

## Model Optimization and Tuning Phase Template

Date	19 july 2024
Team ID	SWTID1720074204
Project Title	prediction and analysis of liver patient data using ml
Maximum Marks	10 Marks

### Model Optimization and Tuning Phase

This phase refines a machine learning model using liver patient data. Key steps include: selecting algorithms (e.g., logistic regression, random forest), tuning hyperparameters, engineering features, and evaluating performance using metrics like accuracy, precision, recall, and F1-score. The optimized model predicts disease risk, identifies high-risk patients, and uncovers disease patterns. Continuous monitoring and retraining ensure model accuracy and effectiveness..

### Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values
Random Forest	<ul style="list-style-type: none"> <li>• n_estimators: Number of trees in the forest</li> <li>• max_depth: Maximum depth of each tree</li> <li>• min_samples_split: Minimum number of samples required to split an internal node</li> <li>• min_samples_leaf: Minimum number of samples required to be at a leaf node</li> <li>• max_features: Number of features to consider when looking for the best split</li> </ul>	<ul style="list-style-type: none"> <li>• n_estimators: 100</li> <li>• max_depth: 8</li> <li>• min_samples_split: 2</li> <li>• min_samples_leaf: 1</li> <li>• max_features: 'sqrt'</li> </ul>
ADA Boost Classifier	<ul style="list-style-type: none"> <li>• n_estimators,</li> <li>learning_rate</li> </ul>	<p>n_estimators often improve performance, but watch for overfitting.</p> <p>Learning rate controls the contribution of each weak learner</p>

Gradient Boosting Classifier	n_estimators, learning_rate, max_depth, min_samples_split, min_samples_leaf	Similar to Random Forest, but learning rate controls the contribution of each tree.
Neural Networks	Number of layers, number of neurons per layer, activation functions, learning rate, optimizer, batch size	Require extensive experimentation and depend on network architecture.

- **Grid Search CV:** Exhaustively searches a specified parameter space.
- Randomized Search CV: **Randomly samples parameter combinations, often more efficient than Grid Search**

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

Model	Baseline Metric	Optimized Metric
Random Forest	0.8516949152542372	0.85
ADA boost classifier	0.7457627118644068	0.75
Gradient boosting classifier	0.8220338983050848	0.82
Randomized search cv	0.8347457627118644	0.84
Grid Search cv	0.8389830508474576	0.84

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

Final Model	Reasoning
Randomized search cv	We saw that after doing RandomizedSearchCV and GridSearchCV, Our accuracy, Precision, Recall, f1-Score doesn't increase