Classification of Heart Disease Within Individuals

MATH2319 Machine Learning

Abhishekh Shankar s3652116 & Eric Bolitho s3455890

Contents

1 Introduction

The objective of this project is to predict whether or not a patient has heart disease indicated as an integer value from 0 to 1 for the absense and presence of heart disease respectively. The dataset was sourced from the UCI Machine Learning Repository[1] (Aha n.d.).

Phase One saw the data being cleaned, with absent values being removed if the total percentage of absent values was under a predefined threshold of 60%, or the entire column being removed if it was over 60%. Phase Two covers model building and analysis with Section Two covering the methodology taken in this phase, Section Three covering ...

2 Methodology

```
#[^1]: https://www.kaggle.com/mirichoi0218/insurance
```

3 Hyperparameter Fine-Tuning

3.1 Naive Bayes

3.2 Random Forest

3.3 K-Nearest Neighbours

We read in the data set using the following code.

```
# Read in the csv file.
#insurance <- read.csv("insurance.csv")
```

4 Evaluation

Initially we fit a full regression model using the code snippet below.

5 Discussion

6 Conclusion

[1] https://archive.ics.uci.edu/ml/datasets/Heart+Disease

References

Aha, David W. n.d. "UCI Machine Learning Repository: Heart Disease Data Set." V.A. Medical Center, Long Beach; Cleveland Clinic Foundation. Accessed April 24, 2019. https://archive.ics.uci.edu/ml/datasets/adult.