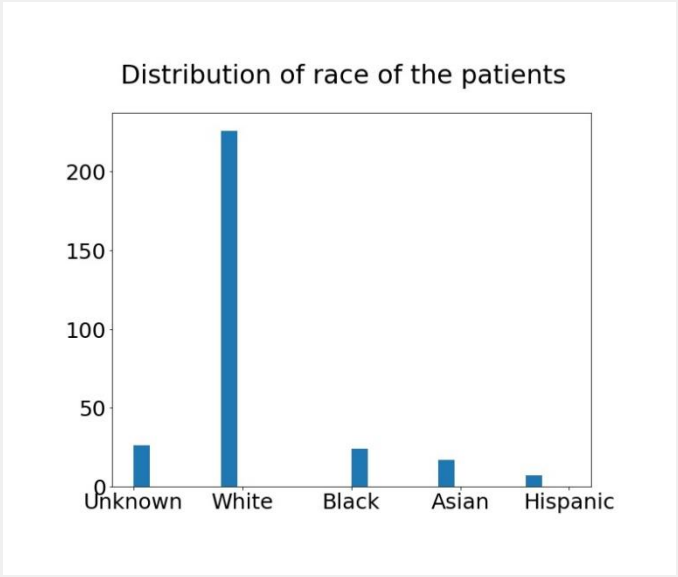
Analysis and Predictive Models: Chronic Kidney Disease (CKD)

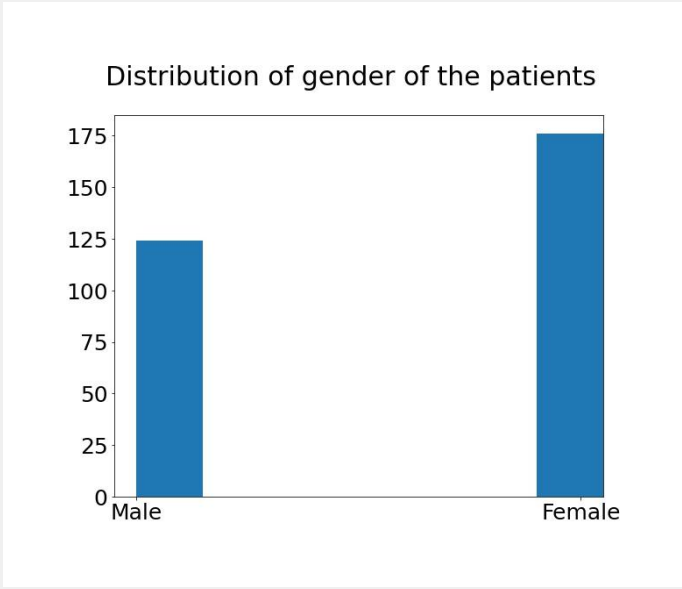
Holmusk Interview Challenge
Jetin E Thomas

