Performance Evaluation

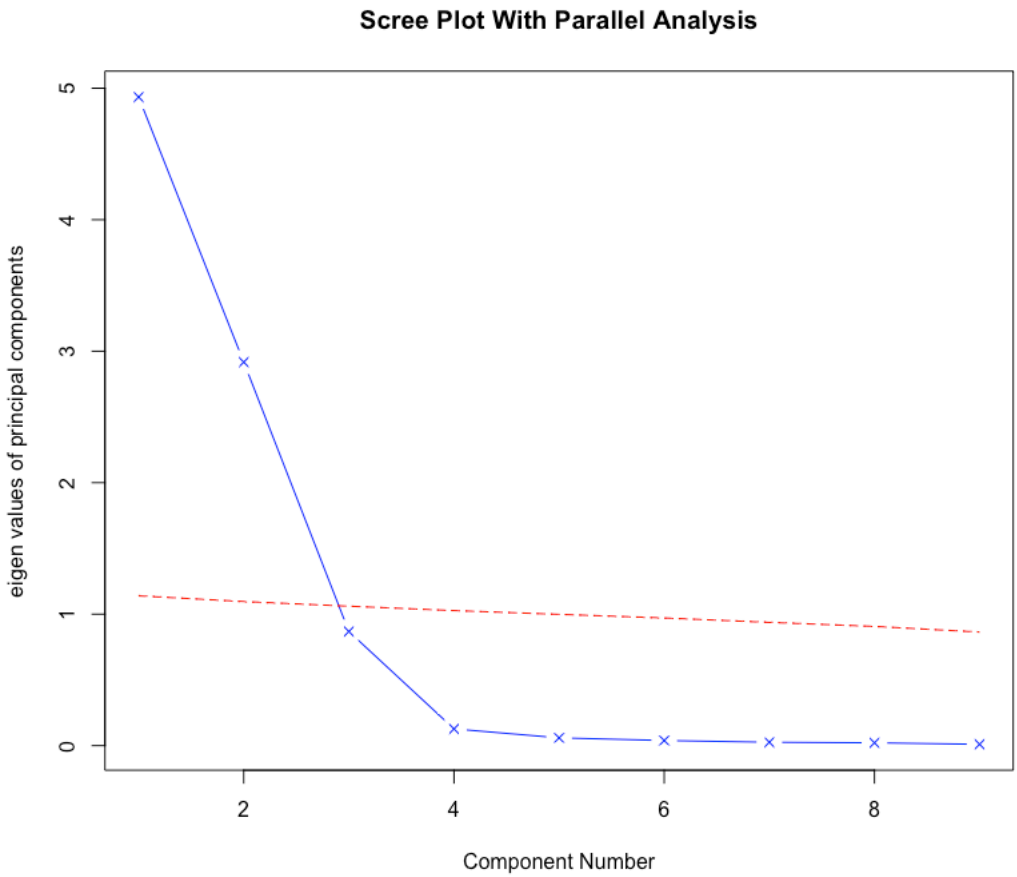
Stock Data

Graphed using Highcharter. Chart type = Candlestick OHLC

Features used for building our model.

RSI: "rsiFNCL","rsiIYF","rsiXLF"
ADX: "adxFNCL","adxIYF","adxXLF"
SAR: "trendfncl","trendiyf","trendxlf"

