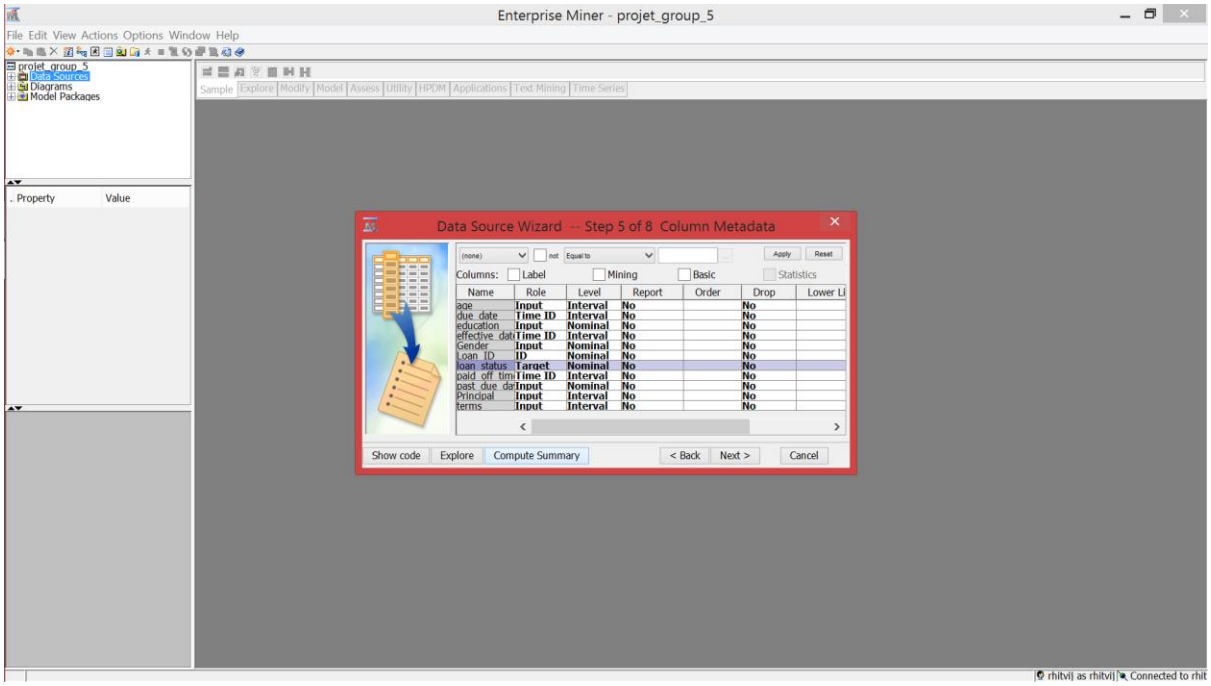
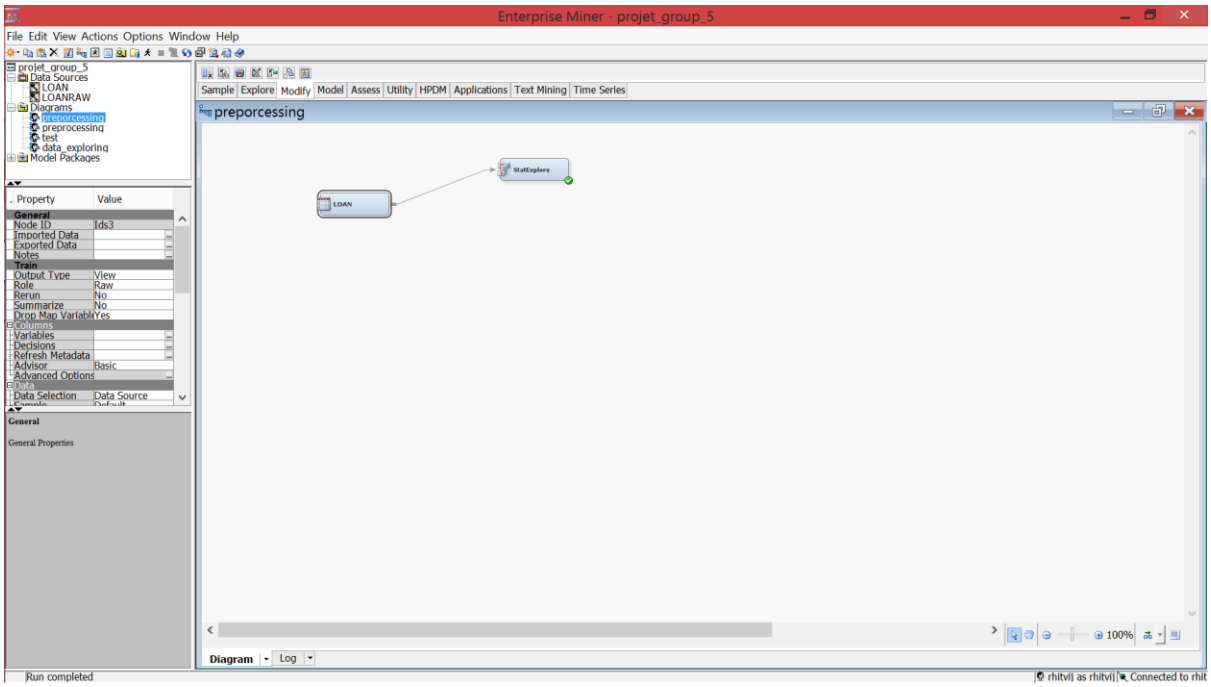
Data Pre Processing

