Phase-1 Abstract : Customer churn analysis is the process of predicting customers who tend to cancel the service (subscription) they receive for various reasons, especially in sectors such as telecommunications, finance and insurance, and determining the necessary operational steps to prevent this cancellation.

Example: Banking Sector

The Problem is based on the domain of the Banking sector where the bank wants to predict the Churn of a customer depending upon the previous data of the customer. By churn it is meant that the bank wants to predict if a customer would be a defaulter in the next quarter depending upon its previous credit history.

**Problem Definition:**

The main problem is to predict if a customer would be credit defaulter or not depending upon the previous data of the customer. A Bank wants to take care of customer retention for its product: savings accounts. The bank wants you to identify customers likely to churn balances below the minimum balance. You have the customers information such as age, gender, demographics along with their transactions with the bank.

**Design Thinking**:

Analaysis Objectives:1.Identifying at Risk Customer, Predictive Accuracy, ROI measurements, Model validation, Continuous Monitoring, Feed back Loop.

Data Collection: Identify the data source ,Data collection ,Labelling Churn, Time period, Data Privacy and Security, EDA, Data Documentation, Data Updates.

Visualization Strategy: EDA, Churn Distribution, Feature Analysis, Time series Analysis, Dashboard Creation, Monitoring.

Predictive Models: Model selection, Model Training, Model Evaluation, Deployment, Monitoring and Maintenance.

Phase-2

**PREDICTION CHURN CUSTOMER**

Some of the techniques used for this purpose include:

* KNN (K-Nearest Neighbors)
* SVM (Support Vector Machine)
* Decision Tree
* Random Forest[**1**](https://www.bing.com/ck/a?!&&p=7c81bf5ea7f35e71JmltdHM9MTY5Njg5NjAwMCZpZ3VpZD0yMmEyNDAzYi01YTYwLTZhMDItMzc0MC01MjE0NWJiMjZiNzMmaW5zaWQ9NTcyMA&ptn=3&hsh=3&fclid=22a2403b-5a60-6a02-3740-52145bb26b73&psq=prediction+churn+customer+what+are+the+machine+learning+techniques&u=a1aHR0cHM6Ly9pZWVleHBsb3JlLmllZWUub3JnL2RvY3VtZW50LzkyOTc1Mjk&ntb=1)

[Amazon SageMaker can be used to **analyze customer churn probability using call transcription and customer profiles**](https://www.bing.com/ck/a?!&&p=dd3aa8c1ad2d8b62JmltdHM9MTY5Njg5NjAwMCZpZ3VpZD0yMmEyNDAzYi01YTYwLTZhMDItMzc0MC01MjE0NWJiMjZiNzMmaW5zaWQ9NTcyMQ&ptn=3&hsh=3&fclid=22a2403b-5a60-6a02-3740-52145bb26b73&psq=prediction+churn+customer+what+are+the+machine+learning+techniques&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS9ibG9ncy9tYWNoaW5lLWxlYXJuaW5nL3ByZWRpY3RpbmctY3VzdG9tZXItY2h1cm4td2l0aC1hbWF6b24tbWFjaGluZS1sZWFybmluZy8&ntb=1)[**2**](https://www.bing.com/ck/a?!&&p=1d441e0924b06babJmltdHM9MTY5Njg5NjAwMCZpZ3VpZD0yMmEyNDAzYi01YTYwLTZhMDItMzc0MC01MjE0NWJiMjZiNzMmaW5zaWQ9NTcyMg&ptn=3&hsh=3&fclid=22a2403b-5a60-6a02-3740-52145bb26b73&psq=prediction+churn+customer+what+are+the+machine+learning+techniques&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS9ibG9ncy9tYWNoaW5lLWxlYXJuaW5nL3ByZWRpY3RpbmctY3VzdG9tZXItY2h1cm4td2l0aC1hbWF6b24tbWFjaGluZS1sZWFybmluZy8&ntb=1). [It can also be used to **build, tune, and deploy an end-to-end churn prediction model using Amazon SageMaker Pipelines**](https://www.bing.com/ck/a?!&&p=932b1e70b98350a9JmltdHM9MTY5Njg5NjAwMCZpZ3VpZD0yMmEyNDAzYi01YTYwLTZhMDItMzc0MC01MjE0NWJiMjZiNzMmaW5zaWQ9NTcyMw&ptn=3&hsh=3&fclid=22a2403b-5a60-6a02-3740-52145bb26b73&psq=prediction+churn+customer+what+are+the+machine+learning+techniques&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS9ibG9ncy9tYWNoaW5lLWxlYXJuaW5nL3ByZWRpY3RpbmctY3VzdG9tZXItY2h1cm4td2l0aC1hbWF6b24tbWFjaGluZS1sZWFybmluZy8&ntb=1)[**2**](https://www.bing.com/ck/a?!&&p=5328b205261be743JmltdHM9MTY5Njg5NjAwMCZpZ3VpZD0yMmEyNDAzYi01YTYwLTZhMDItMzc0MC01MjE0NWJiMjZiNzMmaW5zaWQ9NTcyNA&ptn=3&hsh=3&fclid=22a2403b-5a60-6a02-3740-52145bb26b73&psq=prediction+churn+customer+what+are+the+machine+learning+techniques&u=a1aHR0cHM6Ly9hd3MuYW1hem9uLmNvbS9ibG9ncy9tYWNoaW5lLWxlYXJuaW5nL3ByZWRpY3RpbmctY3VzdG9tZXItY2h1cm4td2l0aC1hbWF6b24tbWFjaGluZS1sZWFybmluZy8&ntb=1).

