

Depi Task 1

1. Define the Problem

- **Objective:** The goal is to identify the main factors driving customer churn and predict which customers are likely to leave in the next 3 months. By understanding these factors, we aim to develop strategies to retain customers.
- **Business Impact:** Reducing churn can improve customer retention, revenue stability, and customer lifetime value for Tele.

2. Collect and Explore Data

- **Dataset Overview:** The dataset includes customer demographics, account details, services used, and churn status. Understanding what each feature represents is key to accurate analysis.
- **Initial Data Analysis:** We would start by examining distributions, correlations, and missing values to gain insight into the structure and quality of the data.
- **Feature Exploration:** Identify potential features that might be strong indicators of churn, such as tenure, payment methods, and contract types.

3. Data Cleaning and Preprocessing

- **Handle Missing Data:** Use imputation techniques or remove irrelevant data entries to ensure data consistency.
- **Encode Categorical Variables:** Convert categorical features (like service type and contract type) into numerical format, suitable for modeling.
- **Feature Engineering:** Create new features, such as 'average monthly expenditure' or 'service bundle count,' which may reveal additional insights.

4. Data Analysis

- **Visualize Key Features:** Visualizations like histograms, box plots, and heatmaps can help reveal patterns in the data.
- **Churn Analysis:** Compare distributions of features for customers who churned versus those who stayed to uncover distinct patterns.
- **Correlation Analysis:** Identify correlations among features to avoid redundancy and improve model performance.

5. Model Building

- **Train/Test Split:** Divide the data into training and test sets to evaluate model performance.
- **Model Selection:** Use classification algorithms such as Logistic Regression, Decision Trees, Random Forest, or XGBoost to predict churn.
- **Feature Importance:** Determine feature importance to identify the key drivers of churn and focus on these areas for targeted interventions.

6. Model Evaluation

- **Metrics:** Evaluate models using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to assess predictive accuracy and ability to identify churn effectively.
- **Model Validation:** Perform cross-validation to ensure robustness and avoid overfitting.
- **Threshold Adjustment:** Adjust the model's probability threshold to balance between false positives and false negatives, considering business priorities.

7. Interpret Results and Generate Insights

- **Feature Impact:** Identify top factors contributing to churn, such as monthly charges, contract type, or service usage. For example, higher churn might correlate with month-to-month contracts or lower service usage.
- **Customer Segmentation:** Segment customers based on churn probability and drivers to personalize retention strategies.
- **Actionable Insights:** Suggest targeted interventions, such as offering discounts to high-risk customers or promoting long-term contracts.

8. Implement and Monitor

- **Deploy Model:** Implement the model in a production environment to make real-time churn predictions.
- **Monitor Performance:** Continuously monitor model performance to ensure accuracy and recalibrate if necessary.
- **Evaluate Retention Strategies:** Measure the effectiveness of retention campaigns and strategies over time, adjusting approaches based on real-time data and outcomes.