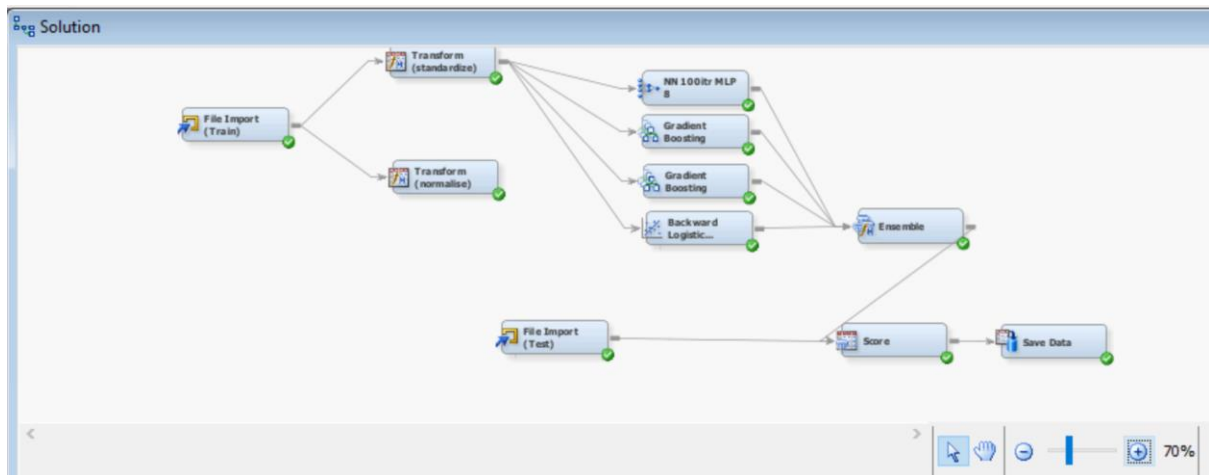# Project 1

We used an ensemble model for this with NN, 2 GB models, and 1 backward logistic regression model

Diagram - Solution

Transform Variable: Standardize

NN:

Train

Variables	...
Continue Training	No
Network	...
Optimization	...
Initialization Seed	1908
Model Selection Criterion	Misclassification
Suppress Output	No

Score

Hidden Units	No
Residuals	Yes
Standardization	Yes

Status

Create Time	11/28/23 3:47 PM

General

General Properties

GB 1:

..	Property	Value
	Variables	...
	Series Options	
	N Iterations	100
	Seed	1512
	Shrinkage	0.1
	Train Proportion	60
	Splitting Rule	
	Huber M-Regression	No
	Maximum Branch	2
	Maximum Depth	5
	Minimum Categorical Size	5
	Reuse Variable	5
	Categorical Bins	30
	Interval Bins	100
	Missing Values	Use in search
	Performance	Disk
	Node	
	Leaf Fraction	0.001
	Number of Surrogate Rules	0
	Split Size	.
	Split Search	

General

Diagram Solution opened

GB 2:

.. Property	Value	
Exported Data		...
Notes		...
Train		
Variables		...
☐ Series Options		
N Iterations	100	
Seed	1512	
Shrinkage	0. 1	
Train Proportion	70	
☐ Splitting Rule		
Huber M-Regression	No	
Maximum Branch	2	
Maximum Depth	5	
Minimum Categorical Size	5	
Reuse Variable	3	
Categorical Bins	30	
Interval Bins	100	
Missing Values	Use in search	
Performance	Disk	
☐ Node		
Leaf Fraction	0. 001	

General

Diagram Solution opened

Sample

Solut

Diagr

Logistic Regression – Backward :

