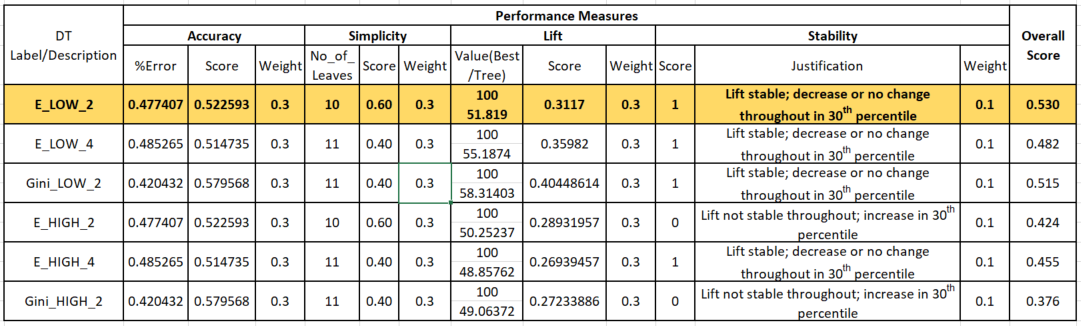
I used SAS EM to run all the algorithms below:

Decision Trees

1 Evaluation

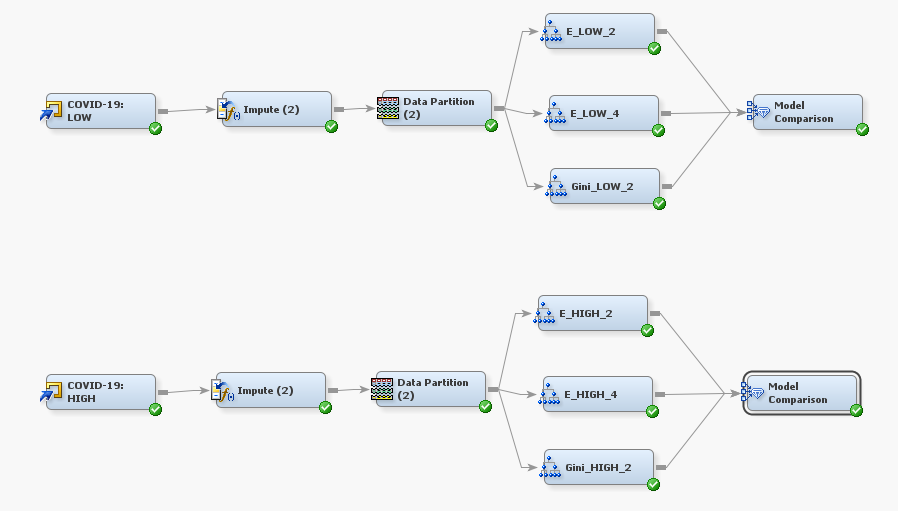
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Measure** | **Description** | **Definition of Value Function** | **Weight** | **Threshold** |
| Accuracy | Proportion correctly classified | 1 – misclassification rate | 0.3 | > 0.50 |
| Simplicity | Number of leaves for the tree when compared with an ideal tree with regard to the number of leaves | Defined by function ScoreSimplicity(NoOfLeaves)  If NoOfLeaves <= 2 or >=13; 0  If NoOfLeaves is >= 3 & <= 4;  (NoOfLeaves - 2)/5-2)  If NoOfLeaves is >= 5 & <= 8; 1  If NoOfLeaves >= 9 & <= 12  (13 -NoOfLeaves)/(13-8) | 0.3 | > 0.00 |
| Lift | Lift value (Cumulative%CapturedResponse) at 3rd percentile | (Tree – Baseline)/(Best – Baseline) | 0.3 | > 0.20 |
| Stability | Stability is binary as determined by the visual inspection of the non-cumulative %Response Lift Chart up to the 3rd percentile | 1 indicating stable model there the lift is stable or decrease throughout and 0 indicating an unstable model where lift increases up to the given nth decile. | 0.1 | > 0.00 |
| Combination Function | Overall Score = wAccuracy\*ScoreAccuracy + wSimplicity\*ScoreSimplicity + wLift\*ScoreLift + wStability \*ScoreStability | | | |

