**Model Development Phase Template**

|  |  |
| --- | --- |
| Date | 12 JULY 2024 |
| Team ID | SWTID1720438678 |
| Project Title | Early prediction of chronic kidney disease. |
| 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)** |
| **Logistic Regression** | Logistic Regression:  Strengths: Simple to interpret, computationally efficient, works well for linearly separable data.  Weaknesses: Might not capture complex relationships between features and the target variable. Can suffer from overfitting if not regularized properly. | random\_state=2  max\_iter=1000 | Accuracy: 0.9875  F1 Score: 0.98  Precision: 1  Recall: 0.98 |
| **Random forest** | Random Forest:  Strengths: Combines multiple decision trees to improve accuracy and reduce variance. Handles missing data and irrelevant features well.  Weaknesses: Can be a "black box" model, making interpretation more challenging. May require more computational resources for training compared to simpler models | n­\_estimators  max\_depth  criterion  random\_state | Accuracy: 0.9625  F1 Score: 0.97  Precision: 1  Recall: 0.94 |
| **Decision tree** | A decision tree is a supervised machine learning algorithm used for classification and regression. It resembles a flowchart with internal nodes representing decisions based on data attributes, branches representing possible outcomes, and leaf nodes representing final predictions. Decision trees are popular for their interpretability but can be prone to overfitting. Techniques like pruning help to mitigate this issue. | criterion='entropy'  random\_state=0 | Accuracy: 0.95  F1 Score: 0.93  Precision: 0.93  Recall: 0.93 |
| **SVM** | SVM is a supervised machine learning algorithm used for classification and regression. It finds the optimal hyperplane to separate data points into different classes. SVMs excel in high-dimensional spaces and are effective for complex patterns. Kernel trick allows SVMs to handle non-linear data by mapping it into higher dimensions. |  | Accuracy: 0.85  F1 Score: 0.83  Precision: 0.72  Recall: 0.97 |