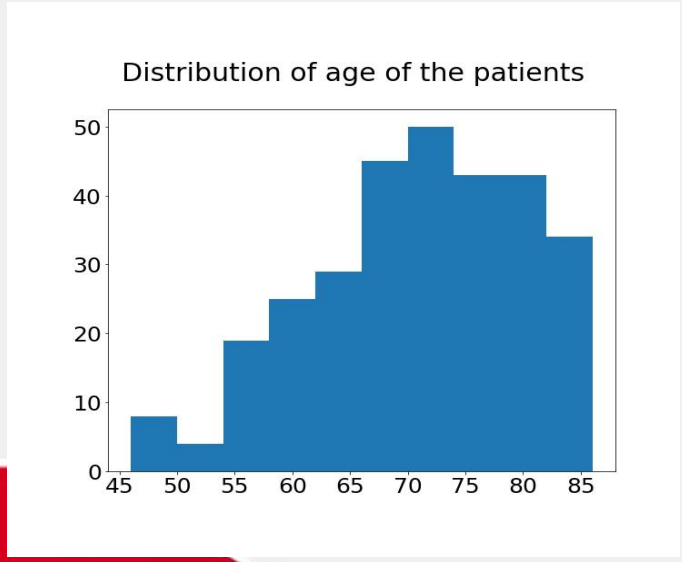
Description of Datasets

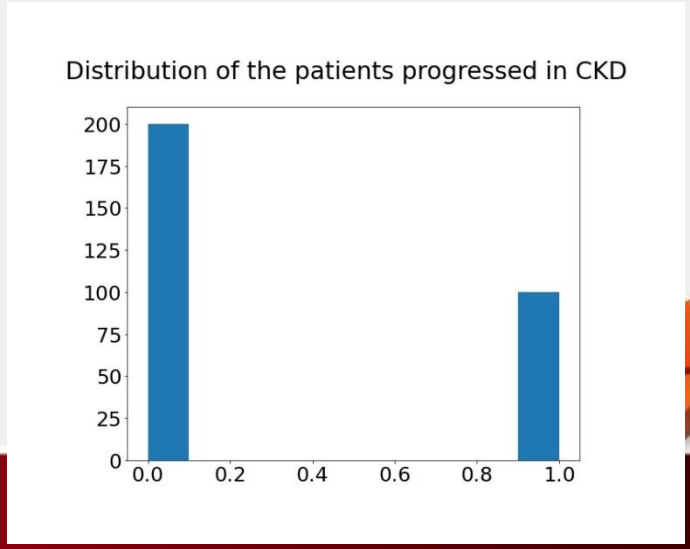
- The Dataset that is analyzed consists of 300 patients who have been diagnosed with chronic kidney disease (CKD). Their demographic information, medications and lab measurements along with their time is given in the dataset.
- The demographic information consists of the race, gender and age of the patients.
- The lab measurements performed on the patients are their serum creatinine count in mg/dl, diastolic blood pressure in mmHg, systolic blood pressure in mmHg, Hemoglobin level in g/dl, glucose level in mmol/l, low density lipoprotein level in mg/dl along with the time of measurement in days.
- The medications given to them along with the starting and ending day of the prescription.
- Additionally, a datasheet telling these patients have progressed in chronic kidney disease or not.

Analysis: Demographic Information









Predictive Model: Lab measurements including time in dataset

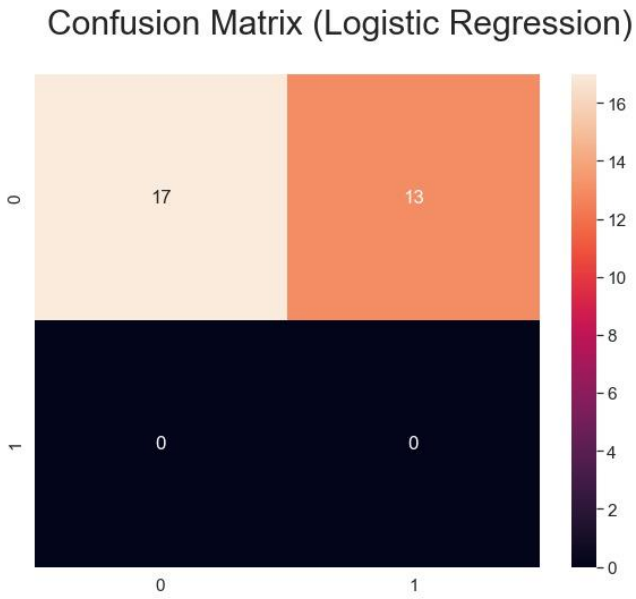
- Neural Network based on Logistic Regression

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_5 (Dense)              (None, 1000)              4000
_________________________________________________________________
dense_6 (Dense)              (None, 1000)              1001000
_________________________________________________________________
dense_7 (Dense)              (None, 1000)              1001000
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 1001
=================================================================
Total params: 2,007,001
Trainable params: 2,007,001
Non-trainable params: 0
_________________________________________________________________
```

Results: Demographic Information

- Logistic Regression

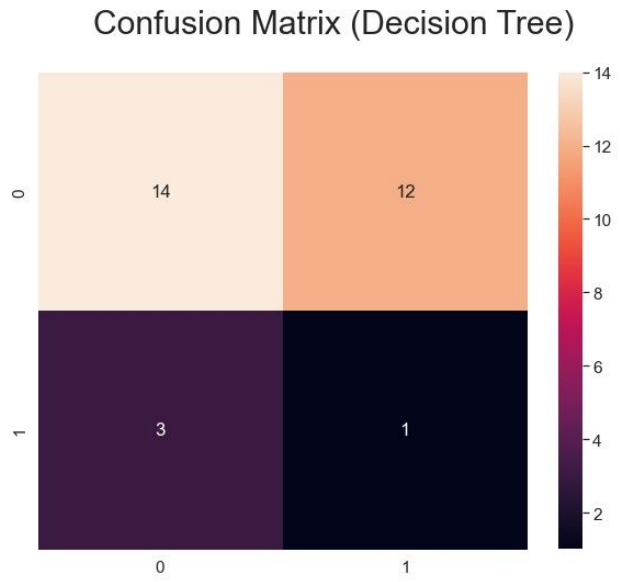
```
----------------------------------------------------------------
Accuracy of Logistic Regression Classifier : 0.5666666666666667
----------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.00      0.00      0.00        13
           0       0.57      1.00      0.72        17

    accuracy                           0.57        30
   macro avg       0.28      0.50      0.36        30
weighted avg       0.32      0.57      0.41        30