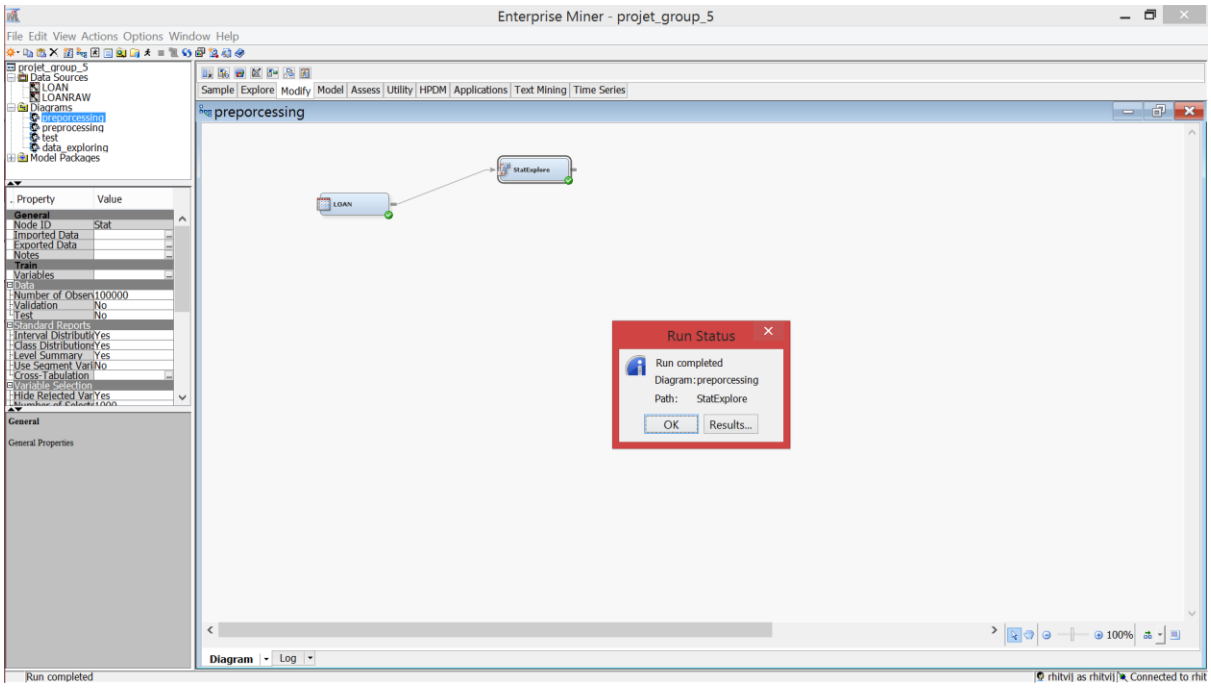
Screen shot number 1



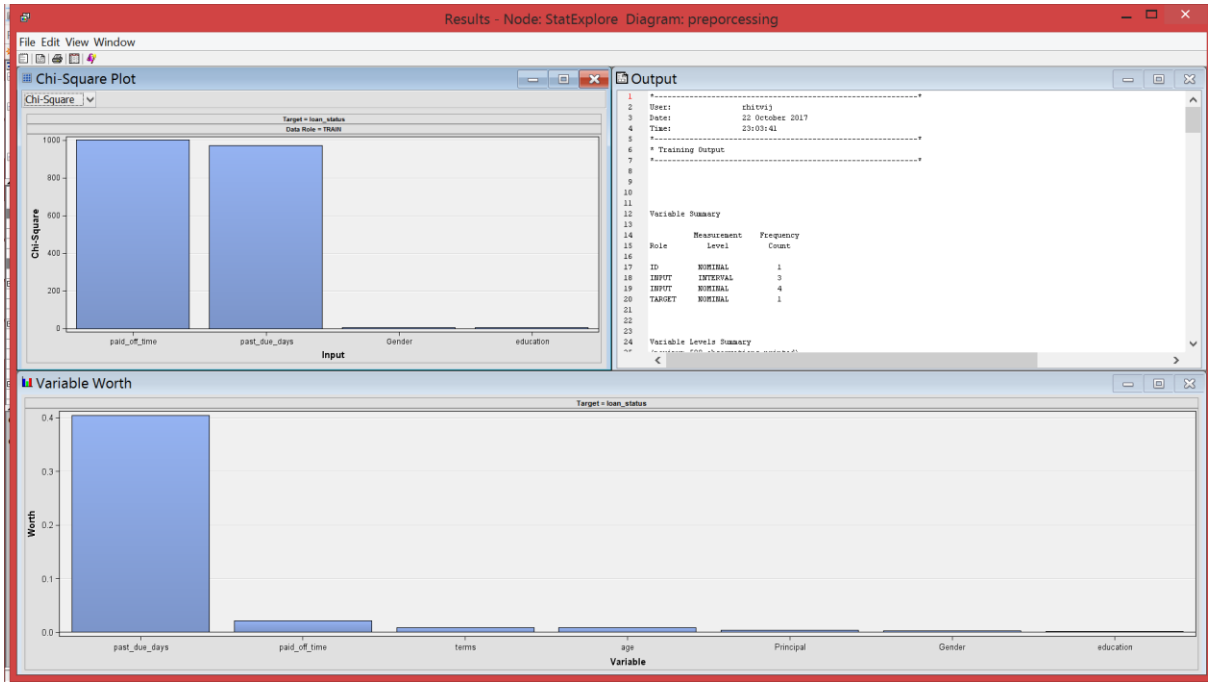
Screen shot number 2



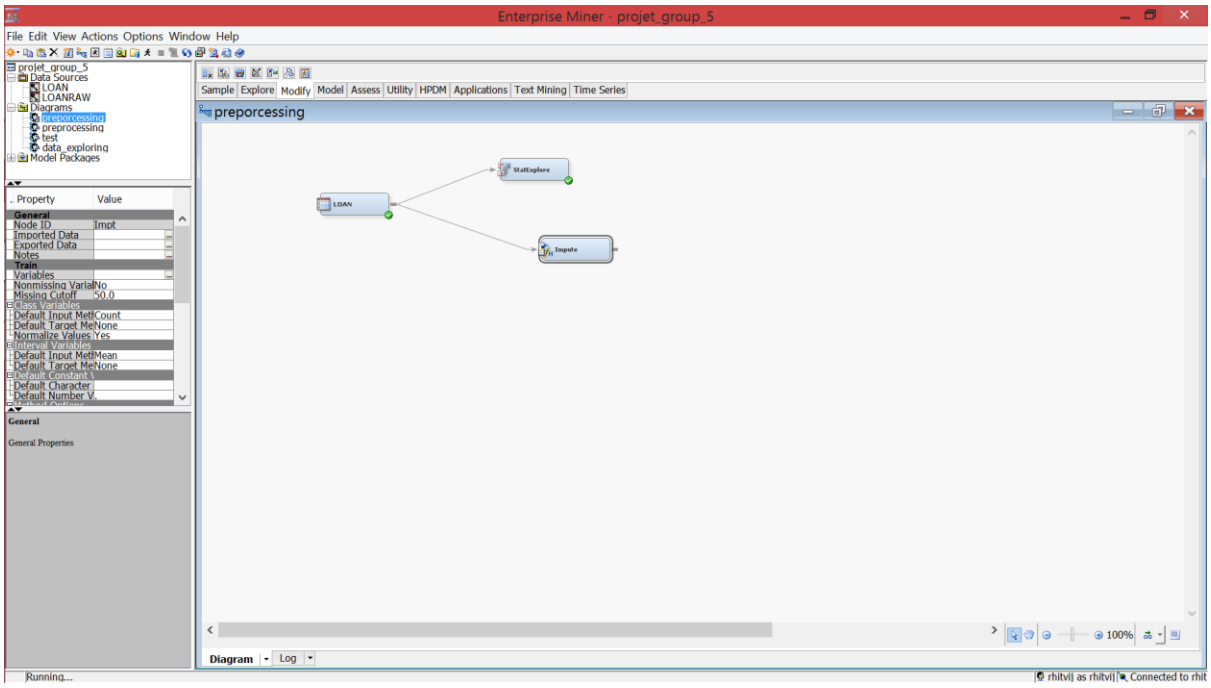
Screen shot number 3



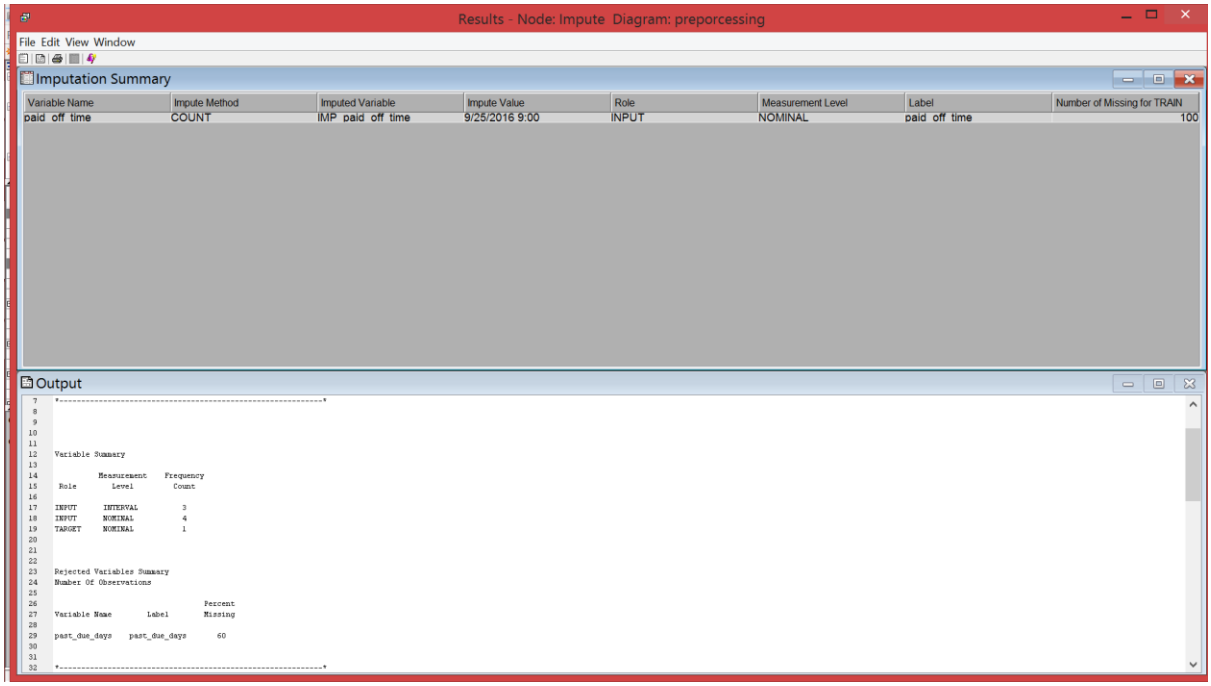
Screen shot 4



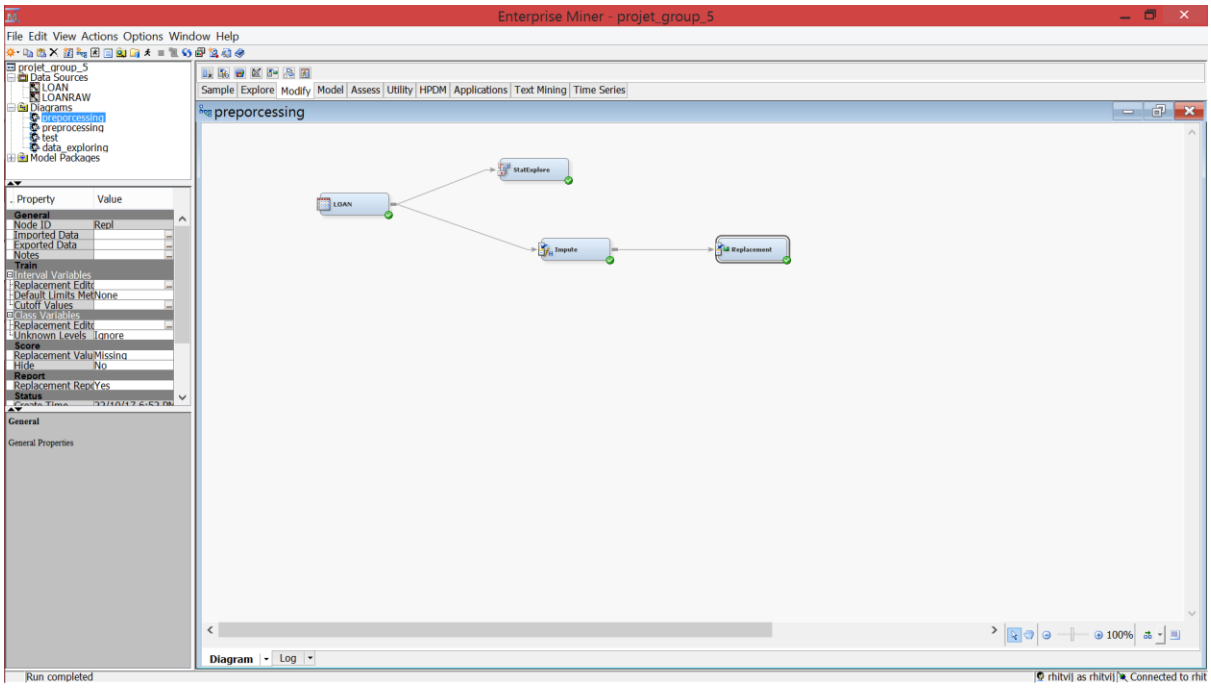
Screen shot 5



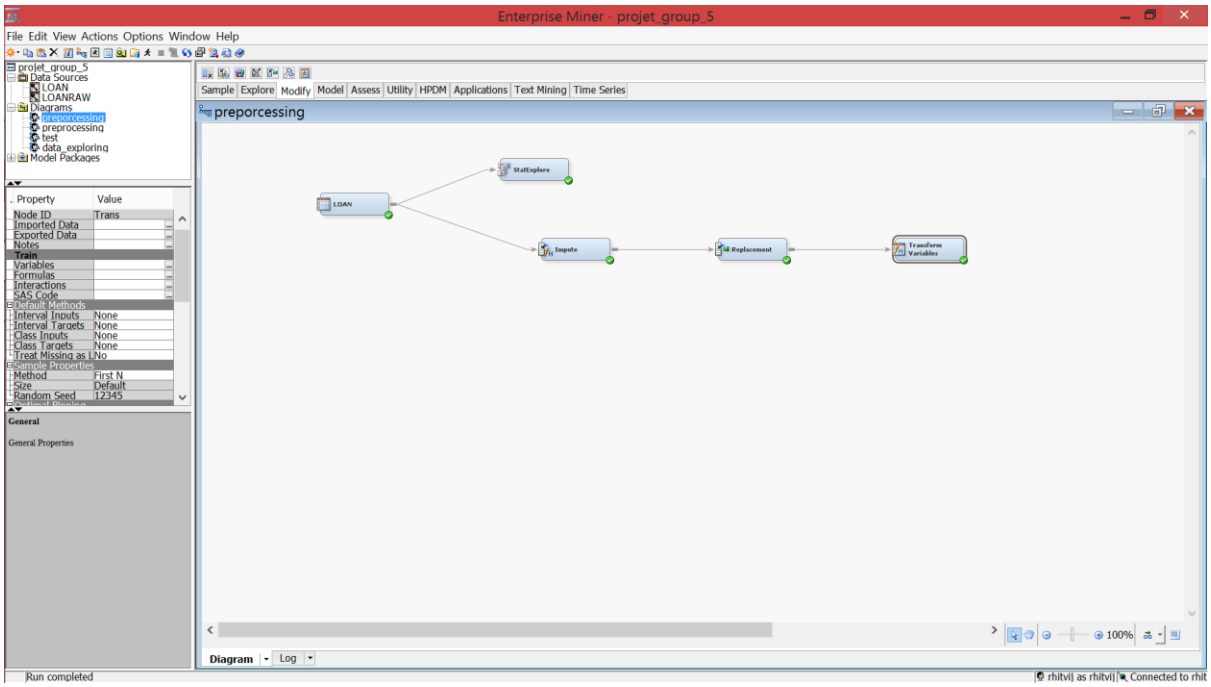
Screen shot 6



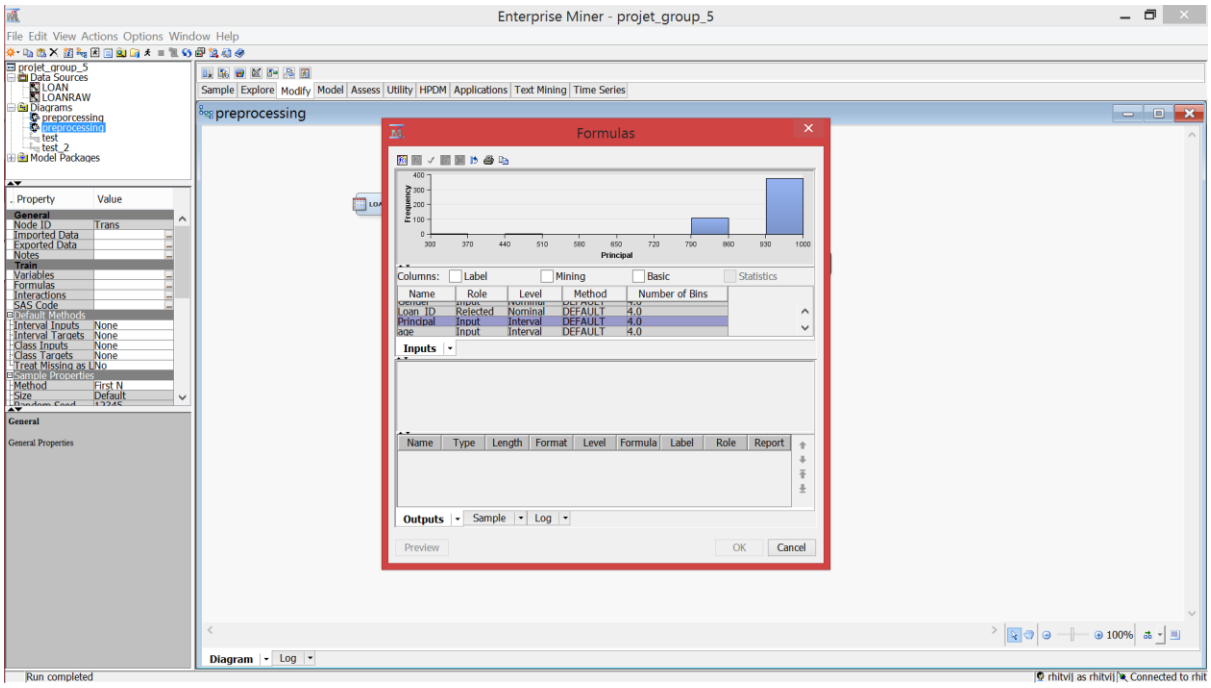
Screen shot 7



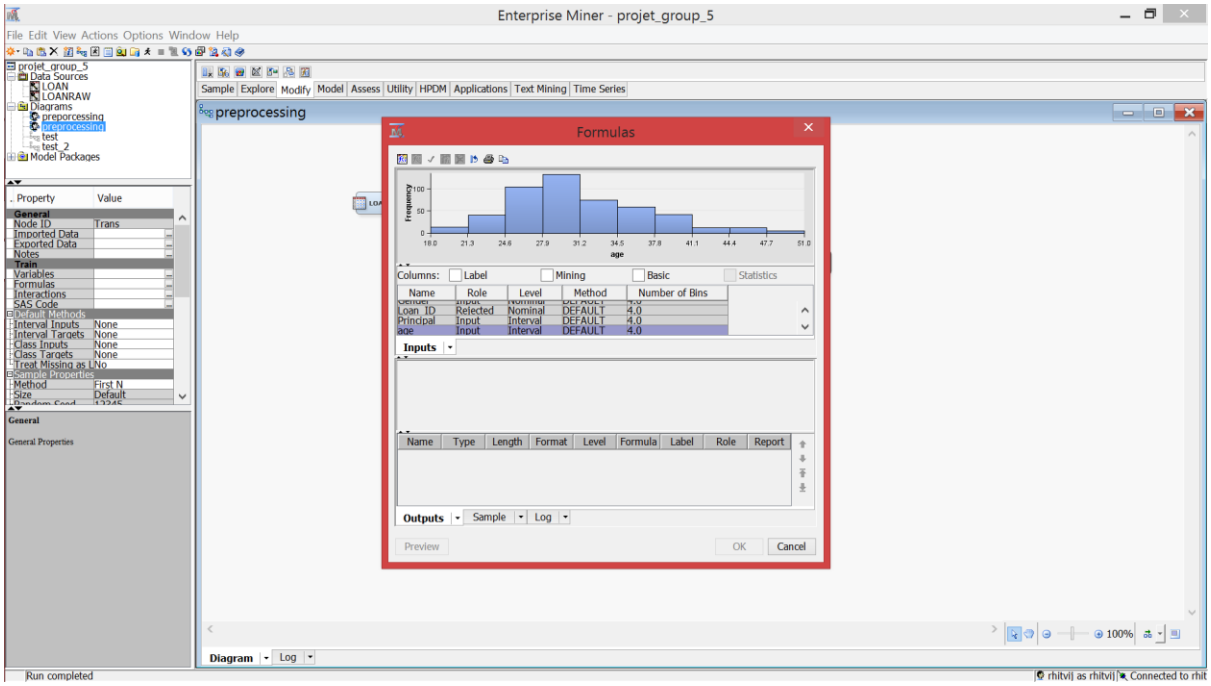
Screen shot 8