2 Summary of Results

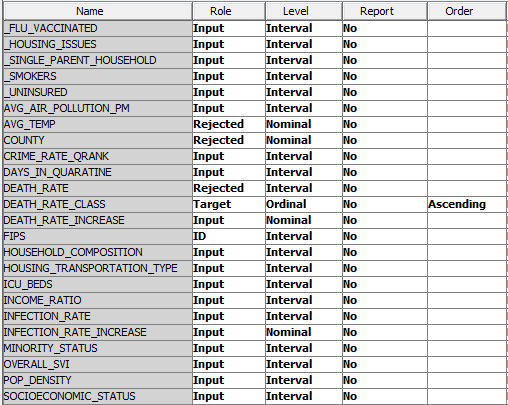


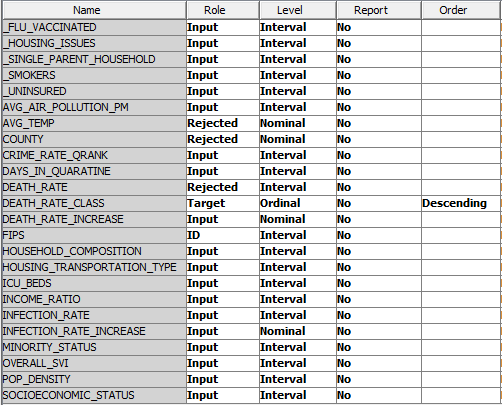
3 Evidence of Experimentation

3.1 Process Flow Diagram

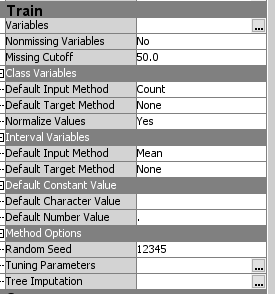


3.2 Variables

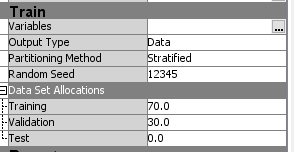


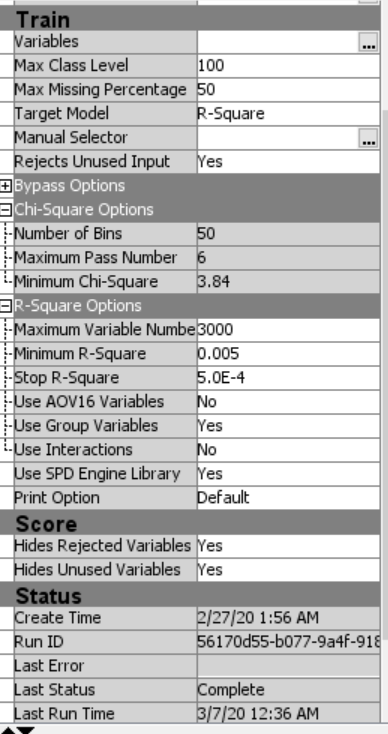


3.3 Impute Node

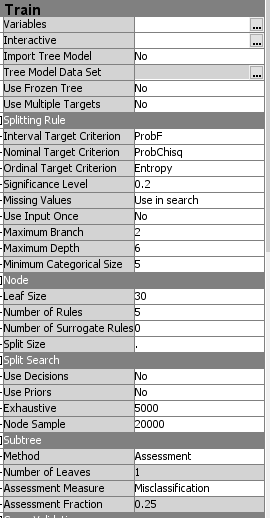
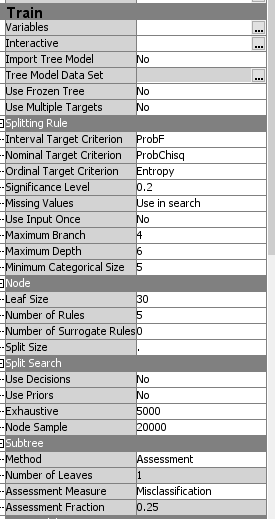


