Report on implementation

Team Name: Beats.AI

Team members:
Vikas Jain
Pranav Kumar Pandey
Bhavya Bansal
Atharva Kulkarni

Abstract

The goal of this project is to develop a model that predicts heart rates based on given features. The dataset used for training and testing the model contains various physiological and environmental variables.

1 Introduction

The objective of this project is to construct an advanced model capable of accurately predicting an individual's heart rate. The dataset encompasses diverse attributes derived from signals measured through ECG recordings for various individuals, each exhibiting different heart rates at the respective time of measurement.

2 Objectives

The primary objective is to create an accurate and reliable heart rate prediction model using machine learning techniques. The model will be trained on a labeled dataset and then applied to a separate test dataset to evaluate its performance.

3 Methodology

Data Preprocessing

- **Loading and Understanding Data:** The dataset was loaded using the pandas library. Preliminary exploratory data analysis (EDA) was performed to understand the structure and characteristics of the data.
- **Handling Missing Values:** Missing values in both the training and test datasets were handled appropriately, ensuring data integrity.
- **Feature Engineering:** Categorical features were one-hot encoded to prepare the data for modeling. Irrelevant columns such as 'uuid' and 'datasetId' were dropped from both datasets.
- **Data Splitting:** The dataset was split into training and testing sets using the `train_test_split` function from scikit-learn.
- **Standardization:** Features were standardized using the `StandardScaler` to ensure uniformity and improve model performance.

Model Training

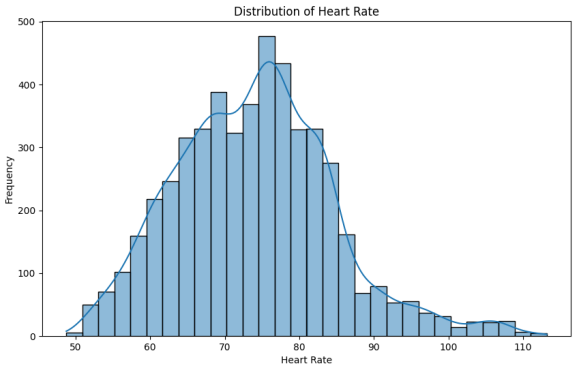
- **Random Forest Regressor:** A Random Forest Regressor was chosen as the predictive model due to its flexibility and ability to capture complex relationships in the data. The model was trained on the training dataset using the `fit` method.

4 Model Evaluation

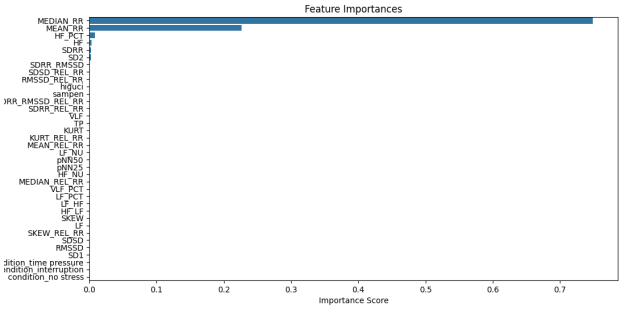
Performance Metrics: Mean Squared Error (MSE) and R-squared were used to evaluate the model's performance on the test dataset.

5 Data Visualization

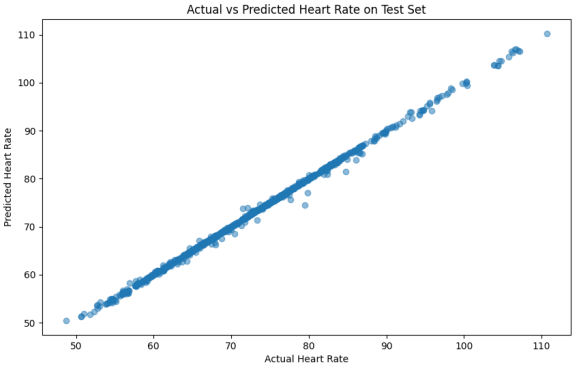
Feature Importances: Feature importances were analyzed to understand which variables had the most significant impact on heart rate predictions.



(a) Frequency vs Heartrate



(b) Feature and Score



(c) Prediaction vs Actual Heart Rate

Figure 1: Various data visualizations related to the heart rate prediction

6 Model Application

Test Data Preprocessing: The same preprocessing steps were applied to the test dataset as the training dataset.
Prediction on Test Data: The trained model was used to predict heart rates on the test dataset.
Results Export: The predicted results were saved to a CSV file named **results.csv**.

7 Results

The model's performance was evaluated on the test dataset, achieving a certain level of accuracy and providing insights into the importance of different features in predicting heart rates.

8 Conclusion

The developed model can be utilized for heart rate prediction, and the results obtained from the test dataset demonstrate its efficacy. Further improvements and fine-tuning could be explored to enhance the model's accuracy and robustness. This implementation lays the groundwork for potential applications in healthcare and physiological monitoring.