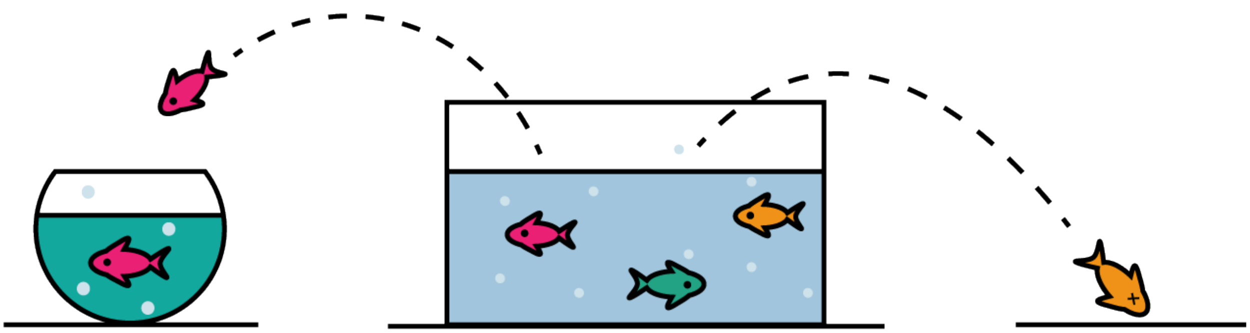
How to Predict Customer Churn

Predict churning customers and take proactive action



PerceptiveAnalytics

Marketing Analytics | Tableau Consulting | Excel Apps
New York • Dallas • San Francisco • Hyderabad
Perceptive-Analytics.com
(646) 583 0001 cs@perceptive-analytics.com

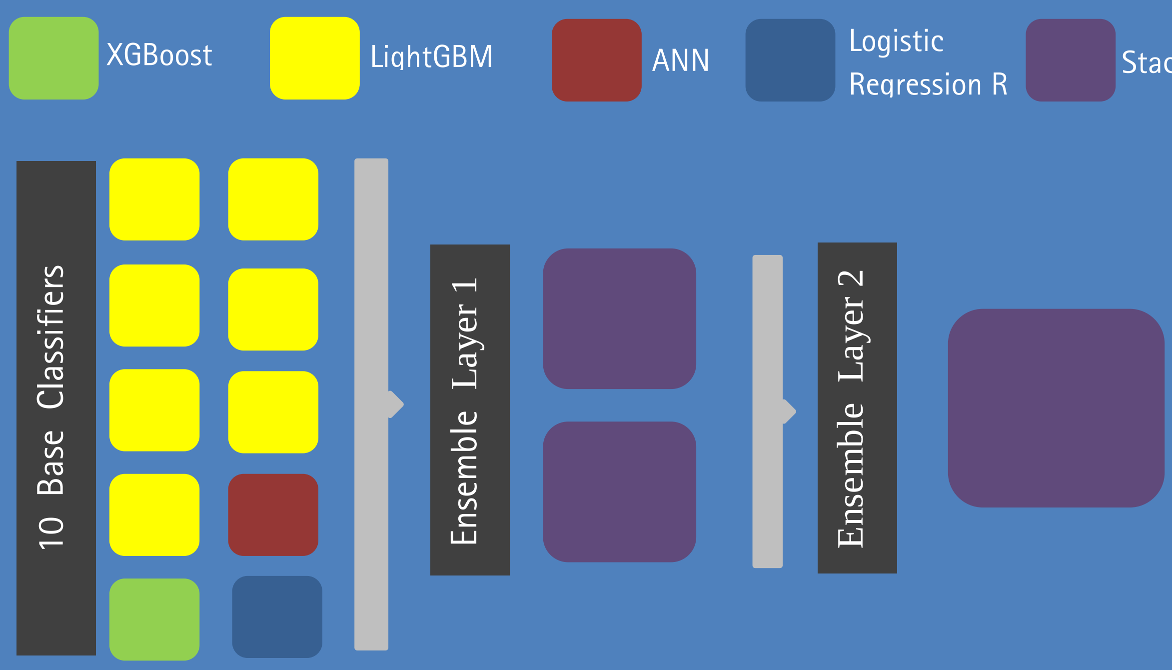
Executive Summary

Predicting customer churn gives the opportunity to stem the leak in revenue base. It has the same impact as making marketing engine more effective. Reducing churn has the following strategic benefits:

- Reduces marketing cost: Acquiring a new customer costs five times or more than that of retaining a customer.
- Provides rentention insights: Churn analysis can provide important cues on retention allowing you to keep a tab on customers changing needs and preferences.
- Foster long-term relationships and loyalty: By acting on insights from churn analysis, you remove bottlenecks and foster long-term customer relationships.

In this churn predicion case study for a music streaming service, we have found that user activity attributes did not identify churning customers but transactional attributes contain potential patterns that help identify customer churn. We developed 10 base models and a two layered ensemble models. The ensemble model was the best and it predicted customers who are likely to churn with an Accuracy of 96% and F1-Score of 86.5%.

Architecture of the Churn Predictive Model



Customer Churn Predictive Model

Company and Data

The data belonged to an online subscription music streaming service called KKBOX, was for 11 years and contained membership data (demographic), transactional data (like payments, renewals, cancellations, etc.) and daily user activity logs (number and percentage of songs heard). A user is said to be churned when she has not made a service subscription transaction within 30 days after current membership expiry date.

Objective

The objective of the project is to predict customer churn and take corrective actions.

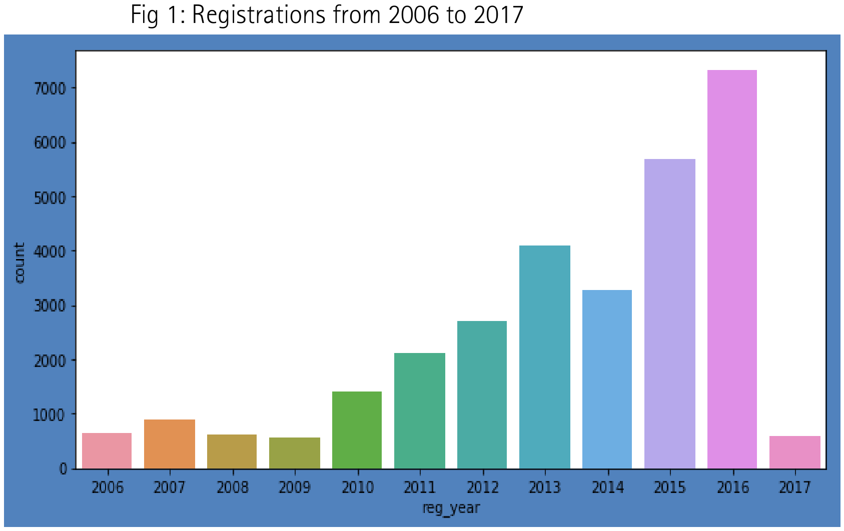
This project is structured in the following steps:

- Gain deeper understanding of customers through Exploratory data analysis (EDA).
- Develop a machine learning model to predict customer churn.
- Identify the factors that caused customers to churn.
- Recommend preventive business actions and decisions that help retain customers.

Exploratory Data Analysis

Through EDA, our goal was to find those attributes/ features that explain churn (EDA visuals below). The following figure shows the number of user registrations from 2006 to 2017.

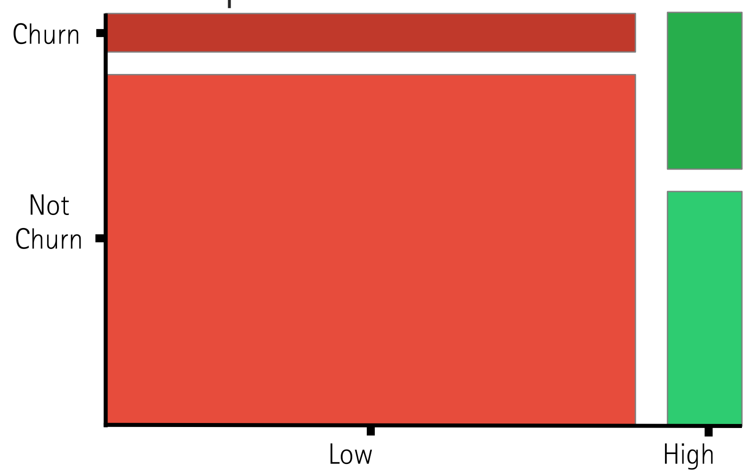
Fig 1: Registrations from 2006 to 2017



The mosaic plots visualizations below show churn / no churn on Y-axis and the area of each block represents the likelihood of churn, with a specific type of attribute defined on X-axis.

