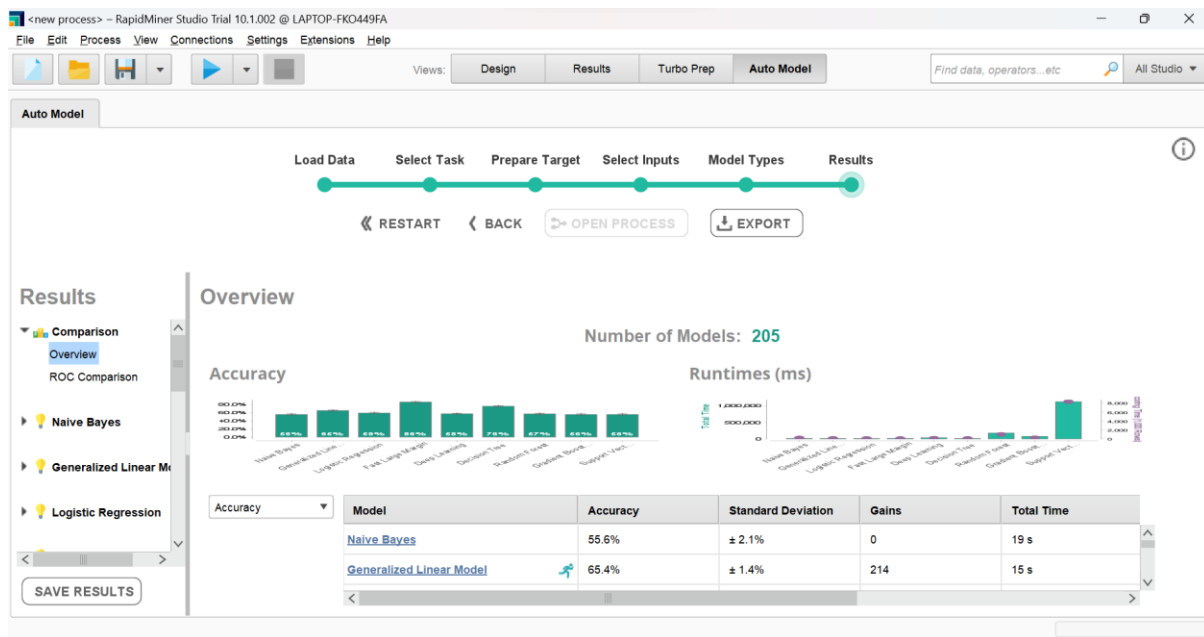
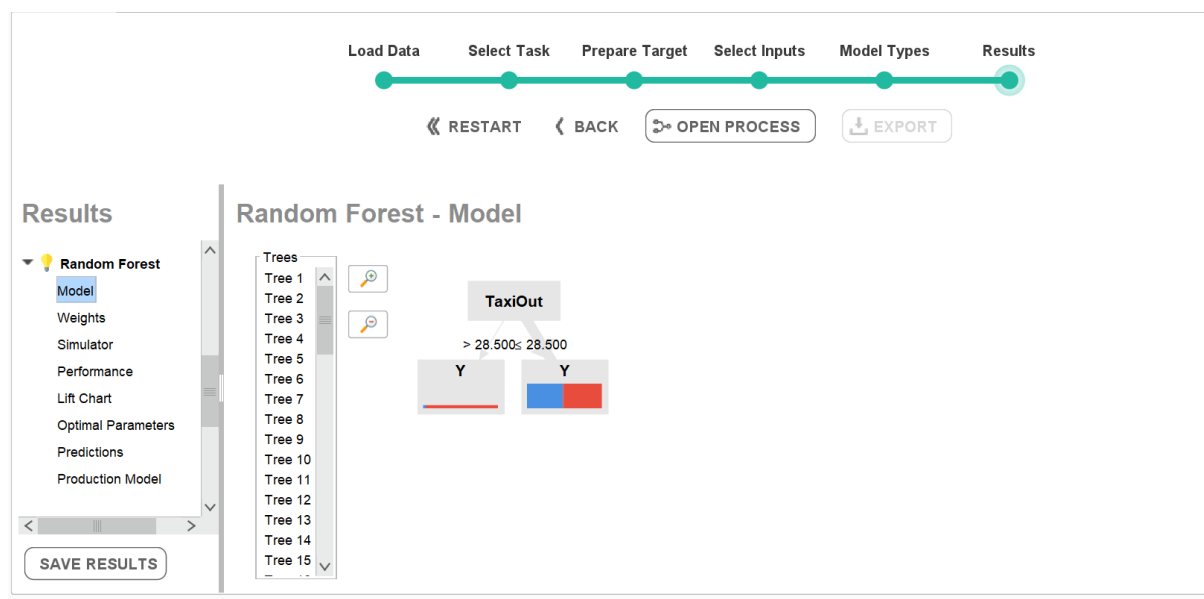
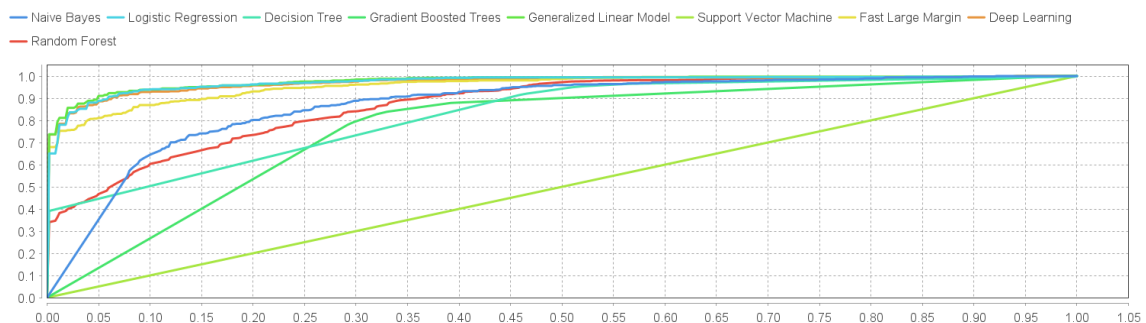


