Using Machine Learning to Predict and Reduce Customer Churn in OTT Platforms: A Practical Approach with Behavioural Data

[1]Bethza Eype
[1]Student
[1]Department of Commerce
[1]St. Joseph's University, Bangalore
[1]bethzaeype@gmail.com

Abstract—**The digital streaming market has become highly competitive with the introduction of Over-the-Top (OTT) platforms such as Netflix, Amazon Prime Video, and Disney+. Customer retention is a major challenge for the digital streaming industry. In this paper, the authors explore how to use Machine Learning (ML) techniques to predict and reduce customer churn from customer behaviour. ML models can accurately identify at-risk users by analysing their behaviour including viewing patterns, time spent and interaction history. But predicting is not enough. There needs to be timely and individual intervention strategies. This paper outlines a practical approach that combines real-time data monitoring, continual model updates, and actionable insights, like targeted content recommendations and promotional offers. This study also highlights the potential of predictive analytics and adaptive retention measures to boost user satisfaction, curb churn, and drive sustainable profitability for OTT service providers.**

I. Objectives

1. To explore how Machine Learning can be used to predict customer churn in OTT platforms.
2. To analyse user behaviour patterns including viewing history, engagement and interactions
3. To determine the factors that lead to customer churn
4. To evaluate different Machine Learning models used for churn prediction
5. To explore the implications of predictive insights and how to turn them into actionable retention strategies
6. To find a viable solution for lowering churn and adapt systems in real time

II. Introduction

Machine learning is a field of Artificial Intelligence that allows a computer to learn from experience and to make decisions or predictions without being explicitly programmed. It does this by extracting patterns from large quantities of data, such as user behaviour, and making predictions based on those patterns. Simply put, machine learning makes it possible for systems to learn and get better over time as they are exposed to more data.

Online streaming services that deliver media directly to the consumer over the internet without the need for cable or satellite services are known as OTT (Over-the-top). And churn is when customers cease using a service or subscribe. So, OTT churn is when customers either cancel their subscription or stop using streaming services, which impacts the business growth and income.

Machine learning can enable OTT to understand user preferences like viewing history, video viewing time and user interactions. This data can be used to identify users who are likely to churn, by analysing this information. This allows companies to proactively engage users with relevant content and offers, leading to better retention and lower churn.

Churn is a major problem for OTT platforms as it impacts both revenue and market share. It is usually cheaper to retain an existing user than it is to acquire a new user due to the expense of marketing and promotions, as well as the cost of onboarding. Companies that do not manage churn may be missing out on profitability for their efforts to attract new customers. So minimising churn is crucial for growth.

III. Literature review

In many sectors, the topic of forecasting customer attrition with machine learning has received a lot of research. A study by Verbeke et al. (2012) found that predictive models based on historical customer data will be able to detect churn patterns and enhance the strategies to retain customers. Research on machine learning for customer churn prediction shows that it is crucial to predict user behaviour to improve customer retention and business performance. Churn models are created from customer data from the past and are used to understand trends and predict customers that are likely to churn from a service. This is even more crucial for OTT service providers like Netflix as customer retention is one of the key factors contributing to their revenues and growth. Making more accurate predictions, however, requires some knowledge of the reasons for leaving, and of who is likely to leave.

Likewise, Huang et al. (2019) point out that behavioural analytics can enhance the accuracy of models. Kumar and Ravi (2021) also points out that watch time and content preferences are the metrics used by users and are good predictors of churn for OTT platforms. As churn increases, these metrics tend to go down.

Table 1: Comparison Between Traditional and Advanced Churn Prediction Systems

Aspect	Traditional Systems	Advanced ML Systems
Data Usage	Uses historical (old) data	Uses real-time data
Model Type	Static models	Continuously updated models
Purpose	Predicts churn only	Predicts churn + suggests actions
Personalization	Limited or generic	High personalization
Actionability	No action after prediction	Proactive retention actions
Adaptability	Does not adapt to changing behavior	Adapts to changing user behavior
Outcome	Lower retention rate	Higher retention rate & business growth

The analysis of user behaviour is essential for OTT services. To build on this, many papers investigate the analysis of user behavior in OTT platforms, where users watch history, watch time, how often the user uses the platform, and the type of content that they watch are considered.

It has been said that if watching time goes down or the interaction to the recommendations is not seen, then there is a possibility of churn. This highlights the need for user behaviour in churn detection and may be helpful when creating churn prediction models.