Visual Analysis of Transactional and User-Activity Features.

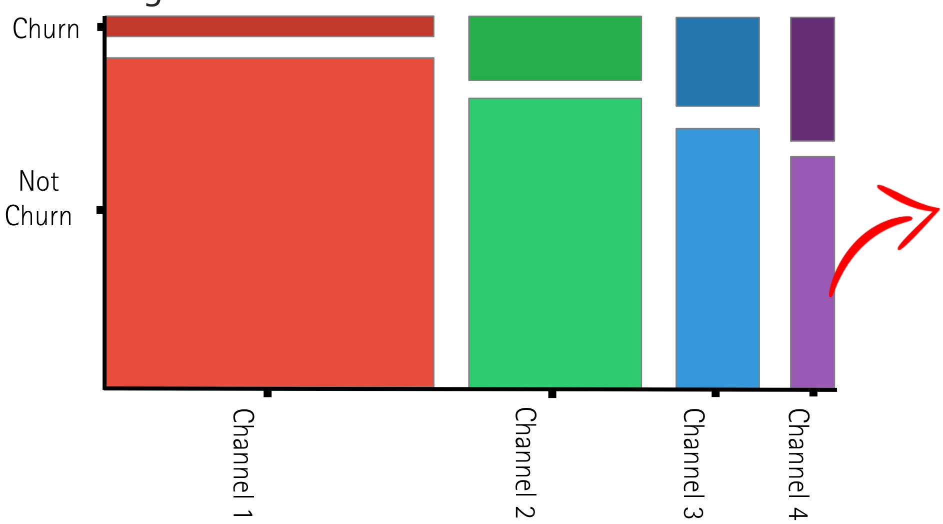
Subscription Plan Prices



Users with **higher** subscription plan prices are more likely to churn than users with lower subscription plan prices.

Fig 2

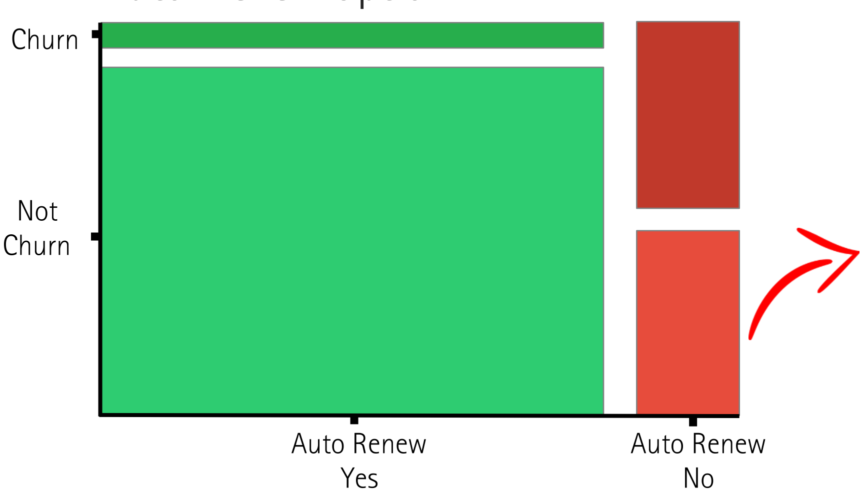
Registered via Other Channels



Users registered via **channel 4** have higher chance to churn than users registered via other channels.