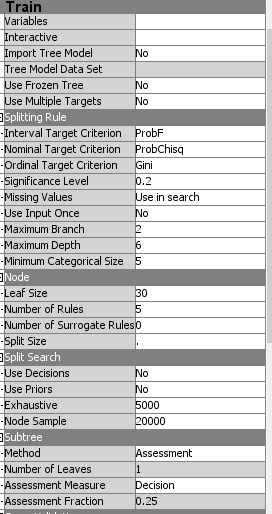
3.4 Data Partition Node





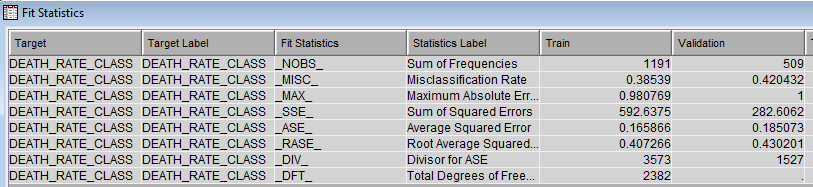
3.5 Decision Tree Nodes

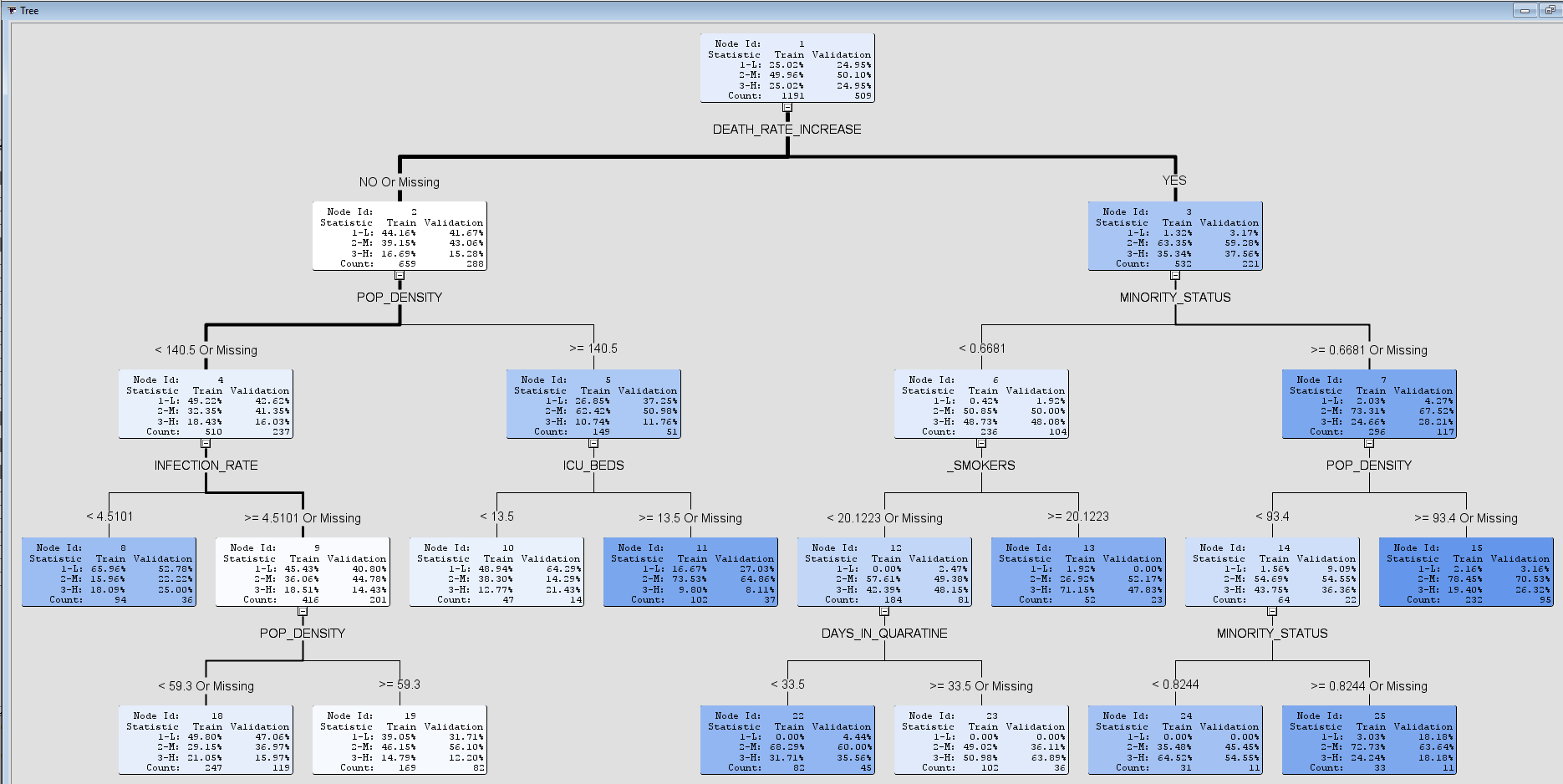


3 Best Results E\_LOW\_4

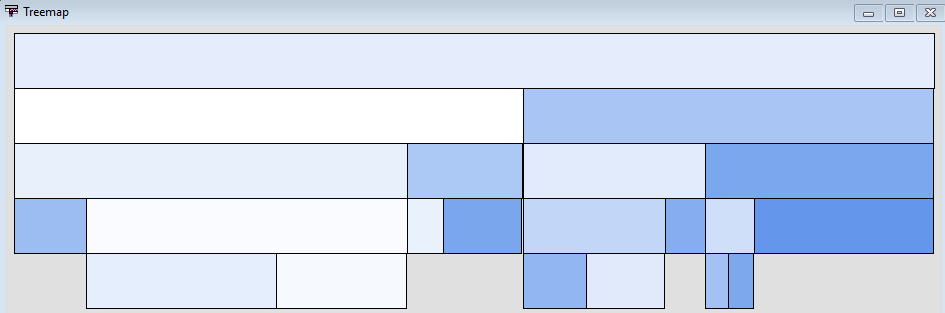
3.5.1 Fit Statistics



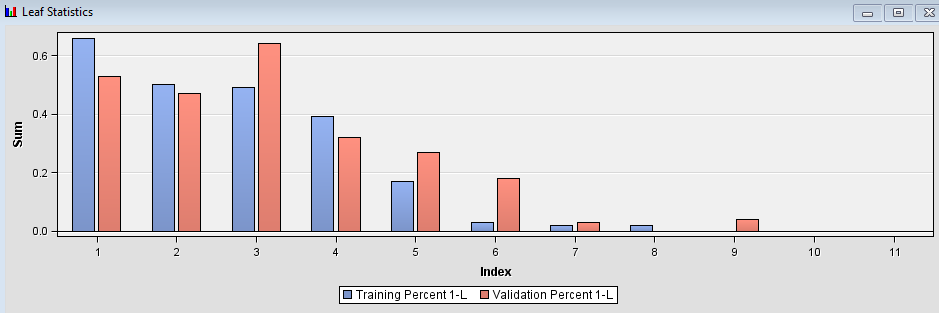
3.5.2 Tree



3.5.3 Treemap



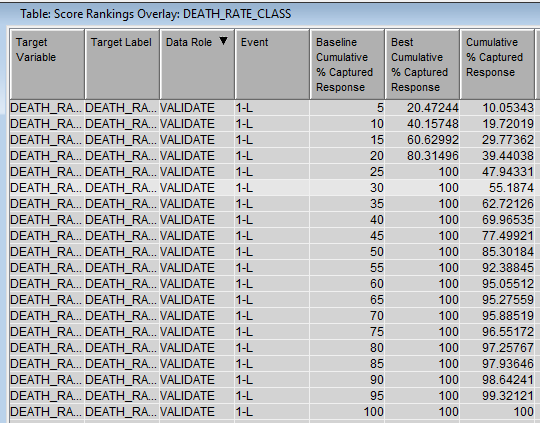
3.5.4 Leaf Statistics



3.5.5 Tree Result, Non-Cumulative %Response Chart



3.5.6 Table: Score Rankings Overlay



3.5.7 Node Rules

\*------------------------------------------------------------\*

Node = 8

\*------------------------------------------------------------\*

if POP\_DENSITY < 140.5 or MISSING

AND INFECTION\_RATE < 4.51006

AND DEATH\_RATE\_INCREASE IS ONE OF: NO or MISSING

then

Tree Node Identifier = 8

Number of Observations = 94

Predicted: DEATH\_RATE\_CLASS=3-H = 0.18

Predicted: DEATH\_RATE\_CLASS=2-M = 0.16

Predicted: DEATH\_RATE\_CLASS=1-L = 0.66

\*------------------------------------------------------------\*

Node = 10

\*------------------------------------------------------------\*

if POP\_DENSITY >= 140.5

AND ICU\_BEDS < 13.5

AND DEATH\_RATE\_INCREASE IS ONE OF: NO or MISSING

then

Tree Node Identifier = 10

Number of Observations = 47

Predicted: DEATH\_RATE\_CLASS=3-H = 0.13

Predicted: DEATH\_RATE\_CLASS=2-M = 0.38

Predicted: DEATH\_RATE\_CLASS=1-L = 0.49

\*------------------------------------------------------------\*

Node = 11

\*------------------------------------------------------------\*

if POP\_DENSITY >= 140.5

AND ICU\_BEDS >= 13.5 or MISSING

AND DEATH\_RATE\_INCREASE IS ONE OF: NO or MISSING