**1. Establish the Business Case**

This step is simply understanding your desired outcome from the ML algorithm. In this case, the final objective is:

* Prevent customer churn by preemptively identifying at-risk customers
* Design appropriate interventions to improve retention

**2. Collect and Clean Data**

The next step is data collection — understanding what data sources will fuel your churn prediction model. Companies [**capture customer data**](https://www.scalr.ai/post/data-capture-services) across their lifecycle through software such as CRM, web analytics, sentiment analysis tools, social listening tools, customer service software, and more.

Building data capture services is one of the easiest and most effective ways to begin collecting data to power your churn prediction model. Turnkey solutions like automated data capture (ADC) can help you leverage your existing software to speed up relevant data collection and apply it to your churn prediction model. ADC eliminates the manual efforts required for data entry and frees up time for your data team to fine-tune your prediction model.

**3. Engineer, Extract, and Select Features**

Feature engineering is a crucial part of the dataset preparation — it helps determine the attributes that represent behavior patterns related to customer interaction with a product or service. Data scientists use feature engineering to assign measurable characteristics to data points that an ML model will process to predict churn probability.

These features could include customer demographics, behaviors (in the mobile phone example, these could be data consumption, calling customer service, using international roaming, etc.), and contextual features that describe other information about a customer like communication preferences, past buying behavior, or birthdays/anniversaries.

Next, feature extraction standardizes the variables (attributes) by only isolating the ones that contain meaningful information in context of the business case (churn). Feature extraction limits data dimensionality (columns representing attributes in a dataset) and only retains helpful data for the business case.