----------------------------------------------------------------
```



Results: Demographic Information

- Decision Tree

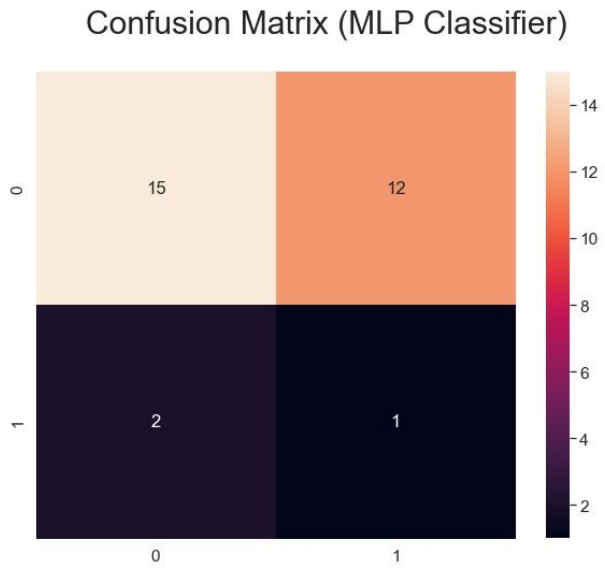
```
------------------------------------------------------------
Accuracy of Decision Tree Classifier : 0.5
------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.25      0.08      0.12        13
           0       0.54      0.82      0.65        17

    accuracy                           0.50        30
   macro avg       0.39      0.45      0.38        30
weighted avg       0.41      0.50      0.42        30

------------------------------------------------------------
```



Results: Demographic Information

- MLP Classifier

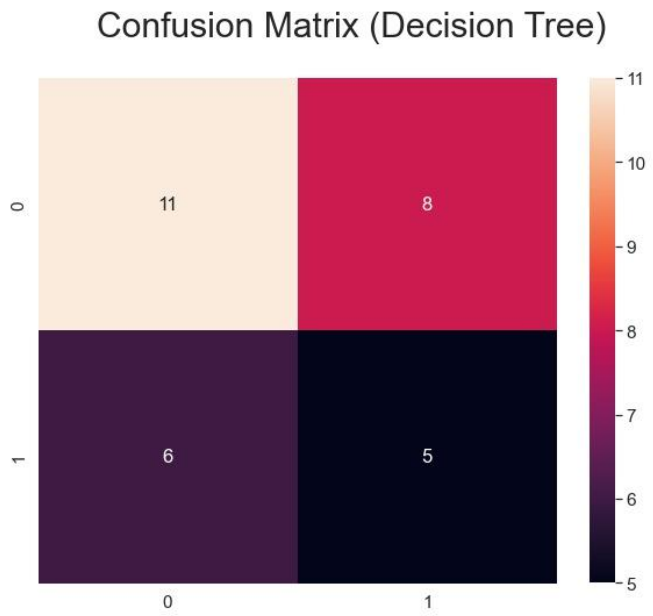
```
----------------------------------------------------------------
Accuracy of MLPClassifier : 0.5333333333333333
----------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.33      0.08      0.12        13
           0       0.56      0.88      0.68        17

    accuracy                           0.53        30
   macro avg       0.44      0.48      0.40        30
weighted avg       0.46      0.53      0.44        30

----------------------------------------------------------------
```



Results: Demographic Information

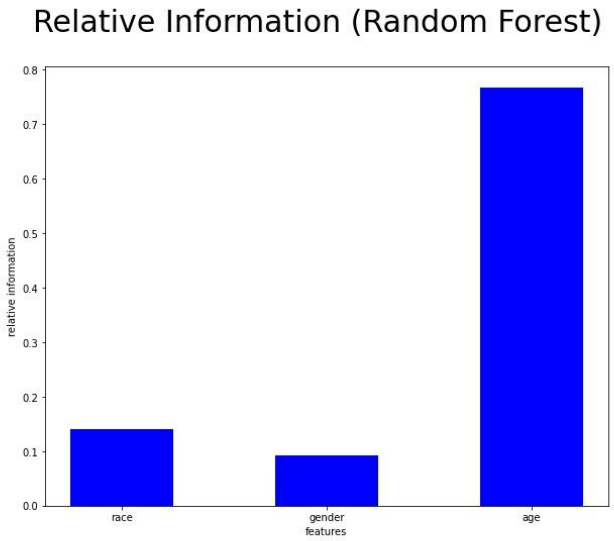
- Random Forest

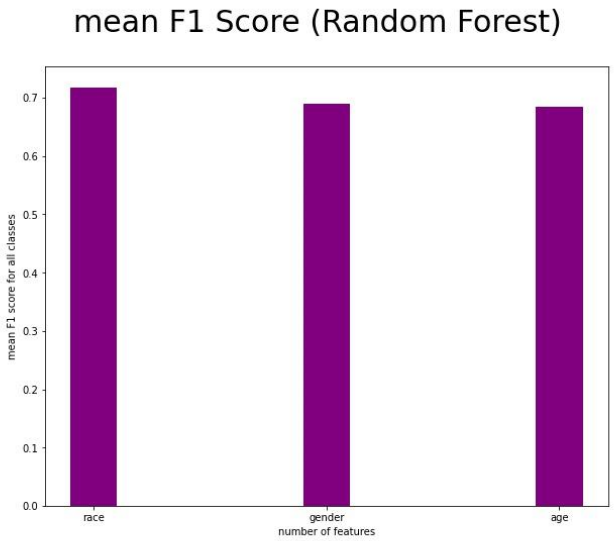
```
----------------------------------------------------------------------
Accuracy of Random Forest Classifier :  0.6333333333333333
----------------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.60      0.46      0.52        13
           0       0.65      0.76      0.70        17

    accuracy                           0.63        30
   macro avg       0.62      0.61      0.61        30
weighted avg       0.63      0.63      0.62        30

