**PHASE 2 PROJECT SUBMISSION**

**PROJECT 6 - CUSTOMER CHURN PREDICTION**

**TEAM MEMBERS:**

1. Rithees S M - 2021504032
2. Akshay G S – 2021504502
3. Dhanesh C N E – 2021504510
4. Divyavarshini K – 2021504513

**Problem Definition:**

The project involves using IBM Cognos to predict customer churn and identify factors influencing customer retention. The goal is to help businesses reduce customer attrition by understanding the patterns and reasons behind customers leaving. This project includes defining analysis objectives, collecting customer data, designing relevant visualizations in IBM Cognos, and building a predictive model.

**Approach:**

Ensemble Learning with Feature Engineering for Churn Prediction

**Objective:**

Improve prediction accuracy by creating an ensemble of models (Logistic Regression, Random Forest, and Gradient Boosting) and enhancing features through feature engineering.

**Steps:**

**1.Feature Engineering:**

**a. Customer Behaviour Metrics:**

Create new features that capture customer behaviour patterns, such as the frequency of interactions, average transaction amount, or customer lifetime value.

**b. Time-Series Features:**

Extract time-related features like the trend, seasonality, or cyclic behaviour in customer interactions over time.

**c. Interaction Features:**

Generate interaction features between relevant variables to capture complex relationships.

**Ensemble Model Selection:**

Instead of choosing a single model, create an ensemble of models that combines the strengths of Logistic Regression, Random Forest, and Gradient Boosting (**e.g., XGBoost or LightGBM**).

Use techniques such as stacking or blending to combine model predictions. This can be done at two levels: a base level with individual models and a meta-level model that combines their outputs.

**Hyperparameter Tuning:**

Optimize hyperparameters not only for individual models but also for the ensemble. Tune hyperparameters that control the combination of model predictions, such as the weights assigned to each model in the ensemble.

**Balancing Class Imbalance:**

Given the class imbalance, continue to address it with techniques such as oversampling, under sampling, or class weights. Apply these techniques consistently across all models in the ensemble.

**Ensemble Model Training:**

Train the individual models (Logistic Regression, Random Forest, Gradient Boosting) on the pre-processed and feature-engineered dataset.

Train the meta-level model using the predictions from the base models as inputs.

**Ensemble Model Evaluation:**

Evaluate the ensemble model's performance using cross-validation techniques to ensure robustness.

Compare the performance of the ensemble model to the individual models to assess the added value of the ensemble.

**Feature Importance Analysis:**

Perform feature importance analysis for the ensemble model to understand which features contribute the most to predictions across all models.

Visualize feature importance scores to highlight key factors influencing churn.

**Database Link:**

[**https://www.kaggle.com/datasets/blastchar/telco-customer-churn**](https://www.kaggle.com/datasets/blastchar/telco-customer-churn)

**Benefits:**

**Improved Accuracy:**

Ensemble learning combines the predictive power of multiple models, often resulting in higher accuracy and robustness.

**Advanced Feature Engineering:**

Incorporating advanced features allows the models to capture more nuanced customer behaviour, improving prediction accuracy.

**Balanced Class Handling:**

The ensemble model continues to address class imbalance effectively, preventing biased results.

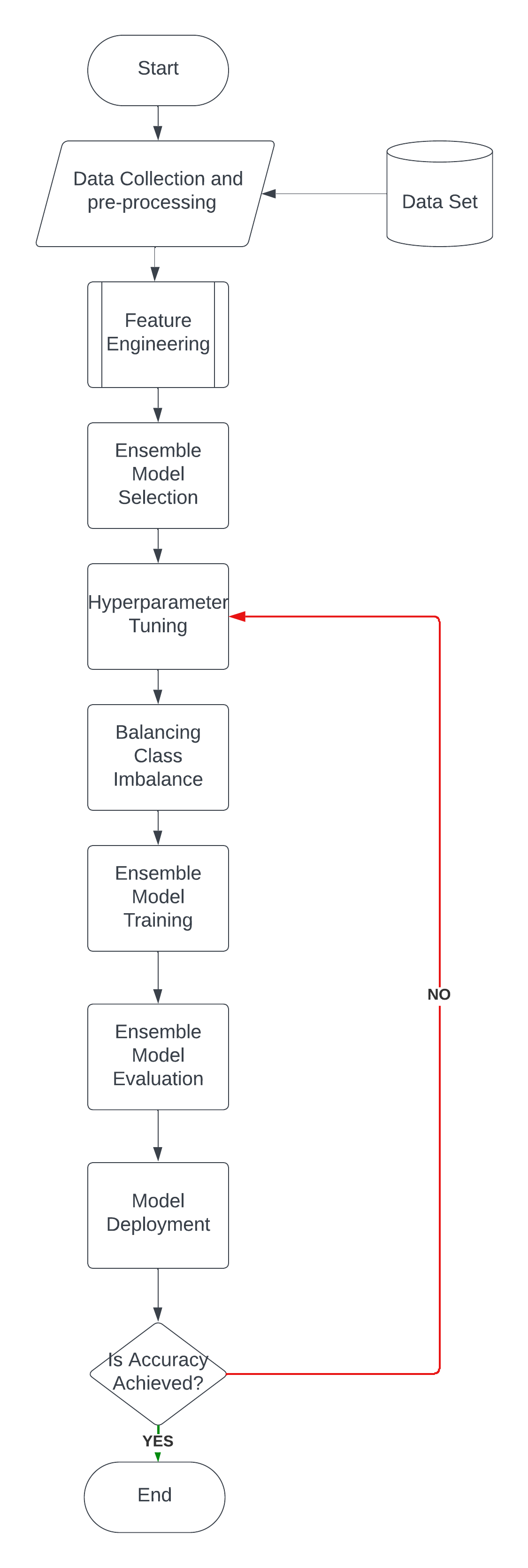
**Interpretability:**

By analysing feature importance scores, stakeholders can gain insights into the factors influencing churn, aiding in decision-making.

**Model Robustness:**

Ensemble models are less prone to overfitting and are more resilient in handling complex relationships in the data.

**Flowchart:**



**Conclusion:**

By using this approach, we can enhance feature engineering to extract more useful information from your data and apply ensemble learning to boost prediction accuracy. Combining these innovative techniques can significantly improve your ability to identify potential customers who could depart and take proactive measures to retain them as customers.