How to effectively predict and increase your trial-to-paid user conversion rates

Index

Introduction

With increasing competition in the OTT and Pay TV space, **customer acquisition has become a more challenging and expensive** proposition for video services.

As it becomes harder to win new customers, it becomes even more important to retain those customers you already have.

The right conversion strategy is a growth accelerator and will have a significant impact on your business profitability.

In this white paper we will explain how Artificial Intelligence algorithms allow video service providers to **build and automatically run more accurate trial conversion prediction models, which** predict future conversion rates based on past audience behavior.

There are good reasons you should use machine learning to predict SVOD conversion rates in your acquisition process.

With more and more information about your video users' behavior, JUMP's machine learning algorithms become more intelligent and can help identify those trial users that are highly likely to convert into paying users. You can then focus your acquisition efforts on this segment, targeting these high-potential conversion candidates.

With today's technology, it's now possible to build vertical SVOD machine learning models specifically designed and implemented for the video industry. This means you will not only identify subscriber clusters with high conversion potential, but you will also be able to pinpoint the leading drivers for such successful conversion rates. You will therefore be able to take proactive steps to ensure it.

In this whitepaper we will show you how to build a machine learning-powered trial conversion prediction model that will learn from your video service historical data so that it can both identify what kinds of users are likely to convert and generate conversion probabilities for current trial users.

By feeding this data into your video service's CRM, customer support personnel will be able to focus on those users who are high-probability/high-value conversion candidates.

What follows is a step-by-step "how-to" guide for building your own model.

1. Collecting the right historical audience data is everything

The first thing we need to build the model is the available data. Historical data you can trust, data that is already cleansed, harmonized and normalized.

If you don't already have a data management platform where all your video service data sources are available and ready to be exploited, then this is first thing you need to address before entering into AI. Without a solid data repository, it is almost impossible to build machine learning models that work. Learn more **here.**

If you do have a working data management strategy for your business, then as you start to build your model you should at a minimum be able to collect data from the following two main blocks of data sources:

User activity

This data is related to the service's user subscription activity. It contains data related to registrations, start and end events related to subscription contracts, status changes (when a user converts from a trial status to a paying status) and product type (standalone/bundle) change events.

Below is an example of the mandatory user activity event data you should collect in order to build a solid trial conversion model.

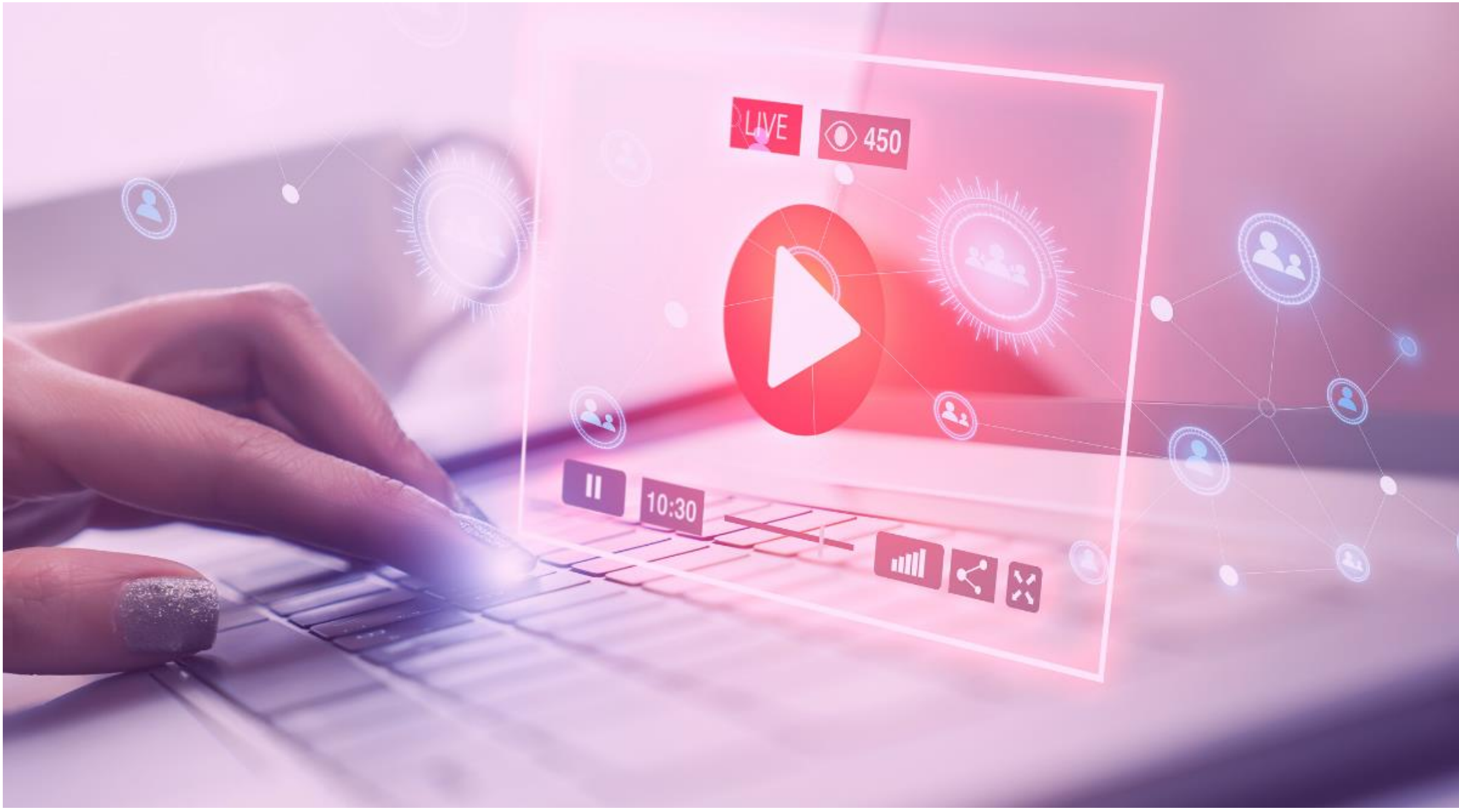
Event Type	Event	Description
Start	Registration	This event represents when a user registers with the video service.
End	Registration	This event represents when a user deletes his/her account from the video service.
Start	Suscription	This event signals the start of a specific commercial package subscription.
End	Suscription	This event signals the end of a specific commercial package subscription
StateChange	TrialSuscription	This event changes the state of a subscription to a trial subscription
StateChange	PayingSuscription	This event changes the state of a subscription to a paying subscription
ProductTypeChange	Bundle	This event signals that the type of subscription has changed to a bundled subscription
ProductTypeChange	StandAlone	This event signals that the type of subscription has changed to a standalone subscription
ProductTypeChange	Other	This event signals that the type of subscription has changed to an undetermined state
Start	Suspension	Indicates when a subscription moves to the suspended status
End	Suspension	Indicates when a subscription ends a suspended status