----------------------------------------------------------------------
```



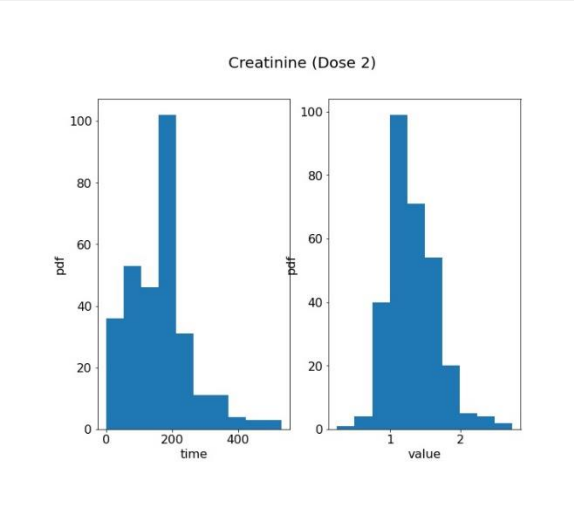
Further Results: Demographic Information

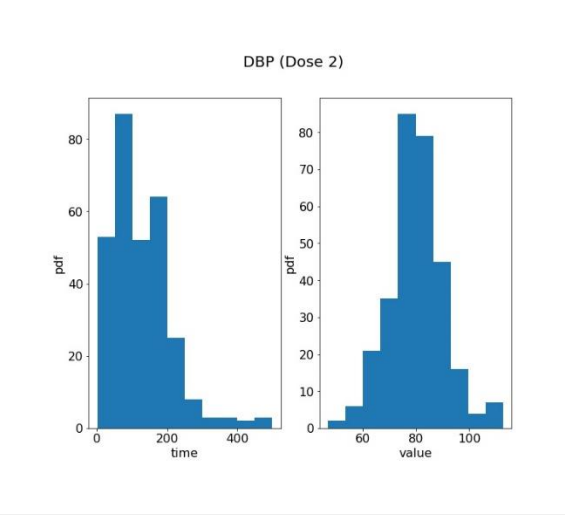
- Random Forest

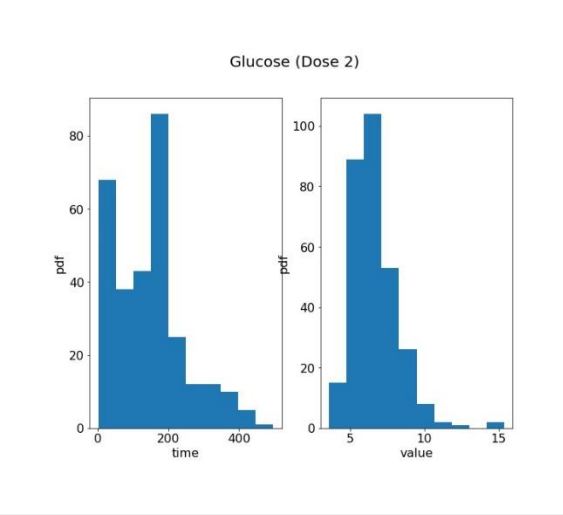


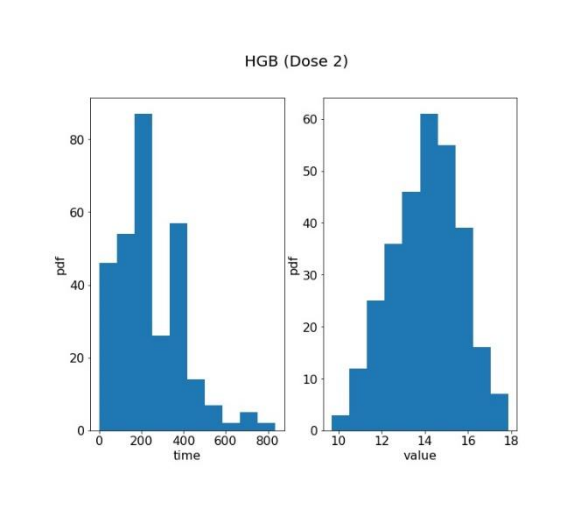


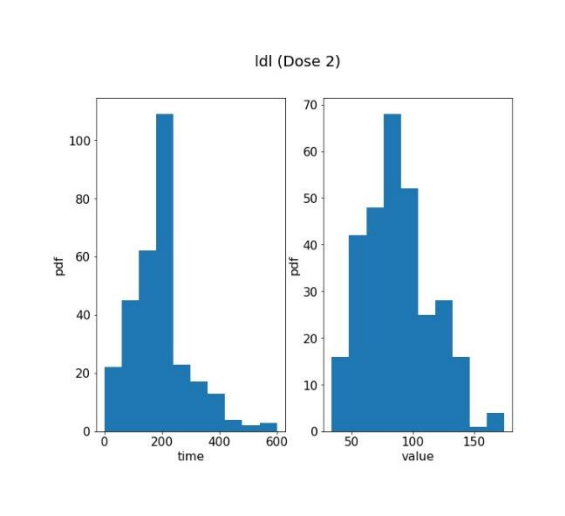
Analysis: Lab measurements including time in dataset

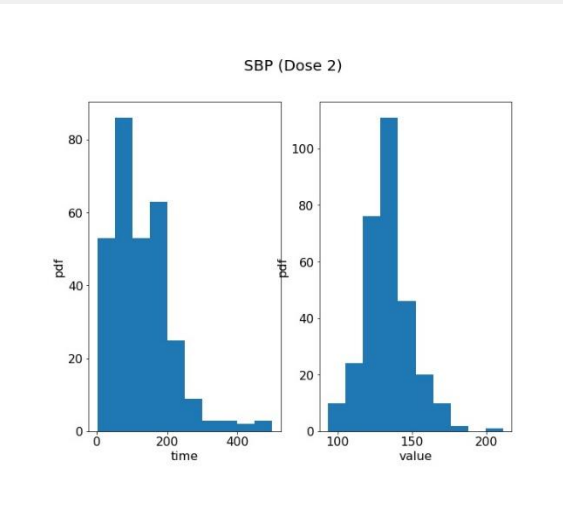




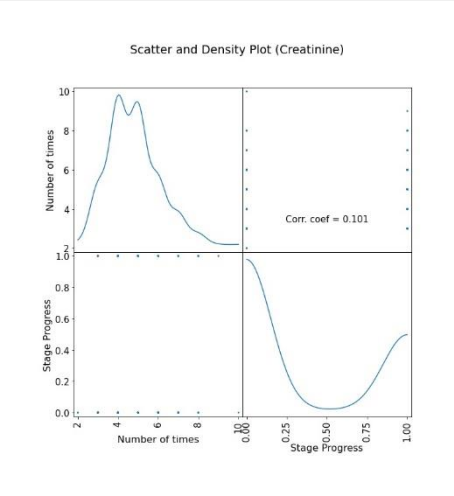


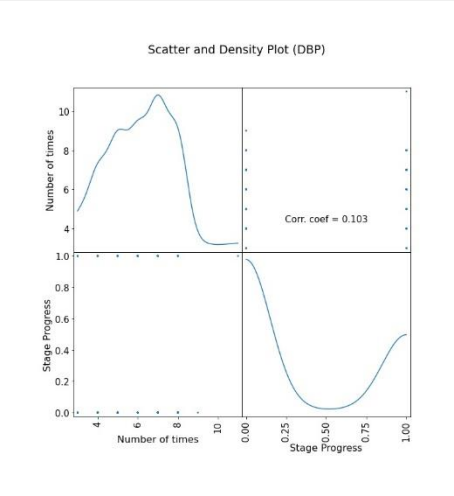


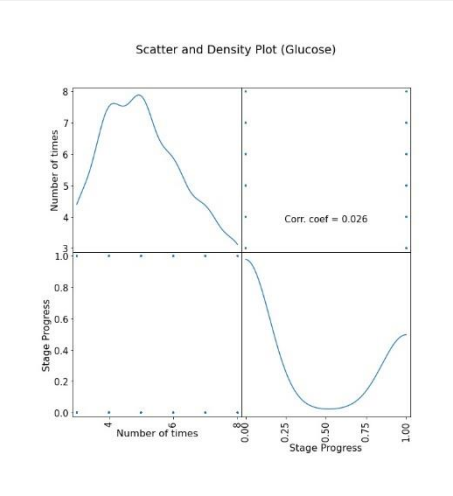


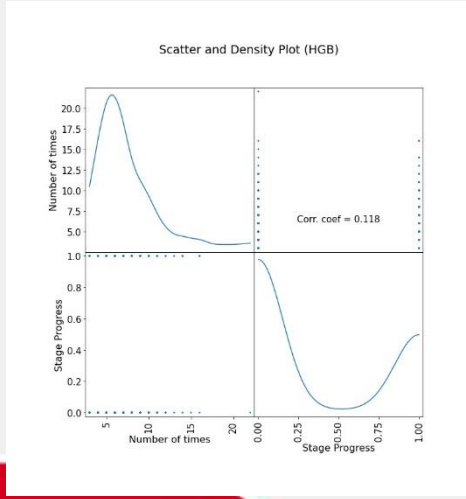


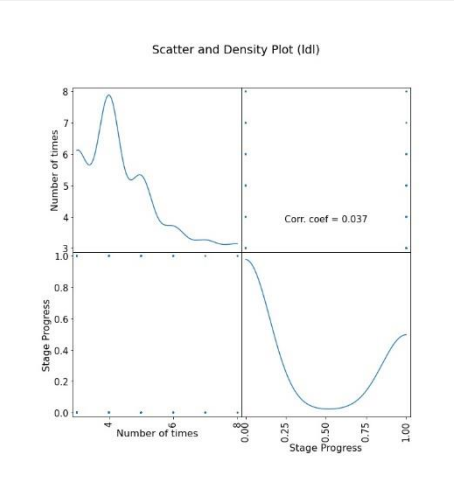
Analysis: Lab measurements including time in dataset

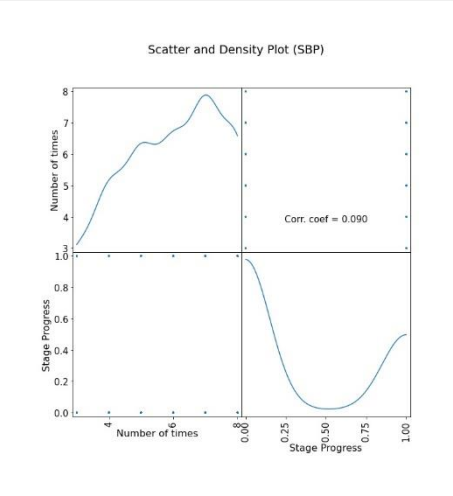












