Introduction to Data Mining and SAS® Enterprise Miner™

CUSTOMER LOYALTY TEAM ▪ Support You Can Count On

Melodie Rush
Customer Loyalty Systems Engineer

Agenda

SAS Enterprise Miner

- SAS® Enterprise Miner™ and data mining overview
- Overview of interface and setting up a flow
  - Running Models
  - Model Comparison
- Rapid Predictive Modeler

BREAK

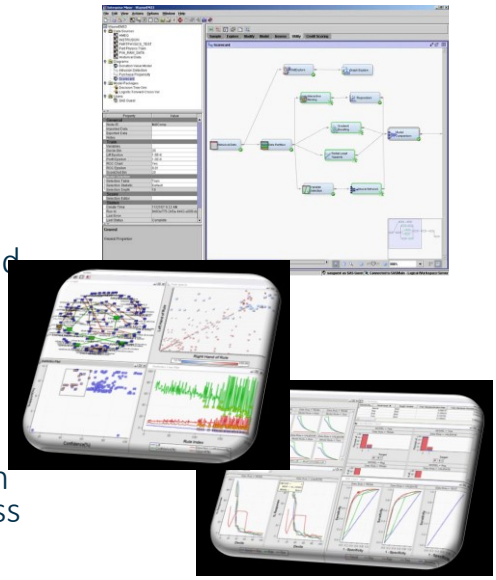
- Integrating models from other sources, including:
  - Champion models
  - Models from outside SAS Enterprise Miner
  - Open source models

What is Data Mining?

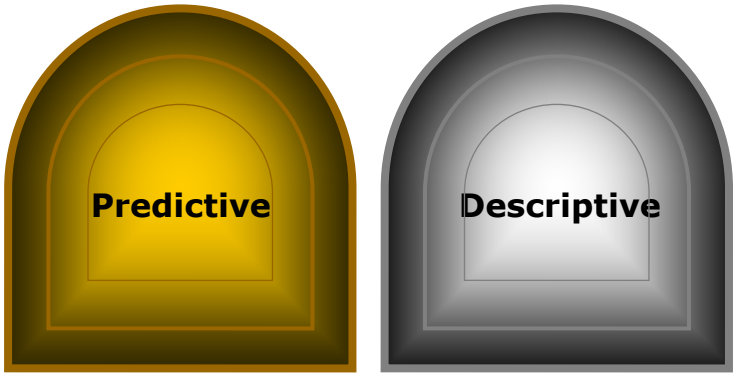
Turning increasing amounts of raw data into useful information

What is SAS® Enterprise Miner™?

- SAS Enterprise Miner is a sophisticated graphical user interface, designed with the specific needs of data miners in mind.
- SAS Enterprise Miner is a data miner's workbench that manages the process and provides a comprehensive set of tools to aid the data miner throughout the essential steps, known by the acronym, SEMMA: Sample, Explore, Modify, Model, Assess.
- SAS Enterprise Miner streamlines the data mining process to create highly accurate predictive and descriptive models based on analysis of vast amounts of data from across an enterprise.



Two Passages into the Data Mine



Descriptive Models

Clustering (Segmentation)

grouping together similar people, things, events

- Transactions that are likely to be fraudulent
- Customers that are likely to have similar behaviors.

Associations

affinity, or how frequently things occur together, and sometimes in what order

- Customers who purchase product A also purchase product B

Predictive Models

Classification models

predict class membership

- 0 or 1: 1 if person responded; 0 otherwise
- Low, Medium, High: a customer's likeliness to respond

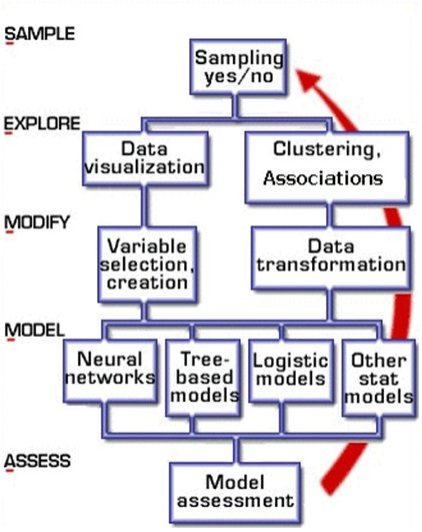
Regression models

predict a number

- $217.56 – Total profit, expense, cost for a customer
- 37 – The number of months before a customer churns

