Prediction of Heart Disease
based on Clinical Factors

By Team 6

Masimukku, Satya Aditya

Nalamothu, Prathyusha

Nandala, Sumana Sree

Krishnagiri Tuppal, Venu Gopalan

Introduction

- Aim of Paper
  - Predict the possibility of a person getting Heart Disease based on Clinical Factors
- Scope
  - Understanding of Occurrence of Cardiac Arrest as per Clinical Factors
- Purpose
  - Heart Attack – increased rate
  - Habits and way of living

Data Collection and Processing

Data Collection

- Based on Paper – *Prediction of Clinical Risk Factors of Diabetes using Multiple Machine Learning Techniques Resolving Class Imbalance*
- Data Source – Kaggle | Cardiac Data NHanes 65c8df6e-2
- 51 Attributes with 37079 Observations
- NHANES – Survey between 1999 to 2015
- Sample of 5000 people/year

Data Processing

- No missing values
- Using 18 out of 51 columns
- Changing Type 3 Diabetes to Type 2
- Finding correlation among variables

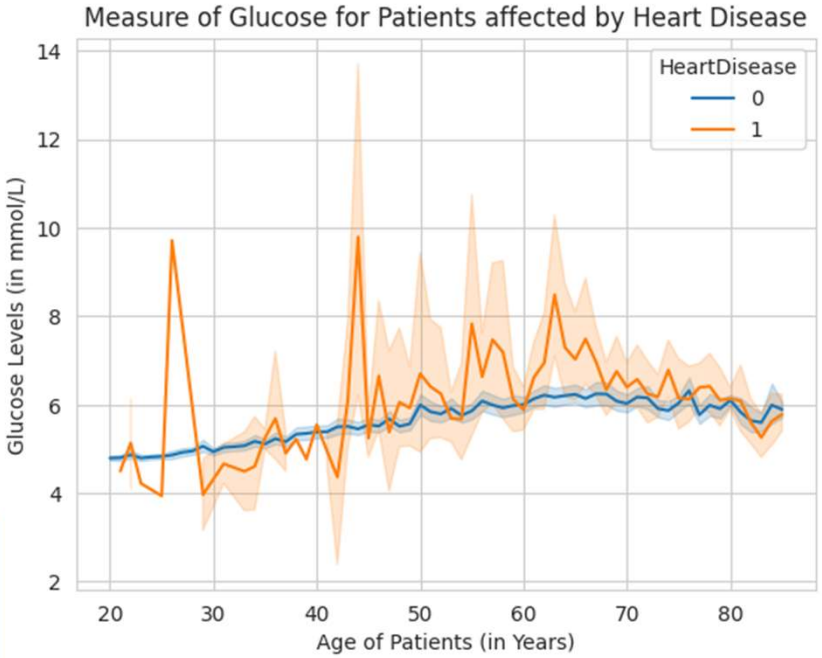
EDA

Descriptive Statistics

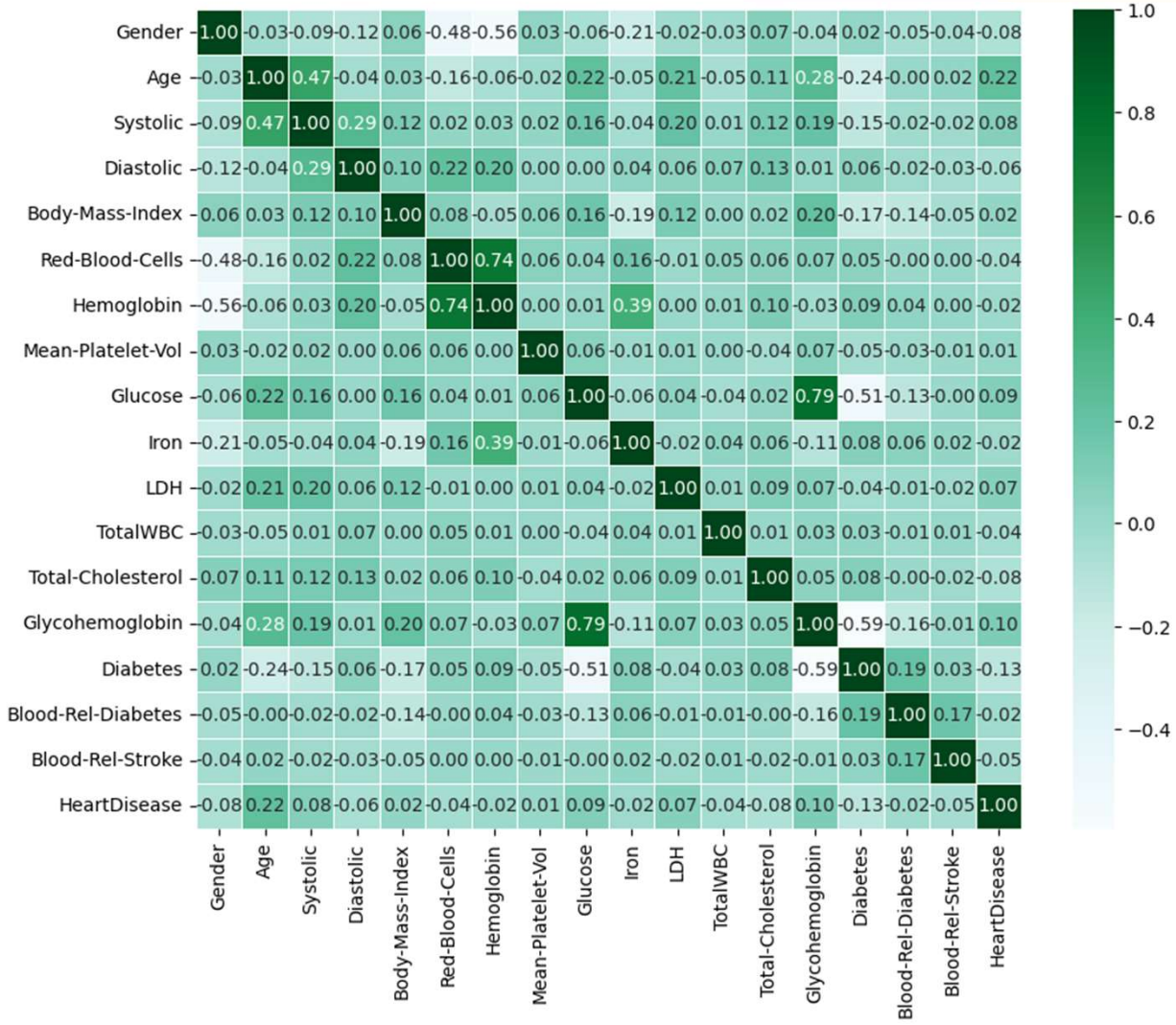
- Gender Ratio – 19032:18047
- Age, 20-85 Years, Avg = 48
- Diabetic Type
  - Type-1: 4144
  - Type-2: 32227
  - Type-3: 708
- BMI – Min: 13.18, Max: 130.21, Mean: 28.23
- Cholesterol - Min: 1.53, Max: 14.09, Mean: 5.08 mmol/L
- Glucose - Min: 1.05, Max: 34.25, Mean: 5.6 mmol/L
- Patients affected by HD - 1508