Predictive Model: Lab measurements including time in dataset

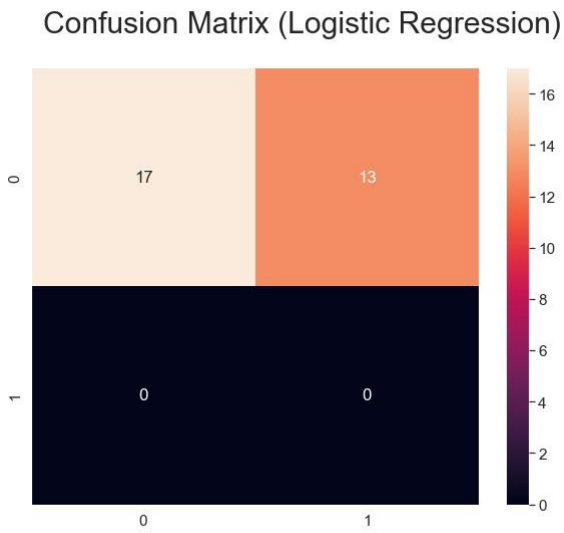
- Neural Network based on Logistic Regression

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_5 (Dense)              (None, 1000)              135000
_________________________________________________________________
dense_6 (Dense)              (None, 1000)              1001000
_________________________________________________________________
dense_7 (Dense)              (None, 1000)              1001000
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 1001
=================================================================
Total params: 2,138,001
Trainable params: 2,138,001
Non-trainable params: 0
_________________________________________________________________
```

