

## Data Collection and Preprocessing Phase

Date	15 july 2024
Team ID	739886
Project Title	Telecom customer churn prediction
Maximum Marks	2 Marks

### Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

#### Data Collection Plan Template

Section	Description
Project Overview	<p>The project focuses on developing predictive analytics capabilities to address customer churn within the telecom industry. By harnessing machine learning techniques, the goal is to accurately forecast which customers are likely to leave a telecom service provider. This involves collecting and preprocessing extensive customer data, performing in-depth exploratory analysis to uncover patterns, and engineering features that enhance prediction accuracy. Multiple machine learning models will be trained and evaluated to identify the optimal approach for churn prediction. The final model will be deployed in a production environment to enable real-time predictions, supported by continuous monitoring and updates to maintain efficacy. The project aims to deliver actionable insights that empower telecom companies to implement targeted retention strategies, ultimately reducing churn rates, improving customer satisfaction, and fostering sustainable business growth.</p>

Data Collection Plan	Kaggle
----------------------	--------

Raw Data Sources Identified	<p>For telecom customer churn prediction, gathering comprehensive raw data sources is crucial to develop accurate predictive models. These sources include customer demographic data, providing insights into age, gender, and location. Usage patterns data encompass details like data usage, call duration, and service frequency, revealing engagement levels. Billing and payment data track monthly charges, payment methods, and history, highlighting financial behaviors. Customer service interactions data capture call volumes, issue resolutions, and response times, reflecting service satisfaction. Contract and subscription details cover plan types, tenure, and changes, indicating loyalty trends. Additionally, customer feedback and satisfaction surveys offer sentiment insights. Historical churn data logs previous instances and reasons for churn, offering patterns. External sources include market trends, competitor actions, and economic indicators, influencing customer decisions. Integrating these diverse data sources ensures a comprehensive understanding of churn drivers, empowering effective retention strategies and service enhancements in telecom operations.</p>
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions

Kaggle Dataset	The columns present are: RowNumber ,CustomerId,surname,creditScore ,Geography ,Gender,Age ,Tenure ,Balance , NumOfProducts,HasCrCard,IsActiveMember EstimatedSalary,Exited.	<a href="https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling">https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling</a>	CSV	300 KB	Public
----------------	---	---	-----	--------	--------