Prediction of Health Condition using Human Activity Recognition

(DA203o)

Abraham Jacob M (abrahamjacob@iisc.ac.in)
Alankar Adarsh (alankara@iisc.ac.in)
Issaac Kommineni (issaack@iisc.ac.in)
Nitin Anand (nitina@iisc.ac.in)

Introduction

- What is HAR ?
  - Human activity recognition (HAR) is to identify specific movement or action of a person based on sensor data

  - HAR information is used in various fields like health, sports etc.
  - Identifying mere activity does not provide much information
- How about predicting health index from HAR?

Problem Statement

- Prior Work
  - Classification of activity based on sensor data
  - Limitations - data sets available were not having the requirements for predicting the health condition

- Our Work
  - Predict health condition of a person from his activities

- Problem statement (HAR and Health Prediction)
  - Use the HAR information to predict the activity using classification models
  - Use duration of activity predicted from classification model to predict health index using regression models

HAR Dataset

- 30 volunteers of age group of 19 to 48
- 561 feature vector with time and frequency domain variables
- 6 target activities
- Captured using accelerometer and gyroscope sensors
- Data captured included 3-axial (X, Y, Z directions) linear acceleration and 3-axial angular velocity at a constant rate of 50Hz (sampling at 20ms)
- Rotation, Jerk, Gravitation, Acceleration etc. of the body
- Estimated attributes like mean, standard deviation, entropy etc.
- Combination of signals and attributes constitute the features of the HAR Data-set
- Limitations
  - Captures 2 minutes of sensor data of each person while performing each activity
  - Can be used for classification

Synthetic Dataset

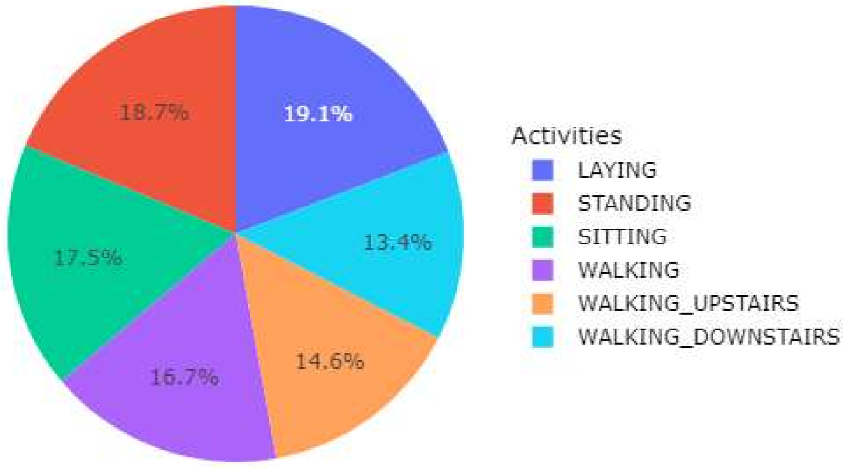
- Extrapolated to 24 hours of activities
- Bias [30] and Weights [0.02, 0.15, 0.08, 0.2, 0.3, 0.25]
- Domain Knowledge

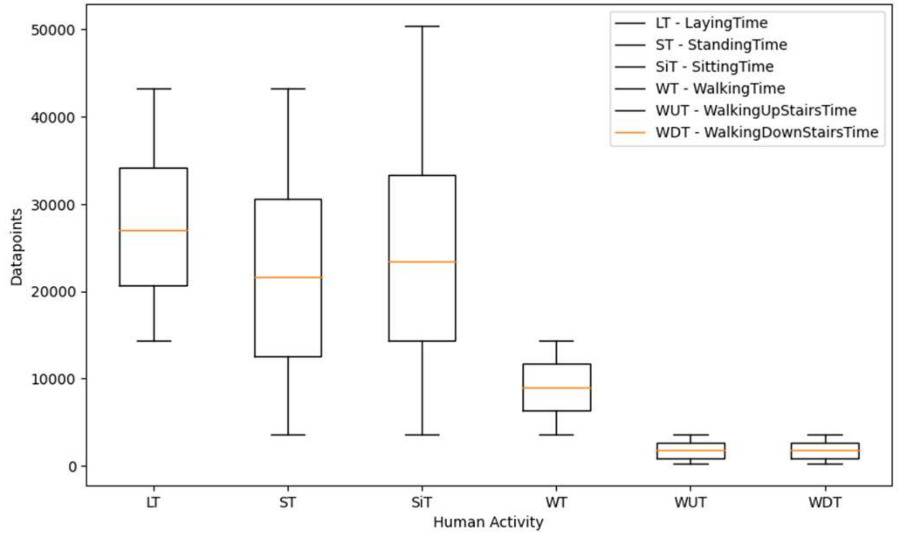
Activity	Minimum Duration	Maximum Duration	Step
Laying	4 hours	12 hours	30 mins
Standing	1 hour	12 hours	30 mins
Sitting	1 hour	14 hours	30 mins
Walking	1 hour	4 hours	15 mins
Walking Upstairs	5 mins	1 hour	5 mins
Walking Downstairs	5 mins	1 hour	5 mins