then

Tree Node Identifier = 11

Number of Observations = 102

Predicted: DEATH\_RATE\_CLASS=3-H = 0.10

Predicted: DEATH\_RATE\_CLASS=2-M = 0.74

Predicted: DEATH\_RATE\_CLASS=1-L = 0.17

\*------------------------------------------------------------\*

Node = 13

\*------------------------------------------------------------\*

if \_SMOKERS >= 20.1223

AND MINORITY\_STATUS < 0.6681

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

then

Tree Node Identifier = 13

Number of Observations = 52

Predicted: DEATH\_RATE\_CLASS=3-H = 0.71

Predicted: DEATH\_RATE\_CLASS=2-M = 0.27

Predicted: DEATH\_RATE\_CLASS=1-L = 0.02

\*------------------------------------------------------------\*

Node = 15

\*------------------------------------------------------------\*

if POP\_DENSITY >= 93.4 or MISSING

AND MINORITY\_STATUS >= 0.6681 or MISSING

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

then

Tree Node Identifier = 15

Number of Observations = 232

Predicted: DEATH\_RATE\_CLASS=3-H = 0.19

Predicted: DEATH\_RATE\_CLASS=2-M = 0.78

Predicted: DEATH\_RATE\_CLASS=1-L = 0.02

\*------------------------------------------------------------\*

Node = 18

\*------------------------------------------------------------\*

if POP\_DENSITY < 59.3 or MISSING

AND INFECTION\_RATE >= 4.51006 or MISSING

AND DEATH\_RATE\_INCREASE IS ONE OF: NO or MISSING

then

Tree Node Identifier = 18

Number of Observations = 247

Predicted: DEATH\_RATE\_CLASS=3-H = 0.21

Predicted: DEATH\_RATE\_CLASS=2-M = 0.29

Predicted: DEATH\_RATE\_CLASS=1-L = 0.50

\*------------------------------------------------------------\*

