**Model Development Phase**

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | SWTID1720077079 |
| Project Title | Wild Blueberry Yield Prediction |
| Maximum Marks | 6 Marks |

**Model Selection Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Description** | **Hyperparameters** | **Performance Metric** |
| Linear Regression | Linear Regression is a simple and interpretable model, but it assumes a linear relationship between the features and the target variable | fit\_intercept=True  copy\_X=True  n\_jobs=None  positive=False |  |
| Decision Tree | It constructs a tree-like model of decisions and their possible consequences. | random\_state=42  max\_depth=5 |  |
| Random Forest | Random Forest is a collection of individual Decision Tree models, where each tree is trained on a random subset of the training data and a random subset of the features. | n\_estimators=100  random\_state=42  max\_depth=5 |  |
| XGBoost | The XGBoost model is a gradient boosting algorithm used for regression tasks, with the objective function set to 'reg:squarederror' to optimize the squared error loss, which is appropriate for regression problems. | max\_depth=5  n\_estimators=100  learning\_rate=0.1 |  |
| SVM Regression | SVM Regression is a supervised learning algorithm used for solving regression problems. It works by finding the best-fitting hyperplane in a high-dimensional feature space that minimizes the error between the predicted and actual target values. | C=100,  epsilon=0.001  gamma='auto'  kernel='linear' |  |