acquiring new users. Streaming platforms, such as Netflix and Hulu, rely heavily on recurring revenue streams, making it essential to understand the factors driving churn and develop effective retention strategies.

Churn analysis provides insights into customer behavior, highlighting areas where interventions such as promotional offers, content curation, or pricing adjustments could improve customer satisfaction and loyalty. As streaming services flourish, customers have more options, increasing their likelihood of switching platforms for better value and experience or remaining with their current one. This study aims to explore the significance of churn in two distinct datasets, providing actionable insights for strategic decision-making in the streaming industry.

## Methodology

The approach to this analysis involved several steps designed to ensure a robust and interpretable model for predicting customer churn in Netflix and Hulu-like streaming services. The primary focus was on utilizing logistic regression due to its simplicity, interpretation, and ability to measure the linear relationships between features and churn outcomes. The first stage involved data acquisition. Two datasets were employed: the Netflix dataset, containing features such as watch hours, tenure, and monthly fees, and the Hulu-like dataset, adapted from a Telco Customer Churn dataset, which included churn labels alongside features like tenure, contract type, payment method, and monthly charges.

Following data acquisition, preprocessing was carried out to prepare the datasets for analysis. Any missing values were adjusted to maintain integrity with our data, and categorical variables in the Hulu-like dataset were transformed into numerical representations using one-hot encoding. Both datasets underwent feature scaling using standardization, ensuring that all features were on comparable scales, an important step for the performance of logistic regression models. The datasets were then split into training and testing sets with an 80:20 ratio, providing a reliable basis for model training and evaluation.

The logistic regression model was chosen for its ability to generate interpretable coefficients, which directly quantify the impact of each feature on the likelihood of churn. This method also enables understanding of the relative importance of predictors, offering practical insights for business strategies. Unlike more complex machine learning models, logistic regression ensures transparency, which is crucial for stakeholders seeking clear explanations for the drivers of churn.

Model evaluation was conducted using multiple metrics, including accuracy, precision, recall, F1-scores, and confusion matrices, to assess the reliability of predictions and to identify any challenges. The coefficients from logistic regression were analyzed to determine feature importance, providing valuable insights into the key factors influencing churn for both datasets. Visualizations were generated to communicate findings effectively.

**Results**

The results of the analysis provided valuable insights into the factors influencing customer churn for both the Netflix and Hulu-like datasets. For the Netflix dataset, the logistic regression model exhibited a high level of accuracy. The most significant predictors of churn for Netflix were tenure and monthly fees. Tenure displayed a strong negative relationship with churn likelihood, suggesting that longer-tenured customers are less likely to churn. Monthly fees showed a smaller positive coefficient, indicating that higher fees slightly increased the likelihood of churn. Watch hours, although included as a feature, were not a significant factor to the model's predictions.

For the Hulu-like dataset, the logistic regression model's performance was less robust due to class imbalance, with an accuracy of 0.67. Precision, recall, and F1-scores reflected this challenge, particularly for the churn class, which was poorly predicted. The key predictor for churn in the Hulu-like dataset was tenure, mirroring the Netflix findings. However, other features such as contract type and payment method were found to have minimal impact on churn predictions, as reflected by their near-zero coefficients.

The factors for both datasets were visualized using bar plots, which highlighted the dominance of tenure in predicting churn. These visualizations provided an intuitive understanding of the relative importance of the features in the model. The results emphasize the effectiveness of logistic regression in identifying churn factors in datasets with clear linear relationships, while also highlighting its limitations when dealing with class imbalance, as observed in the Hulu-like dataset. These findings offer practical takeaways for customer retention strategies, particularly the importance of encouraging long-term customer relationships to reduce churn rates.

**Conclusion**

This study sheds light on the factors influencing customer churn in subscription-based streaming services, offering insights that can guide retention strategies. For Netflix, the analysis identified tenure and monthly fees as key contributors to churn, stressing the importance of encouraging long-term customer engagement and ensuring pricing models remain competitive. In the Hulu-like streaming service, the model's limited performance due to class imbalance underscores the challenges of predicting churn in imbalanced datasets, yet tenure remained a consistent and strong predictor across both platforms.

The findings emphasize the need for streaming services to focus on customer retention tailored to their unique business models. By understanding the drivers of churn, platforms can develop targeted strategies, such as loyalty programs or pricing adjustments. This analysis provides a step toward leveraging data to inform decisions that could improve customer loyalty and reduce churn.