Node = 19

\*------------------------------------------------------------\*

if POP\_DENSITY < 140.5 AND POP\_DENSITY >= 59.3

AND INFECTION\_RATE >= 4.51006 or MISSING

AND DEATH\_RATE\_INCREASE IS ONE OF: NO or MISSING

then

Tree Node Identifier = 19

Number of Observations = 169

Predicted: DEATH\_RATE\_CLASS=3-H = 0.15

Predicted: DEATH\_RATE\_CLASS=2-M = 0.46

Predicted: DEATH\_RATE\_CLASS=1-L = 0.39

\*------------------------------------------------------------\*

Node = 22

\*------------------------------------------------------------\*

if \_SMOKERS < 20.1223 or MISSING

AND MINORITY\_STATUS < 0.6681

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

AND DAYS\_IN\_QUARATINE < 33.5

then

Tree Node Identifier = 22

Number of Observations = 82

Predicted: DEATH\_RATE\_CLASS=3-H = 0.32

Predicted: DEATH\_RATE\_CLASS=2-M = 0.68

Predicted: DEATH\_RATE\_CLASS=1-L = 0.00

\*------------------------------------------------------------\*

Node = 23

\*------------------------------------------------------------\*

if \_SMOKERS < 20.1223 or MISSING

AND MINORITY\_STATUS < 0.6681

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

AND DAYS\_IN\_QUARATINE >= 33.5 or MISSING

then

Tree Node Identifier = 23

Number of Observations = 102

Predicted: DEATH\_RATE\_CLASS=3-H = 0.51

Predicted: DEATH\_RATE\_CLASS=2-M = 0.49

Predicted: DEATH\_RATE\_CLASS=1-L = 0.00

\*------------------------------------------------------------\*

Node = 24

\*------------------------------------------------------------\*

if POP\_DENSITY < 93.4

AND MINORITY\_STATUS < 0.8244 AND MINORITY\_STATUS >= 0.6681

