**Customer Churn Prediction - Approach Summary**

**Data Preprocessing:**

Our approach to predicting customer churn began with data preprocessing, a critical step to ensure the dataset's quality and prepare it for modeling. Key preprocessing steps included:

1. Data Loading: We loaded the provided dataset, which contained customer attributes like 'Age,' 'Gender,' 'Location,' 'Subscription\_Length\_Months,' 'Monthly\_Bill,' 'Total\_Usage\_GB,' and the target variable 'Churn.'

2. Handling Missing Data: We assessed and addressed missing data appropriately, applying techniques such as imputation or removal, depending on the extent and impact of missing values.

**Feature Engineering:**

To enhance the model's predictive capabilities, we performed feature engineering on the limited feature set. Feature engineering involved creating new features and transforming existing ones:

1. Subscription Length in Years: We calculated 'Subscription\_Length\_Years' by dividing 'Subscription\_Length\_Months' by 12, providing insights into customer loyalty.

2. Monthly Bill per GB: 'Monthly\_Bill\_Per\_GB' was calculated by dividing 'Monthly\_Bill' by 'Total\_Usage\_GB,' revealing the cost-effectiveness of customers' plans.

3. Encoding Categorical Variables: We encoded categorical variables like 'Gender' and 'Location' using techniques like one-hot encoding or label encoding, enabling their use in the models.

**Model Selection:**

Our choice of machine learning models was diverse, aiming to explore different approaches. We implemented:

1. Random Forest: A popular ensemble model known for its robustness and ability to handle diverse data.

2. Logistic Regression: A simple yet effective linear model suitable for binary classification tasks.

3. Neural Network (Keras): A deep learning approach using a feed forward neural network architecture.

**Evaluation and Conclusion:**

Despite our varied modeling approaches and thorough data preprocessing and feature engineering, we consistently obtained an accuracy of 0.5, indicating that our models were not effectively capturing the churn patterns in the dataset. Several factors could contribute to this, including data limitations, feature engineering choices, and the potential for class imbalance.

In the conclusion, we emphasized the importance of considering class imbalance, exploring further domain-specific feature engineering, and potentially collecting more data to improve model performance. We also highlighted the significance of using appropriate evaluation metrics aligned with the business context.

This assignment served as a valuable learning experience, demonstrating the complexities and challenges often encountered in real-world machine learning projects, and reinforcing the importance of iterative refinement and domain understanding.