**4. Build a Predictive Model**

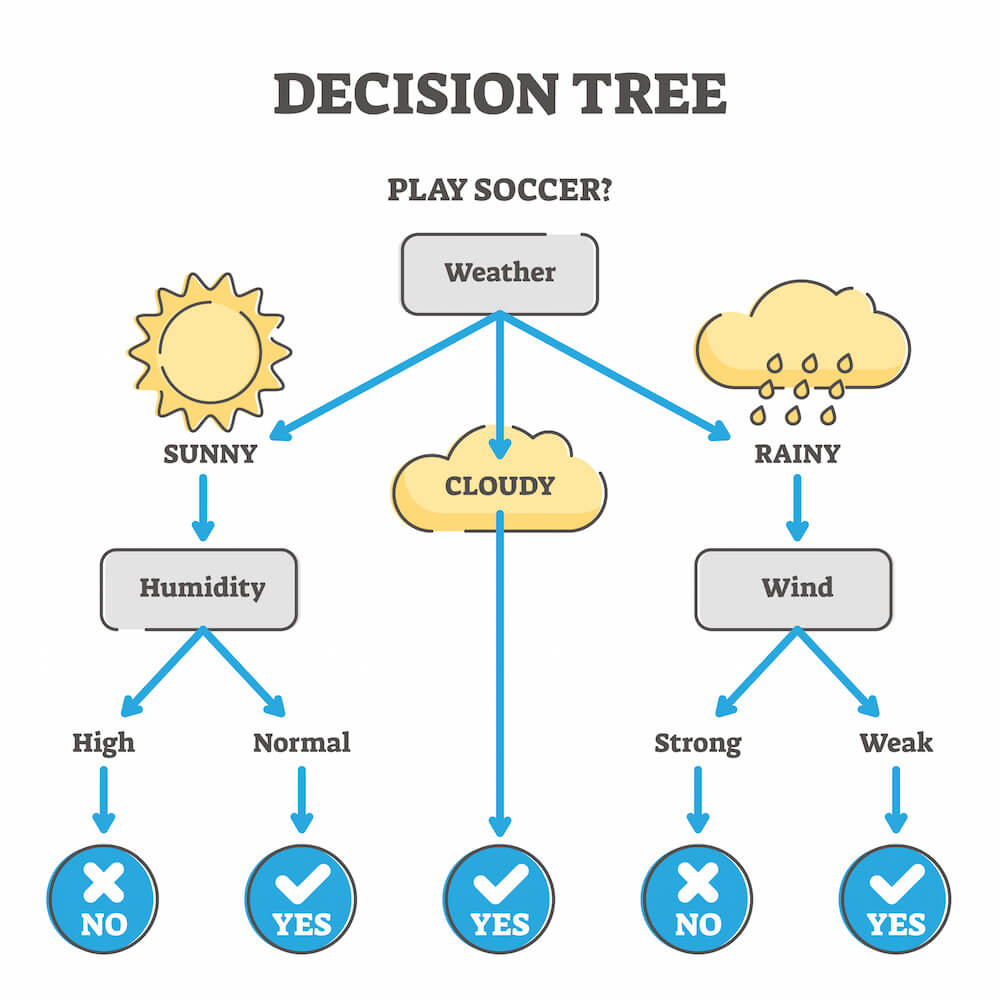
Data analysts typically approach churn prediction using multiple methods such as binary classification, logistic regression, decision trees, random forest, and others.

ML algorithms perform binary classification to slot the attributes of a target variable into two groups on the basis of a classification rule. In this context, the target variable is churn, the outcome of which can be classified as true or false. Binary classification helps us understand which customers churned and which ones stayed on.

Based on this information, data scientists can then run regression analysis to determine the relationship between the target variable (churn) and other data points that influence churn (monthly plan, data consumption, service calls, etc.), in weighted values.

This will provide information on whether variables have a positive or negative relationship with churn. A positive relationship indicates a higher probability for customers to leave and a negative relationship means that customers are less likely to churn.

A decision tree is yet another effective training model for churn prediction. The decision tree model uses available features and splits the data based on features values to provide unique resulting groups. Here’s a simple example of a decision tree:



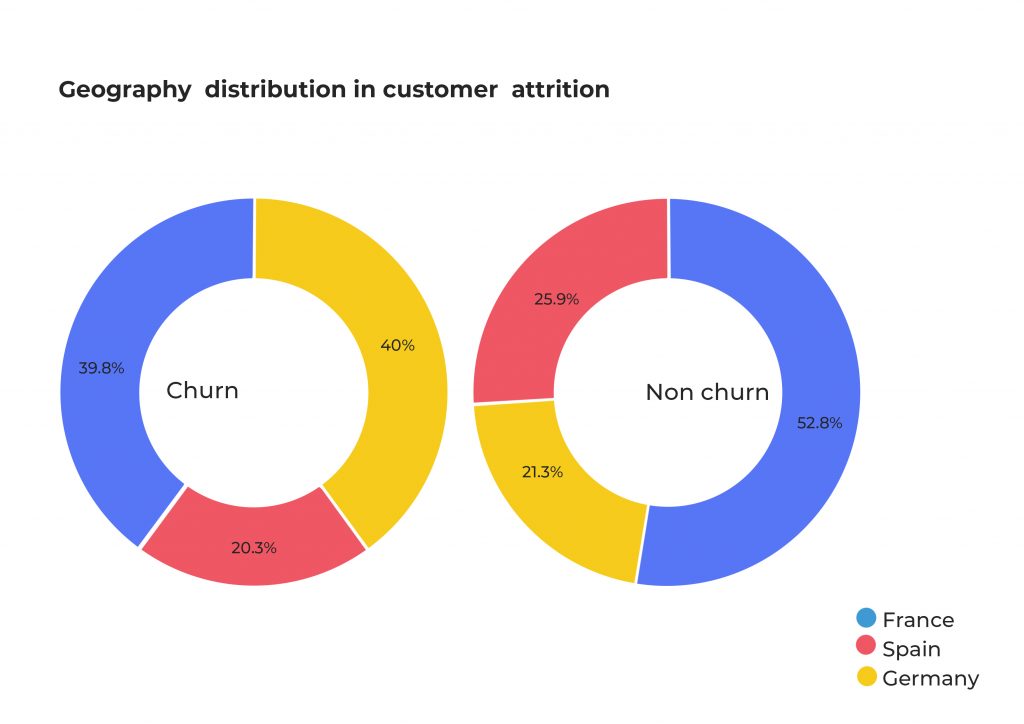
Now, depending on the size of the dataset and the diversity of feature data, you may choose to use multiple decision trees or a Random Forest.