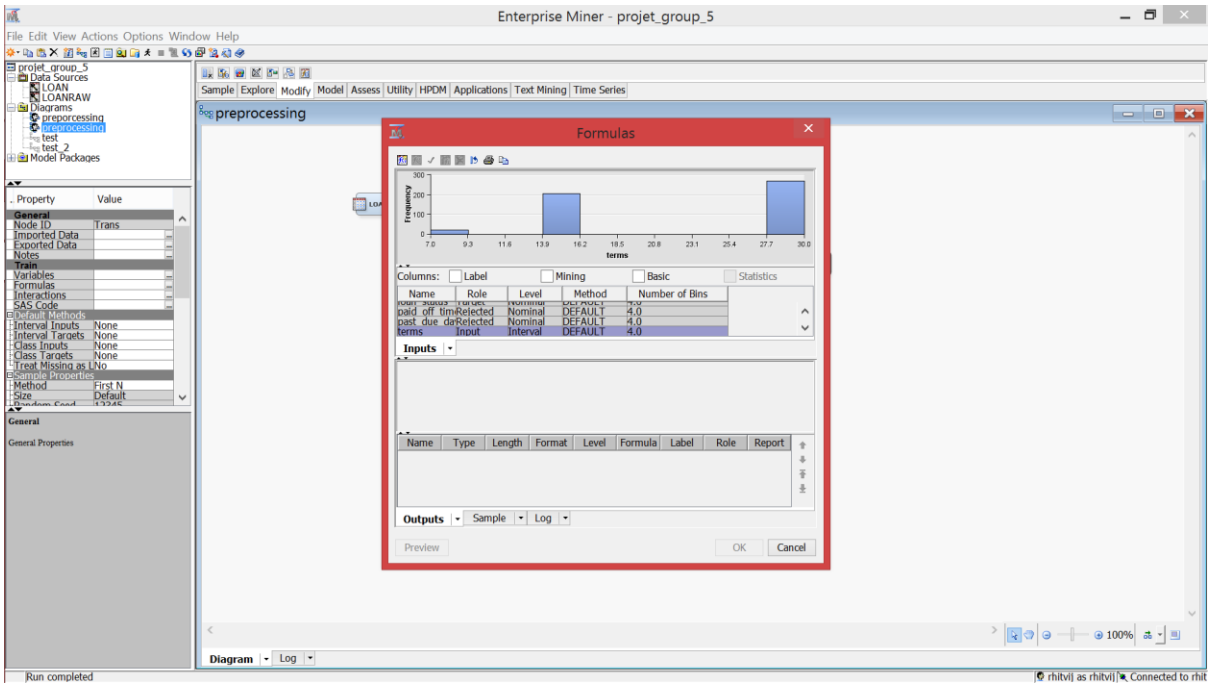
Screen shot 9



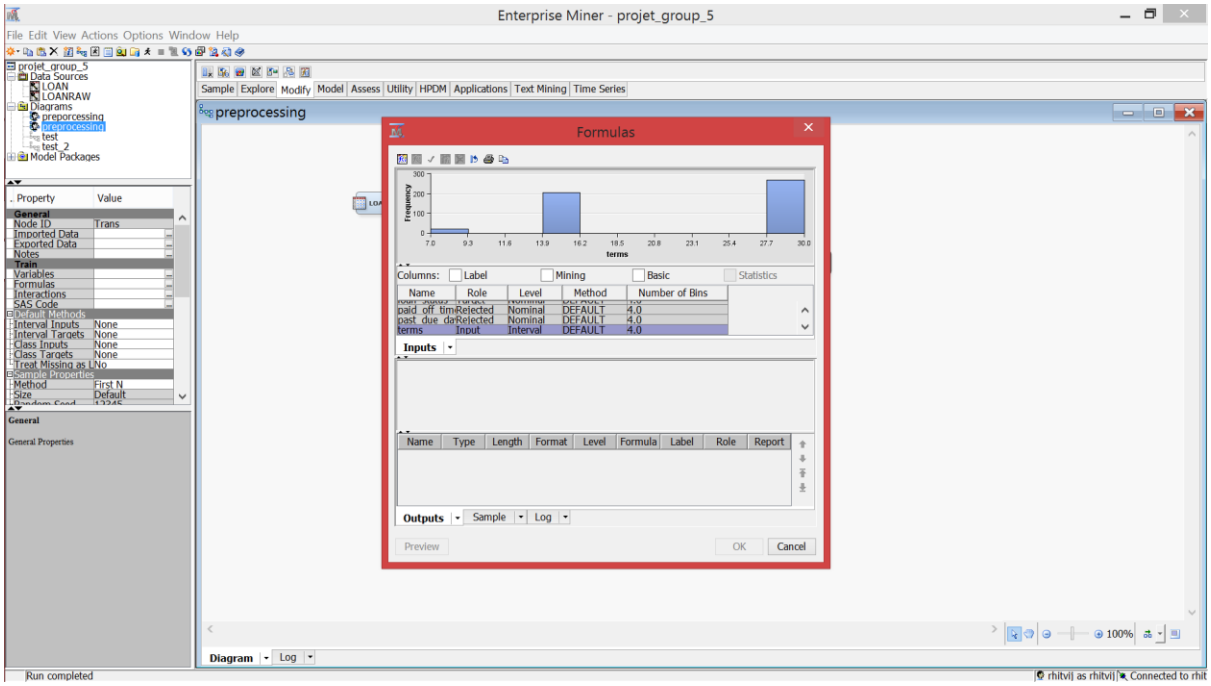
Screen shot 10



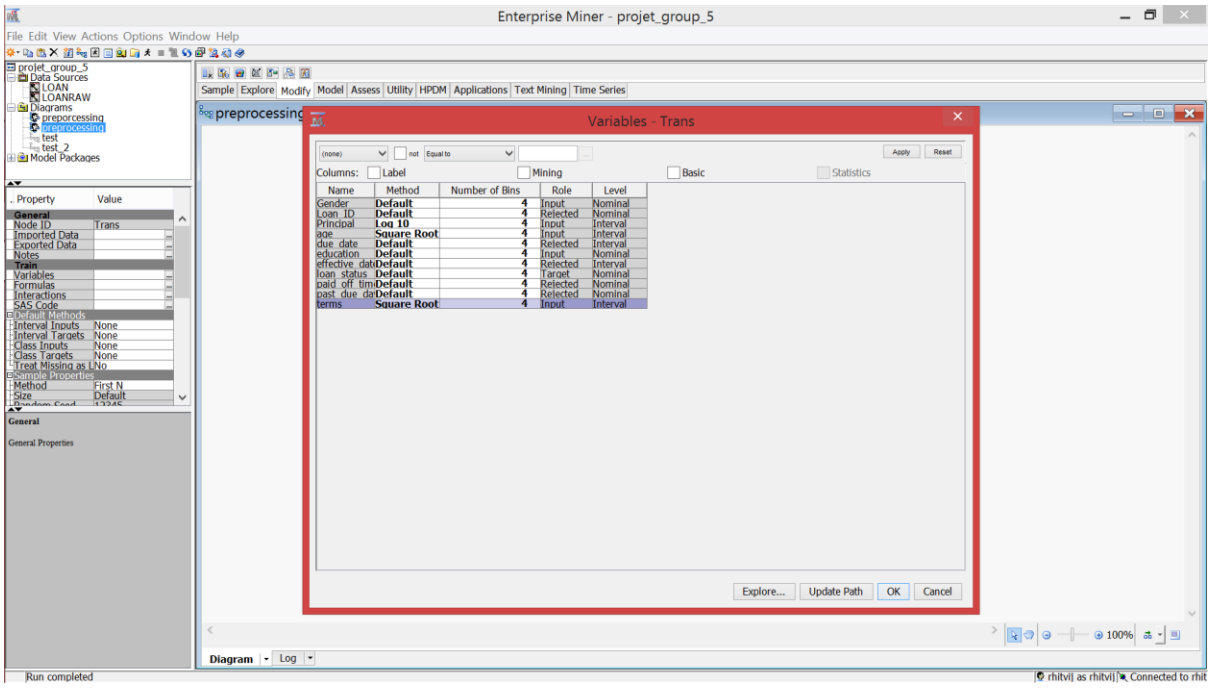
Screen shot 11



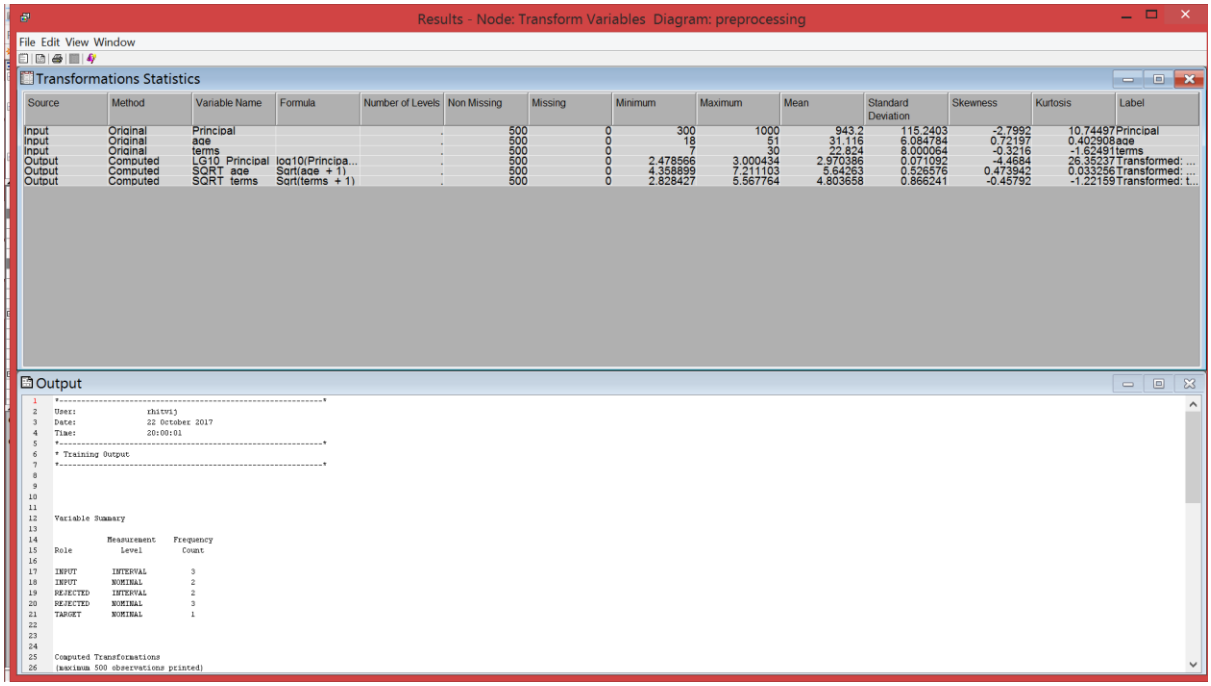
Screen shot 12



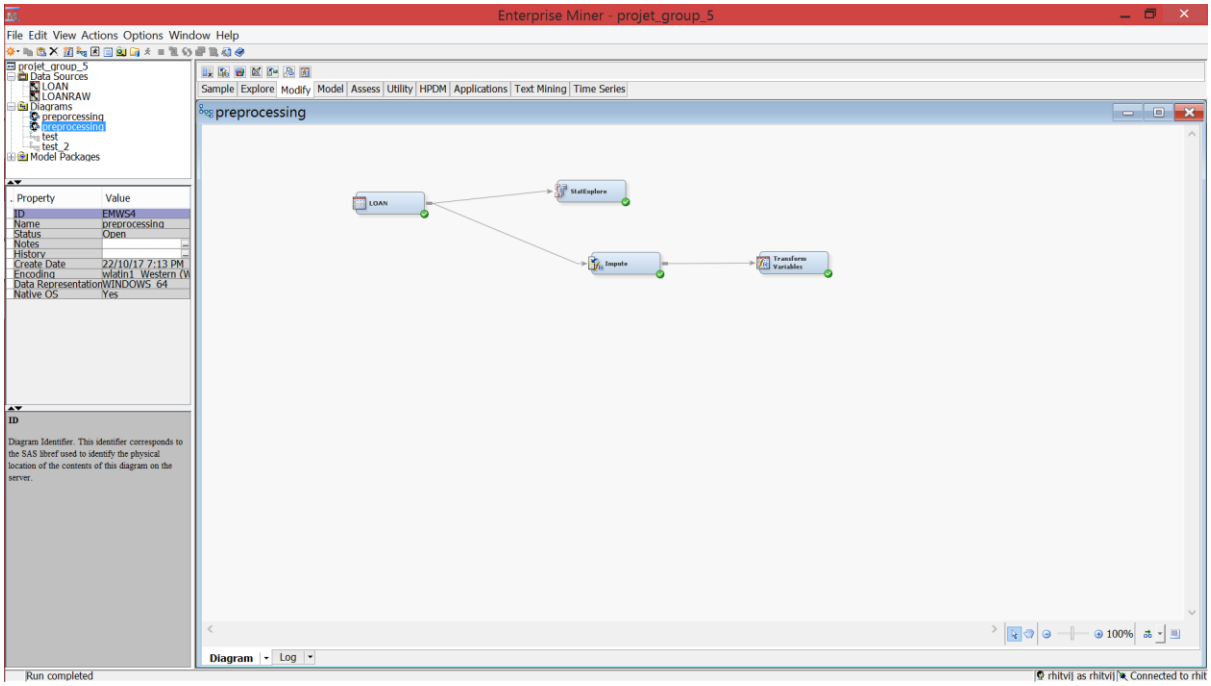
Screen shot 13



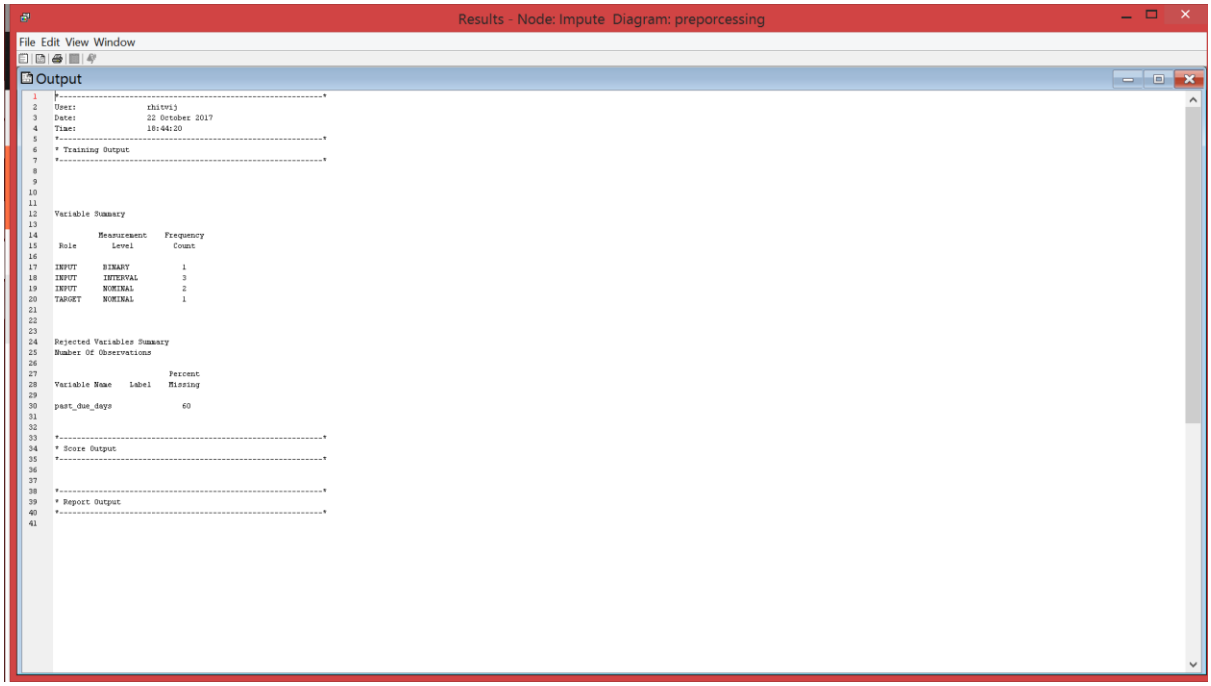
Screen shot 14



Screen shot 15

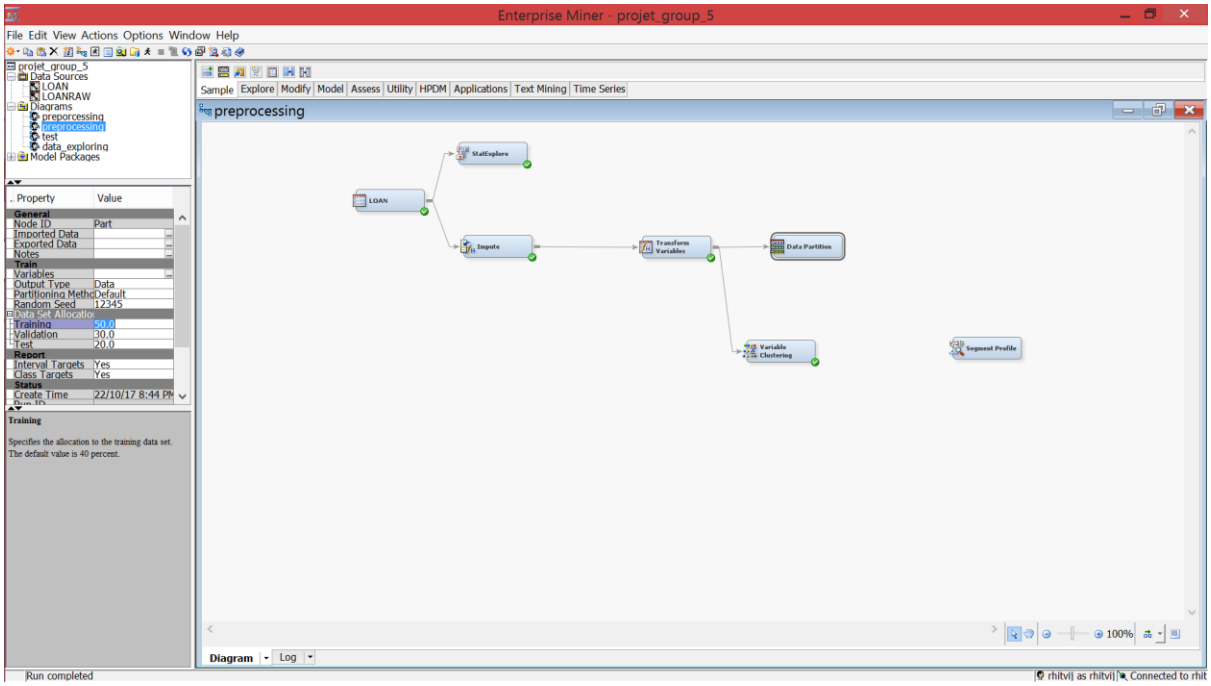


Screen shot 16

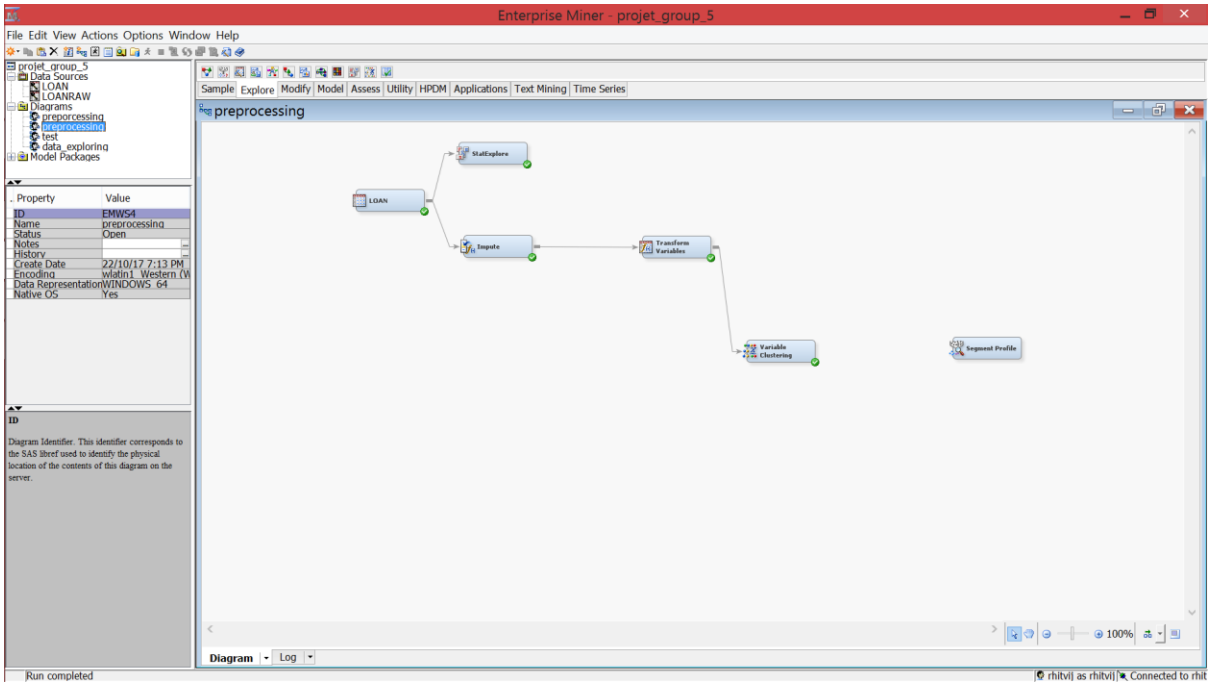