Initially, applied AutoModel to the Main dataset of 4k records:



ROC Comparison



Results

Random Forest

- Model
- Weights
- Simulator
- Performance
- Lift Chart
- Optimal Parameters
- Predictions
- Production Model

SAVE RESULTS

Random Forest - Weights

Attribute	Weight
DepDelay	0.223
ArrDelay	0.207
Year	0.097
TaxiOut	0.061
CRSDepTime	0.060
DepTime	0.040
ActualElapsedTime	0.036
Month	0.034

Random Forest - Simulator

ActualElapsedTime:

AirTime:

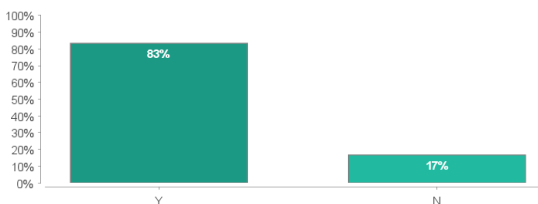
ArrDelay:

ArrTime:

CRSAirTime:

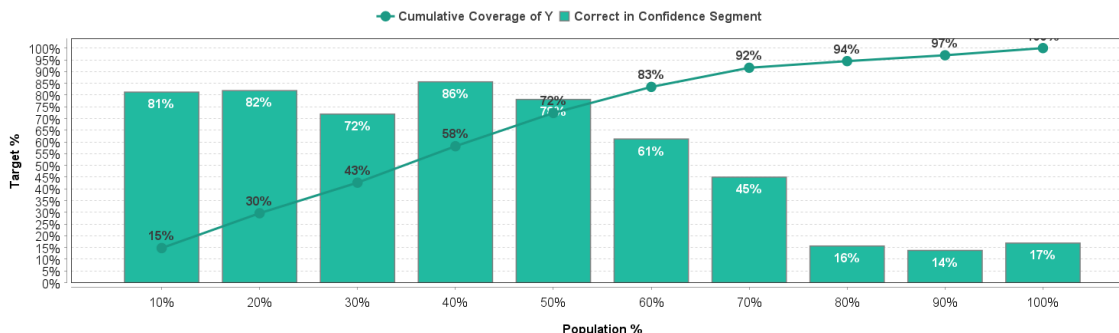
What is [this](#)?

