Model Development Phase Template

Date	16 JULY 2024
Team ID	SWTID1720075199
Project Title	**Early Prediction of Chronic Kidney Disease Using Machine Learning**
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
Decision Tree	A decision tree is a visual model used for decision-making and prediction that resembles a tree structure. It starts with a root node that encompasses the entire dataset and branches out to internal nodes representing decision points based on specific attribute tests. These branches lead to other nodes or leaf nodes, which signify the final decision or classification outcome. Decision trees are commonly used for classification and regression tasks, offering an intuitive way to split data into subsets and make predictions based on input features. They are valued for their simplicity, interpretability, and ability to handle both numerical and categorical data.

Random Forest	A random forest is an ensemble learning method used for classification and regression tasks. It consists of multiple decision trees, known as "forest," which work together to improve the model's accuracy and robustness. Each tree in the forest is trained on a random subset of the dataset and a random subset of the features, which introduces diversity among the trees. During prediction, each tree in the forest makes a decision or prediction, and the random forest algorithm aggregates these individual predictions, often by majority voting in classification tasks or averaging in regression tasks. This approach reduces overfitting and increases the model's generalization ability, making random forests a powerful tool for handling complex datasets with high accuracy.
XGBoost	**XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm used for supervised learning tasks, including classification and regression. It is an implementation of gradient-boosted decision trees designed for speed and performance. XGBoost works by iteratively training a series of decision trees, where each new tree corrects the errors of the previous trees. The algorithm optimizes the model by minimizing a specified loss function, using techniques like gradient descent to update the model parameters. XGBoost includes several enhancements, such as regularization to prevent overfitting, handling of missing data, and parallel processing, which makes it both efficient and effective. It is widely used in competitive machine learning due to its accuracy, flexibility, and ability to handle large datasets with complex patterns.**