Data Exploring

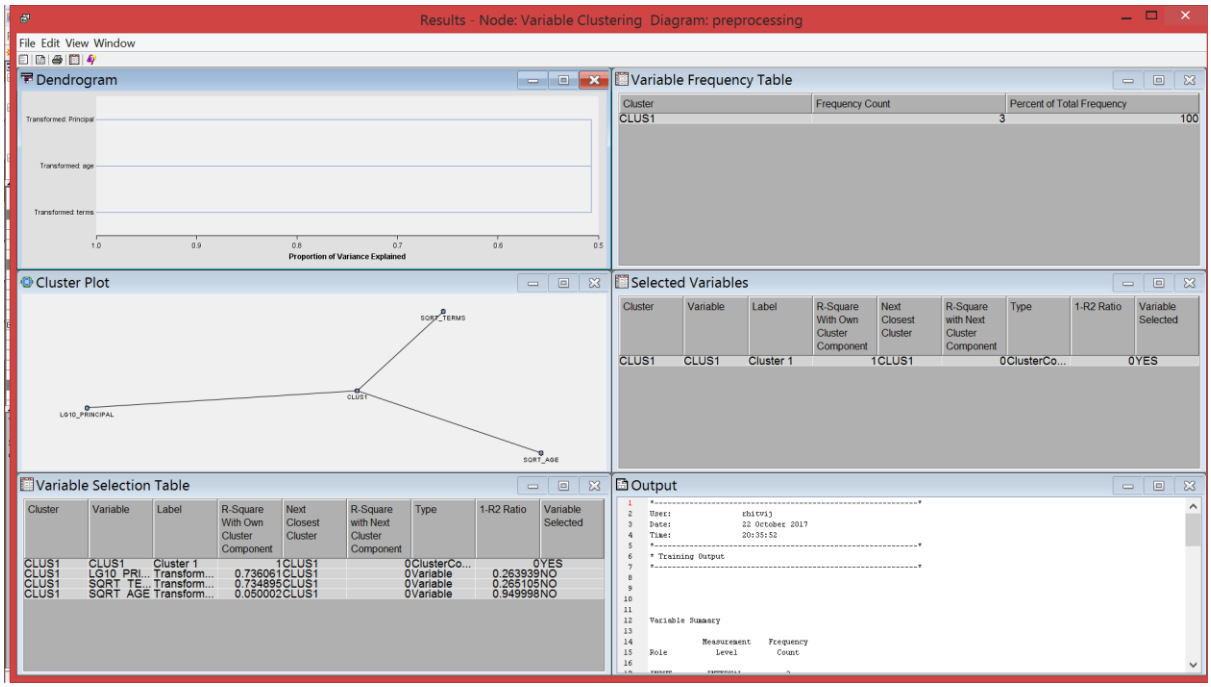
Screen shot number 20



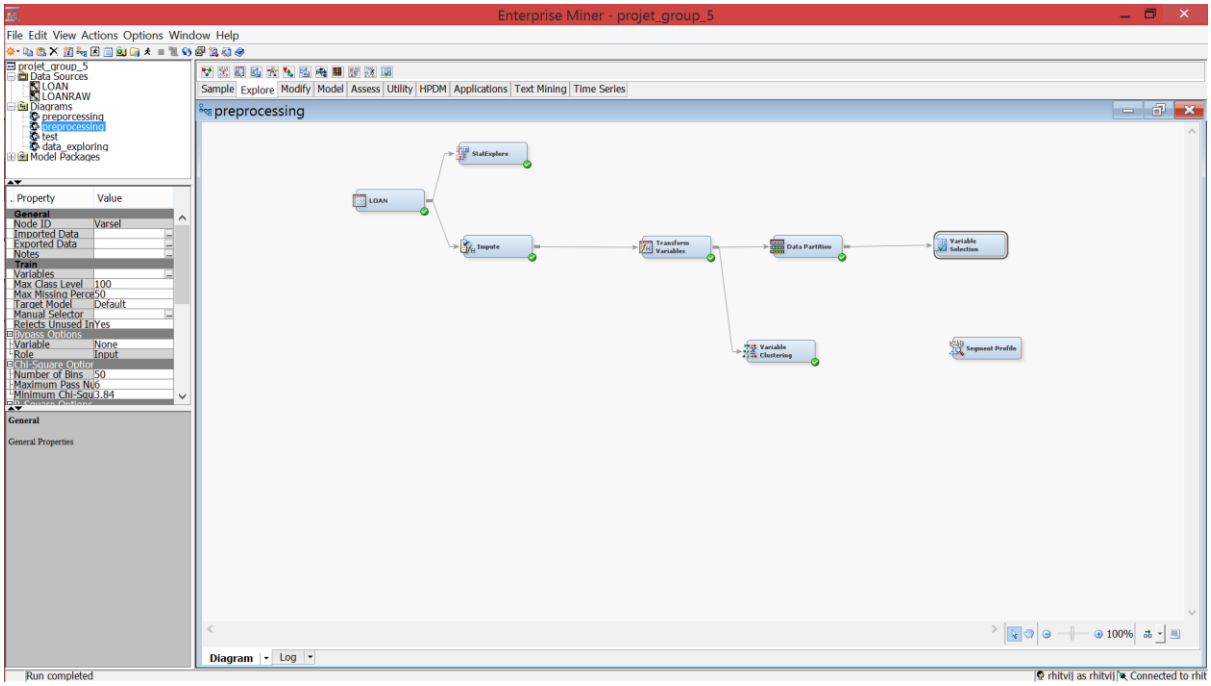
Screen shot number 21



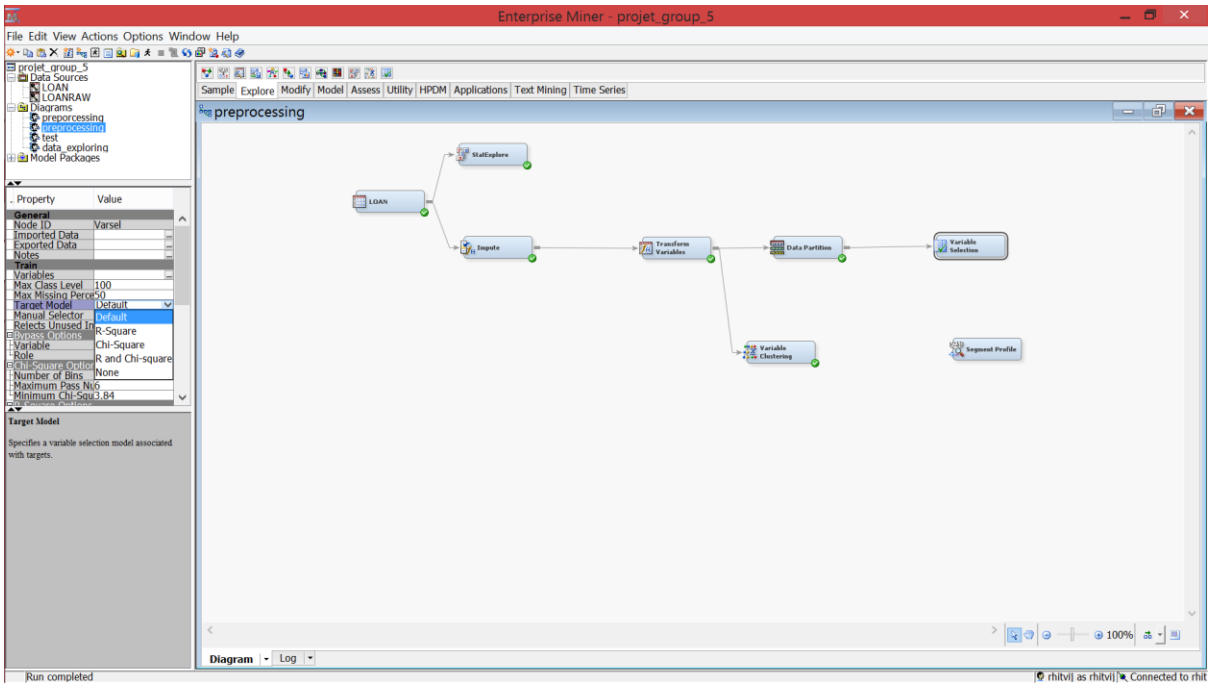
Screen shot number 22



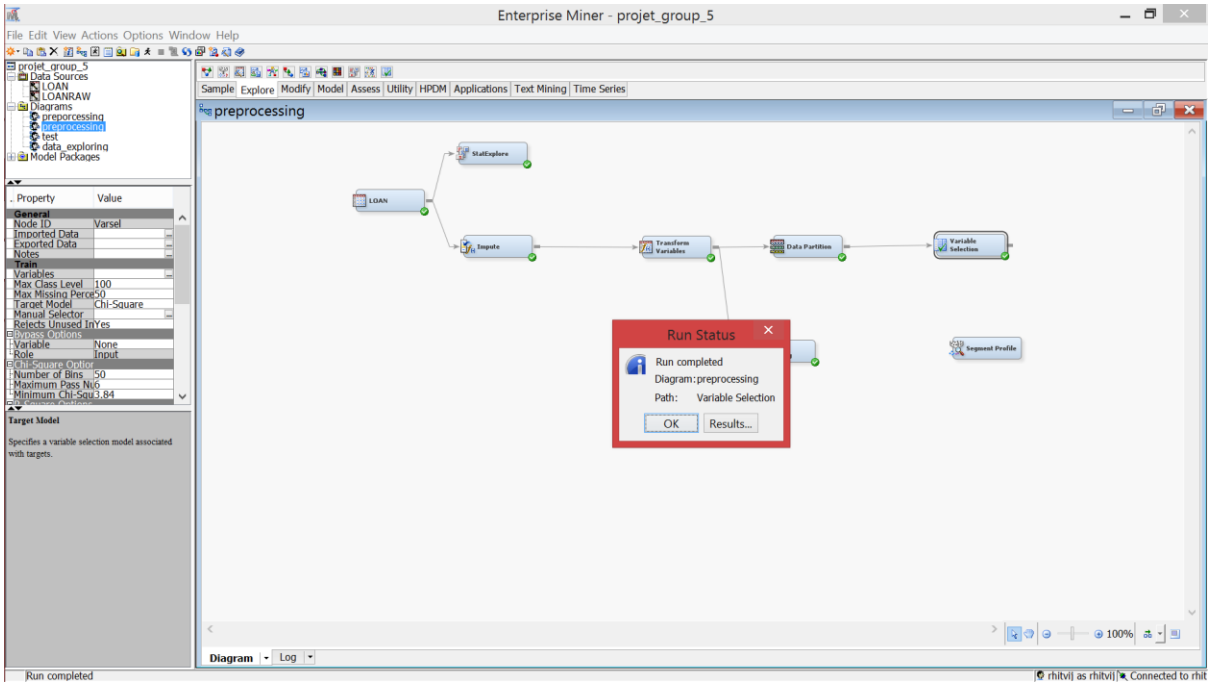
Screen shot number 23



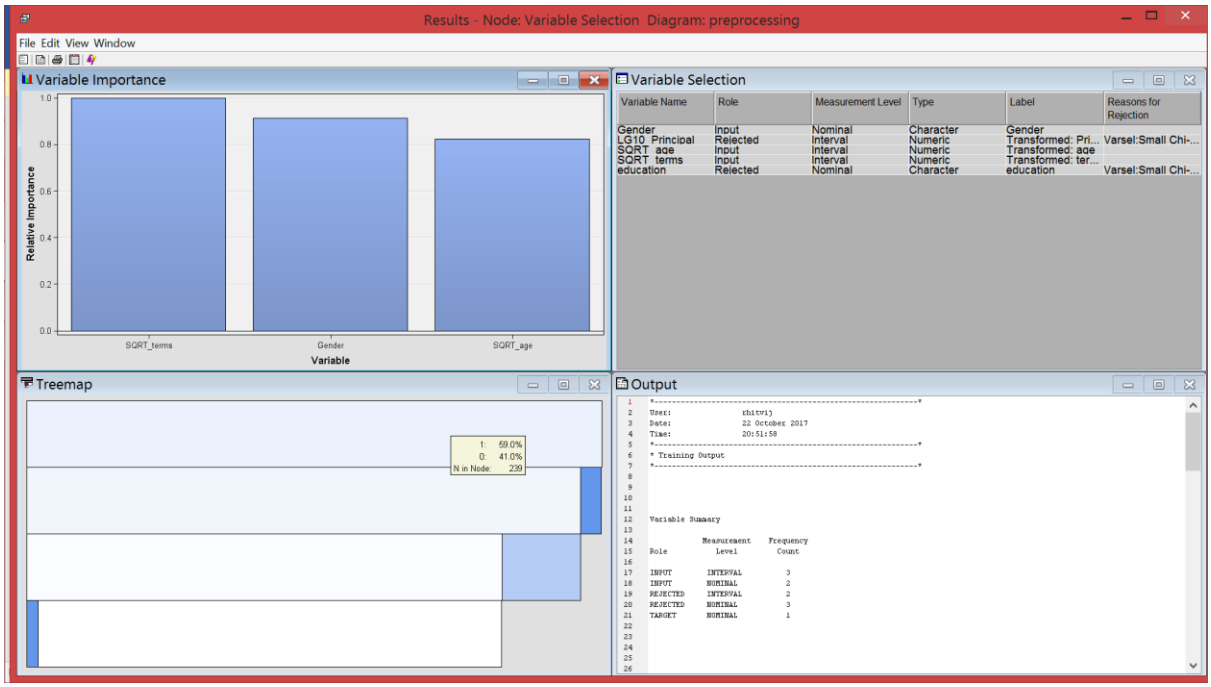
Screen shot number 24



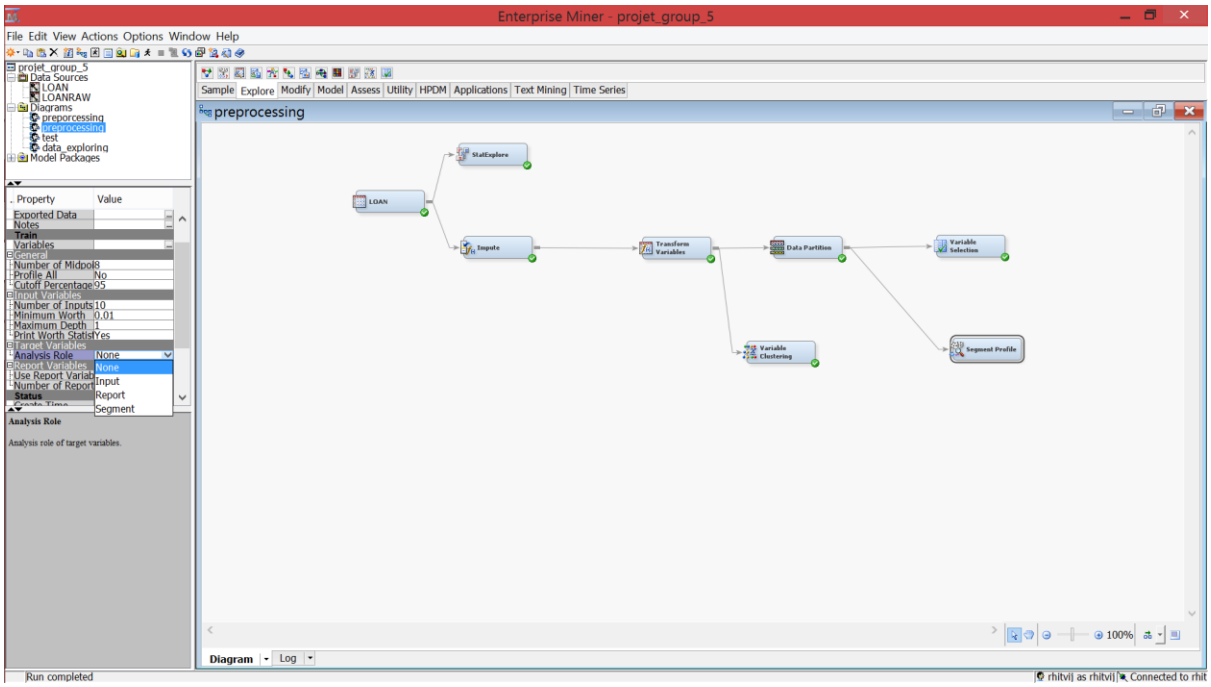
Screen shot number 25



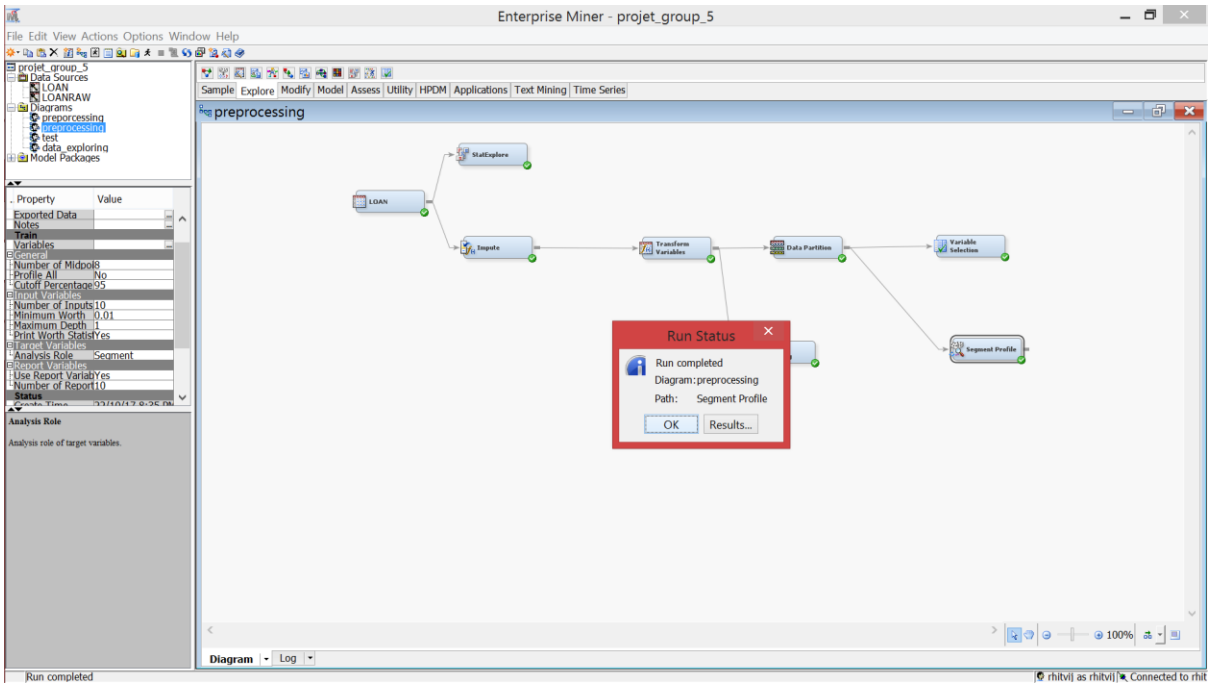
Screen shot number 26



