Model Evaluation & Validation Project Rubric

**Overview**

This rubric is here to help you understand the expectations for how your project will be evaluated. It is the same rubric that the person evaluating your project will use. You should look at the rubric before you begin working on this project and before you submit it.

|  |  |
| --- | --- |
| **Criteria** | **Meets Specifications** |
| **Quality of Code** | |
| **Functionality** | Student’s code runs successfully and produces results similar to those in the PDF. No modifications are made to the template code beyond what is requested without justification. |
| **Statistical Analysis and Data Exploration** | |
| **Statistical Analysis** | All requested statistics for the Boston Housing dataset are accurately calculated. Student correctly leverages NumPy functionality to obtain these results. |
| **Optional:**  **Data Exploration** | Student adequately describes three separate features of the dataset. The corresponding values in the client’s feature set are correctly identified for the chosen features. |
| **Evaluating Model Performance** | |
| **Performance Metric** | An appropriate performance metric is chosen with thorough justification.  The metric is correctly implemented in code. |
| **Splitting the Data** | Student provides a valid reason for why a dataset is split into training and testing subsets for a model.  Training and testing split is correctly implemented in code. |
| **Grid Search** | Student correctly describes the grid search algorithm and briefly discusses its application.  GridSearchCV is properly implemented in code. |
| **Cross-validation** | Student either explains the importance of cross validation or correctly describes how cross-validation is performed on a model. Student identifies why cross-validation is helpful when using grid search.  Modifications beyond the default 3-fold cross-validation for GridSearchCV are reasonably justified. |
| **Analyzing Model Performance** | |
| **Learning Curves:**  **Training Analysis** | Student correctly identifies significant qualities of the training and testing errors as the training set size increases. |
| **Learning Curves:**  **Bias & Variance Analysis** | Student provides analysis for both a max depth of 1 and a max depth of 10. Reasonable justification is given for each graph if the model suffers from high bias or high variance. |
| **Error Curves and**  **Model Complexity** | Student identifies how the training and error curves relate to increasing the model’s complexity. |
| **Best-Guess**  **Optimal Model** | Student picks a best-guess optimal model with reasonable justification using the model complexity graph. |
| **Model Prediction** | |
| **Parameter-Tuned**  **Optimal Model** | Student determines the optimal model from parameter tuning. Student optionally compares this model to the one they chose. |
| **Predicting Selling Price** | Student’s model produces a valid result. The predicted selling price is adequately justified by the calculated descriptive statistics. |
| **Optional:**  **Generalization** | Student thoroughly discusses justification for or against using their model for predicting future selling prices. |