Model Development Phase Template

Date	19 June 2025
Team ID	SWTID1749713922
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.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Random Forest	Ensemble of decision trees; robust, handles complex patterns, reduces overfitting, provides feature importance.	n_estimators: 10 min_samples_split: 80 min_samples_leaf: 40 min_impurity_decrease: 0.05 max_samples: 0.6 max_features: 0.3 max_depth: 7 ccp_alpha: 0.05	Test Accuracy: 0.9250 Precision: 0.9375 Recall: 0.9250 F1 Score: 0.9260 ROC AUC: 0.9933 PR AUC: 0.9889

XGBoost	Gradient-boosted decision trees; efficient, handles imbalanced data and overfitting well.	subsample: 0.9 scale_pos_weight: 1 reg_lambda: 10 reg_alpha: 0 n_estimators: 200 min_child_weight: 7 max_depth: 8 learning_rate: 0.15 gamma: 0.5 colsample_bytree: 0.6	Test Accuracy: 0.9625 Precision: 0.9626 Recall: 0.9625 F1 Score: 0.9624 ROC AUC: 0.9933 PR AUC: 0.9905
LightGBM	Fast gradient boosting framework by Microsoft; uses histogram-based methods for speed and accuracy.	subsample: 0.6 reg_lambda: 0.1 reg_alpha: 0.01 num_leaves: 15 n_estimators: 10 min_split_gain: 1 min_child_weight: 0.001 min_child_samples: 30 max_depth: 2 learning_rate: 0.2 colsample_bytree: 0.5 • class_weight: balanced	Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000 ROC AUC: 1.0000 PR AUC: 1.0000
Logistic Regression	Linear model suitable for binary classification;	• solver: liblinear • penalty: l2 • l1_ratio: 0.5	Test Accuracy: 1.0000 Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000

	interpretable and simple.	• fit_intercept: False • C: 1	ROC AUC: 1.0000 PR AUC: 1.0000
SVM (RBF)	Effective for non-linear data; RBF kernel maps inputs into higher dimensions.	• shrinking: False • kernel: rbf • gamma: 0.1 • cache_size: 500 • C: 100	Test Accuracy: 1.0000 Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000 ROC AUC: 1.0000 PR AUC: 1.0000
Gradient Boosting	Sequentially builds trees; reduces bias and variance.	subsample: 0.6 • n_estimators: 100 • min_samples_split: 80 • min_samples_leaf: 40 • min_impurity_decrease: 0.05 • max_features: log2 • max_depth: 2 • learning_rate: 0.01 • ccp_alpha: 0.0	Test Accuracy: 0.9500 Precision: 0.9500 Recall: 0.9500 F1 Score: 0.9500 ROC AUC: 0.9947 PR AUC: 0.9920
CatBoost	Gradient boosting with categorical feature support; often performs well with minimal tuning.	• subsample: 0.6 • min_data_in_leaf: 40 • learning_rate: 0.2 • l2_leaf_reg: 5 • iterations: 10 • depth: 4 • colsample_bylevel: 0.5 • class_weights: [1, 3]	Test Accuracy: 0.9625 Precision: 0.9659 Recall: 0.9625 F1 Score: 0.9628 ROC AUC: 0.9973 PR AUC: 0.9958

		• border_count: 64 • bagging_temperature: 0	
Decision Tree	Interpretable, simple model; captures non-linear patterns.	min_samples_split: 40 • min_samples_leaf: 40 • min_impurity_decrease: 0.05 • max_features: None • max_depth: 10 • ccp_alpha: 0.05	Test Accuracy: 0.8750 Precision: 0.8883 Recall: 0.8750 F1 Score: 0.8767 ROC AUC: 0.8867 PR AUC: 0.7509
K-Nearest Neighbors	Non-parametric, instance-based learning; sensitive to local patterns.	weights: distance • p: 2 • n_neighbors: 21 • metric: euclidean • leaf_size: 40 • algorithm: auto	Test Accuracy: 1.0000 Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000 ROC AUC: 1.0000 PR AUC: 1.0000
Linear Discriminant Analysis	Projects data to maximize class separability; works well on normally distributed features.	• n_components: None • priors: None • shrinkage: 0.3 • solver: lsqr	Test Accuracy: 1.0000 Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000 ROC AUC: 1.0000 PR AUC: 1.0000
AdaBoost	Adaptive boosting technique; focuses on hard-to-classify points.	• algorithm: SAMME • learning_rate: 0.01 • n_estimators: 10	Test Accuracy: 0.8750 Precision: 0.8883 Recall: 0.8750 F1 Score: 0.8767 ROC AUC: 0.8867 PR AUC: 0.7509

SGD Classifier	Optimizes with stochastic gradient descent; good for large-scale datasets.	• validation_fraction: 0.3 • penalty: l1 • learning_rate: constant • l1_ratio: 0.3 • eta0: 0.01 • early_stopping: True • alpha: 0.01	Test Accuracy: 1.0000 Precision: 1.0000 Recall: 1.0000 F1 Score: 1.0000 ROC AUC: 1.0000 PR AUC: 1.0000