$$HI = B + W1 * LayingTime + W2 * StandingTime + W3 * SittingTime + W4 * WalkingTime + W5 * WalkingUpStairsTime + W6 * WalkingDownStairsTime$$

Data Munging

- Removing special characters from column names
- Data distribution
- Dimensionality Reduction (PCA)
  - 561 -> 157 components
- Data cleaning / Removing outliers
- Addition of noise to Synthetic data



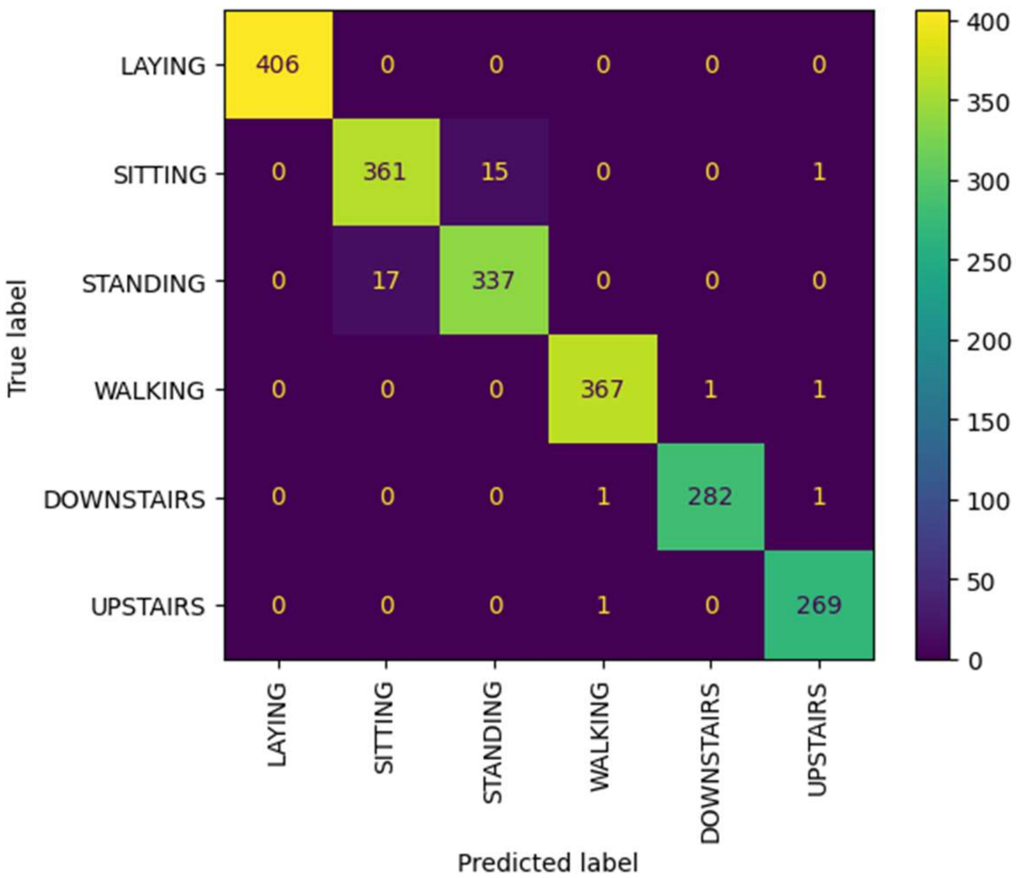


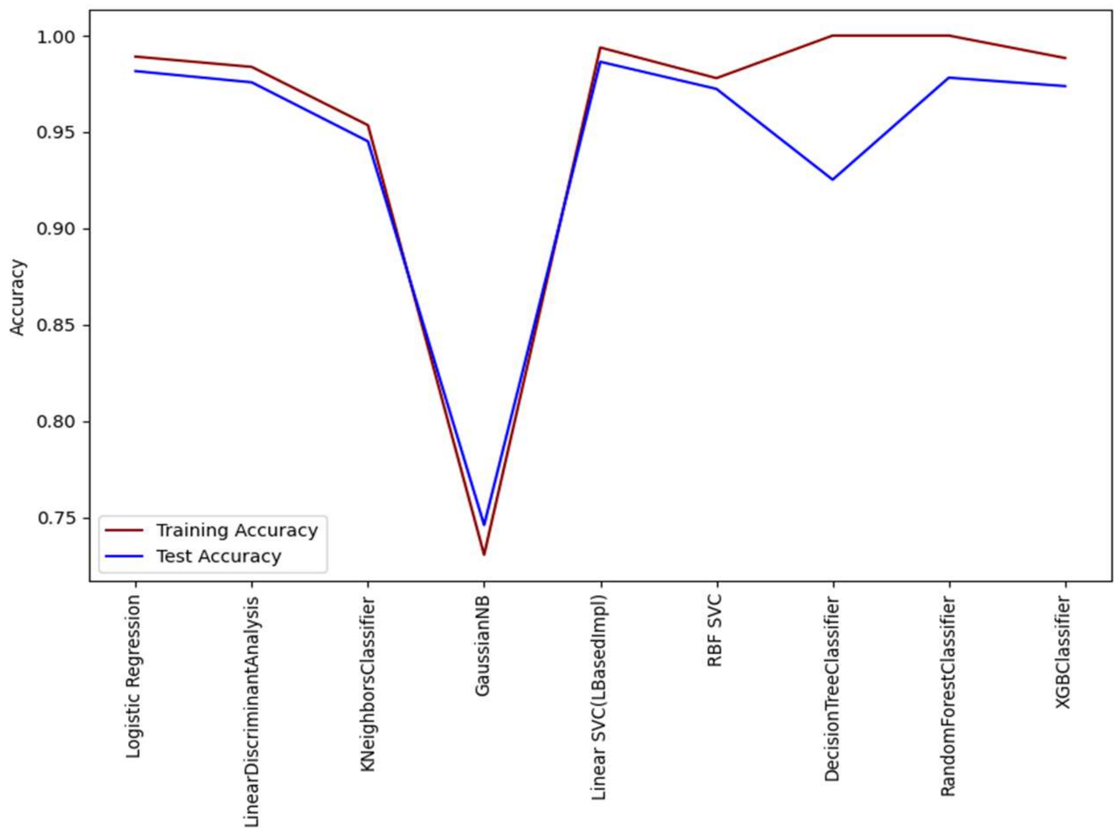
Classification

- Below models were choses as they are
  - Good for high-dimensional data and
  - Capable of handling linear as well as non-linear relation between feature and target