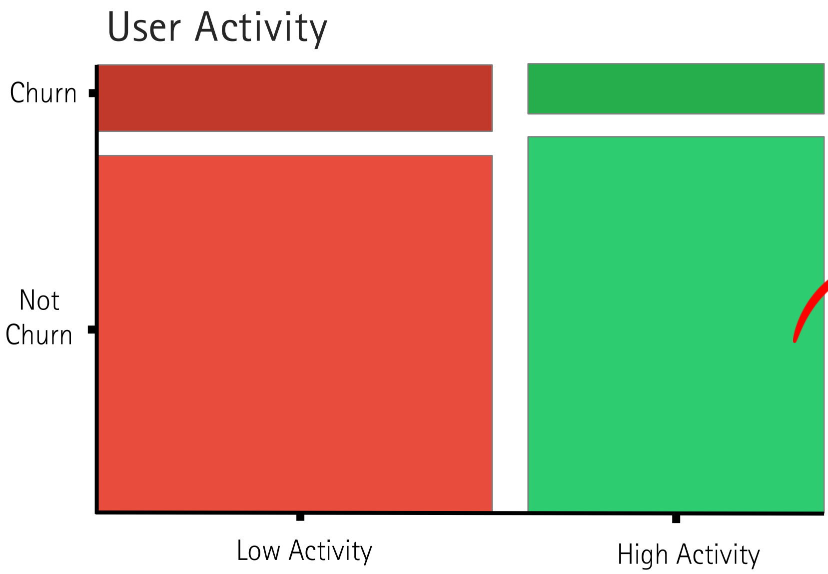
Fig 3

Auto-Renew Option



Users who selected Auto-Renew option during registration are less likely to churn than users who haven't selected auto-renew option.

Fig 4



The chances of a user to churn or not churn remain almost same, irrespective of the user activity of the user. That is why, ***user activity*** *is not a reliable attribute to predict churn.*

Fig 5

Compared to user activity features, transactional data features were able to explain churn better

Feature Engineering

Feature engineering has been done in a way to capture a user's activity and interaction related issues with the music service provider. User's activity related features were categorized as "User Log" category and user's interaction related features as "Transactional" category as shown in Table 2. In order to capture the variations in user's behaviour with time, we have genertaed the transactional and userlog features in different time windows as shown in Table 1.

As user transactions and user logs are time-dependent features, to make it uniform, a cutoff date is set (1 Feb, 2017, after which you want to know if users churn) we analyzed only the data before the cut-off date.

Features

- Membership related features contain the user details like city, age, gender etc. as features and are static with time.
- User log related features capture the activity and listening behaviour of the user.These features are user specific and are dynamic with time.
- Transaction related features contain the details of the user's transactions and describe user sensitivity towards transaction attributes like discounts, payment type, plan price etc.

Table 1

Time Window	Description
Entire history	From beginning to cut-off date
Last month	Last one month before from cut-off date
Day 7	Last one week before from cut-off date
Day 7 - 14	Last Second week before from cut-off date
Day 14 - 21	Last third week before from cut-off date
Day 21 - 28	Last fourth week before from cut-off date
Day 7 - 28	Last Three weeks before from cut-off date
Week 4 - 8	Last week 8 to week 4 before from cut-off date
Week 8 - Month 5	Last month 5 to week 8 before from cut-off date

Feature Generation

We have generated in total 8 different feature sets each falling into one of the above-described features (Transactional, Userlog, Membership) with a combined total of 260 features. Generating such a large number of features through conventional programming would take time and immense computational power, but by taking advantage of advanced NumPy and Pandas functions we were able to drastically reduce computational power requirement and still be highly productive.

Modelling

The modelling approach that we came up with for this case was an ensemble approach, that was built on ten base models from eight different feature sets. Ensemble approach would also help us overcome the class imbalance issue in the data.

- Base Models We built multiple base models like LightGBM, XGBoost, Neural Network and Logistic Regression and are fine tuned to produce optimum performance.
- Ensemble Model On the base models, we have built two layers of Stacking Ensemble Models as described in section Model Architecture.

Modelling Architecture

We have built and tuned 10 different base models on respective feature sets.

