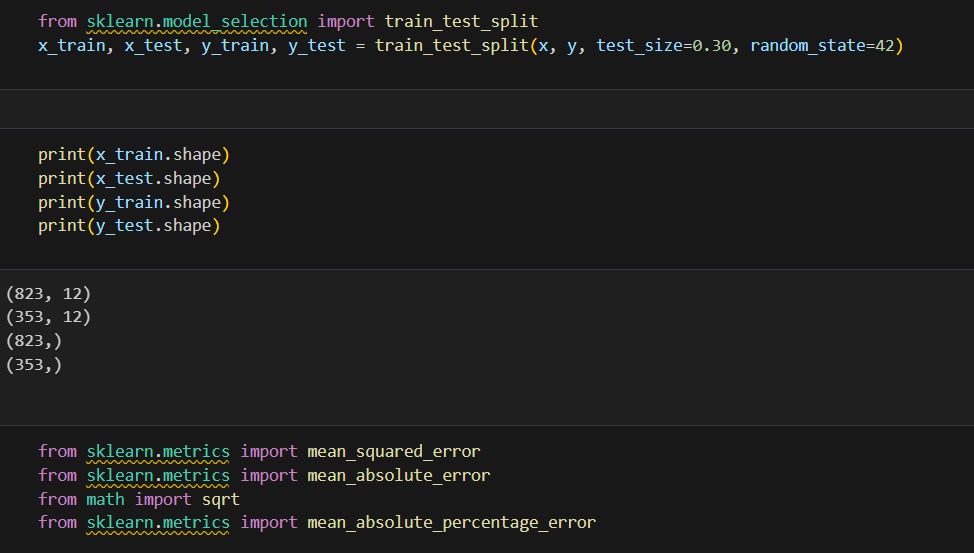
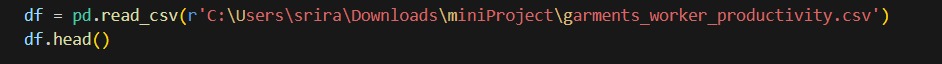
**Model Development Phase Template**

|  |  |
| --- | --- |
| Date | 6th July 2024 |
| Team ID | 739719 |
| Project Title | Garment Workers Productivity Predictions |
| Maximum Marks | 10 Marks |

**Initial Model Training Code, Model Validation and Evaluation Report**

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include a summary and training and validation performance metrics for multiple models, presented through respective screenshots.

**Initial Model Training Code (5 marks):**

|  |  |  |
| --- | --- | --- |
|  | **Summary** | **Training and Validation Performance Metrics** |
| Model 1  **Random Forest Regressor** | **Random Forest Regressor Summary**  **Model Parameters**   * **Number of Trees**: Optimal number of trees determined through hyperparameter tuning. * **Max Depth**: Maximum depth of the trees, optimized to prevent overfitting. * **Min Samples Split**: Minimum number of samples required to split an internal node. * **Min Samples Leaf**: Minimum number of samples required to be at a leaf node.   **Training Process:**   * **Data Preprocessing**: Standardized or normalized input features. * **Bootstrapping**: Random sampling with replacement to create multiple training sets for the trees. * **Feature Selection**: Random selection of features at each split to ensure diverse trees.   **Evaluation Metrics:**   * **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors in the predictions. * **Mean Squared Error (MSE)**: Measures the average of the squares of the errors, penalizing larger errors. * **R² Score**: Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. | FullSizeRender.jpg |
| Model 2  **Gradient Boosting Regressor** | **Gradient Boosting Regressor Summary**  **Model Parameters:**   * **Number of Estimators**: Total number of boosting stages (trees). * **Learning Rate**: Shrinks the contribution of each tree. * **Max Depth**: Maximum depth of the individual regression estimators (trees). * **Min Samples Split**: Minimum number of samples required to split an internal node. * **Min Samples Leaf**: Minimum number of samples required to be at a leaf node. * **Subsample**: Fraction of samples used for fitting the individual base learners.   **Training Process:**   * **Data Preprocessing**: Standardized or normalized input features. * **Initialization**: Starts with an initial prediction, often the mean of the target values. * **Sequential Training**: Each tree is trained on the residuals of the previous trees' predictions. * **Loss Function**: Mean Squared Error (MSE) to minimize the difference between predicted and actual values.   **Evaluation Metrics:**   * **Mean Absolute Error (MAE)**: Average magnitude of the errors in the predictions. * **Mean Squared Error (MSE)**: Average of the squares of the errors, penalizing larger errors. * **R² Score**: Proportion of the variance in the dependent variable that is predictable from the independent variables. | FullSizeRender.jpg |
|  |  |  |

**Model Validation and Evaluation Report (5 marks)**