Model Development Phase Template

Date	25 March 2025
Team ID	SWTID1749641473
Project Title	Early Prediction for Chronic Kidney Disease Detection: A Progressive Approach to Health Management
Maximum Marks	6 Marks

Model Selection Report

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

Dataset Summary:

- Source: chronickidneydisease.csv
- Rows: ~400
- Features: Demographic, clinical, and biochemical indicators (e.g., blood pressure, hemoglobin, albumin)
- Target: class (CKD = 1, Not CKD = 0)
- Preprocessing:
  - Replaced '?' with NaN
  - Dropped irrelevant columns (id)
  - Handled missing values (median for numerical, mode for categorical)
  - Encoded categorical variables using one-hot encoding
  - Split into 80% training and 20% testing sets

Model Description and Performance:

Model	Description	Hyperparameters	Performance Metric
XGBoost	An ensemble-based gradient boosting model, XGBoost optimizes predictive performance using decision trees in sequence. Known for speed and accuracy.	objective='binary:logistic', eval_metric='logloss', random_state=42	Accuracy = 97.5%

Reasons for Model Selection:

- Superior Accuracy: Achieved 97.5% accuracy on the test set, outperforming typical baselines in similar medical datasets.
- Handles Missing Values & Imbalanced Features well through boosting iterations.
- Feature Importance Visualization: Offered interpretable insights into which clinical features most strongly influenced predictions.
- Efficient Training: Trained efficiently on tabular data with high dimensionality due to one-hot encoding.