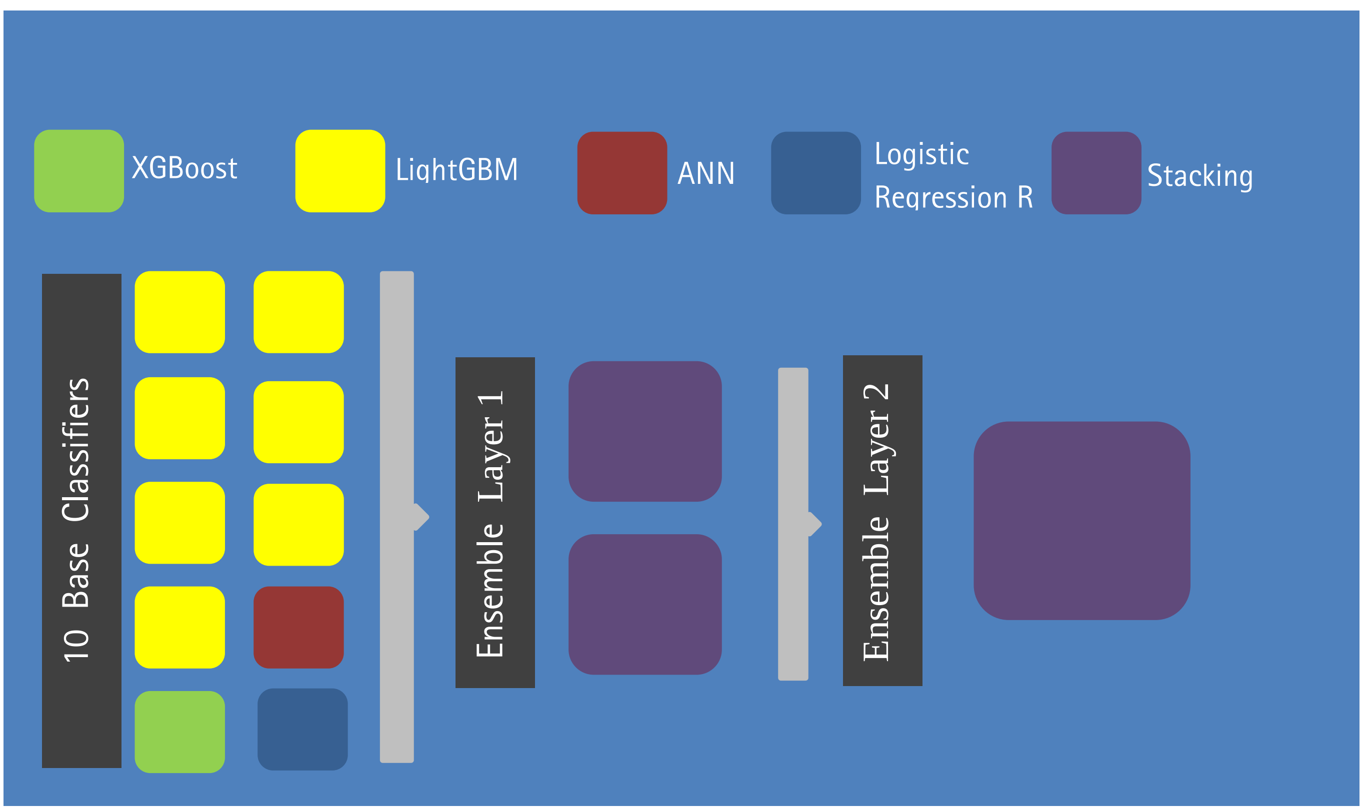
Table 2

Base Model Number	Base Model Name	Feature Set	Feature Set Category
01	LightGBM	F1	Membership
02	LightGBM	F2	User Log
03	LightGBM	F3	Transactional
04	Neural Network	F3	Transactional
05	LightGBM	F4	User Log
06	XGBoost	F5	User Log
07	Logistic Regression	F6	Transactional
08	LightGBM	F6	Transactional
09	LightGBM	F7	Transactional
10	LightGBM	F8	Transactional

Table 3

Ensemble Model Number	Ensemble Layer	Ensemble Technique	Base Models	Ensemble Meta Model
E1	L1	Stacking	1-9	LightGBM
E2	L1	Stacking	5,6	LightGBM
E3	L2	Stacking	E1, E2	Logistic Regression

Figure 7 Modelling Architecture



Model Details.

