

## Model Optimization and Tuning Phase Template

Date	15 March 2024
Team ID	739849
Project Title	Doctor's Salary prediction
Maximum Marks	10 Marks

### Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining neural network 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 (8 Marks):

Model	Tuned Hyperparameters
Linear Regression	<p>The chosen hyperparameters (regularization strength and solver) were fine-tuned to improve the model's ability to generalize from the data.</p> <p>Regularization helps in preventing overfitting by penalizing large coefficients, while the solver 'liblinear' was selected for its efficiency with smaller datasets.</p> <ul style="list-style-type: none"> <li>• <b>Regularization Strength (alpha):</b> Set to 0.1 to prevent overfitting.</li> <li>• <b>Solver:</b> 'liblinear' chosen for its efficiency with small datasets.</li> <li>• The model uses Ridge Regression, which is a type of linear regression with L2 regularization.</li> <li>• <b>Evaluation Metrics:</b> Mean Squared Error (MSE) and R-squared (<math>R^2</math>) are used to assess the model's performance.</li> </ul>
Random Forest	Parameters like the number of trees, maximum depth, and minimum samples split were optimized to balance model complexity and

	<p>performance. Increasing the number of trees generally improves the model's stability and accuracy.</p> <ul style="list-style-type: none"> <li>• <b>Number of Trees (n_estimators):</b> Set to 100 to enhance the model's robustness.</li> <li>• <b>Max Depth:</b> Set to 10 to control the complexity of the trees.</li> <li>• <b>Min Samples Split:</b> Set to 2 to ensure that each split has at least two samples.</li> <li>• <b>Evaluation Metrics:</b> Mean Squared Error (MSE) and R-squared (<math>R^2</math>) are used to assess the model's performance.</li> </ul>
--	---

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

Final Model	Reasoning
Linear Regression	<p>The Linear Regression model was chosen due to its simplicity and interpretability. It provides a clear understanding of the relationship between features and the target variable, which is valuable for explaining the model's predictions. Despite being simpler than more complex models like Random Forest, it achieved reasonable performance metrics (MSE and R-squared), making it a practical choice for predicting doctors' salaries. Its ease of implementation and maintenance further supports its selection as the final model.</p>