A Random Forest is a collection of multiple decision trees, where each individual tree splits out a classification. These classifications are binary in nature, so whichever classification receives the most number of votes, wins. So, if your Random Forest consists of five decision trees, and three of those provide the same classification, your final prediction will be determined by the majority.

**5. Deploy and Monitor**

Once you have developed the model, it needs to be integrated with existing software or serve as the base for a new program or application. You’ll need to pay close attention to the model’s accuracy and performance.

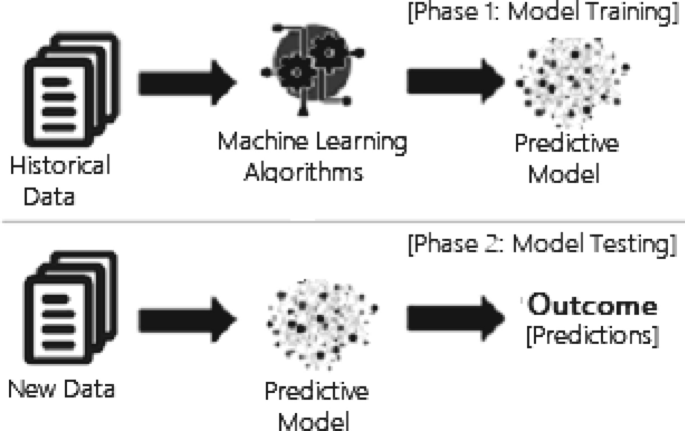
Testing and monitoring model performance to adjust features will help improve the model’s accuracy. From our mobile services example, monitoring and testing could mean logging customer interactions and reviews.

****

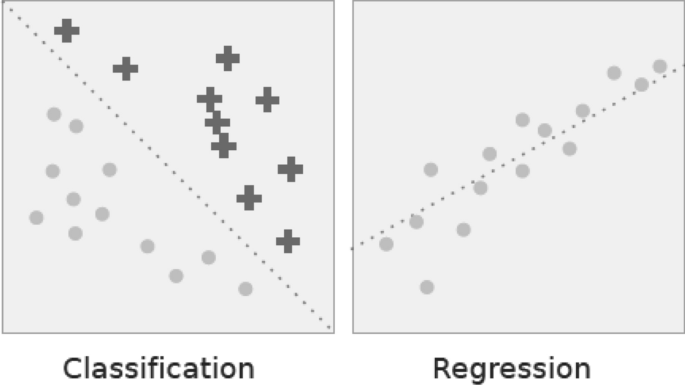
Phase-3

Machine Learning Tasks and Algorithms

In this section, we discuss various machine learning algorithms that include classification analysis, regression analysis, data clustering, association rule learning, feature engineering for dimensionality reduction, as well as deep learning methods. A general structure of a machine learning-based predictive model has been shown in where the model is trained from historical data in phase 1 and the outcome is generated in phase 2 for the new test data.

