CUSTOMER CHURN PREDICTION

Churn rate is a critical metric of customer satisfaction. Low churn rates mean happy customers; high churn rates mean customers are leaving you. A small rate of monthly/quarterly churn compounds over time. 1% monthly churn quickly translates to almost 12% yearly churn. .

Churn much your business is growing over time. If growth is higher than churn, you can say your business is growing. If churn is higher than growth, your business is getting smaller. is a good indicator of growth potential. Churn rates track lost customers, and growth rates track new customers—comparing and analyzing both of these metrics tells you exactly how

Churn rate:

Wikipedia states that the churn rate (also called attrition rate) measures the number of individuals or items moving out of a collective group over a specific period. It applies in many contexts, but the mainstream understanding of churn rate is related to the business case of customers that stop buying from you.

You can classify churn as:

1.Customer and revenue churn
2.Voluntary and involuntary churn

Customer and revenue churn:

Customer churn is simply the rate at which customers cancel their subscriptions. Also known as subscriber churn or logo churn, its value is represented in percentages. On the other hand, revenue churn is the loss in your monthly recurring revenue (MRR) at the beginning of the month. Customer churn and revenue churn aren't always the same. You might have no customer churn, but still have revenue churn if customers are downgrading subscriptions. Negative churn is an ideal situation that only applies to revenue churn. The amount of new revenue from your existing customers (through cross-sells, upsells,   and new signups) is more than the revenue you lose from cancellations and downgrades.

Voluntary and involuntary Churn:

Voluntary churn is when the customer decides to cancel and takes the necessary steps to exit the service. It could be caused by dissatisfaction, or not receiving the value they expected. Involuntary churn happens due to situations such as expired payment details, server errors, insufficient funds, and other unpredictable predicaments.

Importance of customer churn prediction:

The impact of the churn rate is clear, so we need strategies to reduce it. Predicting churn is a good way to create proactive marketing campaigns targeted at the customers that are about to churn.

Churn prediction use cases:

Customer churn prediction is different based on the company's line of business (LoB), operation workflow, and data architecture. The prediction model and application have to be tailored to the company's needs, goals, and expectations. Some use cases for churn prediction are in:

- Telecommunication (cable or wireless network segment),
- Software as a service provider (SaaS),
- Retail market,
- Subscription-based businesses (media, music and video streaming services, etc.),
- Financial institutions (banking, insurance companies, Mortgage Companies, etc.),
- Marketing,
- Human Resource Management (Employee turnover).

Designing churn prediction workflow:

The overall scope to build an ML-powered application to forecast customer churn is generic to standardized ML project structure that includes the following steps:

Module 1:Defining problem and goal
Module 2:Establishing data source
Module 3:Data preparation,exploration,and preprocessing
Module 4:Modeling and testing
Module 5:Deployment and monitoring

1.Defining problem and goal:
It's essential to understand what insights you need to get from the analysis and prediction. Understand the problem and collect requirements, stakeholder pain points, and expectations.
2.Establishing data source:
Next, specify data sources that will be necessary for the modeling stage. Some popular sources of churn data are CRM systems, analytics services, and customer feedback.
3.Data preparation, exploration, and preprocessing:
Raw historical data for solving the problem and building predictive models needs to be transformed into a format suitable for machine learning algorithms. This step can also improve overall results by increasing the quality of data.
4.Modeling and testing:
This covers the development and performance validation of customers churn prediction models with various machine learning algorithms.
5.Deployment and monitoring:
This is the last stage in applying machine learning for churn rate prediction. Here, the most suitable model is sent into production. It can be either integrated into existing software, or become the core of a newly built application.

Dataset:

The sample data tracks a fictional telecommunications company, Telco. It's customer churn data sourced by the IBM Developer Platform, and it's available here. It includes a target label indicating whether or not the customer left within the last month, and other dependent features that cover demographics, services that each customer has signed up for, and customer account information. It has data for 7043 clients, with 20 features.

You can find this entire project on my Github.

Exploratory data analysis (EDA)
Let's critically explore the data to discover patterns and visualize how the features interact with the label (Churn or not).

The dataset has 7043 rows and 21 columns.

There are 17 categorical features:

CustomerID: Customer ID unique for each customer

gender: Whether the customer is a male or a female

SeniorCitizen: Whether the customer is a senior citizen or not (1, 0)

Partner: Whether the customer has a partner or not (Yes, No)

Dependent: Whether the customer has dependents or not (Yes, No)

PhoneService: Whether the customer has a phone service or not (Yes, No)

MultipeLines: Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService: Customer's internet service provider (DSL, Fiber optic, No)

OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup: Whether the customer has an online backup or not (Yes, No, No internet service)

DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service)

TechSupport: Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract: The contract term of the customer (Month-to-month, One year, Two years)

PaperlessBilling: The contract term of the customer (Month-to-month, One year, Two years)

PaymentMethod: The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

Next, there are 3 numerical features:

Tenure: Number of months the customer has stayed with the company

MonthlyCharges: The amount charged to the customer monthly

TotalCharges: The total amount charged to the customer

Churn: Whether the customer churned or not (Yes or No)