Most Likely: Y



Important Factors for Y

Random Forest - Lift Chart

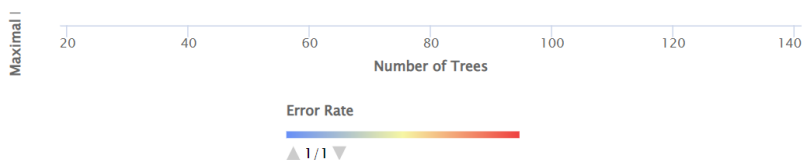


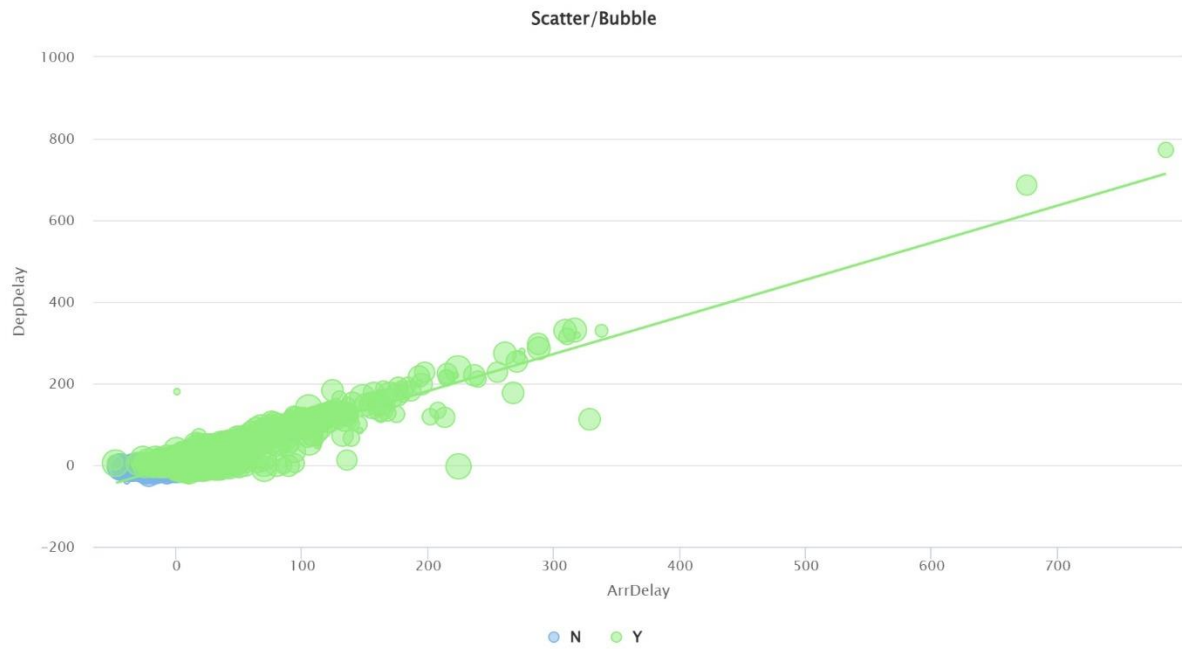
Random Forest - Optimal Parameters

Optimal Parameters

Number Of Trees: 60
Maximal Depth: 2

Error Rates for Parameters