[](https://link.springer.com/article/10.1007/s42979-021-00592-x/figures/3)

A general structure of a machine learning based predictive model considering both the training and testing phase

[](https://link.springer.com/article/10.1007/s42979-021-00592-x/figures/6)

Classification vs. regression. In classification the dotted line represents a linear boundary that separates the two classes; in regression, the dotted line models the linear relationship between the two variables

* *Cybersecurity and threat intelligence:* Cybersecurity is one of the most essential areas of Industry 4.0. [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)], which is typically the practice of protecting networks, systems, hardware, and data from digital attacks [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)]. Machine learning has become a crucial cybersecurity technology that constantly learns by analyzing data to identify patterns, better detect malware in encrypted traffic, find insider threats, predict where bad neighborhoods are online, keep people safe while browsing, or secure data in the cloud by uncovering suspicious activity. For instance, clustering techniques can be used to identify cyber-anomalies, policy violations, etc. To detect various types of cyber-attacks or intrusions machine learning classification models by taking into account the impact of security features are useful [[97](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR97)]. Various deep learning-based security models can also be used on the large scale of security datasets [[96](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR96), [129](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR129)]. Moreover, security policy rules generated by association rule learning techniques can play a significant role to build a rule-based security system [[105](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR105)]. Thus, we can say that various learning techniques discussed in Sect. [Machine Learning Tasks and Algorithms](https://link.springer.com/article/10.1007/s42979-021-00592-x#Sec5), can enable cybersecurity professionals to be more proactive inefficiently preventing threats and cyber-attacks.
* *Internet of things (IoT) and smart cities:* Internet of Things (IoT) is another essential area of Industry 4.0. [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)], which turns everyday objects into smart objects by allowing them to transmit data and automate tasks without the need for human interaction. IoT is, therefore, considered to be the big frontier that can enhance almost all activities in our lives, such as smart governance, smart home, education, communication, transportation, retail, agriculture, health care, business, and many more [[70](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR70)]. Smart city is one of IoT’s core fields of application, using technologies to enhance city services and residents’ living experiences [[132](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR132), [135](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR135)]. As machine learning utilizes experience to recognize trends and create models that help predict future behavior and events, it has become a crucial technology for IoT applications [[103](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR103)]. For example, to predict traffic in smart cities, parking availability prediction, estimate the total usage of energy of the citizens for a particular period, make context-aware and timely decisions for the people, etc. are some tasks that can be solved using machine learning techniques according to the current needs of the people.

import pandas as pd

import numpy as np arrayDataset=pd.read\_csv('mainSimulationAccessTraces.csv')  
x=Dataset.iloc[:,:-2].values  
y=Dataset.iloc[:,12].values

from sklearn.impute import SimpleImputer

imputer=SimpleImputer(missing\_values=np.nan,strategy='constant',verbose=0)

imputer=imputer.fit(x[:,[8]])

x[:,[8]]=imputer.transform(x[:,[8]])imputer1=SimpleImputer(missing\_values=np.nan,strategy='mean',verbose=0)

imputer1=imputer1.fit(x[:,[10]])

x[:,[10]]=imputer1.transform(x[:,[10]])from sklearn.preprocessing import LabelEncoder

labelencoder\_X = LabelEncoder()

for i in range(0,10):

x[:,i] = labelencoder\_X.fit\_transform(x[:,i])

x=np.array(x,dtype=np.float)y=labelencoder\_X.fit\_transform(y)

from sklearn.preprocessing import StandardScaler  
 sc = StandardScaler()  
 x\_train = sc.fit\_transform(x\_train)  
 x\_test = sc.transform(x\_test)

Phase-4

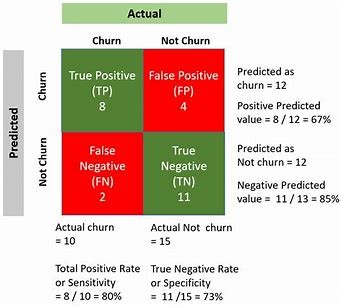
**ANALYSIS OBJECTIVES**

Identify microtrends and subtle patterns that humans can easily miss

* Scale data analysis across thousands or millions of data points simultaneously
* Standardize churn prediction factors and apply them consistently to all customers with less risk of bias and human error
* Reduce resource expenditure on manual and repetitive data analysis work
* Forecast revenue for the year and develop strategies for retaining high-risk customers

