

### CS5228 Team 7 Interim Progress Report

Name	Matriculation Number	Contribution
Chi Xu	A0225775W	Supervised Learning (advanced models such as XGBoost)
Gao Haochun	A0194525Y	Code review for EDA and data preprocessing; Unsupervised Learning (K-means)
Yuan Jinghan	A0296831U	Unsupervised Learning (DBSCAN)
Zhao Peiduo	A0196632X	Data preprocessing, EDA, base notebooks for supervised and unsupervised learning

#### Project Objective:

This project aims to leverage both supervised and unsupervised learning techniques to segment customers based on their usage patterns and predict customer churn. By identifying distinct user clusters, the telecom service provider can develop targeted pricing strategies and personalized advertising campaigns tailored to different customer segments. Moreover, by tracking usage behaviors associated with high churn risk, the company can implement proactive measures such as promotions and service improvements to enhance customer retention.

#### Project progress as of Week 8:

1. Data Preprocessing (Completed): One-Hot Encoding was applied to area\_code, while state was engineered to avoid excessive columns from OHE. Binary encoding was used for other categorical variables. Both standard-scaled and MinMax-scaled datasets were created and cached.
2. Exploratory Data Analysis (EDA) (Completed): Histograms were used to visualize feature distributions, and heatmaps were generated for correlation analysis. Highly correlated features, particularly charge-related columns, were removed, and PCA was applied for dimensionality reduction.
3. Unsupervised Learning – Customer Segmentation (In Progress): Initial clustering attempts were made using K-Means and DBSCAN on both MinMax-scaled and standard-scaled datasets.
4. Supervised Learning – Churn Prediction (In Progress): Logistic Regression, Random Forest, and SVM models were initially trained on both MinMax-scaled and standard-scaled datasets for churn prediction.

#### Task Assignment:

##### By week 10:

- Chi Xu : Supervised Learning (Experiment on XGBoost for supervised learning)
- Gao Haochun: Unsupervised Learning (Kmeans fine-tuning)
- Yuan Jinghan: Unsupervised Learning (DBSCAN fine-tuning)
- Zhao Peiduo: Supervised learning (Logistic Regression, Random Forest and SVM)

##### Week 10 - 12:

- Whole team: Compile findings into an illustrative poster and prepare for the presentation.