According to Gomez-Uribe and Hunt (2015), recommendation systems are crucial because they help Netflix to boost user retention by making recommendations for each user. Recommendation systems are used to recommend content to users according to their interests. Research studies have indicated that by integrating the churn prediction system into the recommendation system, it is possible to increase the retention of users by giving personalised recommendations to the users. One study even adds machine learning to recommendation systems and chatbots to provide personalised retention strategies, such as promotions or content recommendations.

In addition, Burez and Van den Poel (2009) recommend that predictive models be combined with proactive retention for improved business results. Before the development of more sophisticated machine learning techniques, companies relied on traditional methods such as basic data analysis, surveys and manual segmentation. These methods enabled them to learn about general patterns of use, but were not efficient in handling large amounts of data and were less accurate and quick. With the growth of OTT services, these techniques were no longer sufficient and more advanced techniques were used.

Consequently, recent studies are increasingly relying on machine learning techniques like logistic regression, decision trees and random forests for churn prediction. These are capable of processing a significant amount of data and can detect subtle trends in user behavior, resulting in more accurate predictions. Machine learning is a better and scalable solution than traditional methods for processing user data in OTT services.

However, there is still a need for improvement. Most research mainly focuses on predicting which users will churn, but it does not offer specific insights into what to do following the prediction.

As mentioned by Neslin et al. (2006), however, many systems are only concerned with the prediction and do not take any effective action after the prediction, which is a huge research gap.

IV. Understanding about Customer Churn in OTT platforms

Churn happens when a customer stops using a service or when a customer subscribes to a service. For OTT services such as Netflix or Amazon Prime Video, churn happens when users cancel their subscriptions. It has a direct impact on the platform's revenue and growth, and therefore needs to be tackled.

There are two types of churn in OTT platforms. The first is voluntary churn, in which customers opt to cancel their subscription, usually because they are not satisfied with the experience or they simply don't care. The second is involuntary churn, where users are not able to use the service due to factors outside their control, like payment issues or expired credit cards. Voluntary churn can be a problem with the user experience, while involuntary churn could be due to technical glitches or a system outage. There are a number of important reasons for churn in OTT platforms. One of the main reasons is the lack of content selection and users might not be able to see content they are interested in (like film or TV shows) and thus are more inclined to churn.

Also, subscription price plays a significant role, as the user may think that the content is not worth the subscription price, especially when paying for multiple subscriptions. Also, the competition with more attractive content or lower cost may sway users.

Finally, the recommendation algorithms can also play a role in churn if users fail to find content of interest, which can result in dissatisfaction and disengagement.

V. Machine Learning in Churn Prediction

One of the most important factors to consider in retaining customers of over-the-top (OTT) media platforms is predicting customer churn, and machine learning plays a crucial role in this process. Instead of making assumptions, machine learning models watch how users interact with the platform and use that information to identify which ones might be considered "churners. This enables companies such as Netflix to proactively connect with their customers.

Table 2: Types of Data Used in Churn Prediction

Structured Data (Quantitative)	Unstructured Data (Qualitative)
Watch Time	User Reviews
Subscription History	Social Media Comments
Viewing Frequency	Feedback Messages
Login Activity	Ratings & Opinions
Payment Information	Customer Complaints
Device Type	Forum Discussions
Geographic Location	Video/Audio Interactions

Machine learning models use various kinds of data to draw these predictions. These can include user activity (time spent on the platform) and viewing history (what users are interested in). History of searches can also identify user interests and what they may not be able to find and subscription history can give information about payment methods and previous subscription cancellations. This provides a comprehensive view of user activity and enhances the accuracy of predictions.

With this data, there are several machine learning models that can predict churn. Logistic Regression and Decision Trees are simple models that predict whether a user will churn or not and classify users based on their behaviour patterns, respectively. Random Forest is a more complex model that is an ensemble of decision trees to improve performance. These models can be combined to provide OTT services with valuable insights into user behavior and help maximize customer retention.

VI. Research gaps in existing systems

While there is increasing adoption of machine learning techniques in OTT services like Netflix, there are key gaps in its ability to minimise churn. The reason that current systems are not fully effective in handling churn is due to these gaps.

The first is that there is a disconnect between prediction and action. The existing systems are mainly focused on predicting churners but do not suggest what to do with the identified churners. They don't suggest actions like discounts, personalised suggestions or notifications. This implies that systems can detect users who are likely to churn, but they might not do the right thing to stop them from doing so.