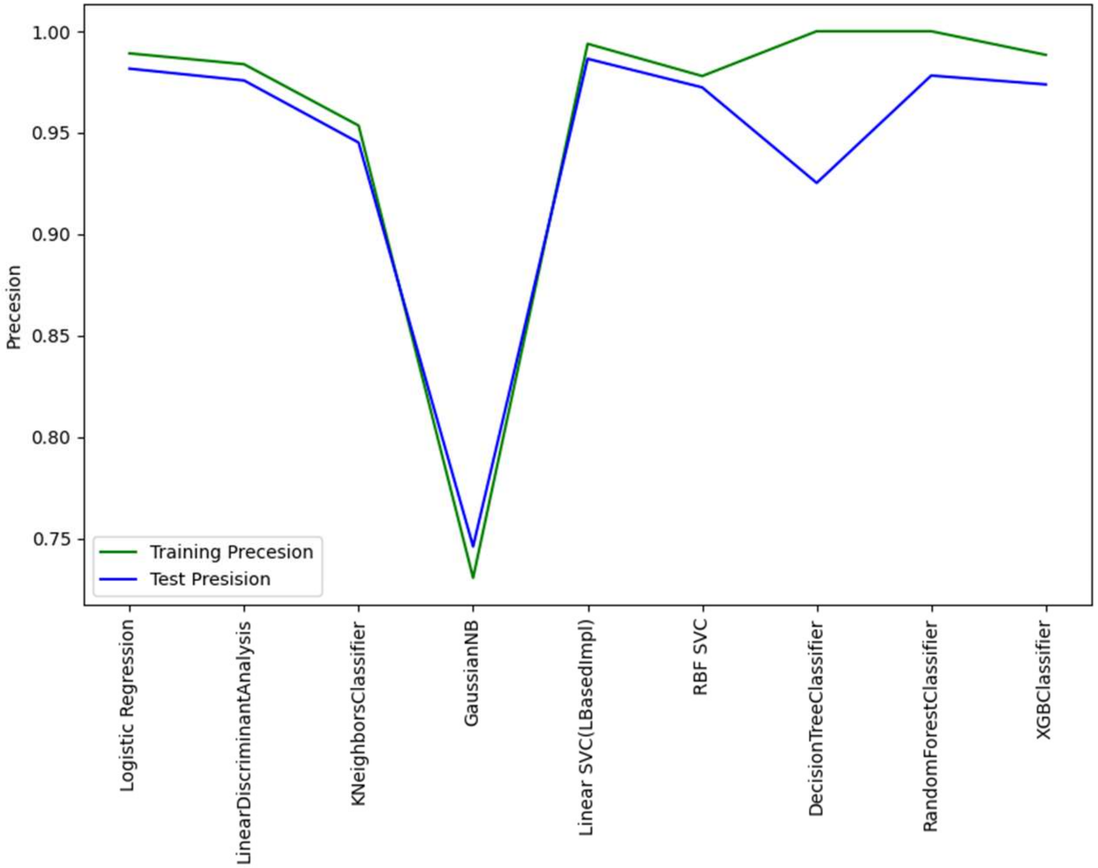
Model	Description
Logistic Regression	Classification by estimating the probability of the target variable
Linear Discriminant Analysis (LDA)	Probabilistic Classification based on Bayes' Rule [by modeling the probability distribution of the features for each class]
K-Nearest Neighbors (KNN)	Classification by finding the common class between K closest neighbors to the input in the feature space
Gaussian Naive Bayes (GaussianNB)	Classification by modeling the features for each class using Gaussian distributions and then using Bayes' rule to calculate probability of each class