Property	Value
Variables	...
⊟ Equation	
Main Effects	Yes
Two-Factor Interactions	No
Polynomial Terms	No
Polynomial Degree	2
User Terms	No
Term Editor	...
⊟ Class Targets	
Regression Type	Logistic Regression
Link Function	Logit
⊟ Model Options	
Suppress Intercept	No
Input Coding	Deviation
⊟ Model Selection	
Selection Model	Backward
Selection Criterion	Cross Validation Misclassifica
Use Selection Defaults	Yes
Selection Options	...
⊟ Optimization Options	
Technique	Default

General

Diagram Solution opened

Ensemble:

..	Property	Value
	General	
	Node ID	Ensmbl
	Imported Data	...
	Exported Data	...
	Notes	...
	Train	
	Variables	...
⊟	Interval Target	
	Predicted Values	Average
⊟	Class Target	
	Posterior Probabilities	Average
	Voting Posterior Probabilitie	Average
	Status	
	Create Time	11/28/23 5:36 PM
	Run ID	e871d916-348e-4a81-bfe4
	Last Error	
	Last Status	Complete
	Last Run Time	11/29/23 3:43 PM
	Run Duration	0 Hr. 0 Min. 33.82 Sec.
	Grid Host	
	User-Added Node	No

Posterior Probabilities

Diagram Solution opened

Score:

Data Mining Project

.. Property	Value
Exported Data	...
Notes	...
Train	
Variables	...
Type of Scored Data	Data
Use Fixed Output Names	Yes
Hide Variables	No
Hide Selection	...
Score Data	
Validation	No
Test	No
Score Code Generation	
Optimized Code	Yes
C Score	No
Java Score	No
Java Package Name	Default
User Package Name	
Report	
Graphical Reports	Yes
Status	
Create Time	11/25/23 4:23 PM