References

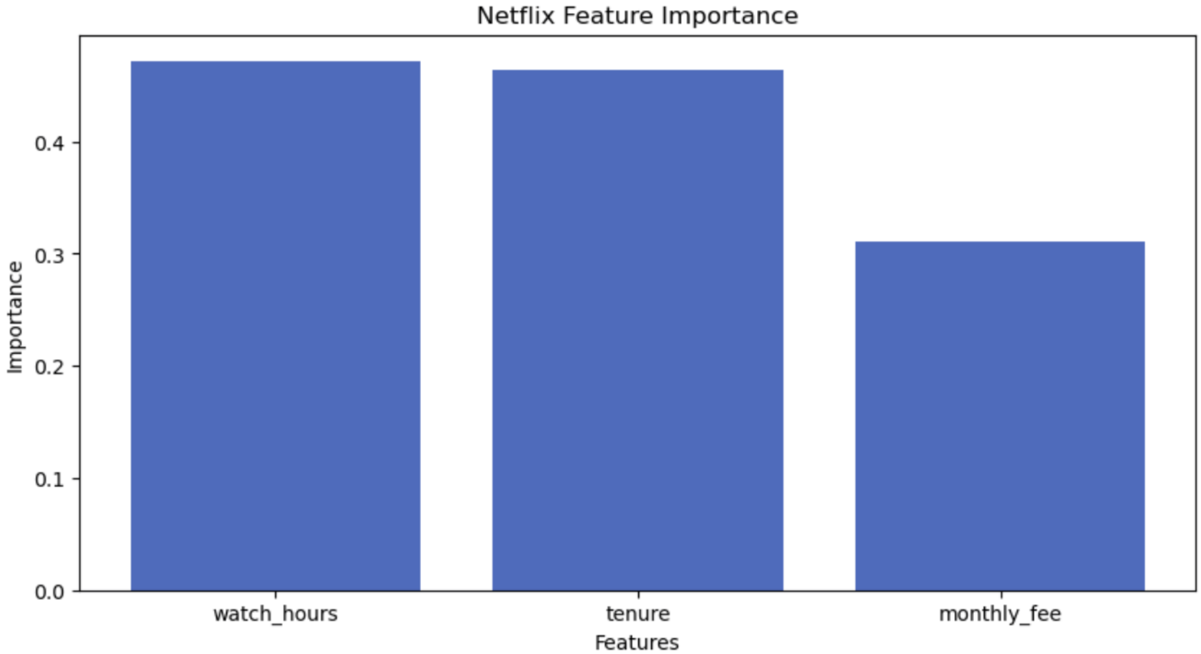

Amin, H. (n.d.). Customer Churn Dataset. Retrieved from https://www.kaggle.com/datasets/hassanamin/customer-churn

Gigasheet. (n.d.). *Netflix Dataset*. Retrieved from https://gigasheet.com/sample-data/netflix-dataset

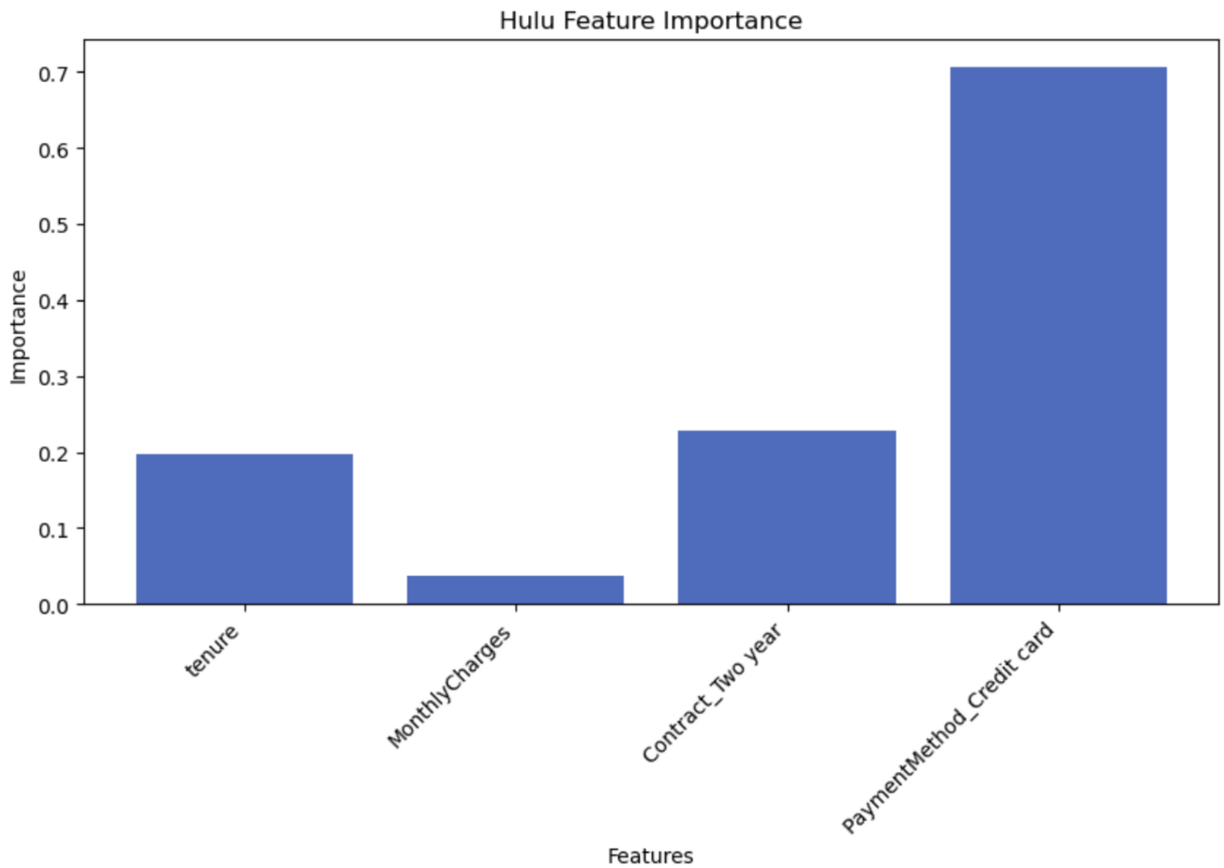
Howarth, Josh. “30+ Cord Cutting Statistics (2024-2027).” *Exploding Topics*, Exploding Topics, 21 Feb. 2024, explodingtopics.com/blog/cord-cutting-stats.

*How Much Is Cable per Month? | Cabletv.Com*, www.cabletv.com/blog/how-much-should-i-pay-for-cable-tv. Accessed 10 Dec. 2024.

Appendix




This graph shows which features most influence customer churn predictions for Netflix. The most important features are the number of hours watched and how long the customer has been subscribed, with subscription fees playing a smaller role.




This graph shows the key features driving churn predictions for a Hulu-like streaming service. The customer's payment method has the most significant impact. Contract type and tenure are less significant and monthly charges appear insignificant.