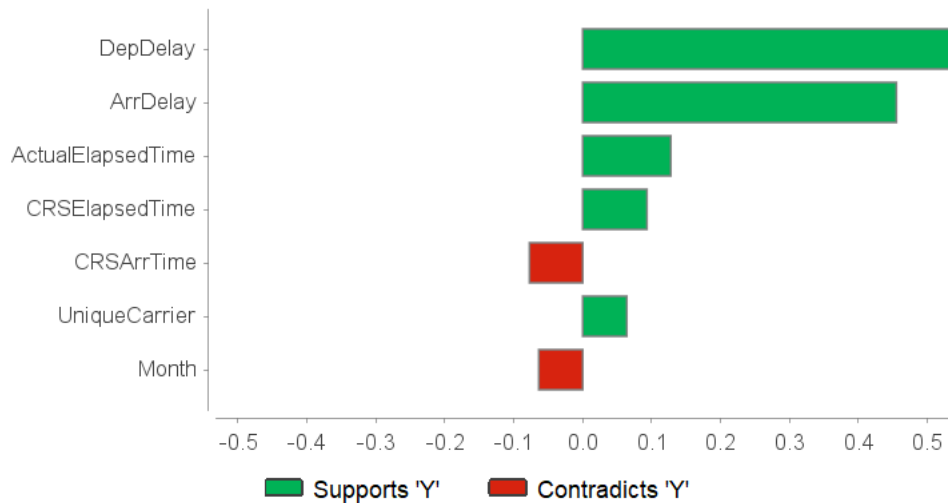




Confusion Matrix

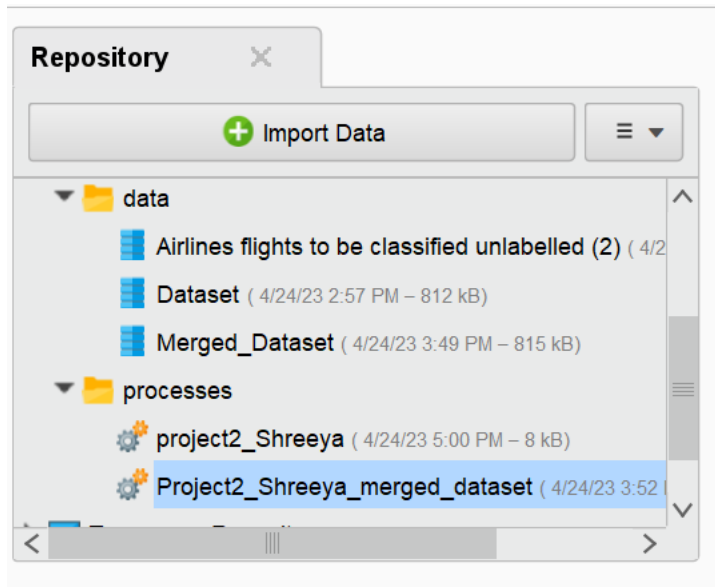
	true N	true Y	class precision
pred. N	16	0	100.00%
pred. Y	492	636	56.38%
class recall	3.15%	100.00%	

Important Factors for Y

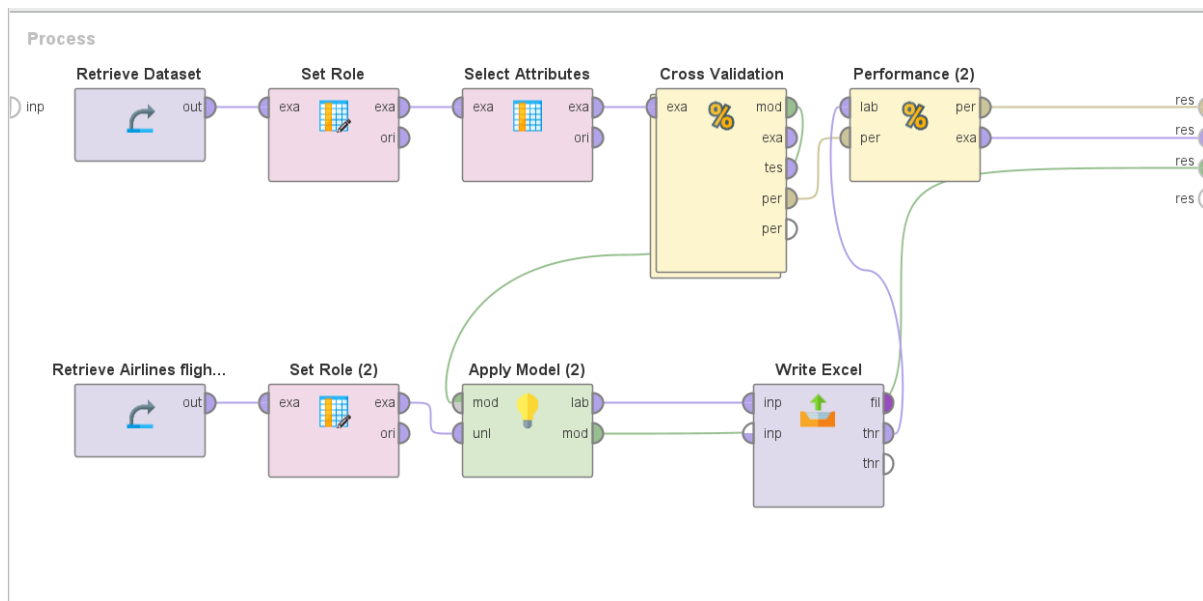


Below are the Data files and processes used.

