

Model Optimization and Tuning Phase

Date	10 July 2024
Team ID	SWTID1720080033
Project Title	Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation:

Model	Tuned Hyperparameters	Optimal Values
Logistic Regression Model	None, as the accuracy of the model was 100% without hyperparameter tuning.	Default Values
Random Forest Model	None, as the accuracy of the model was 100% without hyperparameter tuning.	Default Values
Decision Tree Model	None, as the accuracy of the model was 100% without hyperparameter tuning.	Default Values
Naive Bayes Model	<pre>param_grid = { 'var_smoothing': [1e-9, 1e-8, 1e-7, 1e-6, 1e-5] }</pre>	Best hyperparameters for Gaussian Naive Bayes: {'var_smoothing': 1e-09}

SVM Model	<pre>param_grid = { 'C': [0.1, 1, 10, 100], 'kernel': ['linear', 'poly', 'rbf', 'sigmoid'], 'degree': [2, 3, 4], 'gamma': ['scale', 'auto'], 'coef0': [0, 0.5, 1] }</pre>	The Best hyperparameters for SVM: {'C': 100, 'coef0': 0, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'}
Gradient Boosting Classifier Model	None, as the accuracy of the model was 100% without hyperparameter tuning.	Default Values

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Logistic Regression Model	Accuracy Score of Linear Regression Model: 1.0	Accuracy Score of Linear Regression Model: 1.0
	precision recall f1-score support	precision recall f1-score support
	0 1.00 1.00 1.00 167	0 1.00 1.00 1.00 167
	1 1.00 1.00 1.00 118	1 1.00 1.00 1.00 118
	accuracy 1.00 285	accuracy 1.00 285
	macro avg 1.00 1.00 1.00 285	macro avg 1.00 1.00 1.00 285
	weighted avg 1.00 1.00 1.00 285	weighted avg 1.00 1.00 1.00 285
Random Forest Model	Accuracy Score of Random Forest Model: 1.0	Accuracy Score of Random Forest Model: 1.0
	precision recall f1-score support	precision recall f1-score support
	0 1.00 1.00 1.00 167	0 1.00 1.00 1.00 167
	1 1.00 1.00 1.00 118	1 1.00 1.00 1.00 118
	accuracy 1.00 285	accuracy 1.00 285
	macro avg 1.00 1.00 1.00 285	macro avg 1.00 1.00 1.00 285
	weighted avg 1.00 1.00 1.00 285	weighted avg 1.00 1.00 1.00 285
Decision Tree Model	Accuracy Score of Decision Tree Model: 1.0	Accuracy Score of Decision Tree Model: 1.0
	precision recall f1-score support	precision recall f1-score support
	0 1.00 1.00 1.00 167	0 1.00 1.00 1.00 167
	1 1.00 1.00 1.00 118	1 1.00 1.00 1.00 118
	accuracy 1.00 285	accuracy 1.00 285
	macro avg 1.00 1.00 1.00 285	macro avg 1.00 1.00 1.00 285
	weighted avg 1.00 1.00 1.00 285	weighted avg 1.00 1.00 1.00 285

Naive Bayes Model	<pre> Accuracy Score of Naive Bayes Model: 0.9403508771929825 precision recall f1-score support 0 0.93 0.97 0.95 160 1 0.96 0.90 0.93 125 accuracy 0.94 0.94 0.94 285 macro avg 0.94 0.94 0.94 285 weighted avg 0.94 0.94 0.94 285 </pre>	<pre> Best hyperparameters for Gaussian Naive Bayes: {'var_smoothing': 1e-09} Gaussian Naive Bayes accuracy: 0.9403508771929825 Classification Report for SVM: precision recall f1-score support 0 1.00 1.00 1.00 167 1 1.00 1.00 1.00 118 accuracy 1.00 1.00 1.00 285 macro avg 1.00 1.00 1.00 285 weighted avg 1.00 1.00 1.00 285 </pre>
SVM Model	<pre> accuracy_score of SVM Model: 0.9964912280701754 classification_report: precision recall f1-score support 0 1.00 0.99 1.00 167 1 0.99 1.00 1.00 118 accuracy 1.00 1.00 1.00 285 macro avg 1.00 1.00 1.00 285 weighted avg 1.00 1.00 1.00 285 </pre>	<pre> The Best hyperparameters for SVM: {'C': 100, 'coef0': 0, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'} accuracy_score of SVM: 1.0 Classification Report for SVM: precision recall f1-score support 0 1.00 1.00 1.00 167 1 1.00 1.00 1.00 118 accuracy 1.00 1.00 1.00 285 macro avg 1.00 1.00 1.00 285 weighted avg 1.00 1.00 1.00 285 </pre>
Gradient Boosting Classifier Model	<pre> accuracy_score of Gradient Boosting Classifier Model: 1.0 classification_report: precision recall f1-score support 0 1.00 1.00 1.00 167 1 1.00 1.00 1.00 118 accuracy 1.00 1.00 1.00 285 macro avg 1.00 1.00 1.00 285 weighted avg 1.00 1.00 1.00 285 </pre>	<pre> accuracy_score of Gradient Boosting Classifier Model: 1.0 classification_report: precision recall f1-score support 0 1.00 1.00 1.00 167 1 1.00 1.00 1.00 118 accuracy 1.00 1.00 1.00 285 macro avg 1.00 1.00 1.00 285 weighted avg 1.00 1.00 1.00 285 </pre>

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest Model	<p>The Random Forest Model algorithm contains many decision trees on various subsets of the given dataset and takes the mean of it to improve the accuracy of that dataset. It improves the accuracy and reduces overfitting by combining multiple trees. Random Forest Model also analyses feature importance by looking at how much each feature decreases the impurity (entropy) on average across all trees in the forest. Features that lead to larger reductions in impurity are considered more important.</p>

	<p>In the Anemia Detection Process, a lot of blood parameters are considered, finding the most important features amongst them is vital in order to create a model that will understand the underlying pattern. Detecting Anemia is a safety critical process which will directly result in the overall well being of the patient. Therefore, a model which will analyze multiple patterns and give a good accuracy along with identifying the important features in a must.</p>
--	--