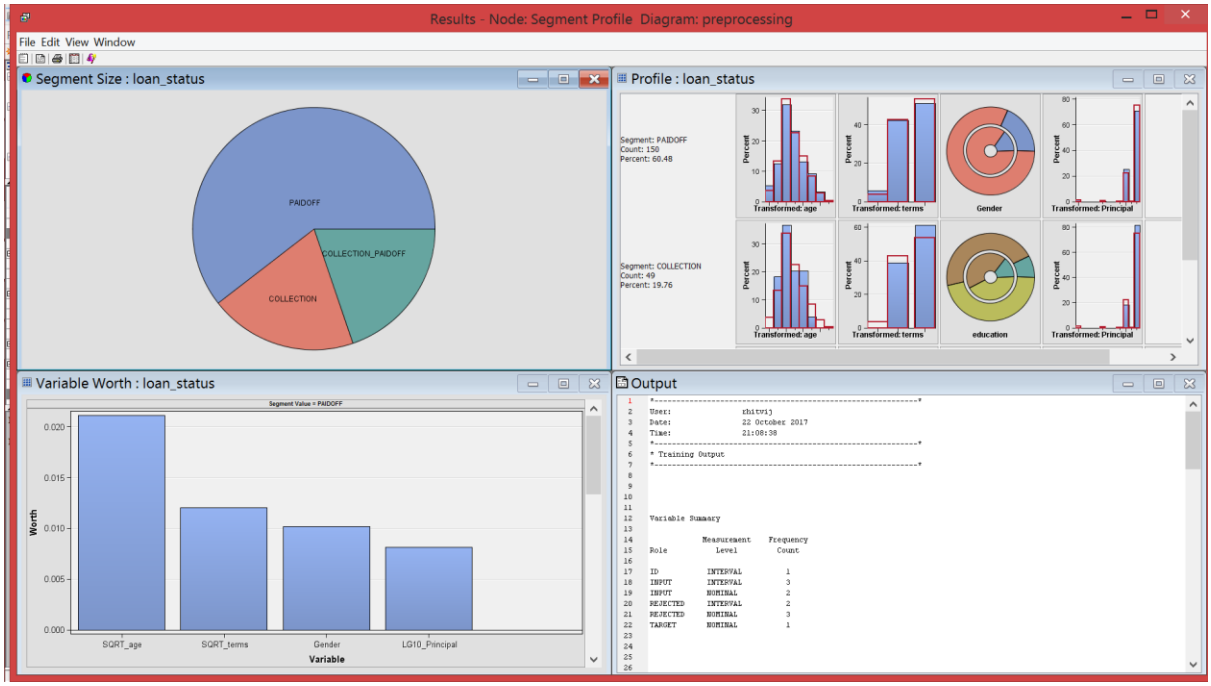
Screen shot number 27



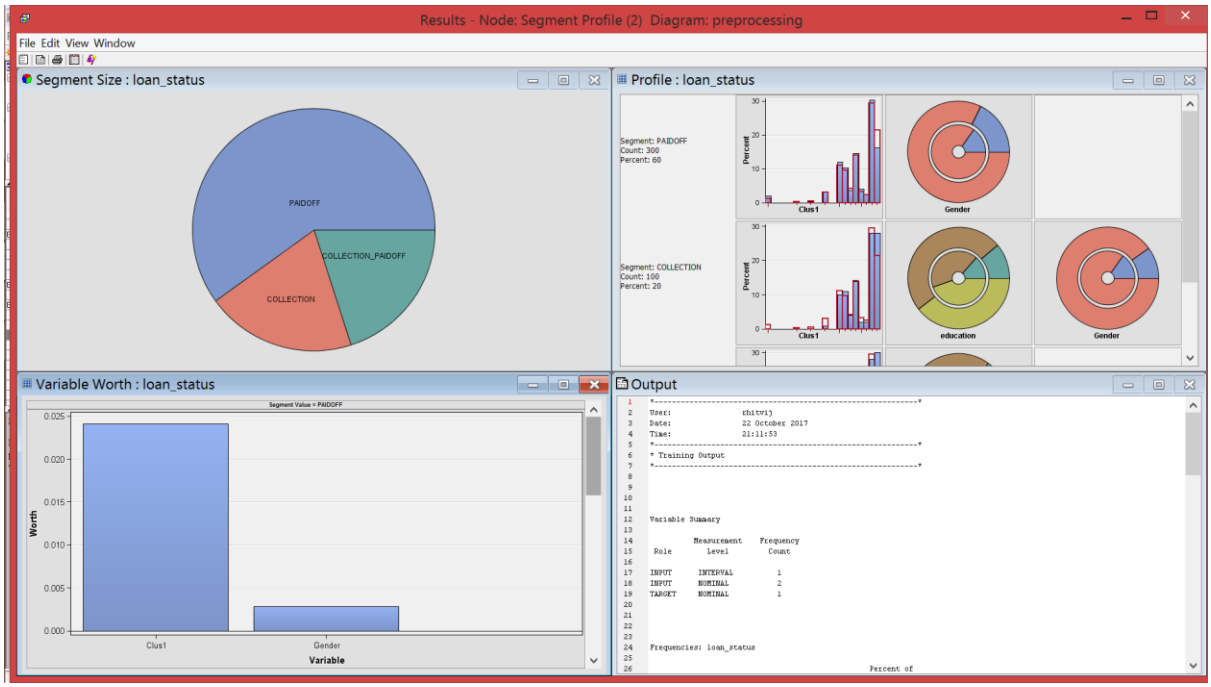
Screen shot number 28



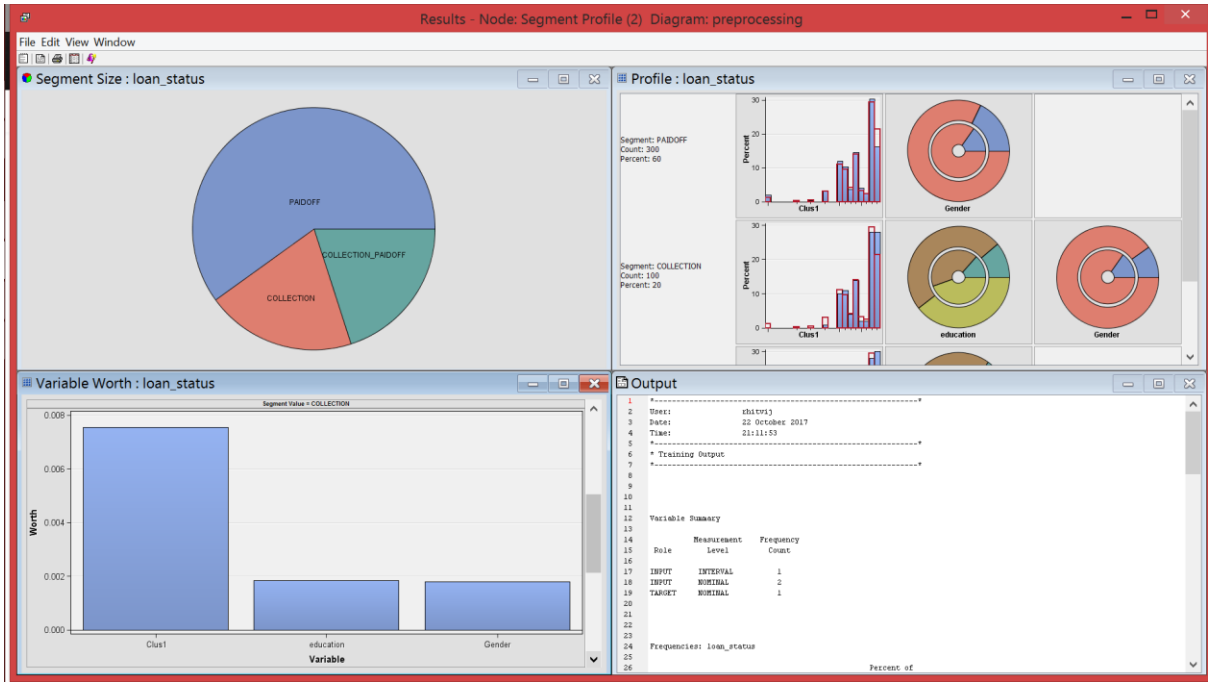
Screen shot number 29



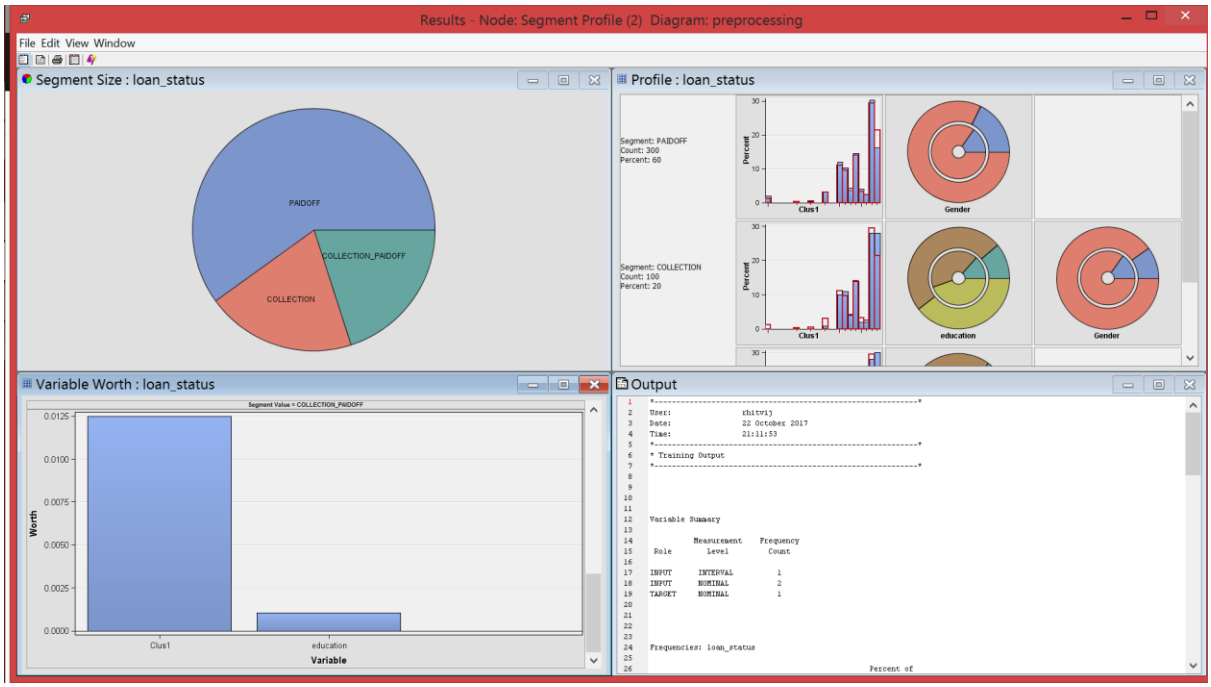
Screen shot number 30



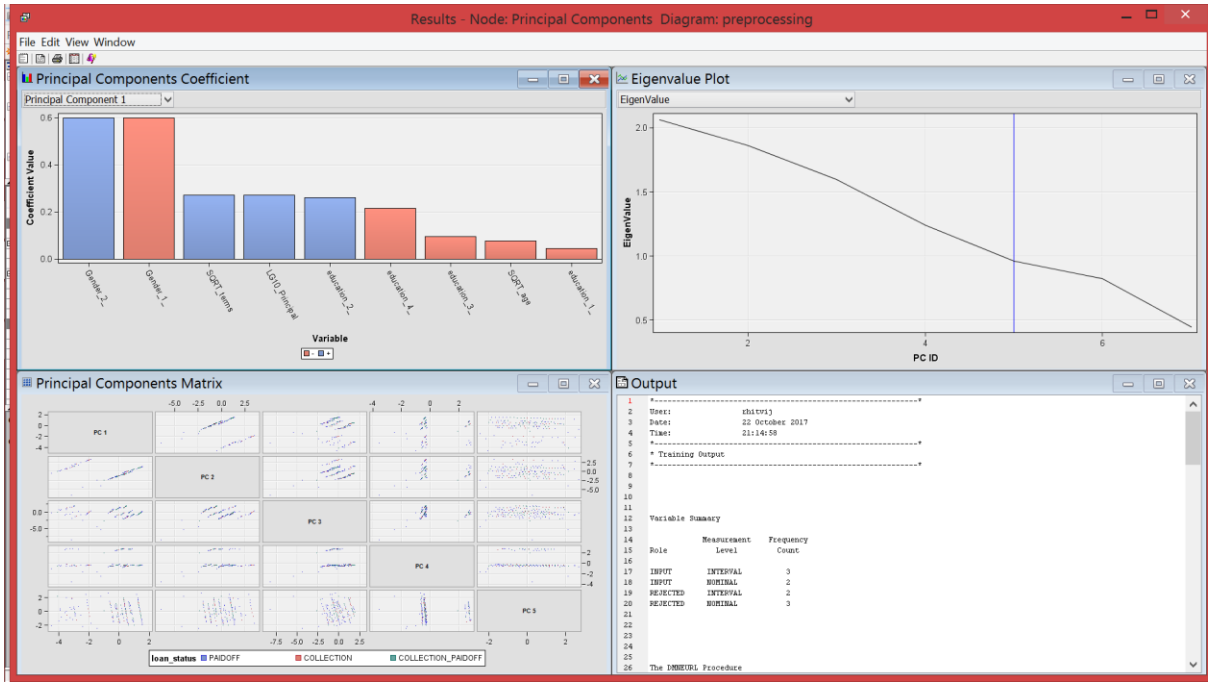
Screen shot number 31



Screen shot number 32

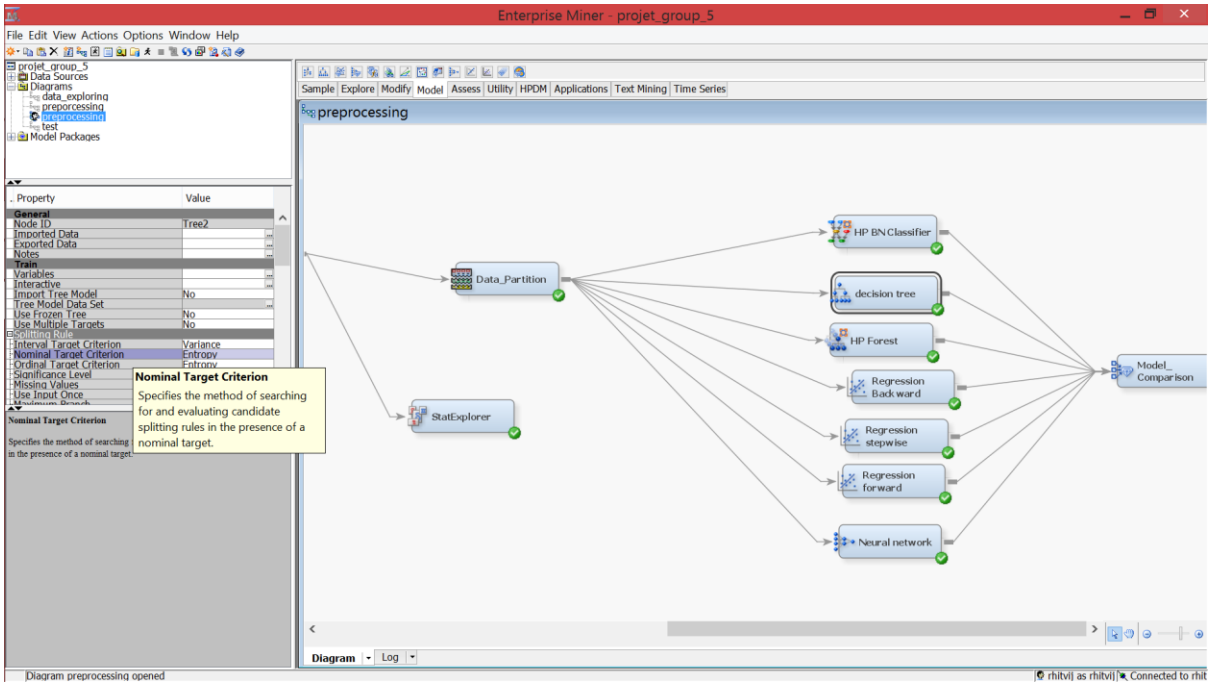


Screen shot number 33

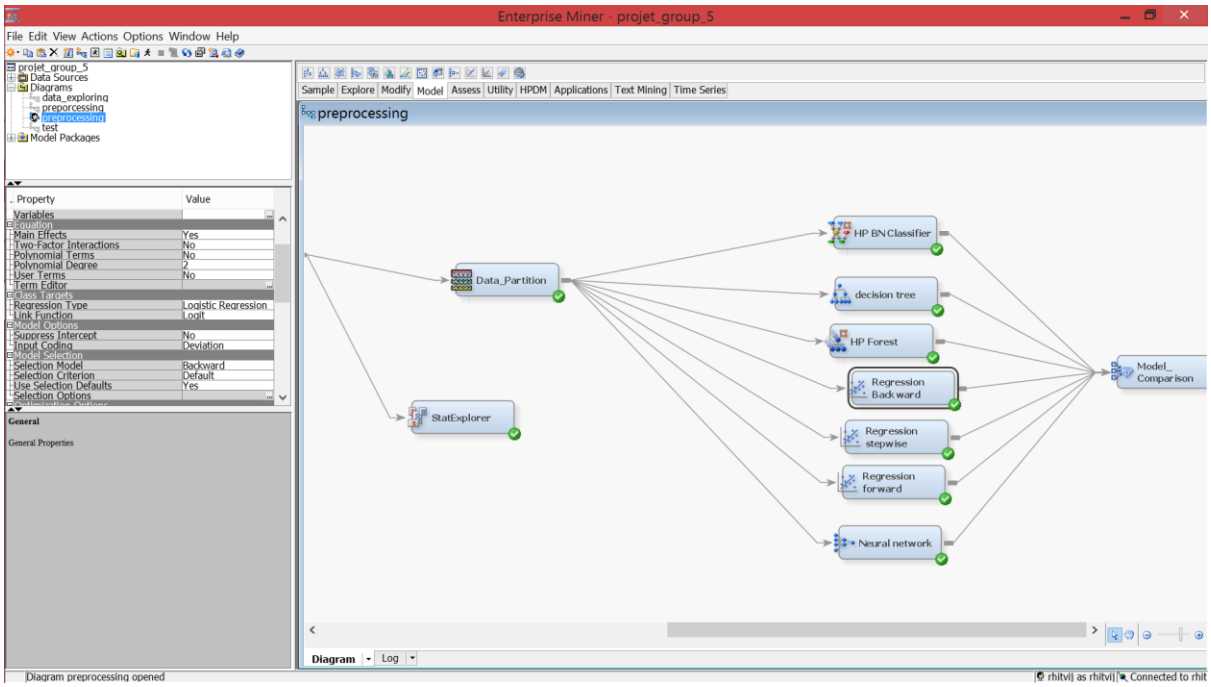


Building model sets.

