AI CURE: Where AI meets health

Team name: Three Musketeers

Team members:

1. T S Akhilesh
2. Karthik Sriram V
3. Keerthisree Sai Narne

Problem statement:

Heart rate is a vital physiological parameter reflecting the frequency of cardiac contractions. Influenced by factors like age, fitness, and autonomic nervous system activity, heart rate serves as a key indicator of cardiovascular health. Monitoring heart rate during activities aids in optimizing exercise intensity and assessing overall well-being. The goal is to construct an advanced model capable of accurately predicting an individual's heart rate.

Abstract:

In the code, we have used a model called the Stacking Regressor Model. We initially import the dataset and perform data preprocessing. Then, we split the data into a train and test set and then train in our custom designed Stacking Regressor Model combined with RidgeCV Estimator. Then, we perform a model evaluation in which we have generated a Mean Absolute Error of 0.35, Mean Squared Error of 0.37 and R-squared value of 1.00.

Dataset used:

The dataset is being extracted from AI Cure: Where AI meets health competition. We have removed the columns uuid, datasetIs, condition as they are irrelevant features. The HR column consists of the Heart rate value to be predicted from the given features.

Proposed solution:

In our proposed solution, we have used a model called the Stacking Regressor Model. The Stacking Regressor used in the provided code is an ensemble learning method that combines the predictions of multiple base regressors to improve overall predictive performance. Stacking involves training multiple individual models and then using another model, called the meta-model or final estimator, to make predictions based on the predictions of the base models. This model is used to predict the Heart rate given the various parameters in the dataset.

Let's break down the components of the Stacking Regressor in this code:

1. Base Estimators:

```
ridge_cv_estimator = RidgeCV()
extra_trees_estimator = ExtraTreeRegressor(
    max_features=0.9500000000000001,
    min_samples_split=10,
    random_state=42
)
```

- **RidgeCV**: This is a linear regression model with built-in cross-validation to find the optimal regularization parameter (alpha). It's suitable for linear relationships in the data.

- **ExtraTreeRegressor**: This is an extremely randomized tree regressor. It builds multiple decision trees with random splits, which helps in capturing complex relationships in the data. The parameters (max_features, min_samples_split) are set to control the tree-building process.

2. Stacking Regressor:

```
stacking_model = StackingRegressor(
    estimators=[
        ('ridgecv', ridge_cv_estimator),
        ('extratrees', extra_trees_estimator)
    ],
    final_estimator=ExtraTreesRegressor(
        max_features=0.9500000000000001,
        min_samples_split=10,
        random_state=42
    )
)
```

- **Estimators**: This parameter takes a list of tuples, where each tuple contains a string identifier and a base estimator. In this case, it includes the RidgeCV model and the ExtraTreeRegressor model.

- **Final_estimator**: This is the meta-model that combines the predictions of the base models. In this code, it's another ExtraTreesRegressor with similar parameters as the individual ExtraTreeRegressor. This meta-model is trained on the predictions made by the base models.

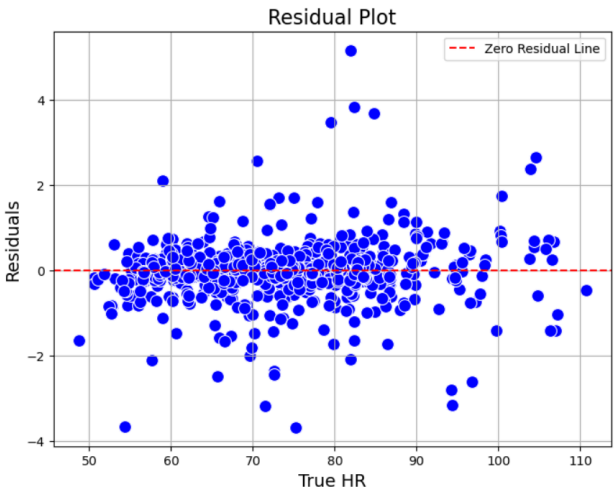
Results obtained:

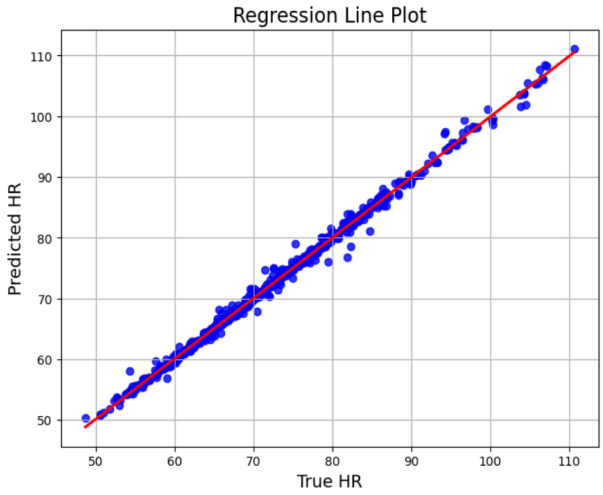
The following results were obtained from the code:

S.No.	Metric Name	Value
1.	Mean Squared Error	0.37
2.	Mean Absolute Error	0.35
3.	R-squared	1.00

Visualizations:

We have plotted a Regression Line Plot and a Residual plot to visualize the results obtained after training the model





Conclusion:

Our proposed Stacking Regressor model has demonstrated excellent performance on the given dataset, yielding impressive results with a Mean Absolute Error (MAE) of 0.35, Mean Squared Error (MSE) of 0.37, and a perfect R-squared value of 1.00. Achieving such high and accurate metrics suggests that the model is well-trained and effective in predicting the target variable.