Support Vector Machine (SVM)	Classification by finding the hyperplane that maximally separates the classes in the feature space
Decision Tree	Classification by recursively splitting the feature space into smaller regions based on the most informative features until each region contains only one class
Random Forest	Ensemble of multiple decision tree
XGBoost	Uses gradient descent to improve accuracy with each decision tree addition

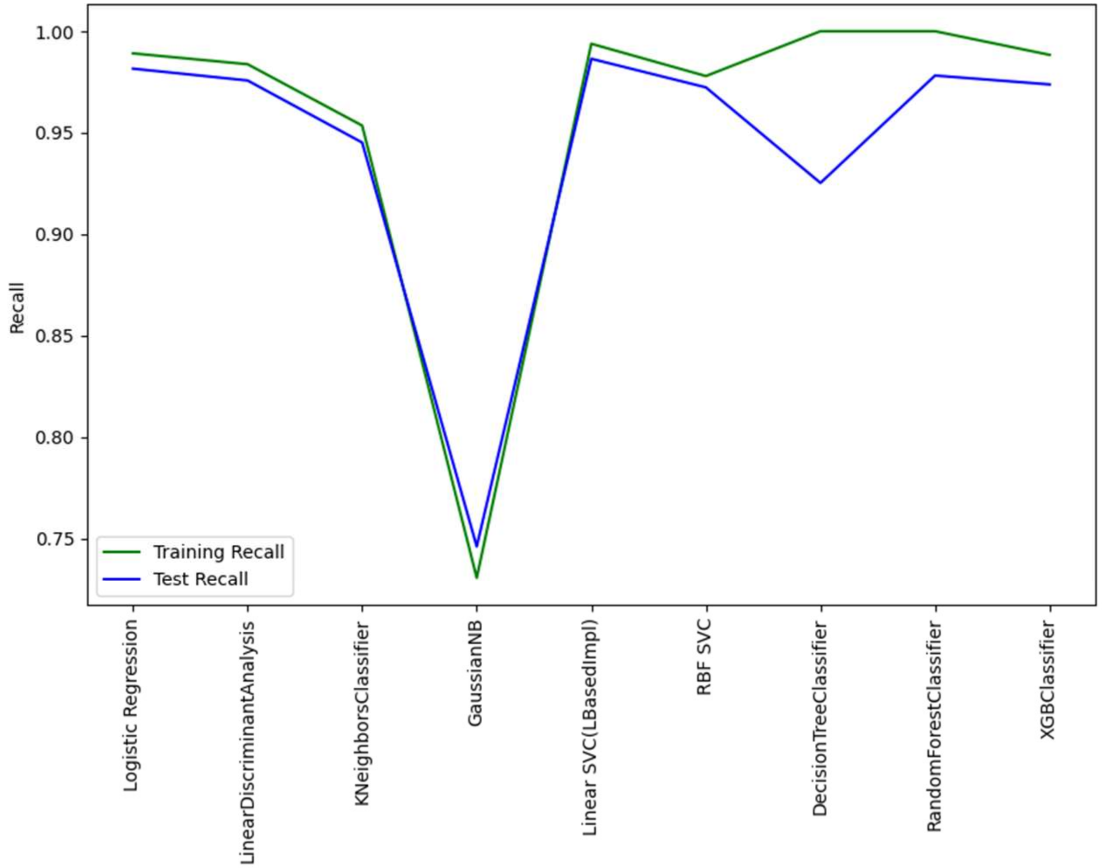
Classification – Confusion Matrix, Accuracy