In the first layer of the ensemble, we built two stacking ensemble models. The first stacking ensemble model (E1) was built on the base models 1 to 9 with a LightGBM as meta-model and the second stacking ensemble model (E2) was built on the base models 6 and 10 with a LightGBM as meta-model.

In the second layer of the ensemble, we built the stacking model (E3) on both the ensemble models in the first layer with a logistic regression as meta-model. All the models were evaluated using 4 to 5 cross-validation folds using the GridSearchCV function of sci-kit learn library, to produce the optimum results for two most important classification metrics F1-score and accuracy.

Results and Discussions

For the ensemble method to perform well, base classifiers should be diverse and independent. In this case, independence is brought through different featuresets and algorithms.

Feature Importance Study

Among the top features, the auto-renew ratio and the is-cancel ratio are of most importance. They are a good indicators of whether or not a customer will churn. It is reasonable because if a user has an auto-renew option, he/she most likely will be using the service in the near future. Similarly, if a user has canceled a particular transaction or has more canceled transactions, he/she might not be using the service in the near feature. Table 5 is the list of top 20 features that have high importance.

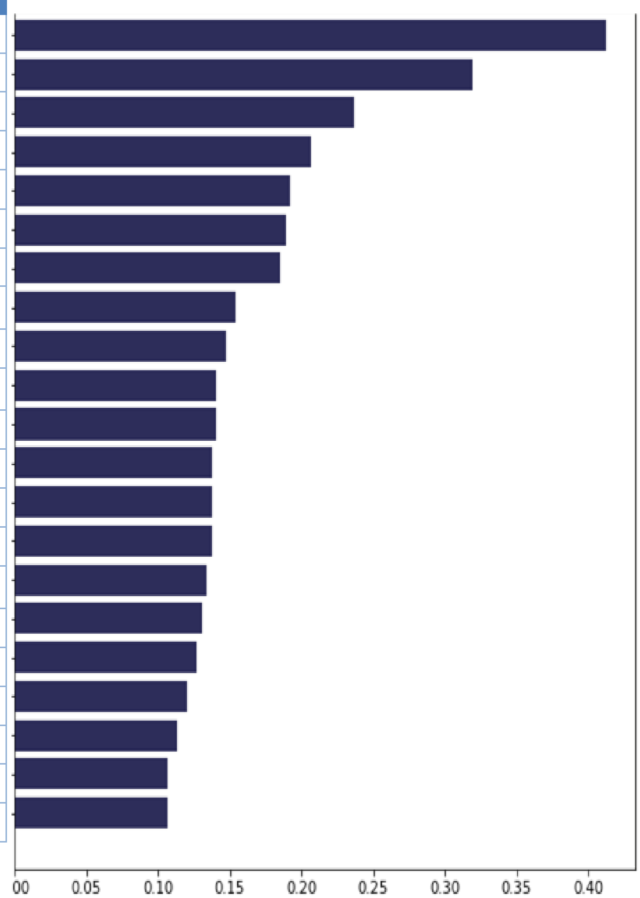
Performances of Models

The final ensemble model has highest accuracy and F1- score of all the base models.

The final ensemble model has highest accuracy and F1- score of all the base models. Notice that the performance of the model 9 was a little below the performance of the final ensemble model.The Ensemble model is more robust because it takes into account many other factors that model 9 didn't and still performs well.

Model	Model Type	Feature Set	Accuracy (%)	F1 – Score (%)
01	LightGBM	Membership	80.0	34.0
02	LightGBM	User Log	59.7	26.3
03	LightGBM	Transactional	85.6	55.8
04	Neural Network	Transactional	88.2	41.6
05	LightGBM	User Log	90.3	46.5
06	XGBoost	User Log	73.7	16.9
07	Logistic Regression	Transactional	85.0	50.4
08	LightGBM	Transactional	81.8	50.3
09	LightGBM	Transactional	95.0	85.1
10	LightGBM	Transactional	86.2	59.6
E3	Final Ensemble	Ensemble data	96.0	86.5