One drawback is that static models are used. Machine learning models are usually not continuously updated, but rather rely on previous data. However, preferences are not static when it comes to OTT. This can lead to a concept drift problem, which is the loss of prediction power of the model. For instance, if a user watched a thriller last month, he is more likely to watch a comedy this month which makes the previous prediction less reliable.

In addition, there is a restriction on the use of data. Current models mostly rely on structured data such as viewing time and subscription data. This information is useful but it will not give the full story. Valuable data like social media reviews, customer complaints and ratings are often overlooked due to their unstructured nature. This information is hard to integrate and hence the usefulness of churn prediction algorithms is restricted.

There are also many systems that are once-trained models. These models are trained only once and they will not update their model to reflect new customer data, so they cannot learn about new trends. Rather, OTT platforms must be able to learn continuously and adjust to new data as it comes in.

These constraints show that machine learning based methods have a great potential but the current churn prediction methods need to be improved by integrating prediction with action, real-time information, multiple data sources and continuous learning.

VII. Findings and Recommendations

In order to overcome these shortcomings of the existing systems, there is a need to provide a more feasible and intelligent solution to reduce churn in OTT services such as Netflix. Rather than just focusing on predictive models, the solution is to develop a system that is dynamic, adaptive and actionable.

First, real-time data monitoring should be put in place, where user data such as viewing trends, search queries and engagement metrics are monitored in real-time. This way, the model can use the new data to make predictions instead of old data.

Second, it should be continually updated to allow it to respond to the evolving user behaviour, and provide more accurate predictions, rather than remaining locked on to the old knowledge.

Moreover, the system should emphasise integration of multiple data streams. It should not be limited to structured data (watch time, subscription data) but should also include unstructured data (user reviews, ratings, feedback). This enables an understanding of the user's needs and more accurate predictions.

More importantly, the system predicts and also provides actions. When a user is likely to churn, the system should take prompt action, by sending a set of personalized recommendations according to the user's preferences, by offering special offers or subscriptions, and by notifying the user of new releases or popular shows. This will help to regain their interest and increase the chance of keeping them.

VIII. Case Based Example

Suppose we have a user of an OTT service (e.g., Netflix) that consumes movies and shows. The system tracks the user's activities, such as the content they watch, how frequently they use the platform, and how much time they spend watching. This is used to train the system on the user's normal usage pattern.

After some time, the user starts to change his/her behavior. They watch less video, spend less time on the website or stop finding new video content. This decrease in use is noticed by real-time monitoring of the user's activity. The machine learning algorithm knows (from historical data of similar users) that this behavior is a precursor to churn and that this user is therefore likely to churn.

Figure 1: Example of Churn Prevention Using Machine Learning

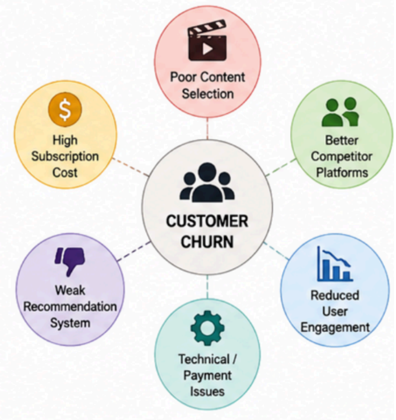
The system acts when the user is marked as a potential churner. It can suggest to the user recommendations like "Top picks for you", based on the user's previous interests. It may also offer a discount, a promotional subscription offer or notify the user of new content of interest to them. The goal of these actions is to re-engage the user and enhance the user experience.

This, in turn, captures the user's interest, brings him back to the site and restarts his subscription. The illustration above shows how machine learning can not only detect churn but also enable OTTs to take action to avoid it.

IX. Challenges

But there are certain difficulties in implementing machine learning on OTT platforms. A major concern is privacy, as platforms deal with a lot of user data, and users need to have trust in the platform if that data is not secure. Another challenge is the high cost of implementation, because machine learning models are costly to create, they need highly skilled personnel, complex technology and continuous maintenance. Then there's the data integration problem where platforms need to rely on a wide range of data types, such as user viewing history, feedback, and other data sources, which can be complex. Finally, there is the problem of the model's accuracy, which can be a problem if the preferences of the users change, so that the algorithms must be continually updated to be able to accurately predict the behavior of the user.