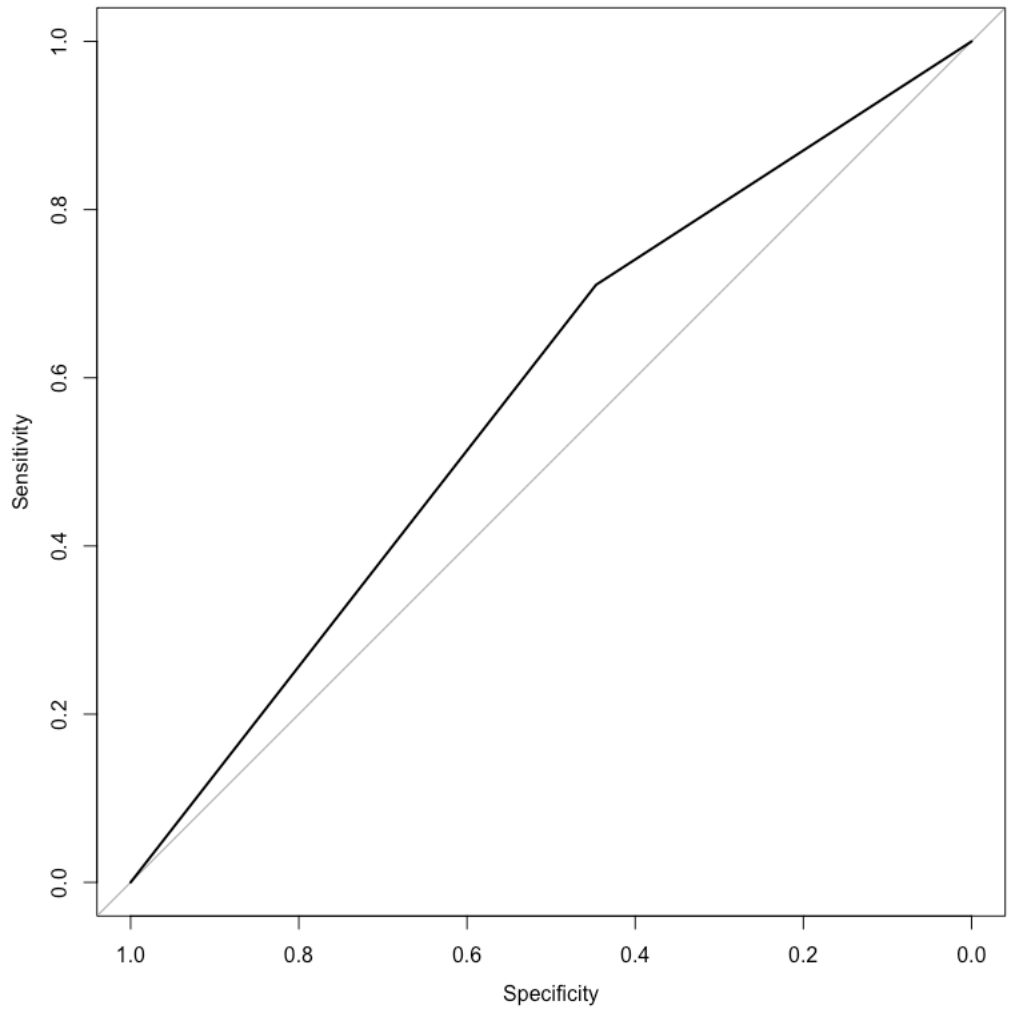
After running PCA as part of our Feature Engineering Process, we arrived at:



Parallel analysis suggests that 2 components suffice.

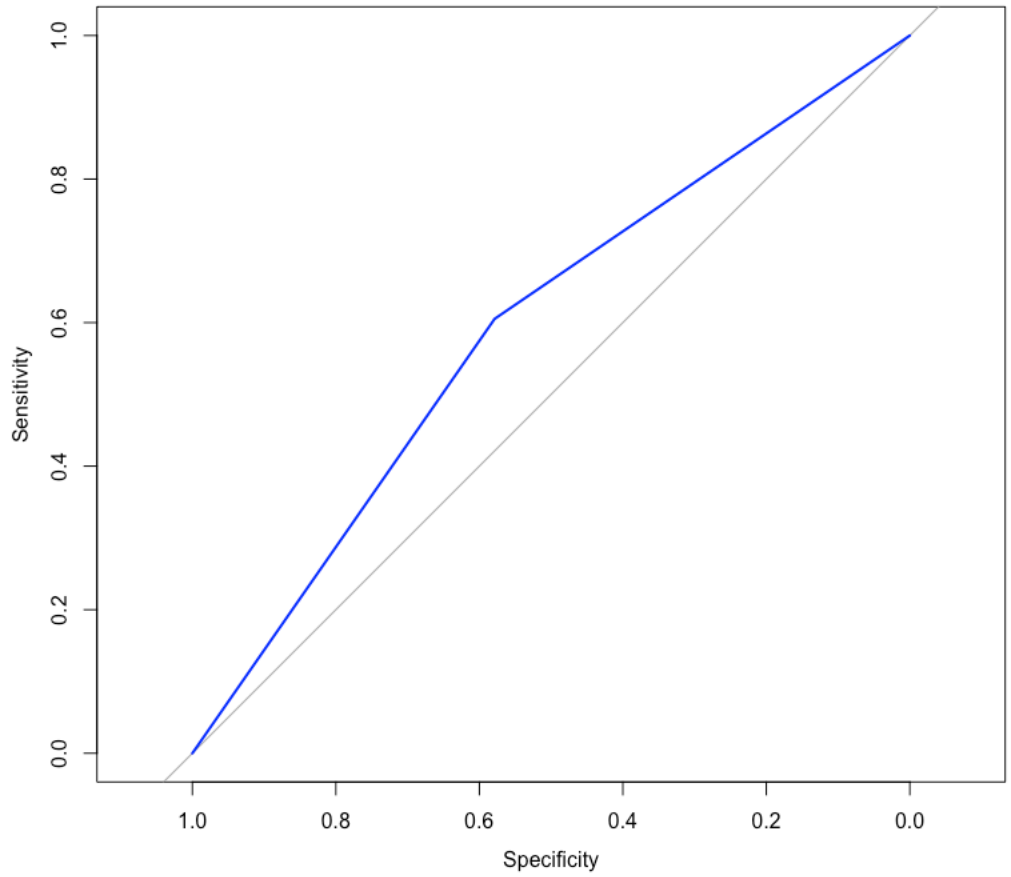
```
Principal Components Analysis
Call: principal(r = model[, -1], nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
             RC1   RC2   h2    u2 com
rsiFNCL     0.93 -0.03 0.86 0.141 1.0
rsiIYF      0.92 -0.02 0.85 0.154 1.0
rsiXLF      0.92 -0.03 0.84 0.158 1.0
adxFNCL.ADX 0.04  0.99 0.97 0.026 1.0
adxIYF.ADX  0.01  0.98 0.96 0.044 1.0
adxXLF.ADX  0.02  0.99 0.98 0.020 1.0
FNCL.Close  0.85  0.16 0.75 0.251 1.1
IYF.Close   0.90  0.05 0.80 0.197 1.0
XLF.Close   0.90  0.09 0.82 0.177 1.0

                       RC1  RC2
SS loadings           4.89 2.95
Proportion Var        0.54 0.33
Cumulative Var        0.54 0.87
Proportion Explained  0.62 0.38
Cumulative Proportion 0.62 1.00

Mean item complexity =  1
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.09
 with the empirical chi square  624.97  with prob <  2.7e-120
```

