XGBoost Model Tuning Notebook ReadMe

Introduction

This notebook focuses on the tuning of an XGBoost regression model for predicting the duration of incarceration (incarcerated_days). The primary goal is to optimize the model's performance by systematically searching for the best hyperparameters using a grid search approach. The hyperparameters under consideration include max_depth, min_child_weight, and gamma.

Dependencies

Ensure that you have the following dependencies installed before running the notebook:

- pandas
- numpy
- matplotlib
- scikit-learn
- xgboost

You can install them using the following commands:

```
pip install pandas numpy matplotlib scikit-learn xgboost
```

Usage

1. Data Preparation:

- Ensure that your dataset (df) contains the necessary columns, with predictors and the target variable (incarcerated_days).
- Update the not_selected list to exclude any columns from consideration.

2. Model Tuning:

- The notebook employs a grid search approach to tune the XGBoost model. The hyperparameters being tuned are max_depth, min_child_weight, and gamma.
- The GridSearchCV function from scikit-learn is used for this purpose.

3. Model Evaluation:

- The modelfit function is defined to facilitate the training and evaluation of the XGBoost model.
- It performs k-fold cross-validation to find the optimal number of boosting rounds and then fits the model using the best parameters.
- The function prints the training and test root mean squared error (RMSE) and displays a bar chart of feature importances.

4. Grid Search Results:

- The notebook outputs the results of the grid search, displaying the best hyperparameters for max_depth, min_child_weight, and gamma.

5. Visualization:

- Feature importance scores are visualized using a bar chart to aid in understanding the impact of different predictors on the model.

How to Use

1. Execute the notebook in a Jupyter environment.
2. Analyze the grid search results to identify the optimal hyperparameters.
3. Use the optimal hyperparameters to configure your final XGBoost model.

Important Note

- Ensure that the input dataset is clean, appropriately formatted, and contains the required predictors and target variable.
- Customize the notebook as needed based on your dataset and analysis requirements.

Happy modeling!