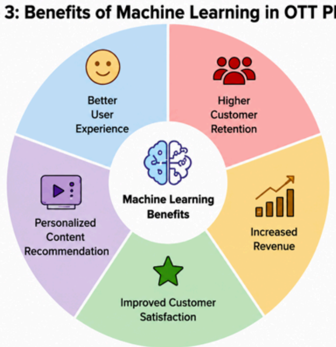
Figure 2: Major Causes of Customer Churn in OTT Platforms



X. The Benefits of Machine Learning

Machine learning is vital to OTT platforms such as Netflix because it can help to personalise and optimise services. For one thing, it improves the user experience because the OTT platform is more aware of what users like and dislike, and can better suggest content to them, making it easier for them to find what they want. This, in turn, will help to boost retention rates because customers are more likely to remain loyal when they receive their desired services. Retention results in higher revenues – it is easier and more profitable to keep customers than to acquire new ones. Moreover, machine learning can be used for personalised content recommendation, which means that the user will be presented with content according to his/her personal taste, thereby enhancing user satisfaction and stickiness.

Figure 3: Benefits of Machine Learning in OTT Platforms



XI. Conclusion

To sum up, Machine learning is an essential tool for OTT companies to gain insight into user behaviour and predict churn. But making predictions is not enough. Machine learning needs to be complemented with actionable and timely interventions such as personalized recommendations, discounts and engagement notifications, to be effective. If predictions are accurate, it may not lead to higher retention without these actions.

Moreover, OTT platforms need to shift from static to dynamic systems that can adapt to users' evolving interests. The interests of the users may vary quickly and the system should be able to learn and adapt itself. This will help to maintain the accuracy of predictions.

Last but not least, when multiple data sources (structured and unstructured) are integrated, user needs can be better understood and decisions made. While there are concerns regarding privacy, cost and complexity, these need to be addressed in the design and implementation of churn management systems.

To conclude, the future of OTT platforms lies in the development of smart, real-time and actionable machine learning systems that can predict churn but can also help prevent it, enhancing the user experience, retention and growth of the platform.

References

[1] Aldowah, H., Rehman, S. U., Ghazal, S., & Umar, I. N. (2025). *A comprehensive survey on customer churn analysis studies*. *Taylor & Francis*. https://doi.org/10.1080/24751839.2025.2528440

[2] Haenlein, M., & Kaplan, A. M. (2019). *Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential*. *Business Horizons*. https://doi.org/10.1016/j.bushor.2019.08.001

[3] Kawade, B. (n.d.). *Role of artificial intelligence and machine learning in decision making for business growth*. *ResearchGate*. https://www.researchgate.net/publication/368976486

[4] Khan, M. A., & Alharbi, F. (2023). *Customer churn prediction in the telecom sector using machine learning techniques*. *ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S2666720723001443

[5] Kumar, R., & Singh, P. (2021). *Customer churn prediction system: A machine learning approach*. *Springer*. https://doi.org/10.1007/S00607-021-00908-Y

[6] Patel, S., & Mehta, R. (2024). *Customer churn prediction: A systematic review of recent advances, trends, and challenges in machine learning and deep learning*. *MDPI*. https://doi.org/10.3390/bdcc7030105

[7] Sharma, V., & Gupta, A. (n.d.). *Customer churn prediction in the banking sector using machine learning-based classification models*. *EBSCO*. https://openurl.ebsco.com/EPDB:gcd:2:12052130/detailv2?sid=ebsco:plink:scholar&id=ebsco:gcd:163271336

[8] Singh, T., & Rao, M. (n.d.). *Enhancing user experience through machine learning-based personalized recommendation systems: Behavior data-driven UI design*. *TechRxiv*. https://doi.org/10.36227/techrxiv.173337559.97825928/v2

[9] Verma, K., & Joshi, R. (n.d.). *Predicting customer churn on OTT platforms: Customers with subscription of multiple service providers*. *Hrčak*. https://hrcak.srce.hr/291221

[10] Wang, Y., & Chen, L. (2018). *Recommendation strategies in personalization applications*. *ScienceDirect*. https://doi.org/10.1016/j.im.2018.01.002

[11] Yılmaz, E., & Demir, H. (n.d.). *The impact of recommendation systems on user experience in digital platforms: Netflix, Amazon Prime Video, and Disney+*. *DergiPark*. https://dergipark.org.tr/en/pub/kiad/article/1647277

[12] Zhang, Q., & Li, X. (n.d.). *The role of machine learning in transforming business intelligence*. *Academia.edu*. https://www.academia.edu/130473471/The_role_of_machine_learning_in_transforming_business_intelligence