**PHASE 1 PROJECT SUBMISSION**

**PROJECT 6 - CUSTOMER CHURN PREDICTION**

**TEAM MEMBERS:**

1. Rithees S M - 2021504032
2. Akshay G S – 2021504502
3. Dhanesh C N E – 2021504510
4. Divyavarshini K – 2021504513

**Problem Definition:**

The project involves using IBM Cognos to predict customer churn and identify factors influencing customer retention. The goal is to help businesses reduce customer attrition by understanding the patterns and reasons behind customers leaving. This project includes defining analysis objectives, collecting customer data, designing relevant visualizations in IBM Cognos, and building a predictive model.

**Database Link:**

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

**Objectives:**

* The primary objective of this project is to predict customer churn, which refers to the rate at which customers stop doing business with a company, and to identify the factors influencing customer retention.
* By understanding the patterns and reasons behind customers leaving, businesses can take proactive measures to retain customers, ultimately leading to increased revenue and customer satisfaction.
* IBM Cognos is the chosen tool for this project, indicating a focus on data analysis, visualization, and predictive modelling.

**Design Thinking:**

1. **Analysis Objective:**

Key factors that contribute to customer churn includes:

* Price and Cost Factors:

Sudden price increases or changes in pricing structures can lead consumers to seek out lower-cost alternatives in the market, driven by their perception of a lack of value for the price paid.

* Competitive Pressure:

This includes aggressive marketing and incentives from competitors and other services with better features or pricing.

* Customer Experience:

Cumilative negative experiences with the brand or product and lack of personalization and engagement may affect loyalty to the organization.

* Product / Service Fit:

Mismatch between customer needs and services offered or failure to adapt or expand the product to meet changing customer needs.

* Customer Satisfaction:

Customer feedback and satisfaction surveys indicating dissatisfaction and consistently low Net Promoter Scores or Customer Satisfaction Scores.

* Customer Segmentation:

Differences in churn factors among different customer segments and identifying specific factors that affect high value or high-risk customer groups.

1. **Data Collection:**

A complete dataset was given containing various 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. **Visualization Strategy:**

To effectively visualize a project centered on forecasting customer churn and improving customer retention through IBM Cognos, commence by integrating and organizing customer data within the platform. Define the essential metrics and key performance indicators (KPIs) linked to churn and retention. Opt for suitable visualization formats, such as employing line charts to dissect time-series trends, bar charts for customer segmentation insights, and maps to unveil geographical patterns. Construct interactive dashboards amalgamating these visualizations to offer stakeholders a comprehensive understanding of customer behaviour. Enhance interactivity by incorporating features like tooltips, filters, and color-coded alerts to underscore pivotal insights. Employ narrative reporting to craft a cohesive data narrative that is adaptable across various devices. Foster collaboration and feedback through comments and annotations, coupled with robust user training. Sustain an iterative approach by continually refining visualizations to align with evolving business requisites based on user input. IBM Cognos equips us with a potent toolkit to craft perceptive and actionable visualizations, bolstering your churn prediction and retention improvement endeavours.

1. **Predictive Modelling:**

* Data Collection and Preprocessing:
  + Data Acquisition: We begin by gathering a comprehensive dataset containing customer information, demographics, usage behavior, and historical interactions.
  + Data Preprocessing: Data quality is paramount. We meticulously preprocess the dataset, addressing missing values, encoding categorical variables, and normalizing numerical features. Additionally, we perform exploratory data analysis (**EDA**) to gain a deeper understanding of the data's characteristics.
* Model Selection:
  + We explore various machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting (**e.g., XGBoost or LightGBM**). Model selection is guided by empirical testing and cross-validation to identify the most suitable approach.
* Hyperparameter Tuning:
  + To optimize model performance, we conduct extensive hyperparameter tuning. This involves fine-tuning model parameters, such as **learning rates, tree depths, and regularization terms**, to achieve the best predictive accuracy.
* Handling Class Imbalance:
  + Given the potential class imbalance in churn prediction, we implement techniques such as **oversampling, undersampling, or using class weights** to address this issue and prevent biased model results.
* Model Deployment:
  + The final model is deployed for real-time or batch predictions, enabling businesses to proactively identify potential churners and take preventive actions.