- **Dataset:** Main dataset with 4k records (1k records each year) for years 2005-2008
- **Airlines flights to be classified unlabelled (2):** Testing dataset provided.
- **Merged_Dataset:** Dataset with initial 4k records and additional predicted 12 records.



The main process (process name: project2_Shreeya)



Views: DESIGN RESULTS LUDO PREP AUTO MODEL

Edit Parameter List: set roles

Edit Parameter List: **set roles**
This parameter defines new attribute roles.

attribute name	target role
Delay	label

Buttons: Add Entry Remove Entry Apply Cancel

Views: DESIGN RESULTS LUDO PREP AUTO MODEL

Select Attributes: select subset

Select Attributes: **select subset**
Click to select the attribute subset.

Attributes

Search

- # Cancelled
- # CarrierDelay
- # DayofMonth
- # DayOfWeek
- # Delay
- # DepDelay
- # Dest
- # Distance
- # Diverted
- # FlightNum
- # LateAircraftDelay
- # Month
- # NASDelay
- # Origin
- # SecurityDelay
- # TailNum
- # TaxiIn
- # TaxiOut

Selected Attributes

Search

- # ActualElapsedTime
- # AirTime
- # ArrDelay
- # ArrTime
- # CRSArrTime
- # CRSDepTime
- # CRSElapsedTime
- # DepTime