Results: Lab measurements including time in dataset

- Logistic Regression

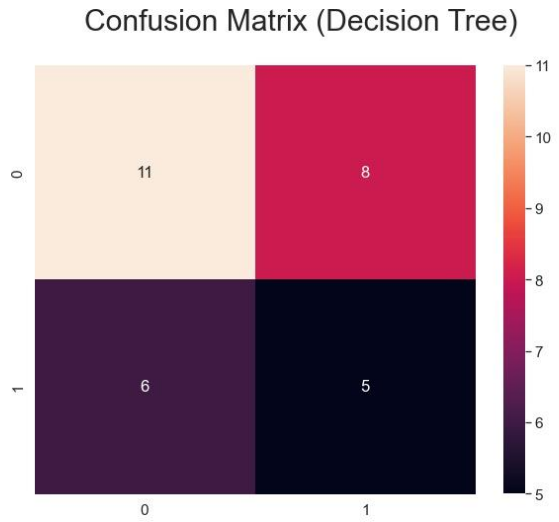
```
--------------------------------------------------------------------
Accuracy of Logistic Regression Classifier : 0.5666666666666667
--------------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.00      0.00      0.00        13
           0       0.57      1.00      0.72        17

    accuracy                           0.57        30
   macro avg       0.28      0.50      0.36        30
weighted avg       0.32      0.57      0.41        30

--------------------------------------------------------------------
```