EDA

Visualization



Correlation Matrix

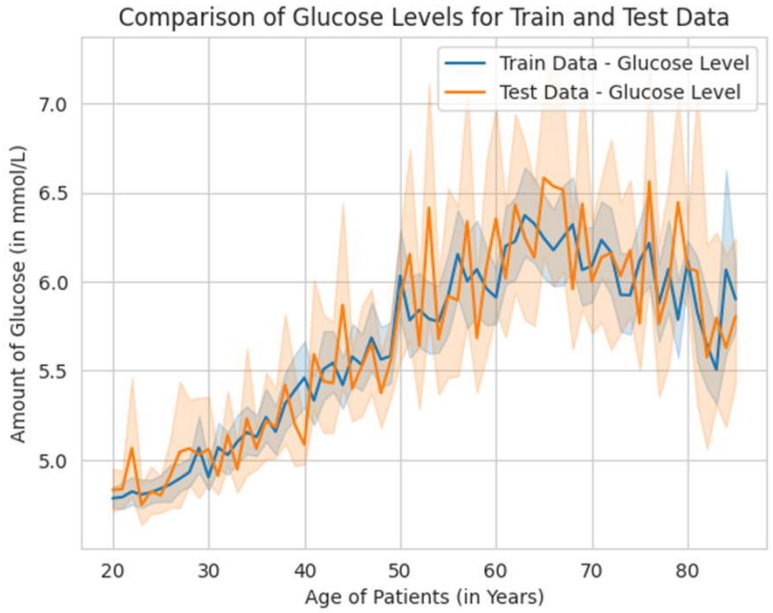


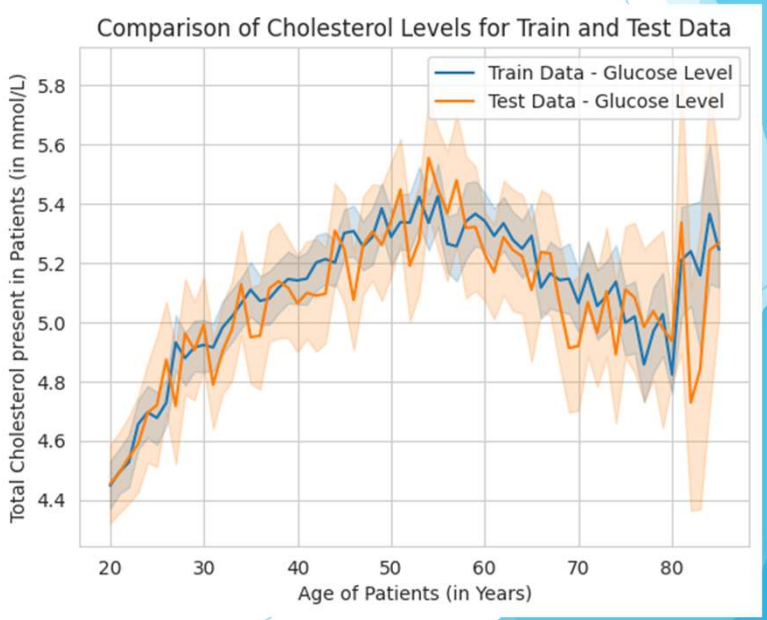
Train and Test Data

- Data is divided into Train and Test data which has following distribution :

```
Train Data Heart Disease Ratio (No : Yes) 28438 : 1225 (0.043)
Test Data Heart Disease Ratio (No : Yes) 7133 : 283 (0.039)
```

- Following graphs show the Distribution of key factors b/w Train and Test Data:





Models and Results

- Data focused on Decision Making
- Models used – Decision Tree Classifier and Random Forest Classifier Algorithms
- No additional parameters passed

- **Decision Tree** achieved – Accuracy 92.5 %

- **Random Forest** achieved – Accuracy 96.2%

Conclusion

- Both Decision Tree and Random Forest provided importance to different parameters as follows:

```
Decision Tree Parameter Importance:
Feature: Total-Cholesterol, Importance: 0.11523590410069175
Feature: Age, Importance: 0.09142817816103668
Feature: Iron, Importance: 0.0863804798454249
Feature: Red-Blood-Cells, Importance: 0.07744491394742013
Feature: Systolic, Importance: 0.07452589703883009
Feature: Glucose, Importance: 0.0724679116538098
Feature: TotalWBC, Importance: 0.07093333668947999
Feature: LDH, Importance: 0.06845846993872529
Feature: Mean-Platelet-Vol, Importance: 0.0673845049850021
Feature: Hemoglobin, Importance: 0.0638699997944746
Feature: Body-Mass-Index, Importance: 0.06364674660011782
Feature: Glycohemoglobin, Importance: 0.05773414934252947
Feature: Diastolic, Importance: 0.05091639656586105
Feature: Gender, Importance: 0.013958477938982139
Feature: Blood-Rel-Stroke, Importance: 0.011443931228267265
Feature: Blood-Rel-Diabetes, Importance: 0.007816916161736045
Feature: Diabetes, Importance: 0.00635378600761085
```

```
Random Forest Parameter Importance:
Feature: Age, Importance: 0.10140663995827869
Feature: Total-Cholesterol, Importance: 0.09706907262853222
Feature: TotalWBC, Importance: 0.07782755931105394
Feature: Body-Mass-Index, Importance: 0.07547053926846037
Feature: LDH, Importance: 0.07319635715471186
Feature: Glucose, Importance: 0.07136210654801821
Feature: Iron, Importance: 0.07100310491640409
Feature: Systolic, Importance: 0.0708207383205424
Feature: Red-Blood-Cells, Importance: 0.06967488549097972
Feature: Hemoglobin, Importance: 0.06430073257719449
Feature: Mean-Platelet-Vol, Importance: 0.06376280730691168
Feature: Glycohemoglobin, Importance: 0.05839882356536948
Feature: Diastolic, Importance: 0.05744023658056408
Feature: Blood-Rel-Stroke, Importance: 0.013962062300892294
Feature: Gender, Importance: 0.0126643479965543
Feature: Blood-Rel-Diabetes, Importance: 0.011485629998966893
Feature: Diabetes, Importance: 0.01015435607656546
```

- Main Parameters that affect : *Age*, *Total-Cholesterol* and *Glucose*
- *Diabetes* is given least importance for both the Algorithms

Future Improvements

- Use more features and focus on features that directly affect Heart
- Implement SVM and other Decision-Making Algorithms
- Use parameters while building algorithms to improve accuracy