AI Cure: Where AI Meets Healing Touch
REPORT

Team Name : Jelly Nuts

Team Member Details :

1. SANJEEV SITARAMAN : LEADER
2. ROUNAK PAUL : MEMBER
3. SOHAM ROY CHOUDHURY : MEMBER

Problem Statement : Heart Rate Prediction using Advanced Models.

Abstract : This project addresses the challenge of accurately predicting an individual's heart rate using the LightGBM model. Leveraging a diverse dataset derived from ECG recordings, our research focuses on informative and useful attributes to determine heart rates at specific moments.

Experiments done to reach the solution : Our experiments involved a systematic exploration of various machine learning models for heart rate prediction using ECG data. The following is a summary of our findings:

- *Linear Regression :*
    - Accuracy: Approximately 90%
    - Observation: Linear regression, while a simple model, demonstrated a reasonable level of accuracy for heart rate prediction.
- *Random Forest Algorithm :*
    - R2 Score: Approximately 94%
    - Observation: The Random Forest algorithm provided improved accuracy compared to linear regression, showcasing its ability to capture non-linear patterns in the data.
- *Neural Network :*
    - Accuracy: Around 96%
    - Observation: A neural network with 7 layers exhibited further enhancement in accuracy, leveraging its capacity to learn intricate patterns in complex datasets.
- *Polynomial XGBoost :*
    - Accuracy: Approximately 98.7%
    - Observation: The XGBoost model, enhanced with polynomial features, demonstrated high accuracy, suggesting its capability to capture nuanced relationships within the data.

- *Feature selection using Correlation Matrix :*
    - Selected Features: 18 features identified through correlation matrix analysis.
    - R2 Score: Substantially improved to approximately 99.76% when used on the existing XGBoost model.
    - Observation: By carefully selecting features based on their correlation with the target variable, we achieved a remarkable boost in prediction accuracy.
- *Feature Confirmation from Medical Professionals :* Features were discussed with Medical Industry professionals and their insights were taken into consideration while finalizing the final 18 features. This collaborative effort ensures that our model aligns with clinical expertise, enhancing its applicability in the medical domain.
- *LightGBM with Feature Selection :*
    - R2 Score: Exceptionally high at 0.9993486547412208
    - Mean Squared Error: 0.06813171786183257
    - Observation: Implementing LightGBM over selected features and using parameters obtained through experimental grid search not only confirmed the effectiveness of feature selection but also resulted in an extraordinary R2 score, emphasizing the robustness and efficiency of the chosen model.

The progressively increasing accuracies highlight the synergies between model complexity, features, and algorithm selection. The outstanding performance of LightGBM, in particular, showcases its prowess in handling the complexity of ECG data for heart rate prediction.

Our Proposed Model :

1. *Data Preprocessing :*
    - The columns namely - datasetId, uuid, condition were dropped.
    - The given training data was normalized.
    - The data was split into training and validation sets.
2. *Evaluation Metrics :*
    - Mean Squared Error [this model] : 0.06813171786183257
    - R-squared value [this model] : 0.9993486547412208
3. *Model Selection :*
    - After rigorous experiments with different solutions, we finalized LightGBM along with selected parameters as the optimal model for the selected features.
4. *Hyperparameters used :*

- Boosting_type : gbdt [Gradient Boosting Decision Trees]
- Objective : regression
- Metric : mse
- Max_depth : 16
- Num_leaves : 18
- Learning_rate : 0.08
- Verbose : 1
- Num_boost_round : 1200, [500 can also be used if cost needs to be reduced which results in a r2 score of 0.9992489344762081]

Future Scopes :

1. *Dynamic model adaptation :* We can explore methods for adapting this very model to continuously changing physiological conditions, as heart rate can vary continuously.
2. *Real - time monitoring :* We plan to develop a Real-time monitoring system that uses a stream of data that flows in as input instead of the static data that we currently have. This may tend to be more useful for medical purposes.
3. *Personalized models :* We plan to provide personalized models for every user/client by considering variations and characteristics of their own data for improving accuracy.
4. *Feature importance :* We can conduct further deep analysis to gain insights on the importance of more such features and try to add those as the selected feature to increase accuracy.

Results :

R-squared (R2) score : 0.9993486547412208

Mean squared error (MSE) : 0.06813171786183257

Example Result Plots :