HARRISBURG UNIVERSITY

ANLY 502 – ANALYTICAL METHODS

PRADEEP PALADUGULA   
Kshitij Deshpande  
Abisheik Mani

## Heart Disease Predictions among Patients

Heart Disease is one of the leading cause of deaths in the world still analysis and diagnosis of Heart Disease among adults both female and male is uncertain due to the difficulty in obtaining clinical data form patients. Such data is needed to accurately predict the Patient health based on the parameters and state whether the patient needs immediate attention and is at High Risk of a Heart Disease. Recent advancement in lab tests some of the parameters are easy to be captured and others are assumed in this case study. We undertook data related to both Male and Female patients from different CARDIALOGY centers.

•. What data do you need?  
 This database contains 76 attributes, but all published experiment refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by  
It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value)

• What work do you plan to do in the project?

Identify Correlation among the predictors.  
Exploratory Data Analysis   
Impute Missing Values  
Remove Outliers  
Perform LOGISTIC Regression and Other Modelling on the test and train data split.

•. Which algorithms/techniques/models do you plan to use/develop? Be as specific as you can.  
Penalized LOGISTIC Regression (LR) with Cross Validation  
Random Forest (RF)

• How will you evaluate what you’ve done?  
Model Performance will be evaluated on the basis of the Prediction of the Heart Disease condition which is unhealthy for the Test Data. One of the key metrics is ROC Curve and accuracy to predict the unhealthy condition of the patient.

## Attributes Description: (# tag indicates variable reference from the list of 76 variables)

1. #3 (age): Age in years

2. #4 (sex): male (1) and Female (0)

3. #9 (cp): Chest Pain type

1: typical angina

2: atypical angina

3: non-anginal pain

4: asymptomatic

4. #10 (trestbps): Resting blood pressure

5. #12 (chol): cholesterol in mg per dl

6. #16 (fbs): fasting blood sugar > 120mg/dl (1: true, 0: false)

7. #19 (restecg): resting ECG

1: value 0: normal

2: value 1: Wave abnormality

3: value 2: showing probable left ventricular hypertrophy

8. #32 (thalach): max heart rate achieved

9. #38 (exang): exercise induced anigina

10. #40 (oldpeak): depression induced

11. #41 (slope): slop of peak exercise

12. #44 (ca): number of major vessels colored by floro scopy

13. #51 (thal): normal fixed reversible defect

14. #58 (num): (the predicted attribute)

Dataset Download  
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

### References:

### Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64,304--310.  [[Web Link]](http://rexa.info/paper/b884ce2f4aff7ed95ce7bfa7adabaef46b88c60c)  David W. Aha & Dennis Kibler. "Instance-based prediction of heart-disease presence with the Cleveland database."  [[Web Link]](http://rexa.info/paper/0519d1408b992b21964af4bfe97675987c0caefc)  Gennari, J.H., Langley, P, & Fisher, D. (1989). Models of incremental concept formation. Artificial Intelligence, 40, 11--61.  [[Web Link]](http://rexa.info/paper/faecfadbd4a49f6705e0d3904d6770171b05041f)