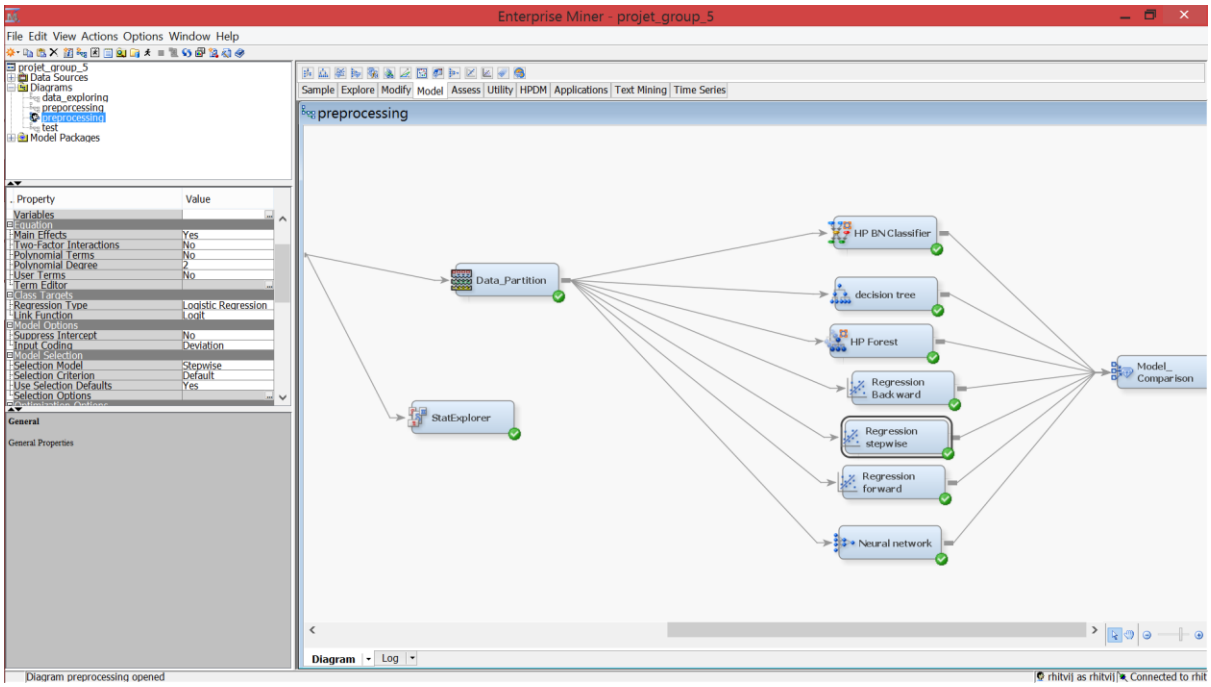
40



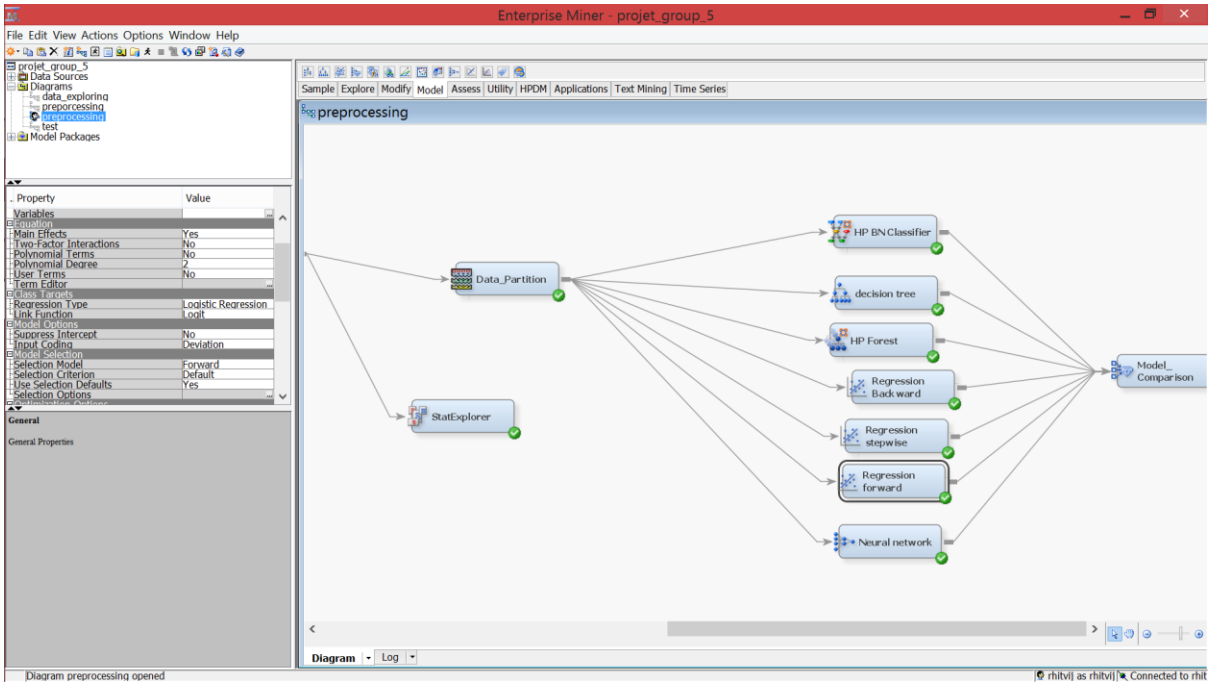
41



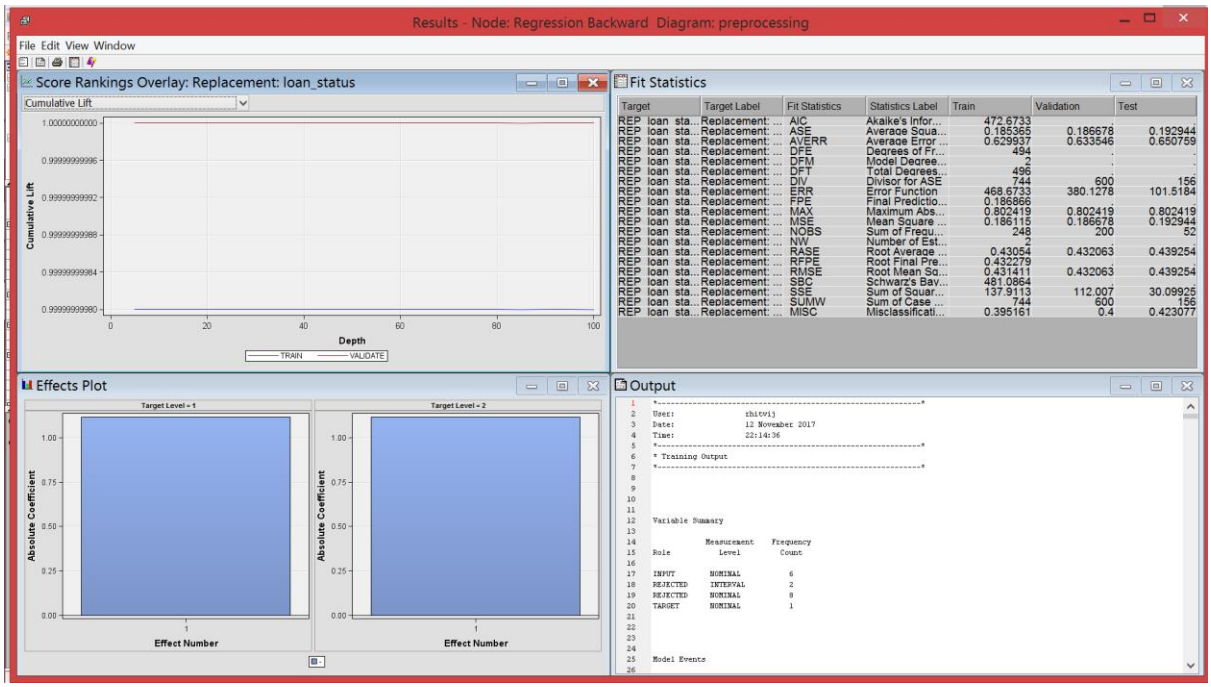
42


Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
preprocessing
test
Model Packages
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
Stepwise
Selection Criterion
Default
Use Selection Defaults
Yes
Selection Options
General
General Properties
Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
preprocessing
Data_Partition
StatExplorer
HP BN Classifier
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
Model_Comparison
Diagram
Log
Diagram preprocessing opened
rhitvi] as rhitvi] Connected to rhit

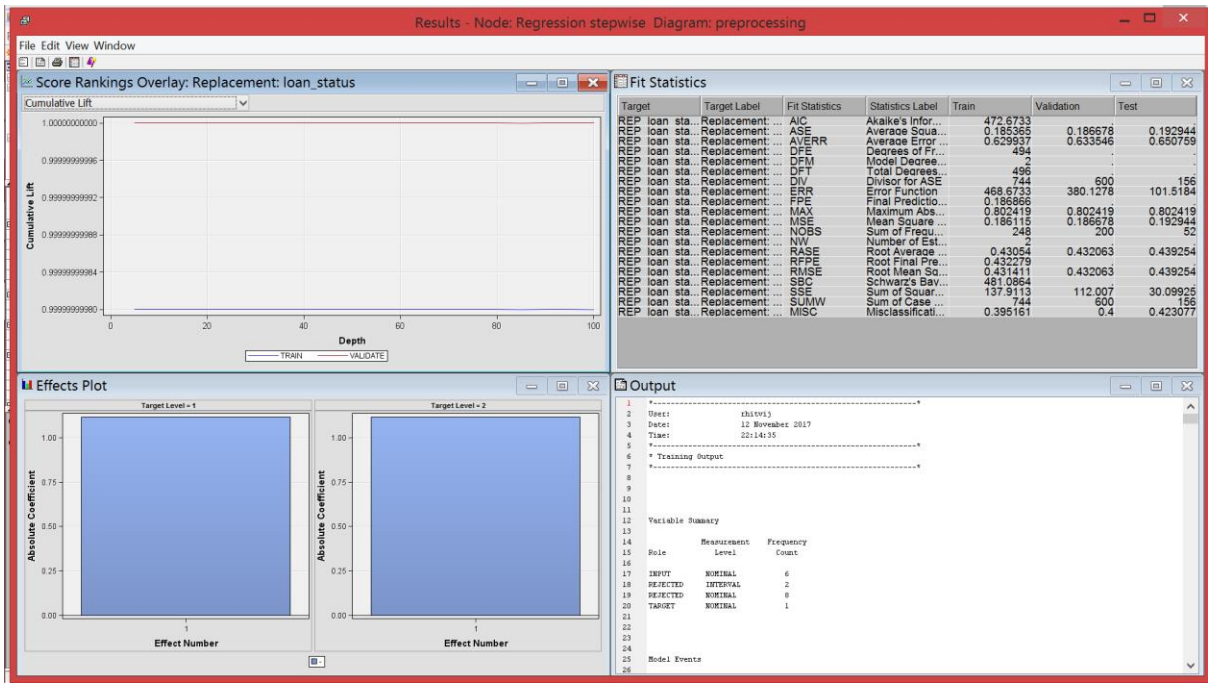
43


Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
preprocessing
test
Model Packages
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
Forward
Selection Criterion
Default
Use Selection Defaults
Yes
Selection Options
General
General Properties
Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
preprocessing
Data_Partition
StatExplorer
HP BN Classifier
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
Model_Comparison
Diagram
Log
Diagram preprocessing opened
rhitvi] as rhitvi] Connected to rhit

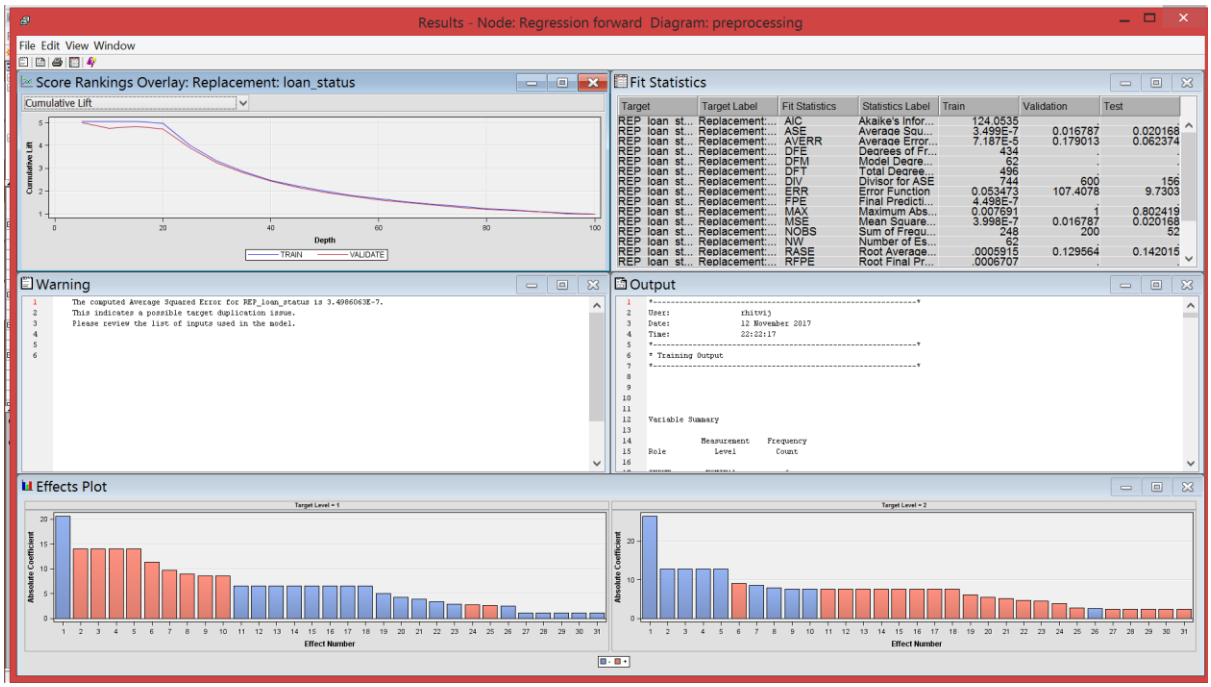
44


Results - Node: Regression Backward Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loan_status
Cumulative Lift
Cumulative Lift
Depth
TRAIN
VALIDATE
Fit Statistics
Target | Target Label | Fit Statistics | Statistics Label | Train | Validation | Test
REP loan sta... Replacement: ... AIC Akaike's Infor... 472.6733
REP loan sta... Replacement: ... ASE Average Squa... 0.185365 0.186678 0.192944
REP loan sta... Replacement: ... AVERR Average Error ... 0.629937 0.633546 0.650759
REP loan sta... Replacement: ... DFE Degrees of Fr... 494
REP loan sta... Replacement: ... DFM Model Degree... 2
REP loan sta... Replacement: ... DFT Total Degrees... 496
REP loan sta... Replacement: ... DIV Divisor for ASE 744 600 156
REP loan sta... Replacement: ... ERR Error Function 468.6733 380.1278 101.5184
REP loan sta... Replacement: ... FPE Final Predictio... 0.186866
REP loan sta... Replacement: ... MAX Maximum Abs... 0.802419 0.802419 0.802419
REP loan sta... Replacement: ... MSE Mean Square ... 0.186115 0.186678 0.192944
REP loan sta... Replacement: ... NOBS Sum of Frequ... 248 200 52
REP loan sta... Replacement: ... NW Number of Est... 2
REP loan sta... Replacement: ... RASE Root Average ... 0.43054 0.432063 0.439254
REP loan sta... Replacement: ... RFPE Root Final Pre... 0.432279
REP loan sta... Replacement: ... RMSE Root Mean Sq... 0.431411 0.432063 0.439254
REP loan sta... Replacement: ... SBC Schwarz's Bay... 481.0864
REP loan sta... Replacement: ... SSE Sum of Squar... 137.9113 112.007 30.09925
REP loan sta... Replacement: ... SUMW Sum of Case ... 744 600 156
REP loan sta... Replacement: ... MISC Misclassificati... 0.395161 0.4 0.423077
Effects Plot
Target Level = 1
Target Level = 2
Absolute Coefficient
Effect Number
Output
User: rhitvij
Date: 12 November 2017
Time: 22:14:36
* Training Output
Variable Summary
Role | Measurement Level | Frequency Count
INPUT NOMINAL 6
REJECTED INTERVAL 2
REJECTED NOMINAL 8
TARGET NOMINAL 1
Model Events