Classification – Precision, Recall





Regression

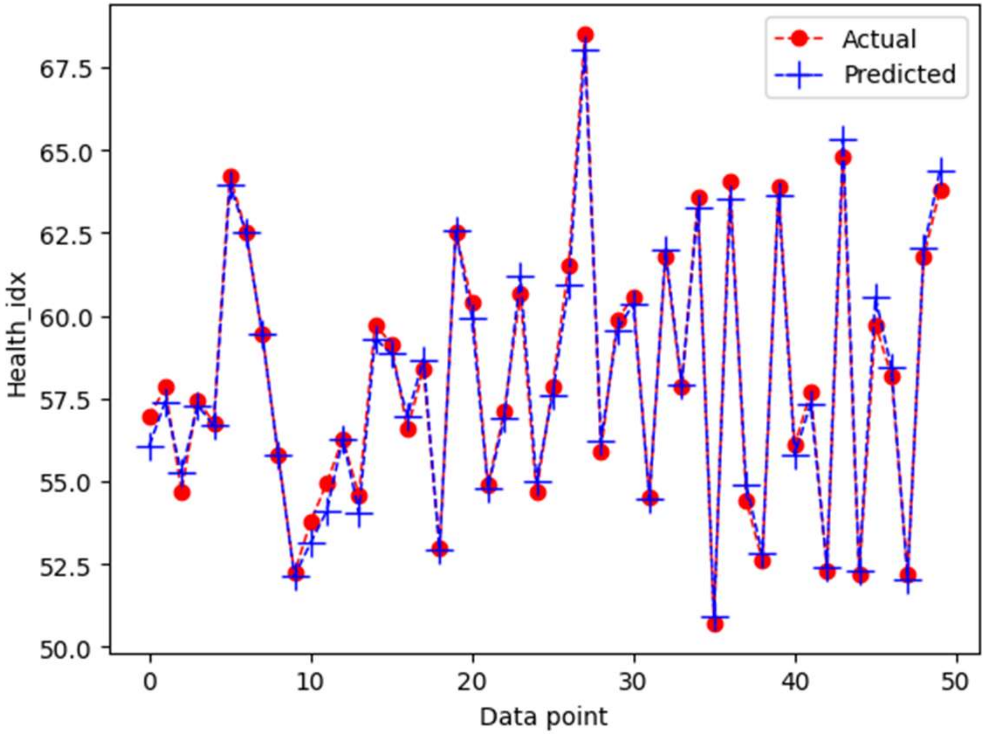
Model	Description
Linear Regression	Establishes liner relationship between dependent & independent variables
Bayesian Ridge	Adds regularization to prevent overfitting by assigning prior probability to coefficients
Ridge	Incorporates L2 regularization to prevent overfitting by adding a penalty term to the sum of squared coefficients
Lasso	Incorporates L1 regularization to prevent overfitting by adding a penalty term to the absolute sum of coefficients
ElasticNet	Combines L1 and L2 regularization to prevent overfitting by adding both penalty terms to the sum of squared coefficients
ElesticNetCV	A model with built-in cross-validation for hyperparameter tuning
Huber Regressor	A model that is robust to outliers by minimizing the sum of absolute errors for small errors and the sum of squared errors for larger errors

