Model Development Phase Template

Date	15 July 2024
Team ID	739761
Project Title	Doctors Annual Salary Prediction
Maximum Marks	5 Marks

Model Selection Report

In the model selection report for future deep learning and computer vision projects, various architectures, such as CNNs or RNNs, will be evaluated. Factors such as performance, complexity, and computational requirements will be considered to determine the most suitable model for the task at hand.

Model Selection Report:

Model	Description
Linear Regression	Linear regression is a fundamental supervised learning algorithm used for predicting a continuous target variable based on one or more input features. It models the relationship between the dependent variable (target) and independent variables (features) by fitting a linear equation to the observed data. **Training MSE**: 1405231186.3519106 **Validation MSE**:3715045452.16928 **R-squared**: 27.269167796800964
Random Forest Regression	An ensemble learning method that constructs multiple decision trees and merges them to get a more accurate and stable prediction. It reduces overfitting, handles large datasets with higher dimensionality, and

	improves accuracy by averaging the results of many decision trees. However, it requires more computational resources and memory. **Training MSE**: 458713655.5555556 **Validation MSE**: 3631440587.5 **R-squared**: 0.289059314547212
Decision Tree Regression	A non-linear model that splits the data into subsets based on the value of input features. It is easy to visualize and understand, handles both numerical and categorical data, and can capture non-linear relationships. However, decision trees can be prone to overfitting, especially with deep trees. **Training MSE**: 0.0 **Validation MSE**:3565125000.0 **R-squared**: 100.0
XGBoost	An optimized implementation of gradient boosting designed to be highly efficient, flexible, and portable. It provides parallel tree boosting to solve many data science problems quickly and accurately. XGBoost is known for its speed and performance and can handle large datasets efficiently. However, it requires careful parameter tuning to achieve optimal performance. **Training MSE**:0.000132921006944444445 **Validation MSE**:3566414555.666992 **R-squared**:99.999999999