Introduction:

The project began with a clear understanding of the business objective: predicting customer inactivity based on the data provided. This involved creating a machine-learning model to forecast the probability of customers refraining from transactions for a 90-day period. The objective was to enable proactive retention strategies. I split the project into three iterations, each building upon the previous iteration to enhance the model's predictive capabilities.

Iteration 1: Clean and fill the data using simple imputation methods. Build 5 classifier models

2. Data Cleaning and Feature Engineering: \*\*

a. \*\*Handling Missing Values: \*\*

- Identified and assessed missing values in the dataset.

- Applied basic imputation techniques, such as filling missing values with the median, to address gaps in the data.

b. \*\*Consistency Check: \*\*

- Ensured data consistency by checking for outliers, anomalies, and unexpected values.

- Addressed any discrepancies to create a more reliable dataset.

c. \*\*Categorical Variable Encoding: \*\*

- Encoded categorical variables to numerical format for compatibility with machine-learning models.

d. \*\*Feature Engineering: \*\*

- Introduced new features like ARPU segment and transaction frequency to provide more relevant information for the predictive models.

- Iteratively refined features to enhance their predictive power.

\*\*3. First Iteration: Baseline Model Building: \*\*

a. \*\*Model Selection: \*\*

- Chose five diverse models (Logistic Regression, RF, XGB, MLP Classifier, ANN Classifier) to establish a baseline performance.

b. \*\*Imputation Strategy: \*\*

- Applied basic imputation techniques to handle missing values.

- Evaluated models with different imputation strategies to identify the most effective approach.

c. \*\*Feature Importance: \*\*

- Extracted feature importance from RF, XGB, and other models to understand the significance of each variable in predicting customer inactivity.

\*\*4. Second Iteration: Exploratory Data Analysis (EDA) and Model Tuning: \*\*

a. \*\*EDA: \*\*

- Conducted EDA to explore relationships between key features and the target variable.

- Identified the most influential features and their impact on customer inactivity.

b. \*\*Model Comparison: \*\*

- Introduced LightGBM to compare model performance on raw versus cleaned data.

- Assessed the strengths and weaknesses of each model to make informed decisions for the subsequent iterations.

c. \*\*Hyperparameter Tuning: \*\*

- Tuned hyperparameters of the chosen models (especially XGB) to optimize performance.

- Evaluated the impact of hyperparameter changes on model accuracy.

\*\*5. Third Iteration: Feature Engineering and Ensembling: \*\*

a. \*\*Feature Engineering: \*\*

- Focused on creating new features based on insights gained from EDA and model tuning.

- Iteratively refined features to enhance predictive performance.

b. \*\*Model Tuning and Ensembling: \*\*

- Continued hyperparameter tuning for XGB and LightGBM.

- Implemented ensembling techniques (e.g., stacking or blending) to combine the strengths of multiple models.

c. \*\*Performance Evaluation: \*\*

- Assessed the final model's performance using appropriate metrics.

- Validated the effectiveness of feature engineering and ensembling in improving predictive accuracy.

\*\*6. Conclusion: \*\*

The three iterations collectively resulted in a robust predictive model for customer inactivity. Each stage contributed valuable insights, from baseline model building to feature engineering and ensembling. The iterative approach ensured a thorough exploration of the dataset and model behaviour.

\*\*7. Recommendations for Implementation: \*\*

- \*\*Real-time Integration: \*\* Implement the final ensemble model for real-time predictions within Sasol's operational systems.

- \*\*Monitoring: \*\* Regularly monitor key features identified during EDA and feature engineering for ongoing analysis of customer behaviour.

- \*\*Integration: \*\* Consider integrating the developed model seamlessly into Sasol's existing systems for practical and continuous use.

\*\*8. Submission: \*\*

The attached document includes the updated CV, detailed code documentation for each iteration, and explanations of data cleaning, feature engineering, and model building steps. The provided solution aligns with Sasol's requirements and aims to contribute significantly to the company's goal of innovating for a better world.