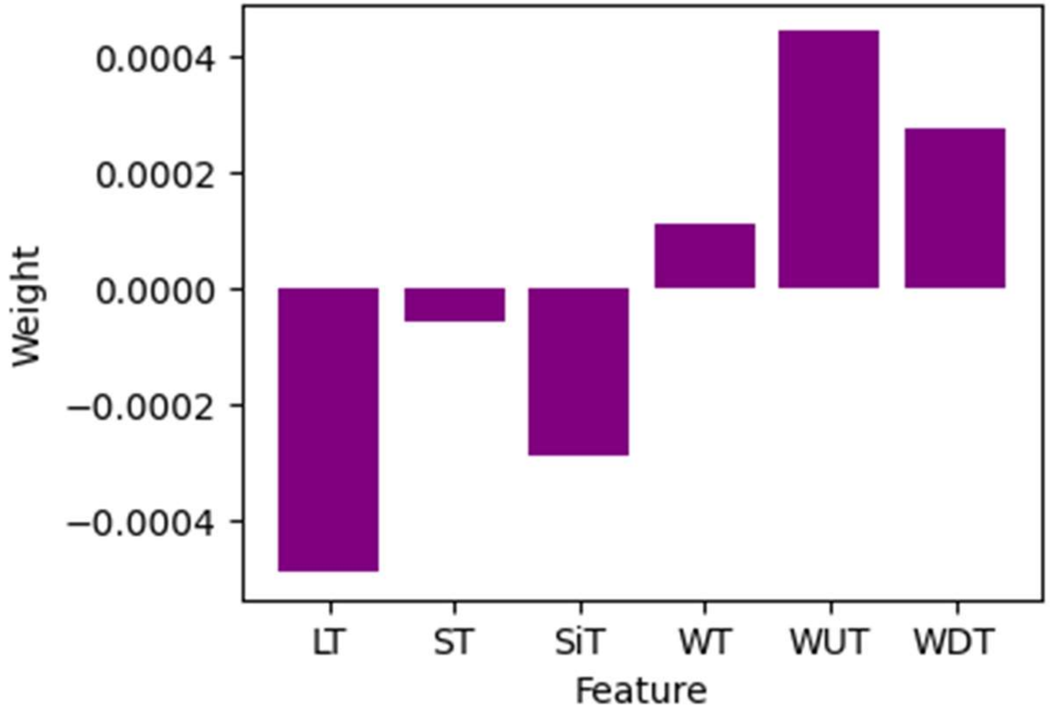
Regression

- Linear relationship between the duration of each activity and HI
- Various regression algorithms
- Model Coefficients