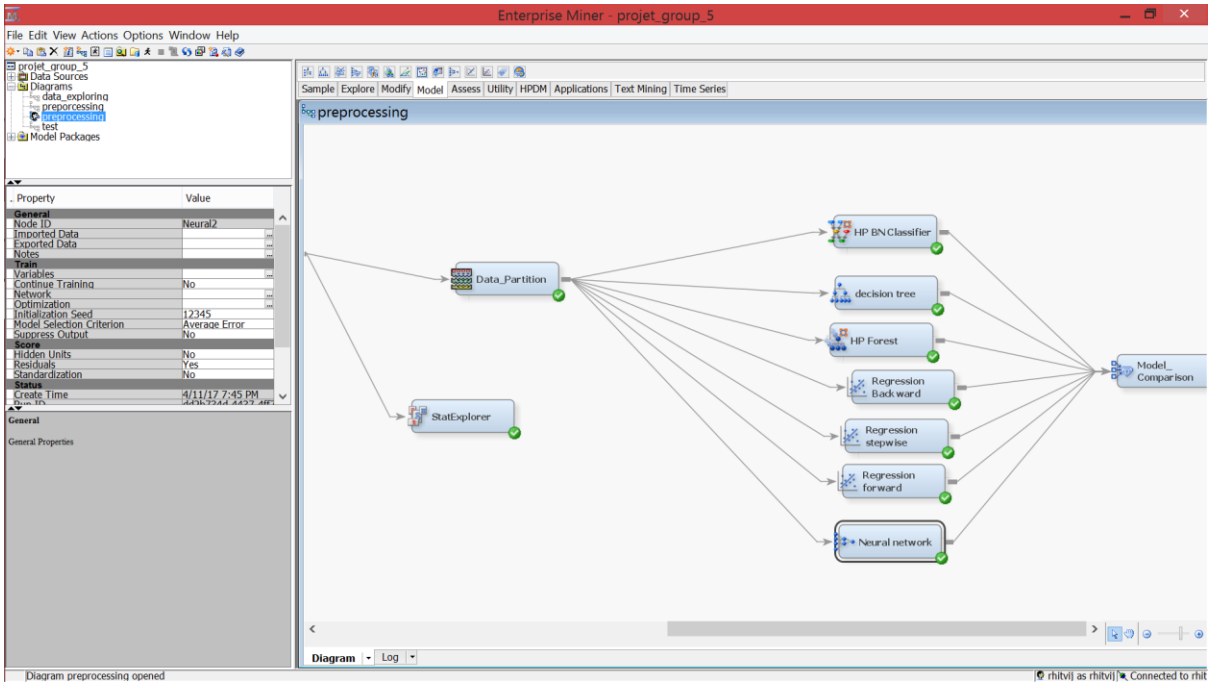
45


Results - Node: Regression stepwise Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loan_status
Cumulative Lift
Cumulative Lift
Depth
TRAIN
VALIDATE
Fit Statistics
Target | Target Label | Fit Statistics | Statistics Label | Train | Validation | Test
REP loan sta... Replacement: ... AIC Akaike's Infor... 472.6733
REP loan sta... Replacement: ... ASE Average Squa... 0.185365 0.186678 0.192944
REP loan sta... Replacement: ... AVERR Average Error ... 0.629937 0.633546 0.650759
REP loan sta... Replacement: ... DFE Degrees of Fr... 494
REP loan sta... Replacement: ... DFM Model Degree... 2
REP loan sta... Replacement: ... DFT Total Degrees... 496
REP loan sta... Replacement: ... DIV Divisor for ASE 744 600 156
REP loan sta... Replacement: ... ERR Error Function 468.6733 380.1278 101.5184
REP loan sta... Replacement: ... FPE Final Predictio... 0.186866
REP loan sta... Replacement: ... MAX Maximum Abs... 0.802419 0.802419 0.802419
REP loan sta... Replacement: ... MSE Mean Square ... 0.186115 0.186678 0.192944
REP loan sta... Replacement: ... NOBS Sum of Frequ... 248 200 52
REP loan sta... Replacement: ... NW Number of Est... 2
REP loan sta... Replacement: ... RASE Root Average ... 0.43054 0.432063 0.439254
REP loan sta... Replacement: ... RFPE Root Final Pre... 0.432279
REP loan sta... Replacement: ... RMSE Root Mean Sq... 0.431411 0.432063 0.439254
REP loan sta... Replacement: ... SBC Schwarz's Bay... 481.0864
REP loan sta... Replacement: ... SSE Sum of Squar... 137.9113 112.007 30.09925
REP loan sta... Replacement: ... SUMW Sum of Case ... 744 600 156
REP loan sta... Replacement: ... MISC Misclassificati... 0.395161 0.4 0.423077
Effects Plot
Target Level = 1
Target Level = 2
Absolute Coefficient
Effect Number
Output
User: rhitvij
Date: 12 November 2017
Time: 22:14:35
* Training Output
Variable Summary
Role | Measurement Level | Frequency Count
INPUT NOMINAL 6
REJECTED INTERVAL 2
REJECTED NOMINAL 8
TARGET NOMINAL 1
Model Events

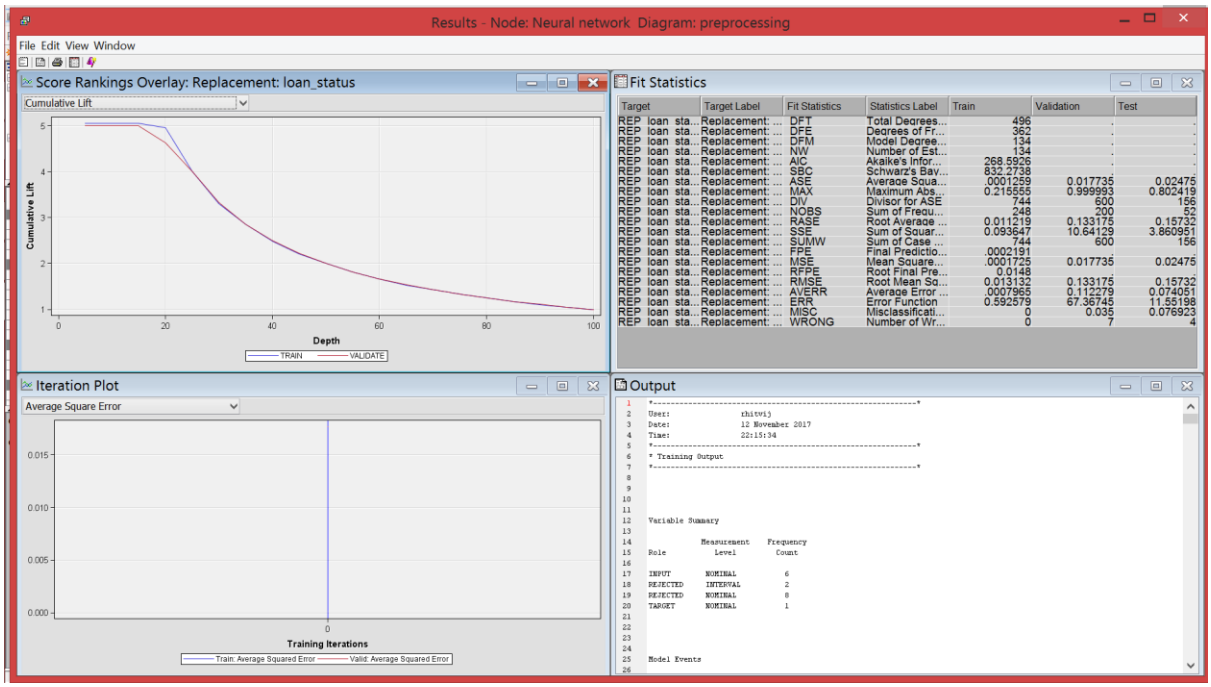
46


Results - Node: Regression forward Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loan_status
Cumulative Lift
Depth
TRAIN
VALIDATE
Fit Statistics
Warning
Output
Effects Plot
Effect Number

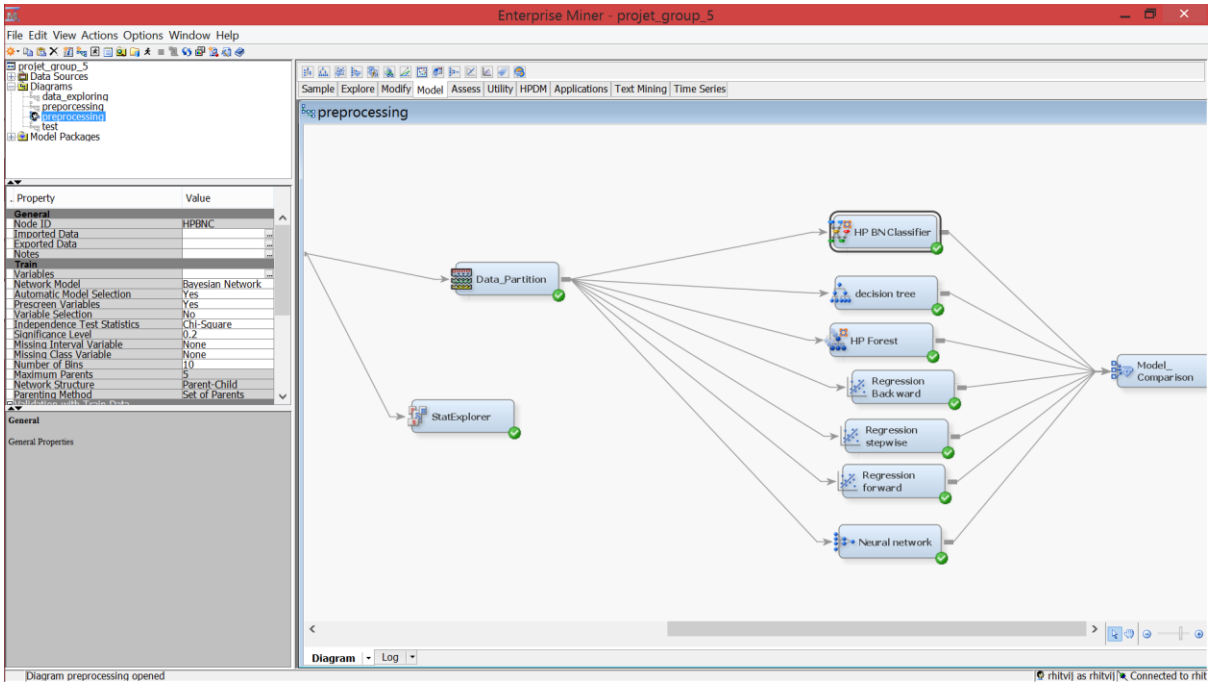
47


Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
test
Model Packages
Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
preprocessing
Data_Partition
StatExplorer
HP BN Classifier
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
Model_ Comparison
Diagram
Log
Diagram preprocessing opened

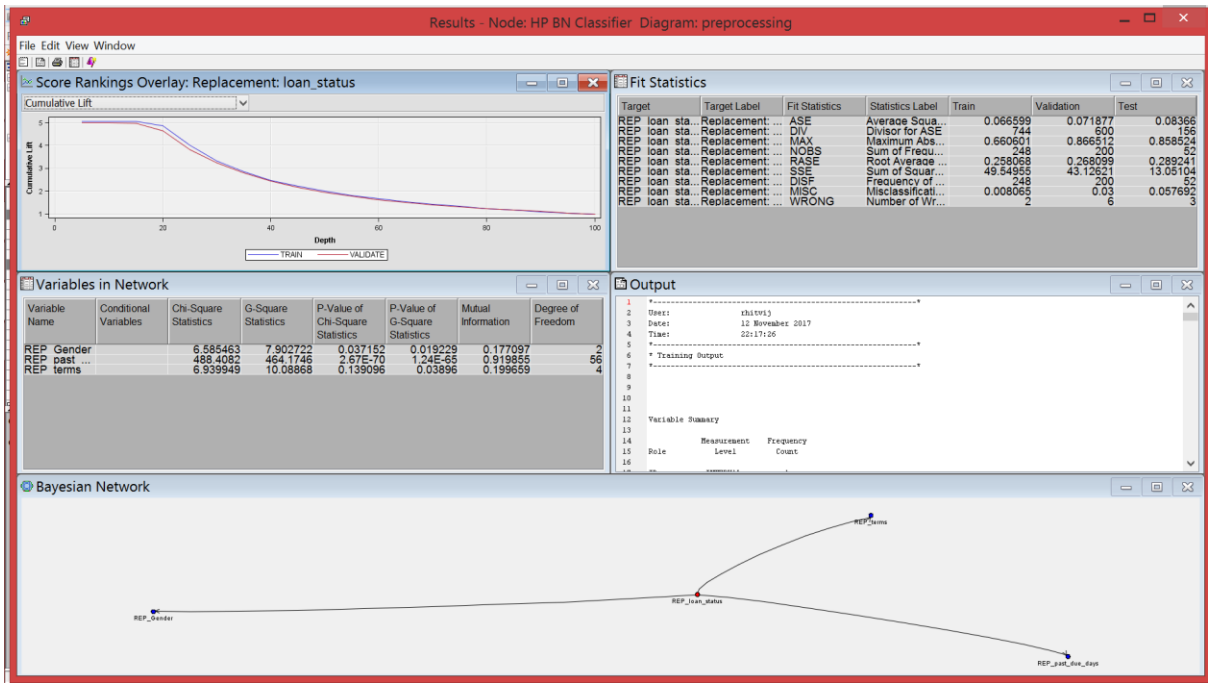
48


Results - Node: Neural network Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loan_status
Cumulative Lift
Depth
TRAIN
VALIDATE
Fit Statistics
Target | Target Label | Fit Statistics | Statistics Label | Train | Validation | Test
REP loan sta... Replacement: ... DFT Total Degrees... 496
REP loan sta... Replacement: ... DFE Degrees of Fr... 362
REP loan sta... Replacement: ... DFM Model Degree... 134
REP loan sta... Replacement: ... NW Number of Est... 134
REP loan sta... Replacement: ... AIC Akaike's Infor... 268.5926
REP loan sta... Replacement: ... SBC Schwarz's Bay... 832.2738
REP loan sta... Replacement: ... ASE Average Squa... .0001259 0.017735 0.02475
REP loan sta... Replacement: ... MAX Maximum Abs... 0.215555 0.999993 0.802419
REP loan sta... Replacement: ... DIV Divisor for ASE 744 600 156
REP loan sta... Replacement: ... NOBS Sum of Frequ... 248 200 52
REP loan sta... Replacement: ... RASE Root Average ... 0.011219 0.133175 0.15732
REP loan sta... Replacement: ... SSE Sum of Squar... 0.093647 10.64129 3.860951
REP loan sta... Replacement: ... SUMW Sum of Case ... 744 600 156
REP loan sta... Replacement: ... FPE Final Predictio... .0002191
REP loan sta... Replacement: ... MSE Mean Square... .0001725 0.017735 0.02475
REP loan sta... Replacement: ... RFPE Root Final Pre... 0.0148
REP loan sta... Replacement: ... RMSE Root Mean Sq... 0.013132 0.133175 0.15732
REP loan sta... Replacement: ... AVERR Average Error ... .0007965 0.112279 0.074051
REP loan sta... Replacement: ... ERR Error Function 0.592579 67.36745 11.55198
REP loan sta... Replacement: ... MISC Misclassificati... 0 0.035 0.076923
REP loan sta... Replacement: ... WRONG Number of Wr... 0 7 4
Iteration Plot
Average Square Error
Training Iterations
Train: Average Squared Error
Valid: Average Squared Error
Output
User: rhitvij
Date: 12 November 2017
Time: 22:15:34
* Training Output
Variable Summary
Role | Measurement Level | Frequency Count
INPUT NOMINAL 6
REJECTED INTERVAL 2
REJECTED NOMINAL 8
TARGET NOMINAL 1
Model Events

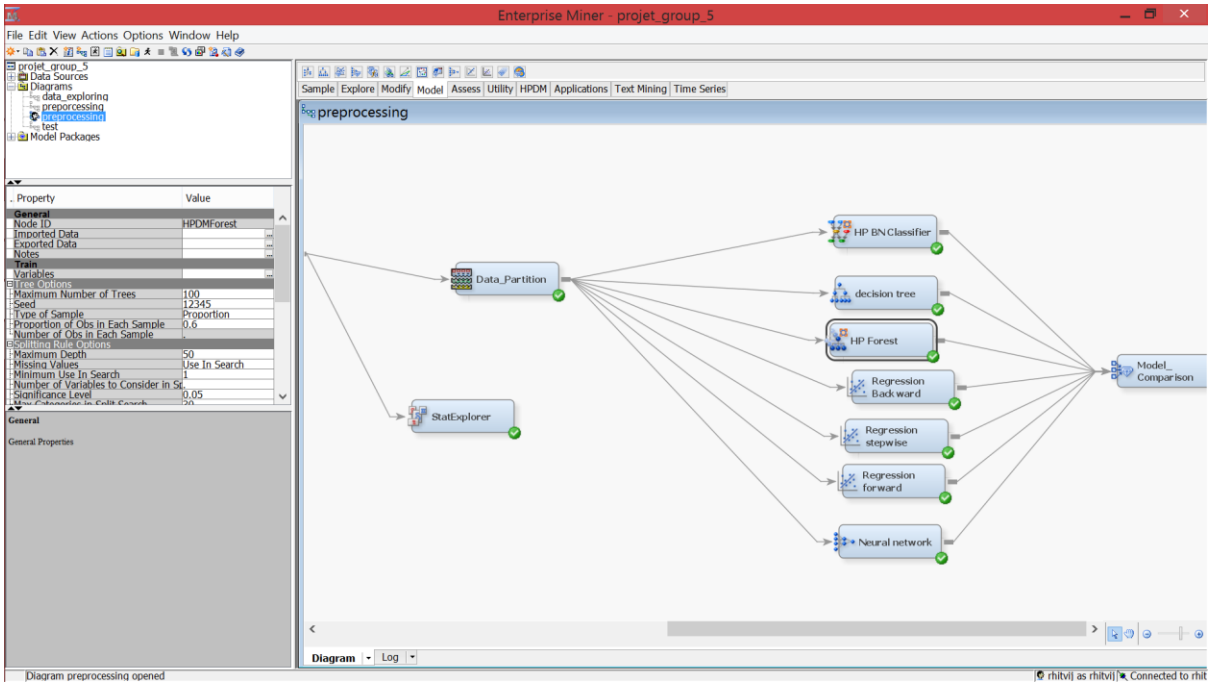
49


Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
test
Model Packages
Property | Value
General
Node ID HPBNC
Imported Data
Exported Data
Notes
Train
Variables
Network Model Bayesian Network
Automatic Model Selection Yes
Prescreen Variables Yes
Variable Selection No
Independence Test Statistics Chi-Square
Significance Level 0.2
Missing Interval Variable None
Missing Class Variable None
Number of Bins 10
Maximum Parents 5
Network Structure Parent-Child
Parenting Method Set of Parents
General
General Properties
Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
preprocessing
HP BN Classifier
Data_Partition
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
StatExplorer
Model_ Comparison
Diagram
Log
Diagram preprocessing opened
rhitvij as rhitvij Connected to rhit