**STEPS TO PROCESS AND CLEAN THE DATA TO ENSURE QUALITY AND ACCURACY**

1. **Data Cleaning**: This step involves removing or correcting any errors, inconsistencies, or inaccuracies in the data. It also includes handling missing data, outliers, and duplicates. This step is crucial as it ensures that the data is reliable and can be used for analysis.
2. **Data Transformation**: This step involves converting the data into a format that is suitable for analysis. It includes tasks such as normalization, aggregation, and feature engineering. This step helps to improve the quality of the data and makes it easier to analyze.
3. **Data Integration**: This step involves combining data from multiple sources into a single dataset. It includes tasks such as matching, merging, and joining datasets. This step helps to improve the completeness of the data and provides a more comprehensive view of the problem.
4. **Data Verification**: This step involves verifying the accuracy and completeness of the data after processing. It includes tasks such as cross-checking, validation, and testing. This step helps to ensure that the processed data is accurate and reliable.
5. **Data Documentation**: This step involves documenting the entire process of data processing, including all the steps taken and decisions made. It helps to ensure that the process is transparent and reproducible.

   
  
 Phase-5

Conclusion: In conclusion, customer churn prediction analysis is helpful for companies trying to move toward a data-driven approach and boost their financial metrics by lowering customer churn. Businesses can proactively address customer concerns and ensure long-term success by conducting regular analyses as a part of AI-based customer service.

Customer churn analysis allows to minimize acquisition costs and increase marketing efficiency, preparing a solid base for future marketing analysis and campaigns. Customer churn analysis opens new opportunities for cross-selling and upselling and serves as one of the starting points for customer-driven product development, keeping customers engaged and loyal over time.

Effective customer churn analysis & prediction

19 September 2023

Author: Valeryia Shchutskaya

Customer churn prediction-s

Content

Customer churn is a measurement that shows how many clients discontinued a service, an application or stopped buying a product during a certain period of time. Churn prediction is one of the most popular applications of machine learning and data science in business.

Although at first, churn analysis was essential for telecoms, now it is applicable for businesses of all sizes, including startups. Marketers with tight budgets are strongly relying on strategies build around churn and retention, considering such strategy the most cost-efficient.

Importance of churn analysis

On average, most company’s business comes from existing customers and acquiring the new ones is 15 times more expensive than retaining existing customers.

Churn analysis is mostly applicable in the following areas:

For subscription-based businesses churn is a critical metric as every customer they lose results in the loss of recurring revenue.

E-commerce companies are highly interested in providing their customers with timely communication without overspending on discounts and special offers for people who were not really about to churn.

Individualized customer retention solution could be a good idea. But being very time and money consuming it’s almost impossible to implement since businesses cannot spend much time on each of their multiple customers. On the other hand, knowing that a particular customer is about to churn, you could direct the retention efforts directly on her.

Improve customer retention with ease

Want to retain more customers but not sure where to start with churn prediction for your business?

Contact us

Why businesses use customer churn prediction

Since the question “what is customer churn analysis”, it’s time to move to why it is leveraged by business across industries.

Businesses employ customer churn analysis to comprehend consumer behavior better, spot opportunities for development, cut expenses associated with client acquisition, boost revenue, and gain a competitive edge. Businesses can proactively address consumer complaints and maintain long-term success by regularly conducting churn analysis. Let’s examine how forecasting client churn might be advantageous.

Better retention strategies

Creating more targeted and effective retention strategies is one of the main advantages of customer churn prediction analysis. By examining customer data, which enables them to spot trends and patterns in consumer behavior, businesses may better understand the needs and preferences of their customers.

Customer churn analysis in retail

Showroomprive.com, a top European E-commerce website, has implemented predictive analytics for effective churn management. Their data science solution involves establishing identification rules for potential churners and assigning a value to each customer at risk of churning. With this data, they have managed to accurately target personalized marketing campaigns to these customers, improving their chances of retention. Implementing churn detection has enabled Showroomprive.com to communicate with potential churners more timely and relevantly, resulting in increased win-back success rates.

Churn prediction in ecommerce

Source: Unsplash

Weak points identification

A further advantage of customer churn analysis is that it enables companies to identify areas where their services might be improved. By examining customer feedback and complaints, businesses can identify typical customer issues or pain areas. Through better addressing consumer demands in their products and services, organizations can increase customer satisfaction and retention rates.

Resultful use case of customer churn analysis in the banking industry

One real-life customer churn analysis case study comes from Barclays, a multinational investment bank and financial services company based in London, UK that leveraged bank customer churn prediction using machine learning. In order to reduce customer turnover in its retail banking sector, Barclays developed a predictive churn model that looked at information such as demographics, transaction history, and behavior patterns.

Low levels of engagement, frequent balance queries, and an abundance of customer service calls were among the primary indicators the model found as being linked to client churn. Utilizing bank customer churn prediction data, Barclays created focused retention measures for clients who were at risk, such as providing tailored incentives and discounts.

Budget saving and higher revenue

Analysis of customer attrition can also aid in cutting back on the expense of acquiring new clients. By identifying the key elements that affect client retention, businesses may create targeted retention strategies that reduce the need for new acquisitions. As a result, client acquisition costs may decrease, and overall profitability may rise.

Technology pros

Effective use of customer churn analysis in the telecom industry

XO Communications, a top-tier communications service provider in the United States, has been focused on elevating its service standards by leveraging customer churn analysis. Let’s view their customer churn analysis example.

For several years, XO has been using vast amounts of customer data, such as demographics, transaction history, calling patterns, and conversations with call center agents, to develop a comprehensive knowledge of each customer. To identify the most predictive variables for voluntary churn, they compiled over 500 data points and used advanced customer churn prediction software to analyze trends and correlations. By applying these insights to its current customer base, XO accurately predicts and proactively engages with the most valuable and profitable customers who carry the highest churn scores.

By revealing deeper insights into customer habits, finding trends, identifying prospective defectors, and enabling proactive actions to retain their most important clients, the adoption of customer churn analysis has saved XO millions of dollars. A major national provider of advanced broadband communications services and solutions, XO has drastically decreased its customer attrition rates by approximately 50% as a result of its use of predictive analytics.

Opportunity to compete

Data science customer analytics can provide companies with a competitive edge. Businesses may set themselves apart from their rivals and draw in and keep more customers by getting to know their clients better and being able to provide customized services and retention methods. This can help businesses establish a strong brand identity and maintain a loyal customer base.

Fruitful customer churn analysis in the insurance industry

In the insurance sector, Avanade—a partnership between Microsoft and Accenture—has used customer churn analysis with Power BI. Avanade has grown its operations and increased revenue due to the data on customer turnover obtained by learning valuable insights from researching customer and policy data.

Churn in insurance

Source: Unsplash

Forecasting and planning

Customer churn analysis can benefit firms, in addition to the ways already mentioned, by aiding in predicting and planning. Businesses can generate more precise sales projections and successful marketing plans by studying customer data to find trends and patterns in customer behavior.

Cross and up sellingIn conclusion, customer churn prediction analysis is helpful for companies trying to move toward a data-driven approach and boost their financial metrics by lowering customer churn. Businesses can proactively address customer concerns and ensure long-term success by conducting regular analyses as a part of AI-based customer service.

How to predict customer churn: churn prediction and analysis steps

The main goal of churn prediction is detecting which customers are likely to cancel a subscription to a service based on how they use it. The question “How to do customer churn analysis” can be answered in three major steps:

Data gathering and preparation

Building the predictive model

Testing and validation of the model on real customers

Each of these steps includes a variety of acquired techniques and smaller steps necessary for the best possible result.

1. Data gathering and preparation

The first step of data gathering includes the process of “feature engineering”. In order to predict churn on a particular customer, he or she is compared to another similar group of customers. Such comparison is based on particular pieces of information called “features” about each group of customers. The features include information about customers’ demographics, interactions history with the service, and other customer information relevant to a particular case. For example, customer’s age, level of education, number of times he logged into the app, time since last login, device type, and etc.

2. Building the predictive model

When building a churn prediction model, one of the most critical steps is to properly define what churn actually is, and how it can be translated into a variable that can be used in a machine learning model. The definition of churn is totally dependent on the business model and can differ widely from one company to another.

In order to effectively analyze and control customer churn, it is important to build an effective and accurate customer churn prediction model. Statistical and data mining techniques are utilized to construct the churn prediction models. The data mining techniques can be used to discover interesting patterns or relationships in the data, and predict or classify the behavior. In other words, it is an interdisciplinary area with a general objective of predicting outcomes and employing sophisticated data processing algorithms to discover mainly hidden patterns, associations, and anomalies in customer data.

3. Testing and validation of the model on real customers

When the model is ready and it is ensured that it takes into account all the special needs of a particular business and its customers, the testing period starts. It normally takes up to a few months. The model is fine-tuned according to the results. Such custom-built models have a solid advantage compared to automatically generated models – they stay very flexible and can be developed according to each company’s growing demands.