**Logistic Regression:**

Model Training:

* Training data was split to determine the accuracy of the model as the given test data had no "creditability" attribute for verification.
* F1 score = 0.798
* Correlation matrix
  + Number of instances

|  |  |  |
| --- | --- | --- |
|  | Bad | Good |
| Bad | 97 | 73 |
| Good | 36 | 354 |

* Proportion of predicted

|  |  |  |
| --- | --- | --- |
|  | Bad | Good |
| Bad | 72.9% | 17.1% |
| Good | 27.1% | 82.9% |

* Proportion of actual

|  |  |  |
| --- | --- | --- |
| A | Bad | Good |
| Bad | 57.1% | 42.9% |
| Good | 9.2% | 90.8% |

* ROC
  + Good
    - No clear distinction for where auc bends

Chart, line chart

Description automatically generated

* Bad
  + No clear distinction for where auc bends

Chart, line chart

Description automatically generated

* F1 Calibration
  + Good
    - Plot shows that logistic regression can accurately predict good creditors around 80% of the time.

Chart, line chart

Description automatically generated

* Bad
  + Logistic regression has a more difficult time predicting bad credit risks at 60%

Chart, line chart

Description automatically generated

Model Testing:

* Logistic regression was trained on the entire training dataset to produce predictions for the given testing dataset.
* F1 score = 0.87608

**Stacking:**

Model Training:

* Training data was split to determine the accuracy of the model as the given test data had no "creditability" attribute for verification
* F1 score = 0.841
* Correlation matrix
  + Number of instances

|  |  |  |
| --- | --- | --- |
|  | Bad | Good |
| Bad | 102 | 68 |
| Good | 16 | 374 |

* Proportion of predicted

|  |  |  |
| --- | --- | --- |
|  | Bad | Good |
| Bad | 86.4% | 15.4% |
| Good | 13.6% | 84.6% |

* Proportion of actual

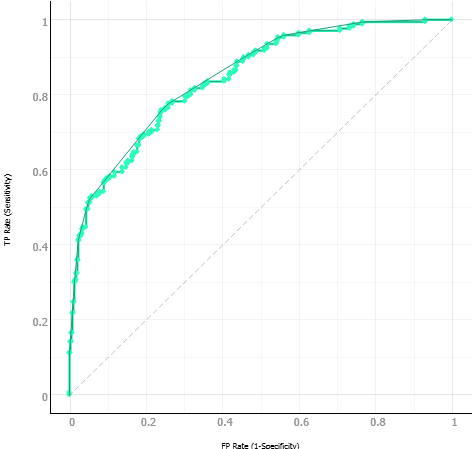
|  |  |  |
| --- | --- | --- |
|  | Bad | Good |
| Bad | 60.0% | 40.0% |
| Good | 4.1% | 95.9% |

* ROC
  + Good

Chart, line chart

Description automatically generated

* Bad



* F1 Calibration
  + Good
    - Shows that with threshold 0.5, accuracy for good creditors is at about 0.8 F1

Chart, line chart

Description automatically generated

* Bad
  + Accuracy at around 0.65

Chart, line chart

Description automatically generated

**Thresholds:**

* + - * F1 calibration plots show the probability thresholds that optimize F1 score

**Louvain Clustering:**

* + - * Clusters identify that aged creditors have the highest credit durations and amounts
      * Most aged creditors are bad creditors
      * Hence most aged creditors have critical accounts