Playback activity

This information includes all the video service's playback sessions.

Below is an example of the mandatory playback activity event data you should collect in order to build a solid trial conversion model.

Field	Type	Description
customerid	String	Unique identifier for customer
brandid	String	Unique identifier for a brand in a customer
subscriberid	String	Unique identifier for a user in a brand
subscriptionid	String	Unique identifier of the subscription package under which the playback session occurred
hourdate	String	Date of the playback session start
daydate	String	Date of the playback session start
monthdate	String	Date of the playback session start
yeardate	String	Date of the playback session start
day	String	Day of month of the playback session start date
month	String	Month of the playback session start date
year	String	Year of the playback session start date
yearmonth	String	Playback session start date
countrycode	String	Country code where the playback session occurred
genres	String	comma separated list of content genres
network	String	content provider for the content
title	String	Original language title of the content
serietitle	String	title of the series in the viewer's native language
episode	String	episode number
season	String	season number
contenttype	String	type of content addressing customer's criteria (Series, Movie, Clip, etc.)
groupcontenttype	String	Content consumption type grouping (SVOD, TVOD, FVOD, LinearTV, etc.)
type	String	Type of event that initiated the playback session: AutoPlay, BingeWatch, PlaybackButtonClicked
playbacktime	Bigint	Duration of playback session in milliseconds (based on the time the content was actually running and viewed by the user)
contentduration	Bigint	Total duration of the content in milliseconds
bufferingtime	Bigint	Total time in seconds that buffering occurred during the playback session
bitrate	Bigint	Average bit rate at which the playback session was streamed
device	Bigint	Device model and brand
devicetype	Bigint	Device type: Web, Mobile, Tablet, SmartTV, STB, etc.
contentid	String	content unique identifier



2. Analysis of converted trial user attributes versus non-converted trial user attributes so you can identify the "signal" (what the data has to say)

Once you have the right data sources with the right level of data quality in your data management platform you can start to create your machine learning model.

As in the case of the data collection for the data management platform, creating a solid model requires an Artificial Intelligence framework where you can create, run and maintain large-scale machine learning models in an efficient and reliable way. Learn why this is so crucial here.

If you already have a reliable AI framework for building large-scale ML models, then you will be able to automate and run a workflow with the minimum number of steps, outlined below.

1- Tuning into the signal by using all user attributes provided in the user activity and playback activity data sources outlined above.

2- Select user attributes from step number 1) or combinations of attributes that demonstrate the signs that data distributions are different between users that converted from trial to paid subscriptions and the ones that did not convert.