General

Sample   Explore

Solution

Diagram

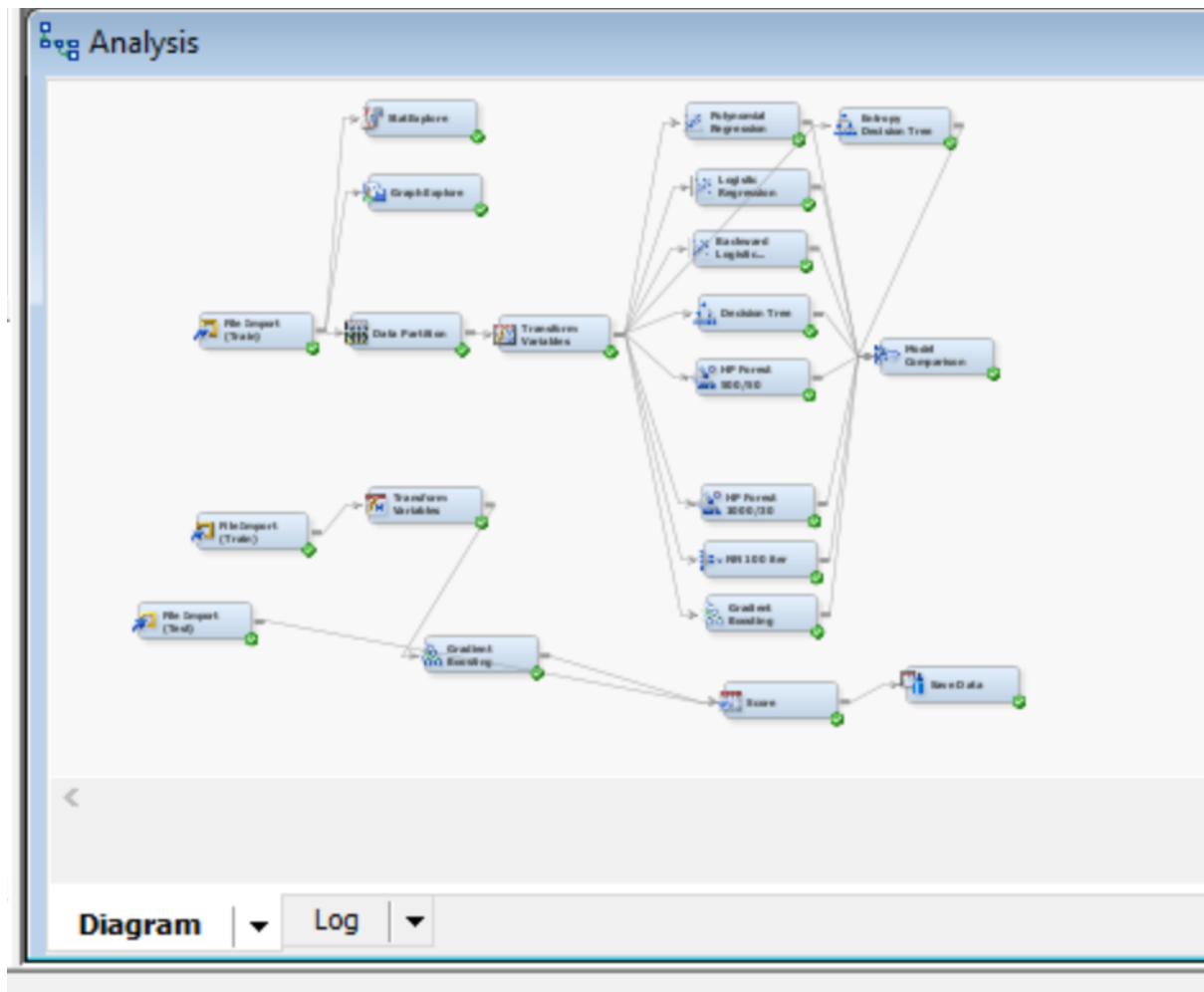Diagram Solution opened

For the second project, we tried many different models and then went ahead which had the highest ROC in validation (gradient boosting)

