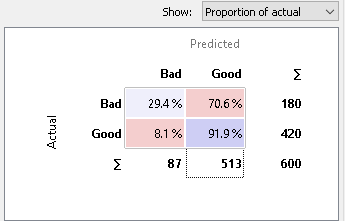
# Predictive Analysis Modelling – Random Forest

Test and Score Results

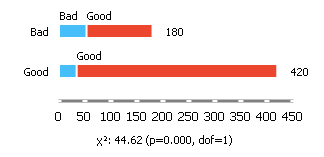
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **AUC** | **CA** | **F1** | **PRECISION** | **RECALL** |
| Random Forest | 0.798 | 0.775 | 0.757 | 0.765 | 0.775 |

Proportion of actual – Confusion Matrix



Same info just whatever way you want to display.

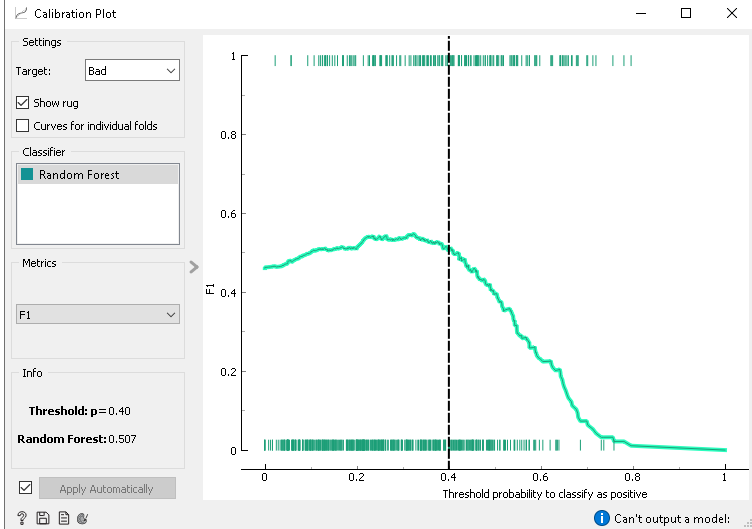
**Prediction Box Plot**

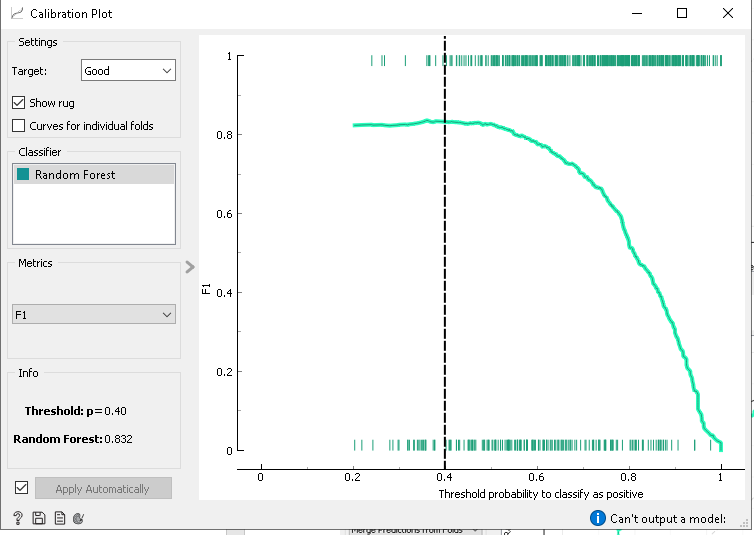


Bad –70.56% correct

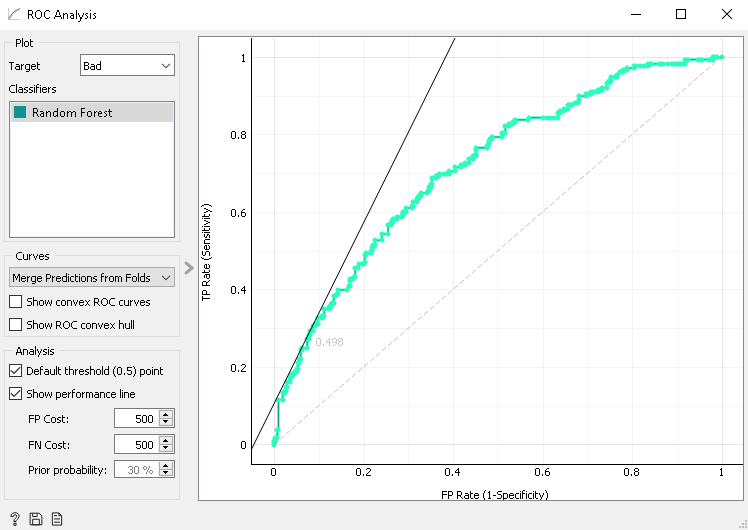
Good – 91.90% correct

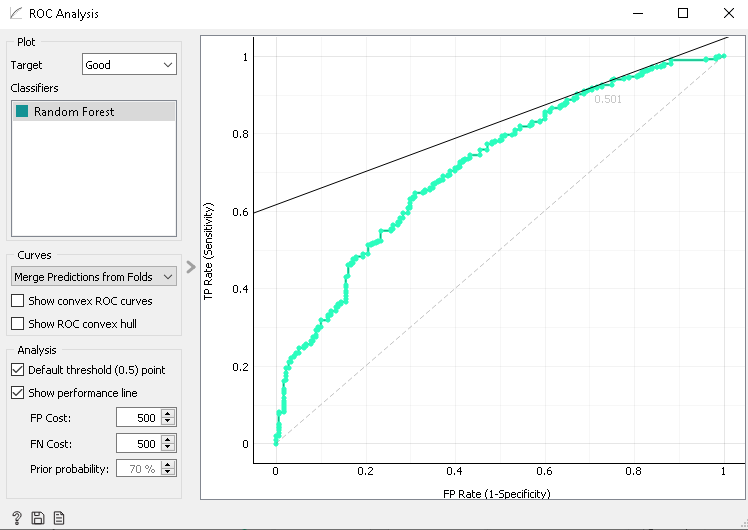
Calibration Plot





|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Target** | **Threshold** | **F1** |
| Random Forest | Bad | 0.4 | 0.507 |
| Random Forest | Good | 0.4 | 0.832 |





Don’t really see too much in ROC analysis

AUC 0.791

**DBSCAN**

DBSCAN does not work with categorical data, no clusters are formed when inputting data straight into scan.

If we use continuize function to change categorical variables into numerical we lose accuracy in our data as the numerical variables do not represent our categories properly.