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

then

Tree Node Identifier = 24

Number of Observations = 31

Predicted: DEATH\_RATE\_CLASS=3-H = 0.65

Predicted: DEATH\_RATE\_CLASS=2-M = 0.35

Predicted: DEATH\_RATE\_CLASS=1-L = 0.00

\*------------------------------------------------------------\*

Node = 25

\*------------------------------------------------------------\*

if POP\_DENSITY < 93.4

AND MINORITY\_STATUS >= 0.8244 or MISSING

AND DEATH\_RATE\_INCREASE IS ONE OF: YES

then

Tree Node Identifier = 25

Number of Observations = 33

Predicted: DEATH\_RATE\_CLASS=3-H = 0.24

Predicted: DEATH\_RATE\_CLASS=2-M = 0.73

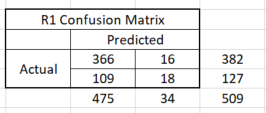
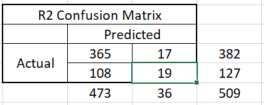
Predicted: DEATH\_RATE\_CLASS=1-L = 0.03

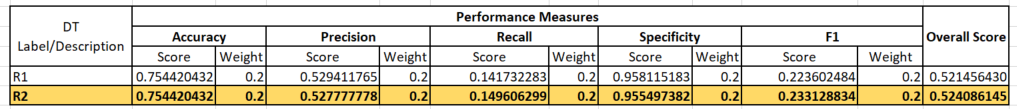
Logistic Regression

1 Evaluation Approach

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Measure** | **Description** | **Definition of Value Function** | **Weight** | **Threshold** |
| Accuracy | Number of correctly classified data instances over the total number of data instances | (TN+TP)/(all cases) | 0.2 | >0.5 |
| Precision | Positive predictive value | TP/(TP+FP) | 0.2 | >0.5 |
| Recall | Ability of classifier to correctly detect true positive rate | TP/(TP+FN) | 0.2 | >0.2 |
| Specificity | Ability of classier to correctly detect true negative rate | TN/(TN+FP) | 0.2 | >0.0 |
| F1 Score |  | 2(Precision\*Recall)/ (Precision+Recall) | 0.2 | >0.2 |
| Combination Function | Overall Score = wAccuracy\*ScoreAccuracy + wSensitivity\*ScoreSensitivity + wSpecificity \*ScoreSpecificity | | | |

