



HEART DISEASE PREDICTION USING MACHINE LEARNING

GROUP 2 PRESENTATION



TEAM MEMBERS

1. Panchol Alier
2. Wanjiku Githu
3. Elizabeth Josephine
4. Faustine Njiriri
5. Pruno Jumah



Introduction

- One of the prominent diseases that affect many people during middle or old age is heart disease, and in many cases it eventually leads to fatal complications.
- One-third of all global deaths are due to heart diseases. Half of the deaths in the United States and in other developed countries are due to heart ailment



Research Question

The diagnostic of heart disease is a challenging task for doctors , given clinical parameters about the patient, can we predict whether the patient has a heart disease or not?



Specific Research Objectives

- We want to identify the relation between the heart disease and the different features or factors
- We will implement various algorithm model classifiers to try to predict if the patient has a heart disease or not.
- What factors contribute the most to a person having a heart disease within the given factors



Data sourcing and cleaning

Data Source = University of California Irvine (UCI)

- Renaming the columns
- Zero duplicates and a few missing values
- Dealing with the missing values, duplicates and anomalies
- Data transformation to numeric to allow further analysis



Data Analysis

- Plotting of univariate and bivariate summaries of the data
- Normalization of the data
- Modelling and classification
- Building and training of neural networks
- Improving of the results using binary classification

Data Visualization and Analysis

Heart Disease Dashboard

Gender

| | |
|--------|--------|
| Female | 17.39% |
| Male | 82.61% |

Filter by Gender

(All)

Filter by Resting ECG Results

(All)

Filter by Target

Less than 50% narrowing

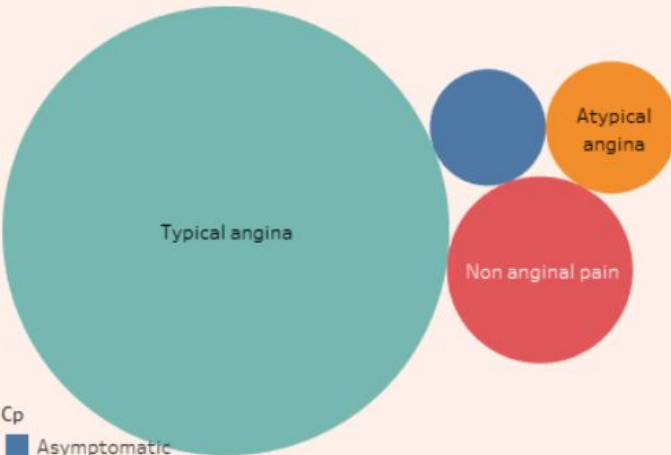
Filter by Thallium Stress Results

(All)

Filter by Exercise Induced Angina

(All)

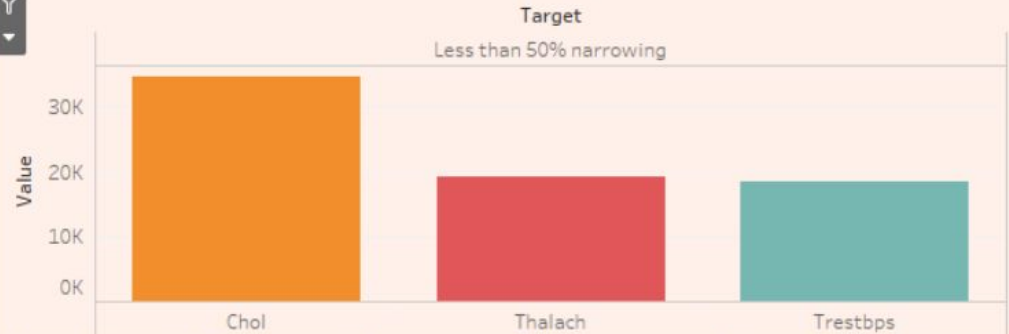
Types of Chest Pain



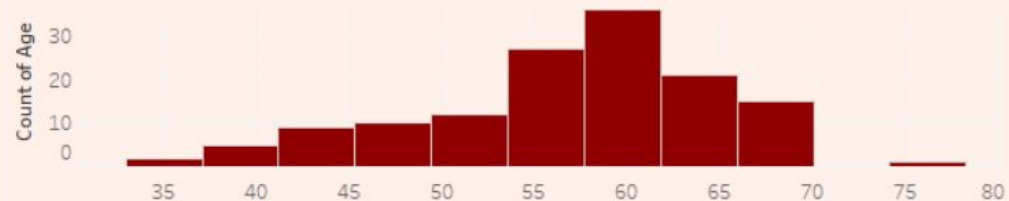
Cp

- Asymptomatic
- Atypical angina
- Non anginal pain

Heart Disease vs. Variables



Distribution of the Age Groups



Data Visualization and analysis

Heart Disease Dashboard

Gender

Female

31.68%

Filter by Gender

Male

68.32%

(All)