After component extraction we proceeded to run our first algorithm which was KNN classification. Using a function to automatically test for accuracy (TP+TN/Total Cases) for up to 20 N, we got a ROC curve optimized for n=8.




```
Area under the curve: 0.5785
```

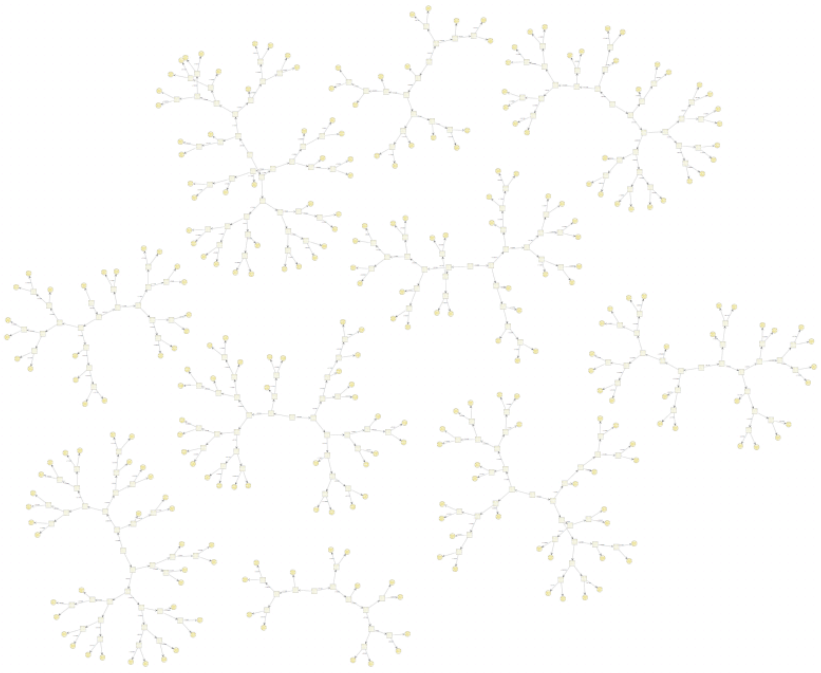
Our second algorithm is an approach called extreme gradient boosting which involves using gradient boosted trees (parallelized for faster computation).



We achieve an AUC Score of: `[1] "AUC Score: 0.591939755047997"`

We can achieve higher scores with xgboost by tweaking the parameters. We are currently in the process of doing so.

Here's the current tree structure we obtained after doing so.



Extracting just the first tree: