**CUSTOMER CHURN PREDICTION PROJECT**

* ***ANALYSIS OBJECTIVE:***

Predicting customer churn is a critical task for businesses, especially in subscription-based or service-oriented industries. To effectively achieve this goal, organizations typically define specific goals such as:

**1. Identification of potential churners:**

* Early detection:

The main goal is to identify customers at risk of defection as quickly as possible. This allows businesses to take proactive steps to retain these customers.

* Segmentation:

Divide your customer base into segments based on likelihood of churn. This segmentation helps you tailor your retention strategy to different customer groups.

**2. Understand the main factors that cause cancellation:**

* Function meaning:

Identify the key factors or characteristics that contribute to customer churn. This includes analysing historical data and performing feature importance analysis using techniques such as machine learning algorithms and statistical analysis.

* Cause analysis:

Go beyond superficial analysis to identify the root reasons for customer churn. This may include qualitative research, customer surveys, or more in-depth data analysis to identify specific issues or pain points.

**3. Model development and evaluation:**

* Model selection:

Choose the right machine learning or statistical model to predict churn based on historical data. Common models include logistic regression, decision trees, random forests, and neural networks.

* Model evaluation:

Establish clear evaluation metrics such as accuracy, precision, recall, F1 score, and ROC AUC to evaluate the performance of your churn prediction model. Cross-validation and holdout validation datasets are often used to ensure the generalizability of the model.

**4. Practical insights and recommendations:**

* Practical Insights:

Translate churn prediction insights into actionable business insights. This could result in suggesting specific strategies to retain high-risk customers or improve overall customer satisfaction.

* Storage Strategy:

Develop and implement retention strategies tailored to different customer segments or individual customers. These strategies include personalized offers, discounts, and improved customer support.

**5. Monitoring and continuous improvement:**

* Real-time monitoring:

Set up a system to monitor customer churn predictions in real time or on a regular basis. This allows for immediate response to potential defectors.

* Feedback Loop:

Establish a feedback loop to continuously improve churn prediction models and retention strategies based on the results and feedback of implemented interventions.

**6. Compliance and ethical considerations:**

Ensure that your churn prediction and retention efforts comply with relevant regulatory and ethical standards, particularly regarding privacy and customer consent. These goals provide a structured approach to predicting customer churn, helping businesses not only identify potential customer churn, but also retain valuable customers and improve overall customer satisfaction.

* ***DATA COLLECTON:***

1. **Web Tracking:**

Customers usually make contact with you through your website. You can collect information about your customers, such as their demographics and location, how they interact with your company and what they do through the website.

Tools like Google Analytics, Mixpanel, Piwik PRO, and Matomo can help you learn about your customer’s interests, referral sources, conversion details, and real-time behavior on your website.

1. **Social Media:**

You may learn a lot about your customers based on their social media interactions with you. Aside from basic customer engagement metrics like likes, comments, and shares, you can learn a lot about your customers by looking at each social media platform’s analytics/insights section.

Through online reputation management (ORM), you can get customer feedback that will help determine how people feel about your brand and products.

1. **Customer feedback and surveys**

Customer feedback and surveys are good ways to find out what your customers are interested in, what they like, and what they prefer. Surveys can help you get information about people’s thoughts and feelings by asking the right questions.

You can find out what people think about your products, services, sales, and marketing through surveys. With Net Promoter Score (NPS), you can determine how much your customers love your products.

* ***VISUALIZATION STRATEGY:***

IBM Cognos Analytics is a powerful business intelligence and analytics software platform that can assist you in visualizing insights and analysing data. It provides a wide range of tools and features to create interactive dashboards, reports, and visualizations.

To showcase factors affecting churn and retention rates, we first make use of the **Telco customer churn sample data module** provided by IBM.

**The data set includes information about:**

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

**1. Dashboard Overview:**

We plan to start with a high-level dashboard that provides an overview of key churn metrics. Visualizations such as:

A line chart showing the churn rate trend over time.

A pie chart or donut chart displaying the current distribution of churned and retained customers.

**2. Customer Segmentation:**

Use segmentation to understand which customer groups are most at risk of churning. Visualize this with:

Bar charts or stacked bar charts showing churn rates for different customer segments. Heatmaps to highlight segments with the highest churn rates.

**3. Churn Drivers:**

Identify the key factors contributing to churn and visualize them through:

Heatmaps, scatter plots, or bubble charts to illustrate correlations between various customer attributes (e.g., age, usage frequency, support interactions) and churn.

Word clouds or word frequency visualizations to highlight common themes in customer feedback associated with churn.

* ***PREDICTIVE MODELLING:***

Before applying any machine learning algorithm, the data needs to be prepared properly, which involves defining the churn event and the time period, collecting and integrating relevant data sources, cleaning and transforming the data, exploring and analysing it, and selecting and creating features.

Churn prediction can be done using several machine learning algorithms, such as logistic regression, decision tree, random forest, k-means, and hierarchical clustering.

**Logistic regression** models the relationship between features and outcomes using a logistic function and is simple and fast, but can be prone to overfitting or underfitting.

**Decision trees** split data into branches based on features and are intuitive and flexible, but can be sensitive to noise, variance, and bias. Random forests combine multiple decision trees and are robust, accurate, and versatile, but complex and slow.

**K-means** assigns customers to one of k clusters based on their distance to the cluster centre and is simple, scalable, and efficient; however, it requires choosing the number of clusters and can be sensitive to initialization and outliers.

After applying the machine learning algorithms, the next step is to evaluate and improve their performance and accuracy. This requires splitting the data into training, validation, and testing sets, choosing an appropriate metric or criteria, comparing and selecting the best algorithm and interpreting and explaining the algorithm.

The algorithm can be optimized by adjusting parameters or selecting features and interpreted by analysing coefficients or importance scores.