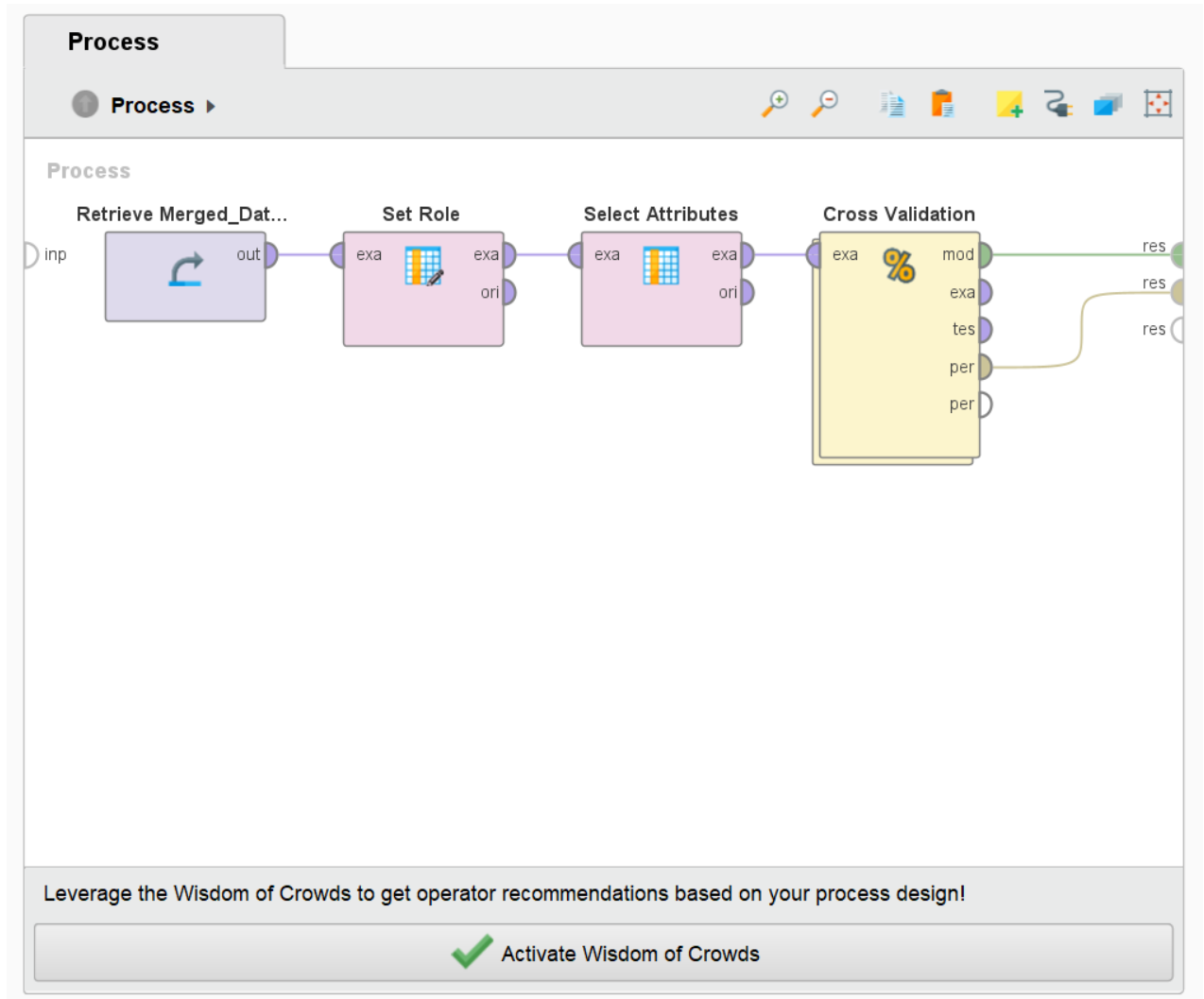
Buttons: Apply Cancel

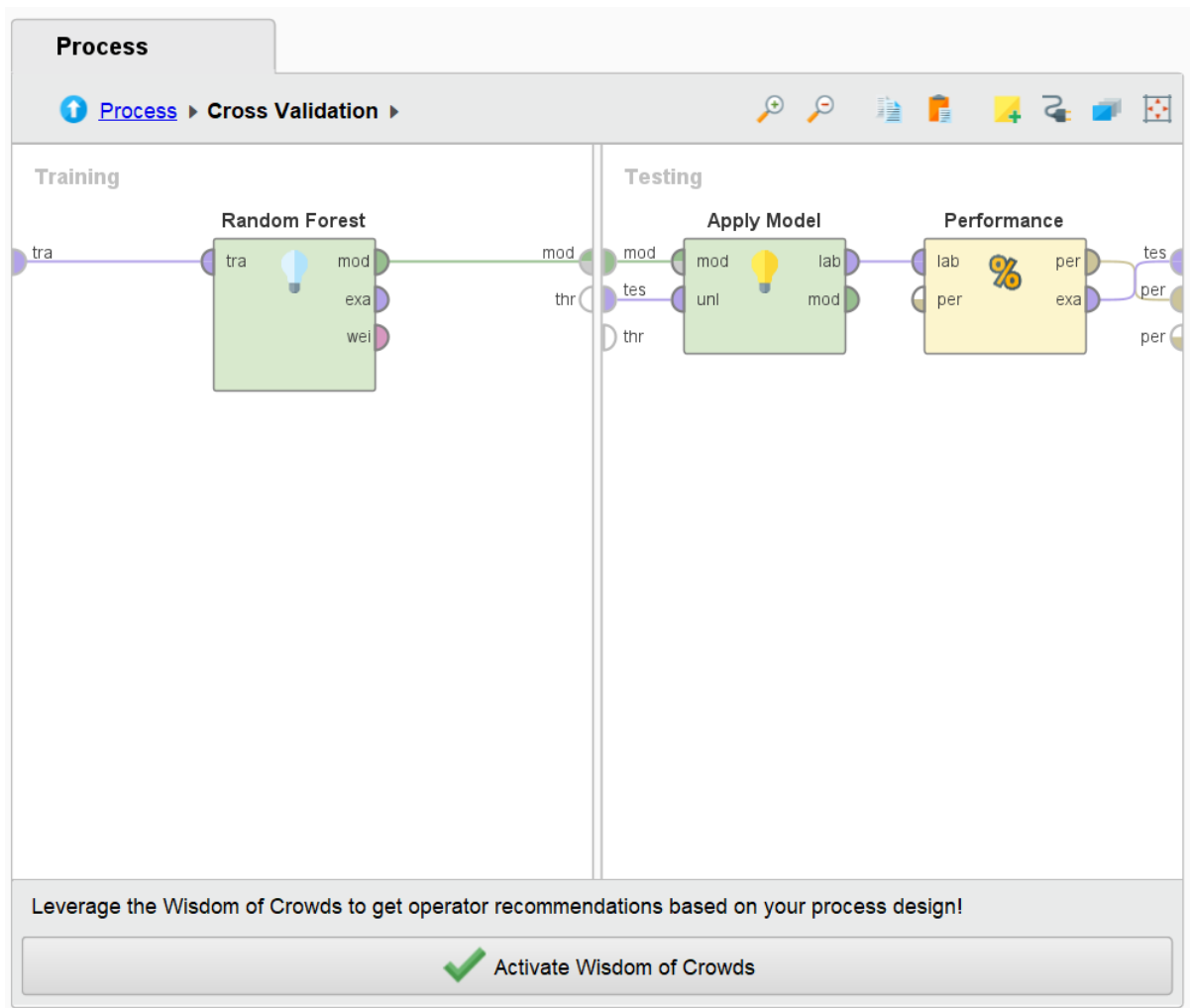
ExampleSet (Apply Model (2))		PerformanceVector (Performance)							
Result History		Random Forest Model (Random Forest)							
<div> <div>Data</div> <div>Statistics</div> <div>Visualizations</div> <div>Annotations</div> </div>	Open in Turbo Prep Auto Model		Filter (12 / 12 examples): all						
	Row No.	Delay	prediction(D...	confidence(...	confidence(...	Year	Month	DayofMonth	DayOfWe...
	1	?	N	0.700	0.300	?	1	30	6
	2	?	Y	0.153	0.847	?	2	6	6
	3	?	N	0.513	0.487	?	3	17	3
	4	?	N	0.701	0.299	?	4	24	6
	5	?	N	0.672	0.328	?	5	1	7
	6	?	N	0.672	0.328	?	6	7	2
	7	?	Y	0.184	0.816	?	7	14	4
	8	?	N	0.672	0.328	?	8	22	1
	9	?	N	0.700	0.300	?	9	12	3
	10	?	Y	0.153	0.847	?	10	14	7
	11	?	Y	0.153	0.847	?	11	6	2
	12	?	N	0.700	0.300	?	12	14	5
ExampleSet (12 examples, 4 special attributes, 29 regular attributes)									