Results: Lab measurements including time in dataset

- Decision Tree

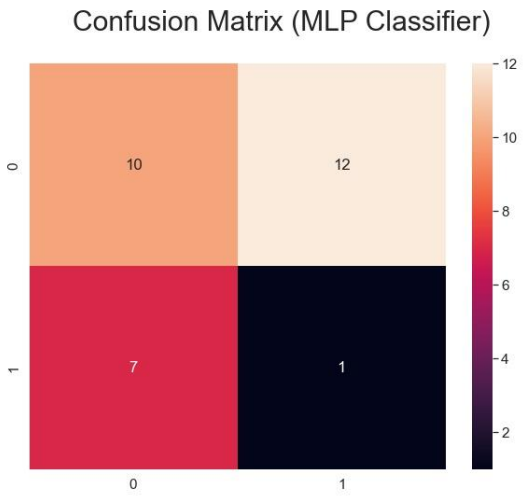
```
-----------------------------------------------------------
Accuracy of Decision Tree Classifier : 0.5333333333333333
-----------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.45      0.38      0.42        13
           0       0.58      0.65      0.61        17

    accuracy                           0.53        30
   macro avg       0.52      0.52      0.51        30
weighted avg       0.53      0.53      0.53        30

-----------------------------------------------------------
```



Results: Lab measurements including time in dataset

- MLP Classifier

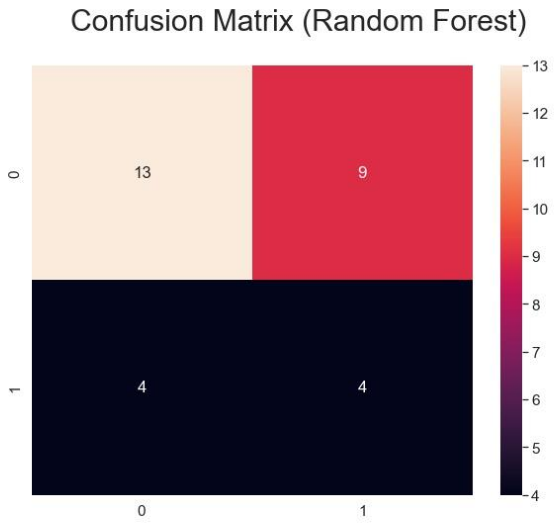
```
------------------------------------------------------------
Accuracy of MLPClassifier : 0.36666666666666664
------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.12      0.08      0.10        13
           0       0.45      0.59      0.51        17

    accuracy                           0.37        30
   macro avg       0.29      0.33      0.30        30
weighted avg       0.31      0.37      0.33        30

------------------------------------------------------------
```



Results: Lab measurements including time in dataset

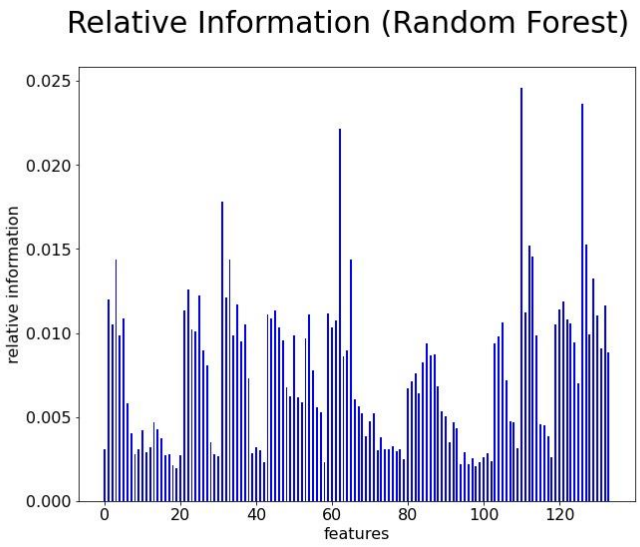
- Random Forest

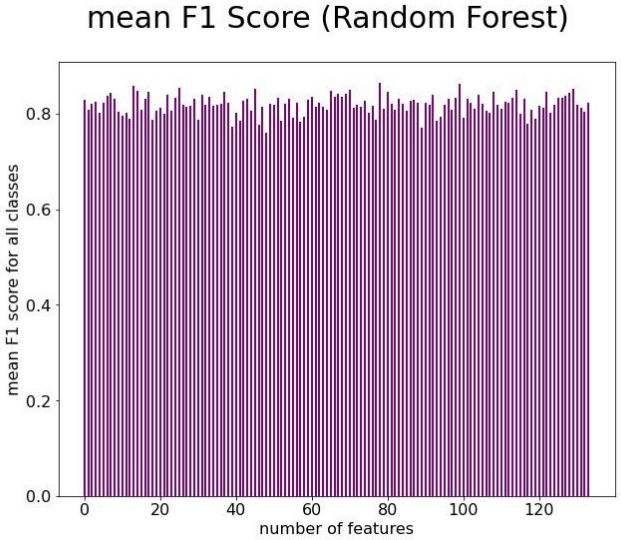
```
------------------------------------------------------------
Accuracy of Random Forest Classifier :  0.5666666666666667
------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.50      0.31      0.38        13
           0       0.59      0.76      0.67        17

    accuracy                           0.57        30
   macro avg       0.55      0.54      0.52        30
weighted avg       0.55      0.57      0.54        30

------------------------------------------------------------
```



