

## Model Development Phase

Date	10 July 2024
Team ID	SWTID1720080033
Project Title	<b>Anemia Sense:</b> Leveraging Machine Learning For Precise Anemia Recognitions
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 Model	Logistic regression is used for predicting the categorical dependent variable using a given set of independent variables. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.	None, as the accuracy of the model was 100% without hyperparameters	<div>Accuracy Score of Linear Regression Model: 1.0</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>1.00</td><td>1.00</td><td>1.00</td><td>167</td></tr><tr><td>1</td><td>1.00</td><td>1.00</td><td>1.00</td><td>118</td></tr><tr><td>accuracy</td><td></td><td></td><td>1.00</td><td>285</td></tr><tr><td>macro avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr><tr><td>weighted avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr></tbody></table>		precision	recall	f1-score	support	0	1.00	1.00	1.00	167	1	1.00	1.00	1.00	118	accuracy			1.00	285	macro avg	1.00	1.00	1.00	285	weighted avg	1.00	1.00	1.00	285
	precision	recall	f1-score	support																													
0	1.00	1.00	1.00	167																													
1	1.00	1.00	1.00	118																													
accuracy			1.00	285																													
macro avg	1.00	1.00	1.00	285																													
weighted avg	1.00	1.00	1.00	285																													

Random Forest Model	Random Forest Model's algorithm is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It improves the accuracy and controls overfitting by combining multiple trees.	None, as the accuracy of the model was 100% without hyperparameters	<div>Accuracy Score of Random Forest Model: 1.0</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>1.00</td><td>1.00</td><td>1.00</td><td>167</td></tr><tr><td>1</td><td>1.00</td><td>1.00</td><td>1.00</td><td>118</td></tr><tr><td>accuracy</td><td></td><td></td><td>1.00</td><td>285</td></tr><tr><td>macro avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr><tr><td>weighted avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr></tbody></table>		precision	recall	f1-score	support	0	1.00	1.00	1.00	167	1	1.00	1.00	1.00	118	accuracy			1.00	285	macro avg	1.00	1.00	1.00	285	weighted avg	1.00	1.00	1.00	285
	precision	recall	f1-score	support																													
0	1.00	1.00	1.00	167																													
1	1.00	1.00	1.00	118																													
accuracy			1.00	285																													
macro avg	1.00	1.00	1.00	285																													
weighted avg	1.00	1.00	1.00	285																													
Decision Tree Model	<p>Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.</p> <p>It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.</p>	None, as the accuracy of the model was 100% without hyperparameters	<div>Accuracy Score of Decision Tree Model: 1.0</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>1.00</td><td>1.00</td><td>1.00</td><td>167</td></tr><tr><td>1</td><td>1.00</td><td>1.00</td><td>1.00</td><td>118</td></tr><tr><td>accuracy</td><td></td><td></td><td>1.00</td><td>285</td></tr><tr><td>macro avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr><tr><td>weighted avg</td><td>1.00</td><td>1.00</td><td>1.00</td><td>285</td></tr></tbody></table>		precision	recall	f1-score	support	0	1.00	1.00	1.00	167	1	1.00	1.00	1.00	118	accuracy			1.00	285	macro avg	1.00	1.00	1.00	285	weighted avg	1.00	1.00	1.00	285
	precision	recall	f1-score	support																													
0	1.00	1.00	1.00	167																													
1	1.00	1.00	1.00	118																													
accuracy			1.00	285																													
macro avg	1.00	1.00	1.00	285																													
weighted avg	1.00	1.00	1.00	285																													
Naive Bayes Model	Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes’ theorem with the “naive” assumption of conditional independence between every pair of features. It is used for classification tasks.	<pre>param_grid = {     'var_smoothing': [1e-9, 1e-8, 1e-7, 1e-6, 1e-5] }</pre> <p>Accuracy Remained the same even after hyperparametric tuning</p>	<p>Before Hyperparametric tuning</p> <div>Accuracy Score of Naive Bayes Model: 0.9403508771929825</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.93</td><td>0.97</td><td>0.95</td><td>160</td></tr><tr><td>1</td><td>0.96</td><td>0.90</td><td>0.93</td><td>125</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.94</td><td>285</td></tr><tr><td>macro avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>285</td></tr><tr><td>weighted avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>285</td></tr></tbody></table> <p>After Hyperparametric tuning</p>		precision	recall	f1-score	support	0	0.93	0.97	0.95	160	1	0.96	0.90	0.93	125	accuracy			0.94	285	macro avg	0.94	0.94	0.94	285	weighted avg	0.94	0.94	0.94	285
	precision	recall	f1-score	support																													
0	0.93	0.97	0.95	160																													
1	0.96	0.90	0.93	125																													
accuracy			0.94	285																													
macro avg	0.94	0.94	0.94	285																													
weighted avg	0.94	0.94	0.94	285																													

			<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	<p>SVM is a powerful classification algorithm that works by finding the hyperplane that best separates the data into classes. It uses support vectors, the data points closest to the hyperplane, and aims to maximize the margin between the classes.</p>	<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>	<p>Before Hyperparametric tuning</p> <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> <p>After Hyperparametric tuning</p> <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	<p>An ensemble technique where multiple weak models (typically decision trees) are built sequentially, with each new model attempting to correct errors made by its predecessors. It's known for its effectiveness in predictive modeling and achieving high accuracy.</p>	<p>None, as the accuracy of the model was 100% without hyperparameters</p>	<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>