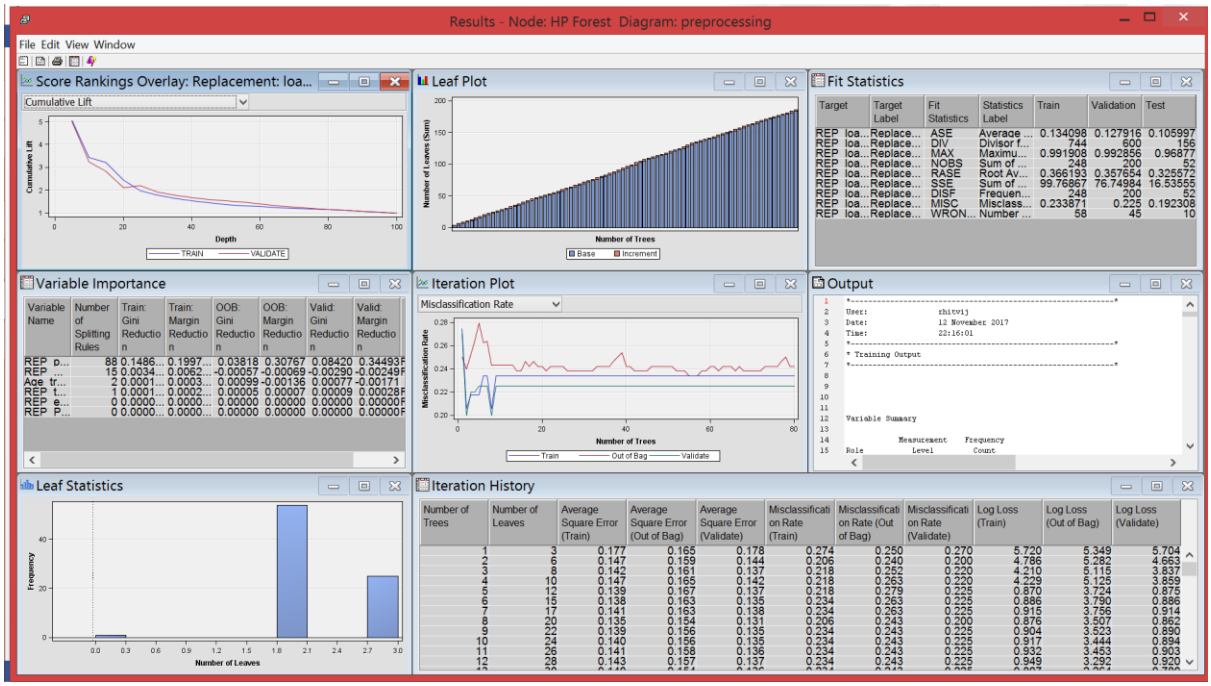
50


Results - Node: HP BN Classifier Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loan_status
Cumulative Lift
Depth
TRAIN
VALIDATE
Fit Statistics
Variables in Network
Output
Bayesian Network
REP_Gender
REP_loan_status
REP_terms
REP_past_due_days

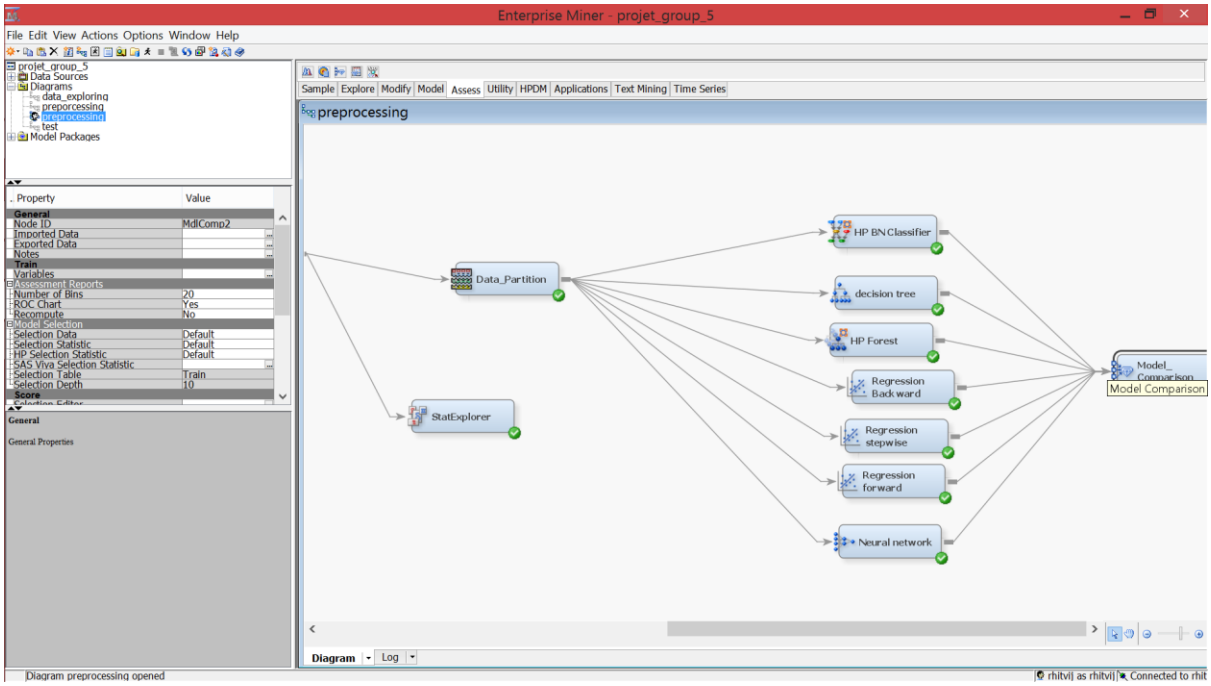
51


Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
test
Model Packages
Sample Explore Modify Model Assess Utility HPDM Applications Text Mining Time Series
preprocessing
Data_Partition
StatExplorer
HP BN Classifier
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
Model_ Comparison
Diagram
Log
Diagram preprocessing opened

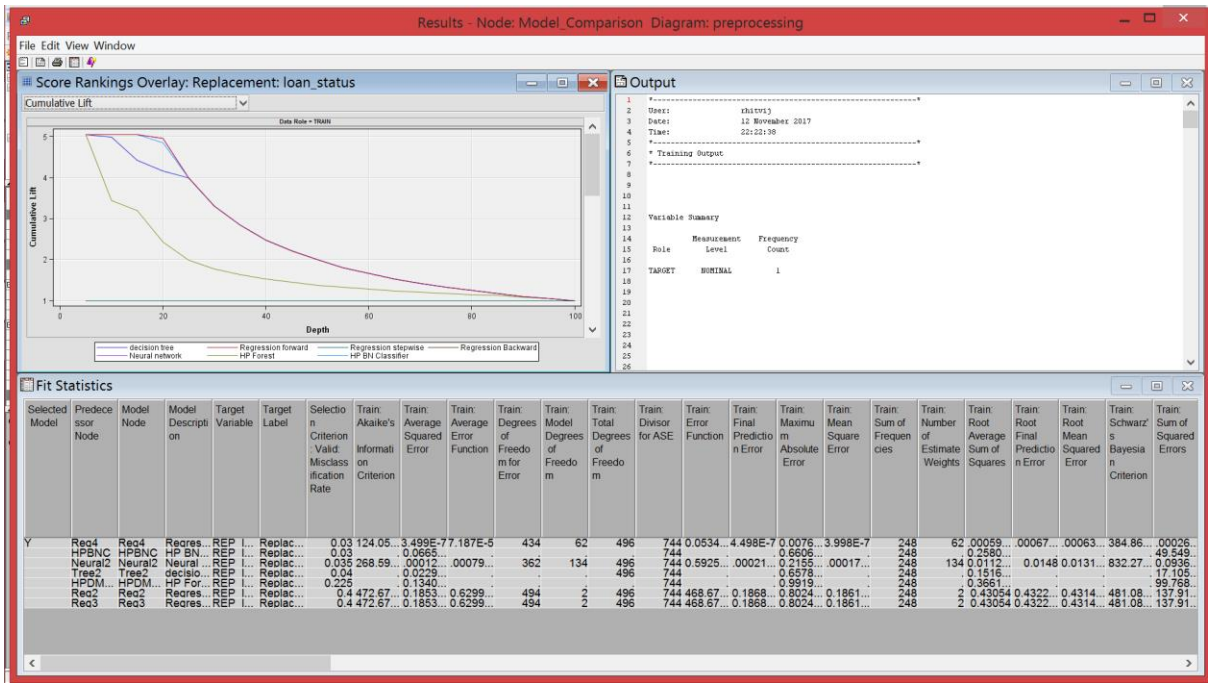
52


Results - Node: HP Forest Diagram: preprocessing
File Edit View Window
Score Rankings Overlay: Replacement: loa...
Cumulative Lift
Depth
TRAIN
VALIDATE
Leaf Plot
Number of Leaves (Sum)
Number of Trees
Base
Increment
Fit Statistics
Target
Target Label
Fit Statistics
Statistics Label
Train
Validation
Test
Variable Importance
Variable Name
Number of Splitting Rules
Train: Gini Reduction
Train: Margin Reduction
OOB: Gini Reduction
OOB: Margin Reduction
Valid: Gini Reduction
Valid: Margin Reduction
Iteration Plot
Misclassification Rate
Number of Trees
Train
Out of Bag
Validate
Output
Leaf Statistics
Frequency
Number of Leaves
Iteration History
Number of Trees
Number of Leaves
Average Square Error (Train)
Average Square Error (Out of Bag)
Average Square Error (Validate)
Misclassification Rate (Train)
Misclassification Rate (Out of Bag)
Misclassification Rate (Validate)
Log Loss (Train)
Log Loss (Out of Bag)
Log Loss (Validate)

53

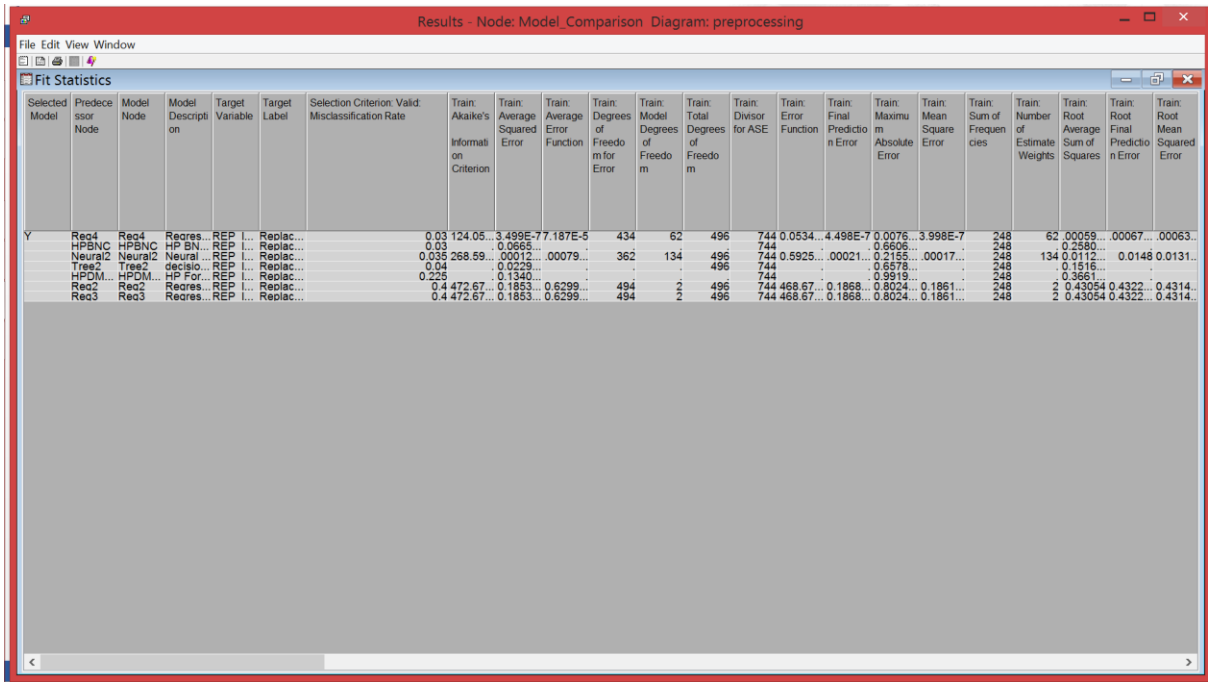

Enterprise Miner - projet_group_5
File Edit View Actions Options Window Help
projet_group_5
Data Sources
Diagrams
data_exploring
preprocessing
test
Model Packages
Property
Value
General
Node ID
MdlComp2
Imported Data
Exported Data
Notes
Train
Variables
Assessment Reports
Number of Bins
20
ROC Chart
Yes
Recompute
No
Model Selection
Selection Data
Default
Selection Statistic
Default
HP Selection Statistic
Default
SAS Viya Selection Statistic
Selection Table
Train
Selection Depth
10
Score
Sample
Explore
Modify
Model
Assess
Utility
HPDM
Applications
Text Mining
Time Series
preprocessing
Data_Partition
StatExplorer
HP BN Classifier
decision tree
HP Forest
Regression Backward
Regression stepwise
Regression forward
Neural network
Model Comparison
Diagram
Log
Diagram preprocessing opened

54



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Total Degrees of Freedom	Train: Divisor for ASE	Train: Error Function	Train: Final Prediction Error	Train: Maximum Absolute Error	Train: Mean Square Error	Train: Sum of Frequencies	Train: Number of Estimate Weights	Train: Root Average Sum of Squares	Train: Root Final Prediction Error	Train: Root Mean Squared Error	Train: Schwarz's Bayesian Criterion	Train: Sum of Squared Errors
Y	Reg4	Reg4	Regres...	REP	I...	Replac...	0.03	124.05...	3.499E-7	7.187E-5	434	62	496	744	0.0534...	4.498E-7	0.0076...	3.998E-7	248	62	.00059...	.00067...	.00063...	384.86...	.00026...
	HPBNC	HPBNC	HP BN...	REP	I...	Replac...	0.03	.	0.0665...	.	.	.	.	744	.	.	0.6606...	.	248	.	0.2580...	.	.	.	49.549...
	Neural2	Neural2	Neural ...	REP	I...	Replac...	0.035	268.59...	.00012...	.00079...	362	134	496	744	0.5925...	.00021...	0.2155...	.00017...	248	134	0.0112...	0.0148	0.0131...	832.27...	0.0936...
	Tree2	Tree2	decisio...	REP	I...	Replac...	0.04	.	0.0229...	.	.	.	496	744	.	.	0.6578...	.	248	.	0.1516...	.	.	.	17.105...
	HPDM...	HPDM...	HP For...	REP	I...	Replac...	0.225	.	0.1340...	.	.	.	.	744	.	.	0.9919...	.	248	.	0.3661...	.	.	.	99.768...
	Reg2	Reg2	Regres...	REP	I...	Replac...	0.4	472.67...	0.1853...	0.6299...	494	2	496	744	468.67...	0.1868...	0.8024...	0.1861...	248	2	0.43054	0.4322...	0.4314...	481.08...	137.91...
	Reg3	Reg3	Regres...	REP	I...	Replac...	0.4	472.67...	0.1853...	0.6299...	494	2	496	744	468.67...	0.1868...	0.8024...	0.1861...	248	2	0.43054	0.4322...	0.4314...	481.08...	137.91...

55



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: Total Degrees of Freedom	Train: Divisor for ASE	Train: Error Function	Train: Final Prediction Error	Train: Maximum Absolute Error	Train: Mean Square Error	Train: Sum of Frequencies	Train: Number of Estimate Weights	Train: Root Average Sum of Squares	Train: Root Final Prediction Error	Train: Root Mean Squared Error
Y	Reg4	Reg4	Regres...	REP	I...	Replac...	0.03	124.05...	3.499E-7	7.187E-5	434	62	496	744	0.0534...	4.498E-7	0.0076...	3.998E-7	248	62	.00059...	.00067...	.00063...
	HPBNC	HPBNC	HP BN...	REP	I...	Replac...	0.03	.	0.0665...	.	.	.	.	744	.	.	0.6606...	.	248	.	0.2580...	.	.
	Neural2	Neural2	Neural ...	REP	I...	Replac...	0.035	268.59...	.00012...	.00079...	362	134	496	744	0.5925...	.00021...	0.2155...	.00017...	248	134	0.0112...	0.0148	0.0131...
	Tree2	Tree2	decisio...	REP	I...	Replac...	0.04	.	0.0229...	.	.	.	496	744	.	.	0.6578...	.	248	.	0.1516...	.	.
	HPDM...	HPDM...	HP For...	REP	I...	Replac...	0.225	.	0.1340...	.	.	.	.	744	.	.	0.9919...	.	248	.	0.3661...	.	.
	Reg2	Reg2	Regres...	REP	I...	Replac...	0.4	472.67...	0.1853...	0.6299...	494	2	496	744	468.67...	0.1868...	0.8024...	0.1861...	248	2	0.43054	0.4322...	0.4314...
	Reg3	Reg3	Regres...	REP	I...	Replac...	0.4	472.67...	0.1853...	0.6299...	494	2	496	744	468.67...	0.1868...	0.8024...	0.1861...	248	2	0.43054	0.4322...	0.4314...