Further Results: Lab measurements including time in dataset

- Random Forest





Predictive Model: Using RNN and LSTM

- Neural Network based on RNN and LSTM

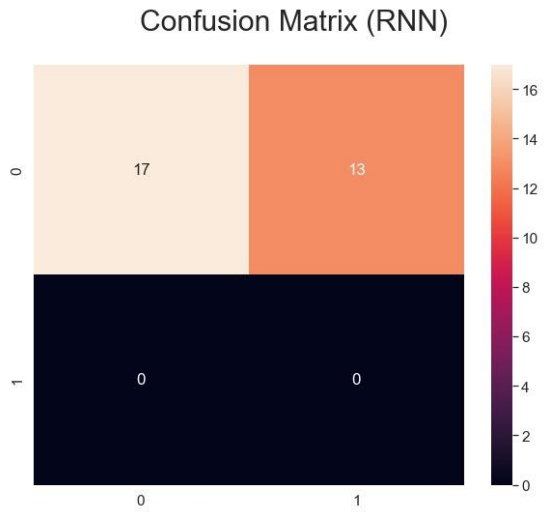
Layer (type)	Output Shape	Param #
simple_rnn_3 (SimpleRNN)	(None, 134, 1)	3
simple_rnn_4 (SimpleRNN)	(None, 134, 1)	3
lstm_2 (LSTM)	(None, 100)	40800
dense_3 (Dense)	(None, 1000)	101000
dense_4 (Dense)	(None, 1)	1001

Total params: 142,807
Trainable params: 142,807
Non-trainable params: 0

Results: Using RNN and LSTM

- RNN and LSTM

```
------------------------------------------------------------
Accuracy of RNN Classifier : 0.5666666666666667
------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.00      0.00      0.00        13
           0       0.57      1.00      0.72        17

    accuracy                           0.57        30
   macro avg       0.28      0.50      0.36        30
weighted avg       0.32      0.57      0.41        30

------------------------------------------------------------
```



Predictive Model: Using CWRNN

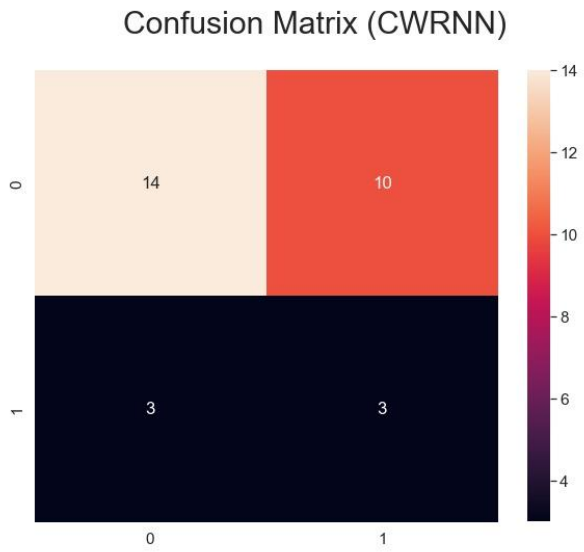
- Neural Network based on CWRNN

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
clockwork_simple_rnn_2 (Cloc (None, 137)               4907
_________________________________________________________________
dense_9 (Dense)              (None, 1000)              138000
_________________________________________________________________
dense_10 (Dense)             (None, 1)                 1001
=================================================================
Total params: 143,908
Trainable params: 143,908
Non-trainable params: 0
_________________________________________________________________
```

Results: Using CWRNN

- CWRNN

```
------------------------------------------------------------------
Accuracy of CWRNN Classifier : 0.5666666666666667
------------------------------------------------------------------
Classification report :
              precision    recall  f1-score   support

           1       0.50      0.23      0.32        13
           0       0.58      0.82      0.68        17

    accuracy                           0.57        30
   macro avg       0.54      0.53      0.50        30
weighted avg       0.55      0.57      0.52        30

------------------------------------------------------------------
```



Predictive Model: Lab Measurement placed at the timeth entry

- Neural Network based on CNN

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	(None, 3632, 1)	0
conv1d_4 (Conv1D)	(None, 3632, 64)	256
batch_normalization_4 (Batch	(None, 3632, 64)	256
re_lu_4 (ReLU)	(None, 3632, 64)	0
conv1d_5 (Conv1D)	(None, 3632, 64)	12352
batch_normalization_5 (Batch	(None, 3632, 64)	256
re_lu_5 (ReLU)	(None, 3632, 64)	0
conv1d_6 (Conv1D)	(None, 3632, 64)	12352
batch_normalization_6 (Batch	(None, 3632, 64)	256
re_lu_6 (ReLU)	(None, 3632, 64)	0
global_average_pooling1d_2 (	(None, 64)	0
dense_2 (Dense)	(None, 2)	130

Total params: 25,858
Trainable params: 25,474
Non-trainable params: 384

Results: Lab Measurement placed at the timeth entry

```
------------------------------------------------------------
Accuracy from CNN Classifier :  0.4333333373069763
------------------------------------------------------------
```