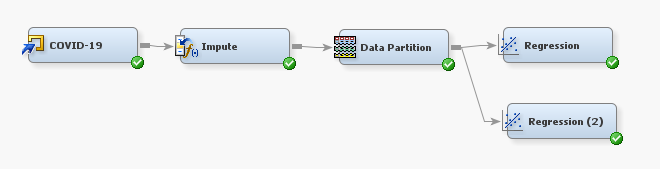
2 Summary of Results

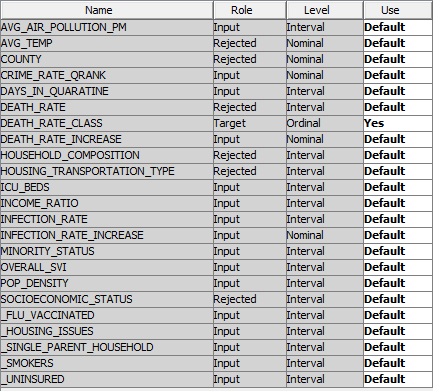


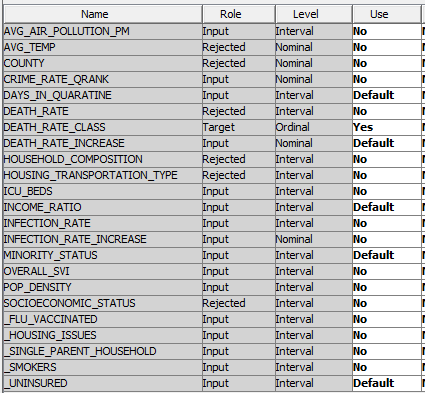
2 Evidence of Experimentation

2.1 Process Flow Diagram

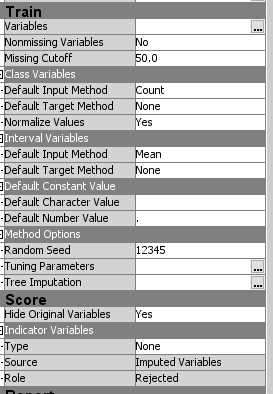


2.2 Variables

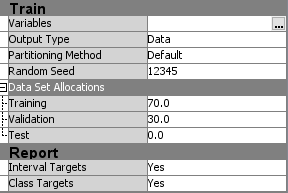




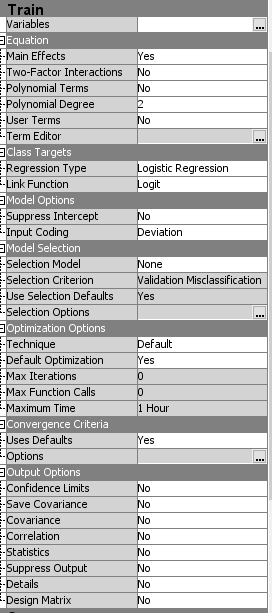
2.3 Impute Node



2.4 Data Partition Node

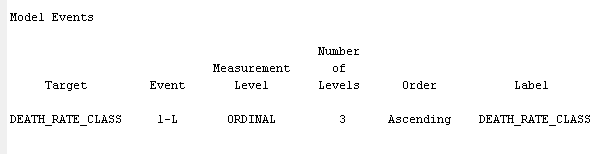


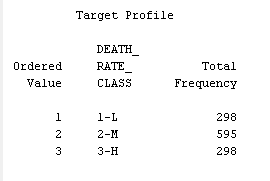