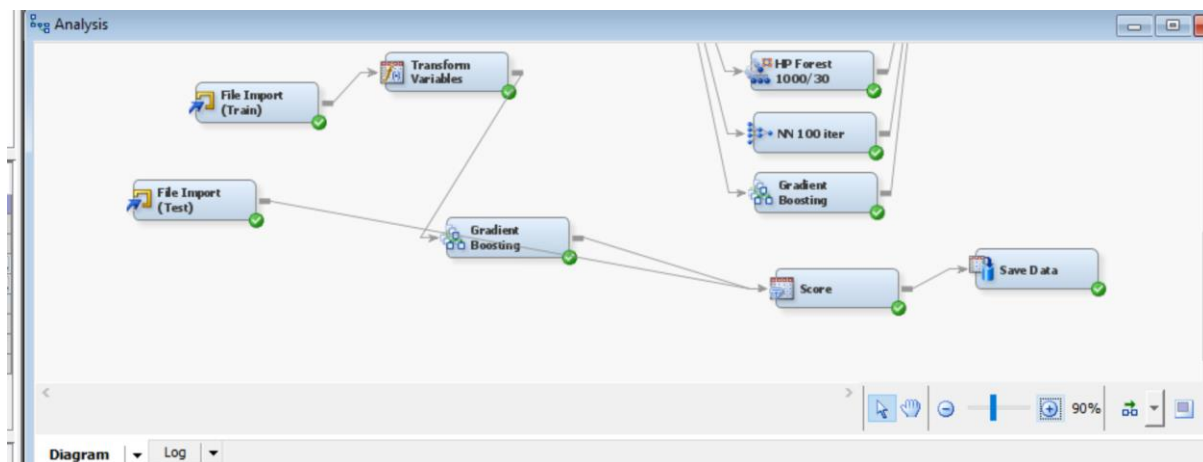**This was the final model selected:**

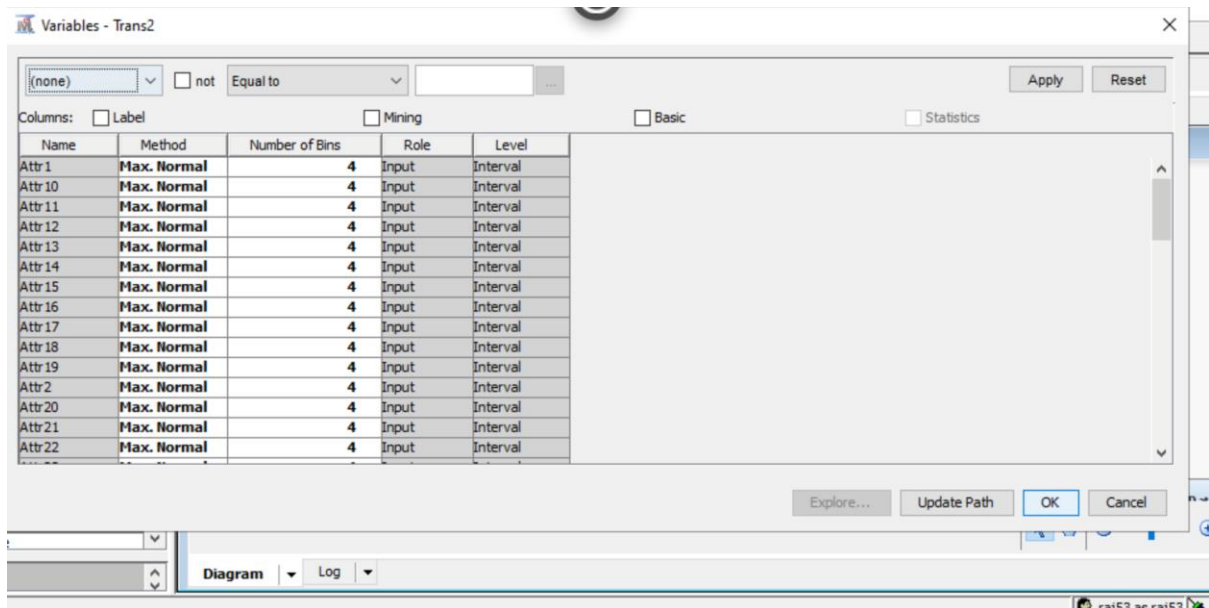Results - Node: Model Comparison Diagram: Analysis

File Edit View Window

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Roc Index	Train: Sum of Frequencies	Train: Sum of Case Weights Times Freq	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Tra Av Sq Err
Y	Boost	Boost	Gradient Bo...	class		0.913	8000	16000	0.009875	0.99259	116.6336	
	HPDMForest	HPDMForest	HP Forest 5...	class		0.896	8000	.	0	0.496	46.90946	
	Neural2	Neural2	NN 100 iter	class		0.892	8000	16000	0.015625	0.995647	216.2475	
	HPDMFore...	HPDMFore...	HP Forest 1...	class		0.891	8000	.	0.00025	0.545	47.9278	
	Reg2	Reg2	Backward L...	class		0.876	8000	16000	0.018125	0.999999	252.625	
	Reg	Reg	Logistic Re...	class		0.875	8000	16000	0.01825	0.999999	253.0986	
	Reg3	Reg3	Polynomial ...	class		0.861	8000	16000	0.0205	0.999239	287.6408	
	Tree2	Tree2	Entropy De...	class		0.776	8000	.	0.0165	0.999107	237.3799	
	Tree	Tree	Decision Tr...	class		0.761	8000	.	0.018875	0.998715	273.7841	

raj53 as raj53

Transformation – Max Normal

9

GB paramaters:

Property	Value
General	
Node ID	Boost2
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
☐ Series Options	
N Iterations	120
Seed	1512
Shrinkage	0.1
Train Proportion	60
☐ Splitting Rule	
Huber M-Regression	No
Maximum Branch	2
Maximum Depth	5
Minimum Categorical Size	5
Reuse Variable	3
Categorical Bins	30
Interval Bins	100
Missing Values	Use in search

General

Diagram Analysis opened