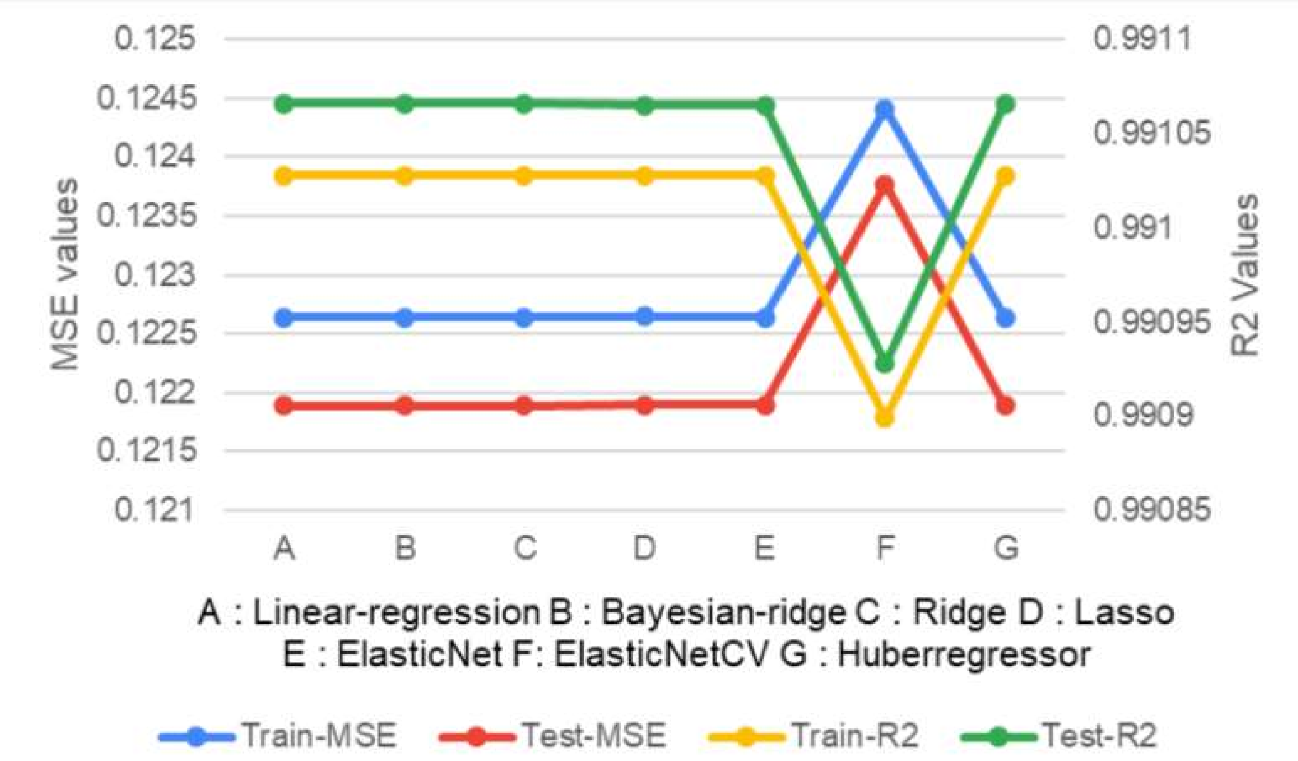
Model	Bias	W1	W2	W3	W4	W5	W6
Linear Regression	78.000	-4.889e-04	-5.556e-05	-2.888e-04	1.109e-04	4.449e-04	2.775e-04
Bayesian Ridge	78.000	-4.889e-04	-5.556e-05	-2.888e-04	1.109e-04	4.449e-04	2.775e-04
Ridge	77.627	-4.846e-04	-5.124e-05	-2.845e-04	1.152e-04	4.492e-04	2.818e-04
Lasso	59.747	-2.776e-04	1.557e-04	-7.751e-05	3.221e-04	6.553e-04	4.879e-04
ElasticNet	59.238	-2.717e-04	1.616e-04	-7.164e-05	3.280e-04	6.616e-04	4.942e-04
ElasticNetCV	73.329	-4.330e-04	0.0	-2.330e-04	1.64e-04	4.73e-04	3.05e-04
Huber Regressor	6.269e-08	4.14e-04	8.47e-04	6.14e-04	1.01e-03	1.35e-03	1.18e-3

Feature Importance





MSE, R2


0.125
0.1245
0.124
0.1235
0.123
0.1225
0.122
0.1215
0.121
MSE values
0.9911
0.99105
0.991
0.99095
0.9909
0.99085
R2 Values
A
B
C
D
E
F
G
A : Linear-regression B : Bayesian-ridge C : Ridge D : Lasso
E : ElasticNet F: ElasticNetCV G : Huberregressor
Train-MSE
Test-MSE
Train-R2
Test-R2

Challenges & Future Extension

- Challenges
  - o High dimensional dataset
  - o Mismatch in HAR dataset from different sources
  - o Round the clock sensor data
  - o Dependency of Health Index on other factors
    - ✓ Smoking, Eating Habits etc.
    - ✓ Age, Ethnicity etc.
- Future Extension
  - o Other sensor information (Pressure monitor, ECG sensors)
  - o Habits (drinking, smoking, food)
  - o Age, gender, ethnicity etc.

Inspirational Links

- https://www.kaggle.com/datasets/arashnic/har-1
- https://www.neuraldesigner.com/solutions/activity-recognition
- https://www.mdpi.com/1424-8220/10/2/1154/htm
- https://www.kaggle.com/code/fahadmehfoooz/human-activity-recognition-with-neural-networks/input