Feature Name**
Is Cancel ratio month 5- week 8
Is Cancel ratio week 8 - week 4
Is Cancel ratio week 4 - week 1
Is Cancel ratio month 1
Is cancel ration entire history
Auto renew ratio m5 - w8
Auto renew ratio w8 – w4
Auto renew ratio w4 – w1
Auto renew ratio w1
Auto renew ratio m1
Auto renew ratio entire history
Pay ratio
Length of first transaction
whether latest transaction canceled
User's payment status in last month
Does last transaction has auto renew
Payment method of last transaction
Length of last transaction
Min days bw trans & cutoff dates m5-w8
Min days bw trans & cutoff dates w8-w4
Min days bw trans & cutoff dates w4-w1



Interaction features performed better than activity features in predicting customer churn

Analysis

Among the base classifiers that were built the highest F1-score was 80 percent and lowest F1-score was 20 percent. The base classifiers built on transactional feature sets have performed well. This means that the models were able to identify underlying patterns to differentiate between churned and non-churned customers in transactional data. The base classifiers that were built on the user log data did not perform as well, this simply means that the activity of the user on the music service provider platform does not predict customer churn. The transaction attributes like discounts offered by the service provider, subscription plans, subscriptions price, subscription period, mode of transaction, ease of transaction etc., are the deciding factors to know whether a customer will churn. User's activity features are not predictive of customer churn but the interaction factors between user and service provider during transactions have played key role in customer churn. This sort of situation of churn independence on the activity of the user will usually occur when the market of online music service is occupied with several best service providers. Due to rapid digitalization and internet growth in recent years, all service providers are saturated in terms of the quality of the music they can improve and most of them fall at the same level, hence leading the users not to choose among the best service providers on the basis of quality of music.

Conclusion

We can predict churn by analyzing behavioral and transactional data of customers. Mining their data, to find patterns that are specific to customers who have already churned, can help you identify those customers, who are likely to churn. In order for the firm to increase the retention rate and reduce the churn rate, the firm has to focus on the customer satisfaction and making user interactions (transaction attributes), more user friendly and hassle free for its customers.

Churn prediction can be used within your business. It is one of the key components in determining lifetime value of customers (CLV). It will help us understand and compare chrun rates in different customer segements (high CLV segment and low CLV segment), and inturn we can take customized decisions to retainhighly profitable customers and also not to spend more then required resources on retaining customers who are least profitable to the business.

**Is Cancel Ratio – The ratio of cancelled transactions in mentioned time frame; Auto Renew Ratio – Ratio of transactions where auto renew is selected; Pay Ratio – Ratio of sum payment made to sum actual total price for all transactions; Length of a Transaction – Days from transaction date to cutoff date; Payment Status – Paid or Free Trial or converted to pay; Min days between trans & cutoff – Of all the transactions in mentioned time, total days from latest transaction to cutoff date.

About Perceptive Analytics

Perceptive Analytics is a Data Analytics Company recognized as a Top 10 Emerging Analytics Company. It is the winner, Fidelity Data Challenge in which 54 analytics companies participated. It also received an award at Netflix Hackathon at Tableau Conference, 2018. The clients we served include Morgan Stanley, Johnson & Johnson, Amex, Wells Fargo, and PepsiCo to name a few. We work with clients as their analytics department or alongside internal teams to deliver long-term competitive advantage.

Winner, Fidelity Investments Data Challenge

Top 10 Emerging Analytics Companies to Watch

Advanced Analytics

- Customer Analytics
- Marketing mix modeling
- Attribution modeling
- Customer segmentation
- Response modeling
- Churn modeling
- Next basket analysis

Business Intelligence & Reporting

- Visualizations & Dashboards
- ETL (Extract, Transform and Load)
- Data warehousing
- Cloud computing (AWS, Azure, Google Cloud)

Schedule Free Consulting Call

Contact
Chaitanya Sagar, CEO
cs@perceptive-analytics.com
+1 (646) 583 0001

Marketing Analytics | Tableau Consulting | Excel Apps
New York • Dallas • San Francisco • Hyderabad
Perceptive-Analytics.com