2.5 Regression Node

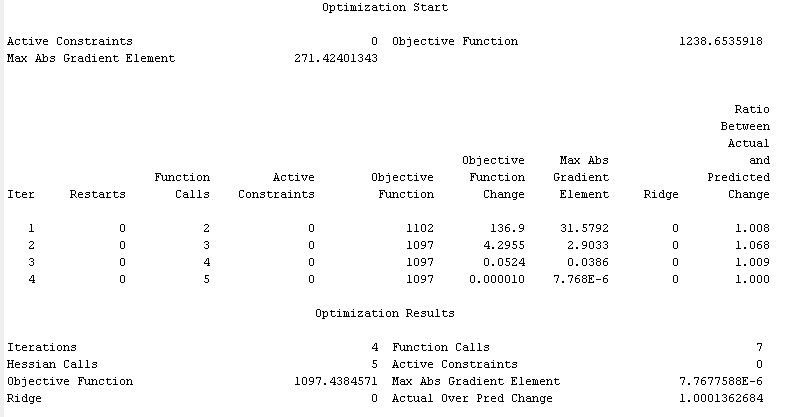


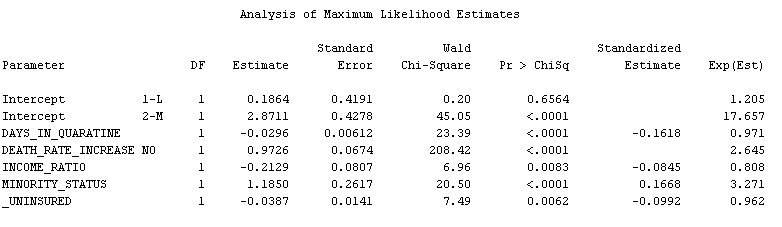
3 Results

3.1 Output

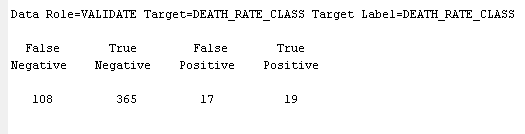




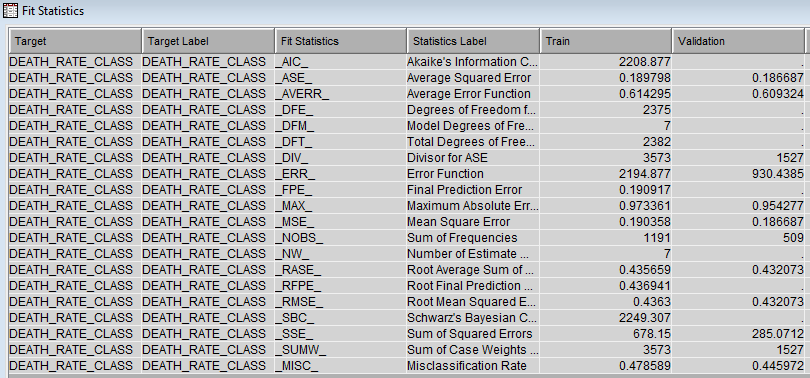




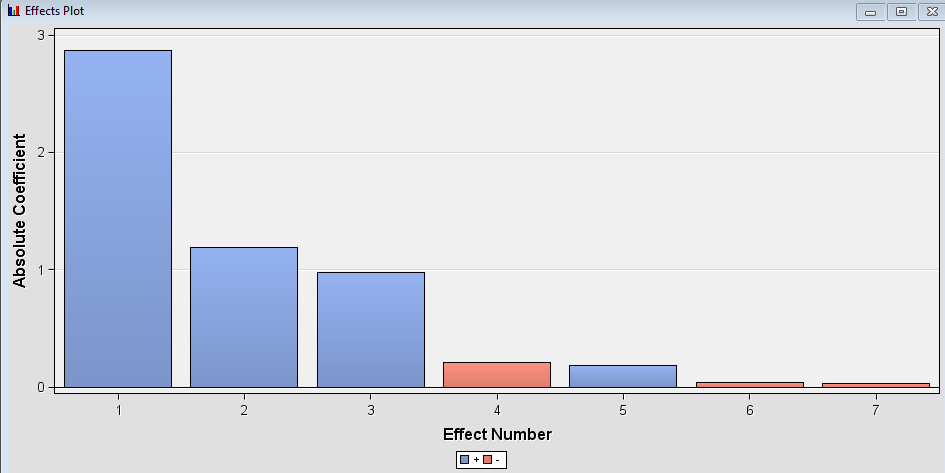
3.2 Confusion Matrix



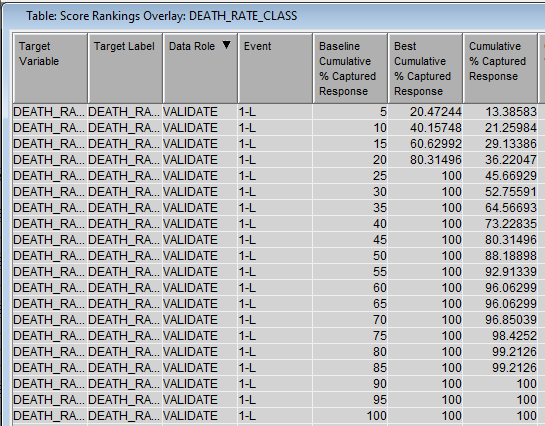
3.3 Fit Statistics



3.3 Effect Plot



3.4 Table: Score Rankings Overlay



3.5 Graph: Score Rankings Overlay %Response