3- Create features to feed the model meaning, using the user attributes where signal was detected. This is a very important process and will determine the predictive power of your model. Automating feature selection is a crucial part of the process because the potential feature – user attribute combinations could both vary extensively and grow exponentially, thus necessitating the automation of this process; it would be a near impossible manual endeavor.

Example attributes that we have seen to have predictive insight for trial conversion models are:

- Tenure
- Number of devices used
- Number of TV devices used
- Number of mobile devices used
- Engagement index
- Consumption-type affinity index
- Content-type affinity index
- Content genre affinity index
- Engagement index acceleration/deceleration
- Content-type affinity index acceleration/deceleration
- Content genre affinity index acceleration/deceleration
- Gender
- Age
- Geographic location

3 . Try as many algorithms as possible to create your trial conversion model and take a careful look at the performance metrics

Test different models using a year’s worth of data to train each model. Then test it on new, untested data from the next 6-month period.

Some modeling techniques you should try (at a minimum) include:

1. Logistic regression
2. Random forest
3. Gradient boost tree
4. Recurrent neural network
5. A combination of several models/techniques

An important and oftentimes undervalued task is the selection of one or more models that best perform for your business strategy.

Selecting the best performance metrics for the optimization strategy based on your business goals is not an easy task and requires in-depth discussions between your data science team and/or provider and the business managers.

An example of the analysis you should do is below:

(DISCLAIMER: You need to be familiar with basic performance metric concepts to fully understand the example below)

1. Are we looking to align our retention strategy with maximizing our video service profit? ➡ *BUSINESS IMPACT PRIORITIZATION* ➡ *FOCUS ON PROFIT MATRIX*

- Pros: Full alignment between the conversion model and the real business performance
- Con: It could be difficult to define economic impact of true positives, true negatives, false positives, false negatives

2. Looking to identify in advance as many future converted users as possible? ➡ *FOCUS ON RECALL PRIORITIZATION*

- Pros – Prioritization of number of real converted user identification
- Cons – More non-converted users will be impacted with conversion marketing actions

3. *Looking to minimize the impact of retention marketing actions on currently non-identified future churners while maximizing advance identification of churners?* — *FOCUS ON A PRECISION/RECALL BALANCE PRIORITIZATION (F1-SCORE)*

- Pros – More users that are impacted with conversion marketing actions are more likely to be converted users
- Cons – Lower level of real converted user pre-identification

Again, if you use an AI framework to try different models, assessing their performance and/or assembling (combining) several of them to select the ones with the best performance metrics, the task will be easier. If you do this manually, the time and effort required to produce reasonable results could outweigh the benefits.

Preparing your results for actions

Finally, it is very important that the outcome of your model can be easily used by your acquisition team in order to make a business impact.

Clustering the audience based on a trial conversion scoring will certainly help better focus the actions that should be taken per user group or cluster.

Using a clustering model (e.g.: K-Means) to group users based on their trial conversion scoring and assign traits to each cluster based on average values from all features is an effective way to focus on those segments most likely to convert. So, it provides you the information you need to feed your CRM and by extension, your customer support personnel, so they can then focus on high-probability/high-value conversion candidates.

Conclusions

- Artificial Intelligence algorithms can help providers **build and automatically run more accurate trial conversion prediction models, which** can predict future conversion rates based on past audience behavior.

- By feeding this information into your video service CRM, customer support personnel can then focus on high-probability/high-value trial conversion candidates which will undoubtedly increase your conversion rate, making the most of your acquisition budget.

- You need a data management platform where you have all your video service data sources available and ready to be exploited; otherwise, it is an almost impossible task to build machine learning models that work. Learn more **here.**

- Creating solid machine learning models requires you have an Artificial Intelligence framework to create, run and maintain large scale machine learning models and associated workflows in an efficient and reliable way. Learn why this is crucial **here.**

- Once you are able to build up your models, selecting the best ML model performance metrics for your optimization strategy, based on your business goals, is not an easy task. It will require in-depth discussion between data scientists and business managers.

- It is crucial that the output from your ML model can be easily used by your acquisition team so they can make a meaningful business impact. Audience clustering based on trial user conversion scoring will definitely help better focus those actions that should be taken per user group or cluster.

About JUMP

JUMP is an advanced business analytics solution that uses data to help video service providers gain valuable insights about their audience and content performance, predict churn to retain users, identify the clusters that video users belong to, and personalize the video experience. Jump is built on the foundations of cutting-edge big data and artificial intelligence technology customized for the video industry.

Our vision is that data and its effective use will be the new competitive advantage in the next phase of the video industry. Nowadays only big players like Netflix, Amazon, and Google use cutting-edge data technologies to compete in the video market: retaining customers and increasing revenues.

Do you want to learn more?

Contact Us:

info@jumptvs.com
www.jumptvs.com

Stay tuned

If you want to receive upcoming **whitepapers** from our **series**

Contact Us for more info or advice:

www.jumptvs.com
info@jumptvs.com

JUMP