**COGNITIVE COMPUTING**

**CUSTOMER CHURN PREDICTION**

**TEAM TRIANGLE**

Kushali Gupta (102317148)

Kashika Jindal (102317154)  
Manya Jindal (102317253)

1. **Introduction & Problem Statement**

Customer churn, where users stop using a service, leads to revenue loss and business instability. This project aims to build a machine learning model to predict churn based on customer behaviour, enabling businesses to take proactive steps to retain at-risk customers. The dataset we are using is Telco Customer Churn which includes details like contract type, tenure, monthly charges, and payment method. The target variable, "Churn", shows whether a customer stayed or left.

**Why is this important?**

* **Cost Savings**: Retaining customers is far cheaper than acquiring new ones.
* **Better Customer Experience**: Early churn prediction allows personalized offers and improved support.
* **Increased Revenue**: Lower churn means higher customer lifetime value (CLV) and sustainable growth.
* **Smarter Decisions**: Businesses can use data-driven insights to optimize marketing and retention strategies.

1. **Dataset Overview**

The dataset contains 7,043 customer records with the following attributes:

* **Contract Type** – Type of subscription (Month-to-month, One year, Two years).
* **Monthly Charges** – The amount a customer pays per month.
* **Payment Method** – Mode of payment (Credit Card, Electronic check, etc.).
* **Tenure** – The number of months a customer has stayed with the company.
* **Total Charges** – The total amount paid by the customer.
* **Internet Service Type** – Type of internet service used (DSL, Fiber optic, None).
* **Customer Support Interactions** – Number of times a customer contacted support.
* **Churn** – The target variable (Yes/No), indicating whether a customer left the service.

1. **Technology Stack**

**Libraries Used:**

* **Pandas** – For dataset handling and preprocessing
* **NumPy** – For numerical operations
* **Matplotlib & Seaborn** – For data visualization
* **SkLearn** – Used for train-test splitting, Decision Tree, Random Forest, and evaluation metrics
* **XGBoost** – Used for gradient boosting classification
* **Pickle** – Used for saving trained models

**Machine Learning Models Used (from SkLearn & XGBoost):**

* **Decision Tree Classifier** – Helps identify key factors leading to churn by making simple, rule-based decisions.
* **Random Forest Classifier** – Improves accuracy and reduces overfitting by combining multiple decision trees.
* **XGBoost Classifier** – Boosts performance by learning from errors, making it the most effective model in our project.

1. **ML Model Implementation & Evaluation**

* The dataset was preprocessed by handling missing values, converting text-based categories into numbers, scaling numerical features, and splitting the data into 80% training and 20% testing.
* Three machine learning models—Decision Tree, Random Forest, and XGBoost—were trained to predict customer churn.
* Model performance was evaluated using key metrics such as Accuracy, Confusion Matrix, Precision, Recall, F1-score, and ROC-AUC to determine the most effective model.

1. **Results & Insights (Visualizations & Metrics)**

* XGBoost performed the best, achieving the highest accuracy among all models.
* Random Forest also performed well, making fewer incorrect predictions than the Decision Tree.
* The main factors influencing customer churn were contract type, monthly charges, and tenure. Customers with month-to-month contracts and higher monthly bills were more likely to leave.
* Confusion Matrix & ROC Curve showed that XGBoost had the best balance between correctly predicting churn and minimizing false alarms.
* Feature importance analysis helped identify which customer details had the most impact on churn.

1. **Challenges & Future Improvements**

**Challenges:**

* **Data Quality & Availability –** Incomplete, inconsistent, or outdated data can affect model performance. Ensuring high-quality, real-time data is crucial for accurate predictions.
* **Churn Imbalance** – Churn datasets are often imbalanced, with fewer churn cases compared to non-churn. This can lead to biased models that fail to detect actual churners.
* **Feature Engineering Complexity** – Identifying meaningful features from raw telecom data requires domain expertise.
* **Changing Customer Behaviour** – Customer preferences evolve due to market trends, new competitors, and pricing changes. A static model may become ineffective over time.
* **Privacy & Compliance Issues** – Telecom data is sensitive, and regulatory constraints like GDPR can limit data collection and model deployment.

**Future Improvements:**

* **Advanced Machine Learning & Deep Learning** – Exploring deep learning models (e.g., LSTMs for sequential data) and reinforcement learning can improve predictive accuracy.
* **Customer Sentiment & Social Media Analysis** – Integrating text and sentiment analysis from customer complaints, reviews, and social media can provide deeper insights into churn behaviour.
* **Self-Learning Models** – Implementing continuous learning (online learning) ensures the model adapts to changing customer behaviours dynamically.

1. **Conclusion & Learnings**

**Conclusion**

Customer churn prediction is essential for telecom service providers to retain customers and maintain profitability. By analyzing customer behaviour, usage patterns, and demographic data, machine learning models can effectively predict churn and enable proactive retention strategies. Implementing predictive analytics helps telecom companies improve customer satisfaction, reduce churn rates, and optimize their marketing efforts.

**Learnings**

**Data Preprocessing is Essential –** Handling missing values, encoding categorical data, and scaling features are crucial for better model performance.

**Balancing Data Helps** – Since fewer customers leave than stay, techniques like **SMOTE** or class weighting improve model accuracy.

**Models Need Updates** – Customer behaviour changes over time, so models must be monitored and retrained to stay accurate.

1. Reference

* <https://www.kaggle.com/datasets/blastchar/telco-customer-churn?resource=download>
* <https://docs.python.org/3/>
* <https://docs.python.org/3/library/pickle.html>
* <https://xgboost.readthedocs.io/en/stable/>