**Model Development Phase Template**

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720013031 |
| Project Title | Prediction and Analysis of Liver Patient Data Using Machine Learning |
| 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 is a linear model used for binary classification tasks. It estimates the probability that a given input belongs to a particular class by fitting a logistic function (sigmoid) to the linear combination of input features. It's simple, interpretable, and works well for linearly separable data. | random\_state=42 | Accuracy score = 76% |
| K neighbors  Classifier | The KNeighbors Classifier (k-NN) is a non-parametric, instance-based learning algorithm. It classifies a data point based on the majority class among its k-nearest neighbors in the feature space. The value of k determines the number of neighbors considered. | n\_neighbors=6, weights='uniform',  algorithm='kd\_tree',  leaf\_size=20 | Accuracy score = 77% |
| Random  Forest  Classifier | .Random Forest Classifier is an ensemble learning method that combines the predictions of multiple decision trees to improve accuracy and prevent overfitting. Each tree is trained on a random subset of the data and features, and their results are aggregated for final classification. This approach leverages the diversity of the individual trees to enhance robustness and accuracy. | n\_estimators=500,criterion='entropy',random\_state=18 | Accuracy score = 76% |
| SVC | Support Vector Classifier (SVC) is a supervised learning model that constructs a hyperplane or set of hyperplanes in a high-dimensional space to classify data points. SVC aims to maximize the margin between different classes, making it robust to outliers and effective for high-dimensional data . | kernel="rbf",random\_state=100,gamma='auto',verbose=2,decision\_function\_shape='ovo' | Accuracy score = 78% |