Filter by Resting ECG Results

(All)

Filter by Target

(All)

Filter by Thallium Stress Results

(All)

Filter by Exercise Induced Angina

(All)

Types of Chest Pain

Typical angina

Atypical angina

Non anginal pain

Cp

- Asymptomatic
- Atypical angina
- Non anginal pain
- Typical angina

Heart Disease vs. Variables

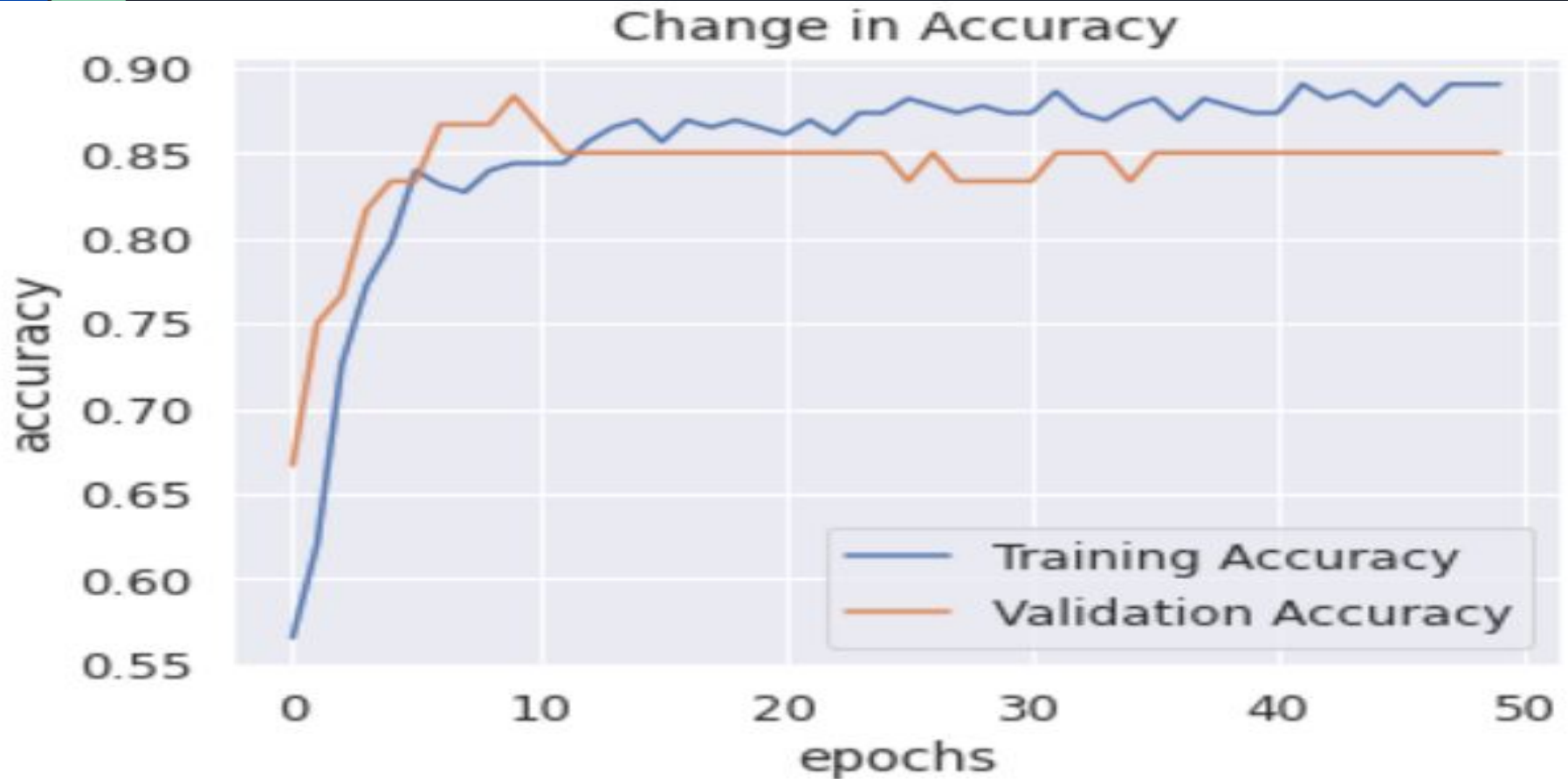


Distribution of the Age Groups

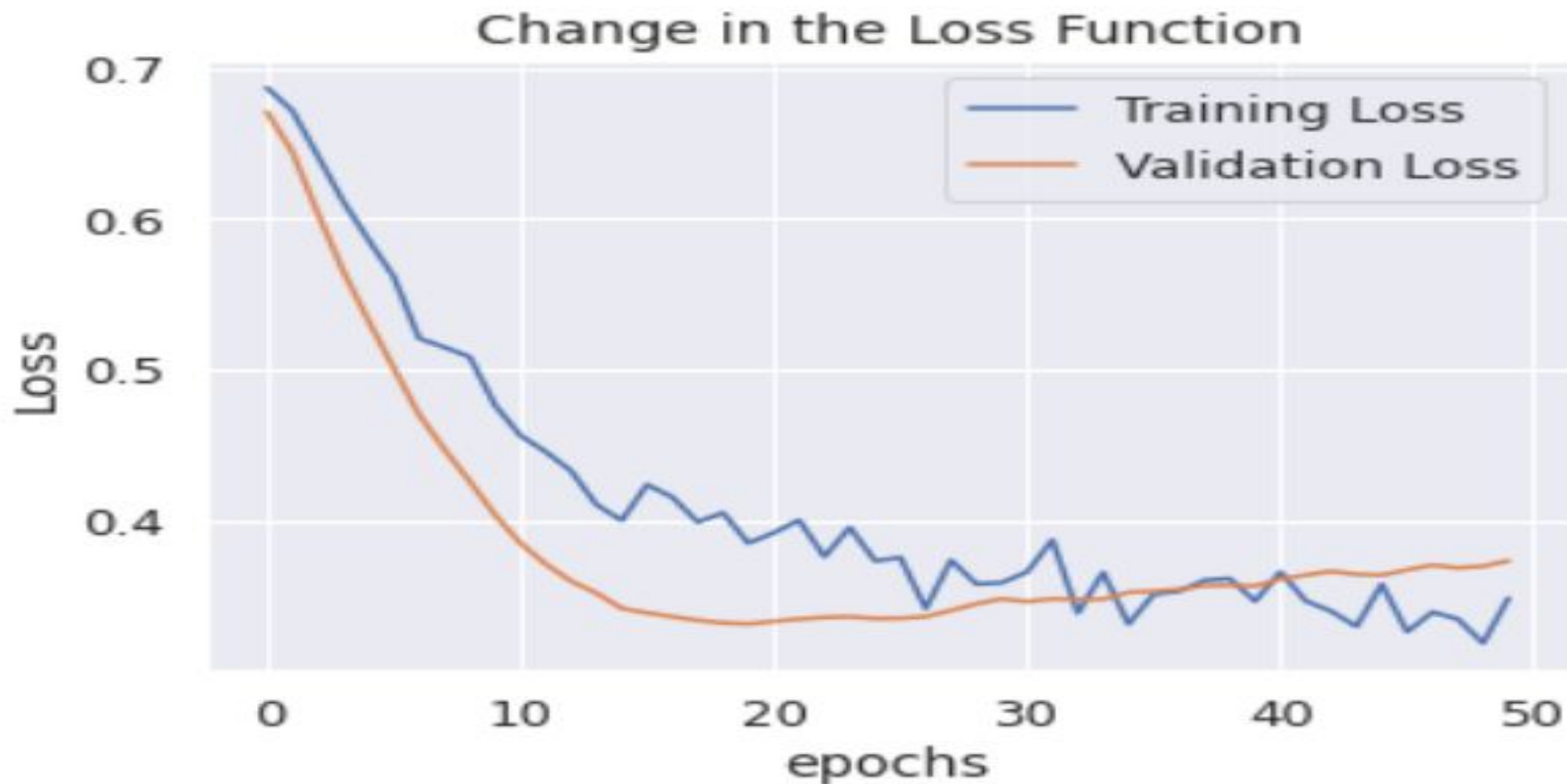



Activate Windows

Data Visualization and Analysis



Visualization cont



- 
- From the above results, we notice that the accuracy on the training set has consistently improved with ascending epoch. However, the accuracy on the validation set was quite unstable. This implies that our baseline model is great on the training data but could suffer from instability on completely new data. As such, the baseline model should be further optimized using Convolutional Neural Networks (CNN) to improve val_accuracy stability and overall accuracy score.



Conclusion

- 7 models were implemented in the analysis of this dataset
- The performance of each model recorded and accuracies compared

KNN = 0.5

RF = 0.57

Ad = 0.53

GB = 0.55

SVM = 0.5

Neural Network = 0.83



Recommendations

- More analysis of this dataset is needed to make a sound conclusion, because the 14 features in the dataset are not the only factors to predict heart disease
- But if the doctor were to implement the model we will advise them to use a neural network model because of its high accuracy of prediction for patients with heart disease.



THANK YOU