


## Model Optimization and Tuning Phase Report

Date	15 July 2024
Team ID	740683
Project Title	Doctors Annual Salary Prediction
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 (6 Marks):

Model	Tuned Hyperparameters
Linear Regression	 <pre> [20] g_train_pred(x)  array([[2791.8001807],        [2808.5407019],        [1677.4714724],        [2880.13481885],        [1267.38881743]])  [21] r2_score(y_train,y_train_pred)*100  67.4614207775856  [22] Mean square error for training data mean_squared_error(y_train,y_train_pred)  3485.2711381511 </pre>

Random Forest	<pre> [18] y_train_pred = rf.predict(x_train)       y_test_pred = rf.predict(x_test)  [19] r2_score(y_train, y_train_pred)*100       88.11481511617%  [20] mean_squared_error(y_train, y_train_pred)       47796.710777777775 </pre>
Decision Tree	<pre> [17] y_train_pred[0]       array([1016., 1076., 1016., 1006., 1700.])  [18] y_test_pred[0]       array([1700., 1106., 1506., 1006., 1506.])  [19] r2_score(y_train, y_train_pred)*100       100.0  [20] mean_squared_error(y_train, y_train_pred)       0.0 </pre>
XGBRegressor	<pre> [46] y_train_pred = xg_reg.predict(x_train)       y_test_pred = xg_reg.predict(x_test)  [47] r2_score(y_train, y_train_pred)*100       88.88888888888889  [48] mean_squared_error(y_train, y_train_pred)       4.888888888888889  [49] r2_score(y_test, y_test_pred)*100       88.17827011886014 </pre>

## Performance Metrics Comparison Report (2 Marks):

Linear Regression	<pre> [27] accuracy for with testing data (linear regression)       r2_score(y_test, y_test_pred)*100       27.10010776666667  [28] mean square error for testing data       mean_squared_error(y_test, y_test_pred)       371566.5012100288 </pre>

Random Forest	<pre> [31] r2_score(y_test, y_test_pred)*100 27.81281080767088  mean_squared_error(y_test, y_test_pred) 218408.13881127779 </pre>
Decision Tree	<pre> [41] r2_score(y_test, y_test_pred)*100 30.20411647088034  [42] mean_squared_error(y_test, y_test_pred) 350332.5 </pre>
XGBRegressor	<pre> from sklearn.metrics import mean_squared_error  mean_squared_error(y_test, y_test_pred) 218408.13881127779  [32] r2_score(y_test, y_test_pred)*100 30.170607011288079 </pre>

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

Final Model	Reasoning
Decision Tree	Decision trees can be a good starting point for predicting doctors' annual salaries due to their interpretability and ability to handle non-linear relationships. By carefully tuning hyperparameters and evaluating performance, you can build a robust model. For better generalization, consider using ensemble methods like random forests or gradient boosting if decision trees alone do not provide satisfactory results.