ExampleSet (Apply Model (2))		PerformanceVector (Performance)																	
Result History		Random Forest Model (Random Forest)																	
<div> <div>Performance</div> <div>Description</div> <div>Annotations</div> </div>	Criterion	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View																	
	precision	precision: 99.72% +/- 0.39% (micro average: 99.73%) (positive class: Y)																	
	recall																		
	AUC (optimistic)																		
	AUC																		
	AUC (pessimistic)																		
		<table> <tr> <th></th><th>true N</th><th>true Y</th><th>class precision</th></tr> <tr> <td>pred. N</td><td>1791</td><td>381</td><td>82.46%</td></tr> <tr> <td>pred. Y</td><td>5</td><td>1823</td><td>99.73%</td></tr> <tr> <td>class recall</td><td>99.72%</td><td>82.71%</td><td></td></tr> </table>			true N	true Y	class precision	pred. N	1791	381	82.46%	pred. Y	5	1823	99.73%	class recall	99.72%	82.71%	
	true N	true Y	class precision																
pred. N	1791	381	82.46%																
pred. Y	5	1823	99.73%																
class recall	99.72%	82.71%																	

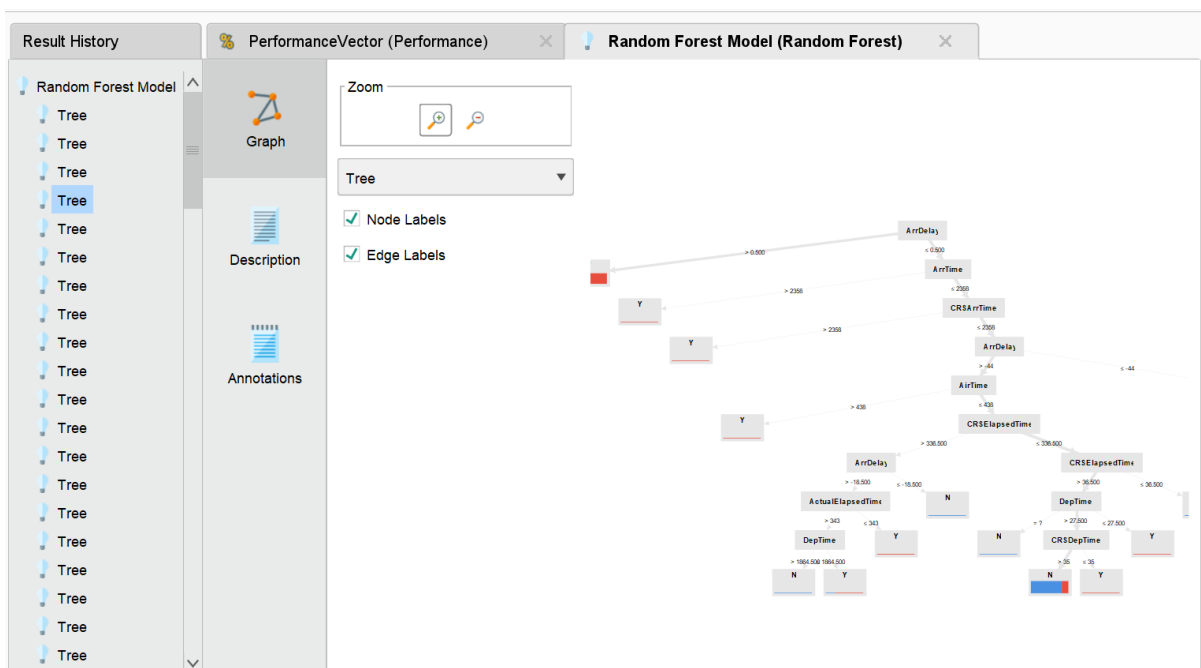
ExampleSet (Apply Model (2))		PerformanceVector (Performance)																	
Result History		Random Forest Model (Random Forest)																	
<div> <div>Performance</div> <div>Description</div> <div>Annotations</div> </div>	Criterion	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View																	
	precision																		
	recall	recall: 82.71% +/- 2.54% (micro average: 82.71%) (positive class: Y)																	
	AUC (optimistic)																		
	AUC																		
	AUC (pessimistic)																		
		<table> <tr> <th></th><th>true N</th><th>true Y</th><th>class precision</th></tr> <tr> <td>pred. N</td><td>1791</td><td>381</td><td>82.46%</td></tr> <tr> <td>pred. Y</td><td>5</td><td>1823</td><td>99.73%</td></tr> <tr> <td>class recall</td><td>99.72%</td><td>82.71%</td><td></td></tr> </table>			true N	true Y	class precision	pred. N	1791	381	82.46%	pred. Y	5	1823	99.73%	class recall	99.72%	82.71%	
	true N	true Y	class precision																
pred. N	1791	381	82.46%																
pred. Y	5	1823	99.73%																
class recall	99.72%	82.71%																	

The second process to get accuracy, precision and recall value of the merged data





The output of second process:



Views: Design Results Turbo Prep Auto Model Find

Result History

PerformanceVector (Performance)

Random Forest Model (Random Forest)

Performance

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Description

Annotations

Table View

Plot View

accuracy: 90.40% +/- 1.08% (micro average: 90.40%)

	true N	true Y	class precision
pred. N	1799	380	82.56%
pred. Y	5	1828	99.73%
class recall	99.72%	82.79%	

Views: Design Results Turbo Prep Auto Model Find d

Result History

PerformanceVector (Performance)

Random Forest Model (Random Forest)

Performance

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Description

Annotations

Table View

Plot View

precision: 99.72% +/- 0.39% (micro average: 99.73%) (positive class: Y)

	true N	true Y	class precision
pred. N	1799	380	82.56%
pred. Y	5	1828	99.73%
class recall	99.72%	82.79%	

Views: Design Results Turbo Prep Auto Model Find d

Result History

PerformanceVector (Performance)

Random Forest Model (Random Forest)

Performance

Criterion

accuracy

precision

recall

AUC (optimistic)

AUC

AUC (pessimistic)

Description

Annotations

Table View

Plot View

recall: 82.79% +/- 1.78% (micro average: 82.79%) (positive class: Y)

	true N	true Y	class precision
pred. N	1799	380	82.56%
pred. Y	5	1828	99.73%
class recall	99.72%	82.79%	