**Model Optimization and Tuning Phase Template**

|  |  |
| --- | --- |
| Date | 9 JULY 2024 |
| Team ID | 739661 |
| Project Title | Anemiasense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 10 Marks |

**Model Optimization and Tuning Phase**

Optimizing and tuning machine learning models is crucial for enhancing the accuracy, reliability, and robustness of Anemiasense in recognizing anemia with precision. This phase focuses on refining model performance through systematic adjustments of hyperparameters and techniques tailored to the dataset characteristics and model requirements.

### Hyperparameter Tuning Documentation (6 Marks):

|  |  |  |
| --- | --- | --- |
| **Model** | **Tuned Hyperparameters** | **Optimal Values** |
| Logistic Regression | Regularization: L2 penalty (C) | C = 0.1 |
| Random Forest | Number of trees, Maximum features, Minimum samples per leaf. | Number of trees = 150, Max features = sqrt, Min samples leaf = 2 |
| Decision Tree | Maximum depth, Minimum samples per leaf | Max depth = 10, Min samples leaf = 5 |
| Gaussian Naive Bayes | No hyperparameters to optimize | N/A (No tuning required) |
| Gradient Boosting Classifier | Learning rate, Number of trees, Maximum depth | Learning rate = 0.05, Number of trees = 200, Max depth = 3 |

### Performance Metrics Comparison Report (2 Marks):

|  |  |  |
| --- | --- | --- |
| **Model** |  | **Optimized Metric** |
| Logistic Regression |  |  |
| Random Forest |  |  |
| Decision Tree |  |  |
| Gaussian Naive Bayes |  |  |
| Gradient Boosting Classifier |  |  |

### Final Model Selection Justification (2 Marks):

|  |  |
| --- | --- |
| **Final Model** | **Reasoning** |
| Gradient Boosting Classifier | After extensive experimentation and hyperparameter tuning, the Gradient Boosting Classifier emerged as the optimal choice for several reasons: |
| 1. Performance Metrics: | |  | | --- | | - **Highest Optimized Accuracy:** Through rigorous cross-validation and  hyperparameter tuning, the Gradient Boosting Classifier consistently achieved the  highest accuracy among the tested models. This indicates its ability to correctly  classify anemia cases with high precision. |  |  |  | | --- | --- | |  |  |  |  |  | | --- | --- | |  | - **Highest F1 Score:** F1 score, which balances precision and recall, is crucial in  medical diagnostics where both false positives and false negatives can have  significant consequences. The Gradient Boosting Classifier demonstrated the highest  F1 score after tuning, indicating robust performance across multiple evaluation  metrics. | |
| 2. Ensemble Learning Benefits: | |  | | --- | | - **Robustness to Overfitting:** Gradient Boosting combines multiple weak learners  (usually decision trees) sequentially, focusing on instances that previous models  misclassified. This ensemble method helps mitigate overfitting and enhances  generalization ability, crucial for reliable anemia recognition across diverse datasets. |  |  |  | | --- | --- | |  |  |  |  |  | | --- | --- | |  | - **Effective Handling of Complex Relationships:** Anemia classification can  involve intricate relationships between various clinical features. Gradient Boosting  effectively captures these complexities through its iterative learning process, thereby  improving model accuracy compared to simpler models. | |
| 3. Practical Considerations: | |  | | --- | | - **Scalability and Deployment:** Gradient Boosting, while computationally intensive  during training, can be efficiently deployed in production environments. Its  predictive power and the ability to handle large datasets make it suitable for real-time or batch processing scenarios typical in healthcare settings. |  |  |  | | --- | --- | |  |  |  |  |  | | --- | --- | |  | - **Interpretability:** While not as straightforward as simpler models like logistic  regression, Gradient Boosting can still provide insights into feature importance,  aiding clinicians in understanding which clinical factors contribute most to anemia  diagnosis. | |
| 4. Industry Standard: | |  | | --- | | - **Widely Adopted in Healthcare:** Gradient Boosting techniques are well-established in medical diagnostics and have shown success in various healthcare applications, making them a reliable choice backed by industry adoption and research support. |  |  |  | | --- | --- | |  |  | |
| Conclusion: | Based on its superior performance in accuracy, F1 score, robustness, and practical suitability for deployment, the Gradient Boosting Classifier is selected as the final model for "Anemiasense." Its capabilities align closely with the project's objectives of achieving precise and reliable anemia recognition through machine learning. |