Heart Failure Outline

1. Import Libraries
2. Import Dataset and Inspect Data
   1. Read file and print a sample of 10 entries
   2. Print the shape of the data
   3. Investigate the types of data
   4. Describe all columns
   5. Boolean categorical variables
   6. Check for missing values
3. Exploratory Data Analysis
   1. Investigate the distributions of feature values
   2. Distribution of death event by age
   3. Distribution of age by gender
   4. Investigating correlation
      1. Seaborn PairPlot
   5. Exploration of covariance
      1. Covariance Matrix
   6. Investigating variable distributions
      1. Distribution of death event by age and gender
      2. Distribution of diabetes by death event and age
      3. Distribution of smoking by death event and age
      4. Distribution of serum creatine and death event by gender
      5. Distribution of high blood pressure by death event and age
      6. Distribution of serum sodium by death event and gender
      7. Distribution of anaemia by death event and age
      8. Distribution of creatinine phosphokinase by death event and gender
      9. Distribution of ejection fraction by death event and gender
4. Exploration of ejection fraction
   1. What’s normal
   2. Correlation between serum sodium levels and ejection fraction
   3. Correlation between serum sodium levels and death event
   4. Correlation between ejection fraction and death event
   5. Correlation between serum creatinine and death event
   6. Correlation between ejection fraction and smoking
   7. Correlation between ejection fraction and platelets
   8. Correlation between ejection fraction and creatinine phosphokinase
   9. Correlation between ejection fraction and serum creatinine
   10. Correlation between serum creatinine and age
5. Linear regression model between all variables and death event
   1. Predicting ejection fraction based on serum sodium
      1. Subset data
      2. Create train test split
      3. Verify data shapes
      4. Instantiate linear regression model and fit to data
      5. Use model to form predictions
      6. Calculate R Squared and other metrics
      7. Create data frame of actual and predicted values
      8. Sample of results data frame
      9. Plot actual versus predicted ejection fraction
      10. Cross validation
6. Modeling
   1. Model 1: Decision tree entropy model
      1. Subset data
      2. Create train tests split
      3. Instantiate decision tree entropy model
      4. Use model to make predictions
      5. Visualize decision tree
      6. Model evaluation metrics
   2. Model 2: Decision tree gini model
      1. Instantiate decision tree entropy model
      2. Use model to make predictions
      3. Visualize decision tree
      4. Model evaluation metrics
      5. Conclusion
   3. Model 3: Entropy model with less features
      1. Evaluating the features based on covariance
      2. Subset data
      3. Create train tests split
      4. Instantiate decision tree entropy model
      5. Use model to make predictions
      6. Visualize decision tree
      7. Model evaluation metrics
      8. Plot confusion matrix
      9. Calculate and plot ROC
   4. Model 4: Random forest with 4 prime features