SEMMA

Data Mining With SAS® Enterprise Miner™

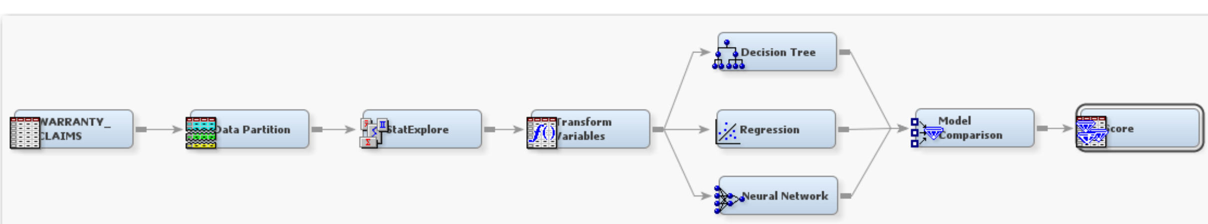


Complete List of SAS Enterprise Miner nodes in 14.2

SAMPLE	Append; Data Partition; File Import; Filter; Merge; Sample; Input Data
EXPLORE	Association; Cluster; Graph Explore; Variable Clustering; DMDB; MultiPlot; Market Basket; StatExplore; Link Analysis; Path Analysis; Variable Selection; SOM/Kohonen
MODIFY	Drop; Impute; Interactive Binning; Principal Components; Replacement; Rules Builder; Transform Variables
MODEL	Decision Tree; AutoNeural; Regression; Neural Network; Partial Least Squares; Dmine Regression; DM Neural; Ensemble; Rule Induction; Gradient Boosting; LARS; MBR; Two Stage; Model Import
	Incremental Response; Survival Analysis; Credit Scoring*; TS Correlation; TS Data Prep; TS Dimension Reduction; TS Decomp.; TS Similarity; TS Exponential Smoothing
	HP Explore; HP Bayesian Network; HP Regression; HP Transform; HP Impute; HP Variable Selection; HP Neural; HP Forest; HP Decision Tree; HP Data Partition; HP GLM; HP SVM; HP Cluster; HP Principal Components
ASSESS	Cutoff; Decisions; Model Comparison; Score; Segment Profile
UTILITY	Control Point; End Groups; Start Groups; Open Source Integration; Reporter; Score Code Export; Metadata; SAS Code; Ext Demo; Save Data; Register Metadata; SAS Viya Code

SEMMA

A Glimpse of SAS® Enterprise Miner™

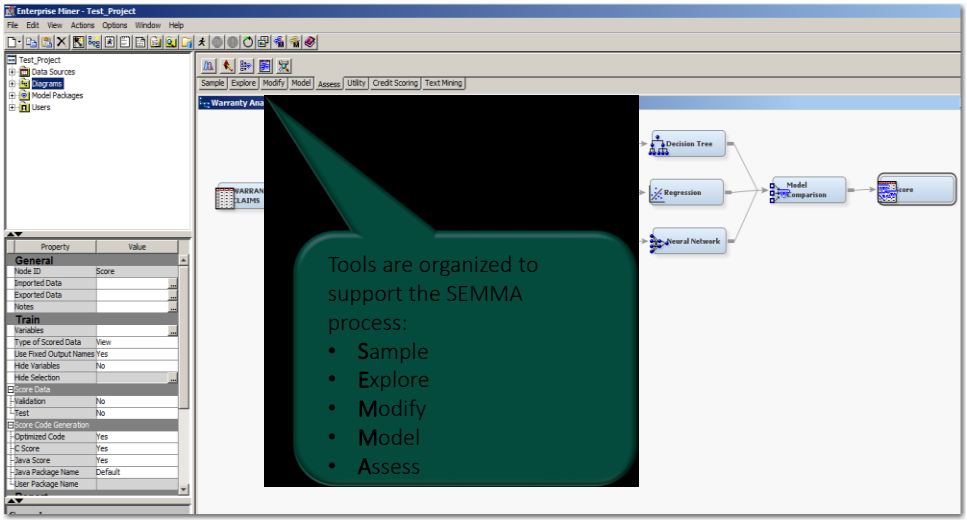


SAMPLE EXPLORE MODIFY MODEL ASSESS and SCORE

Sample **E**xplore **M**odify **M**odel **A**ssess

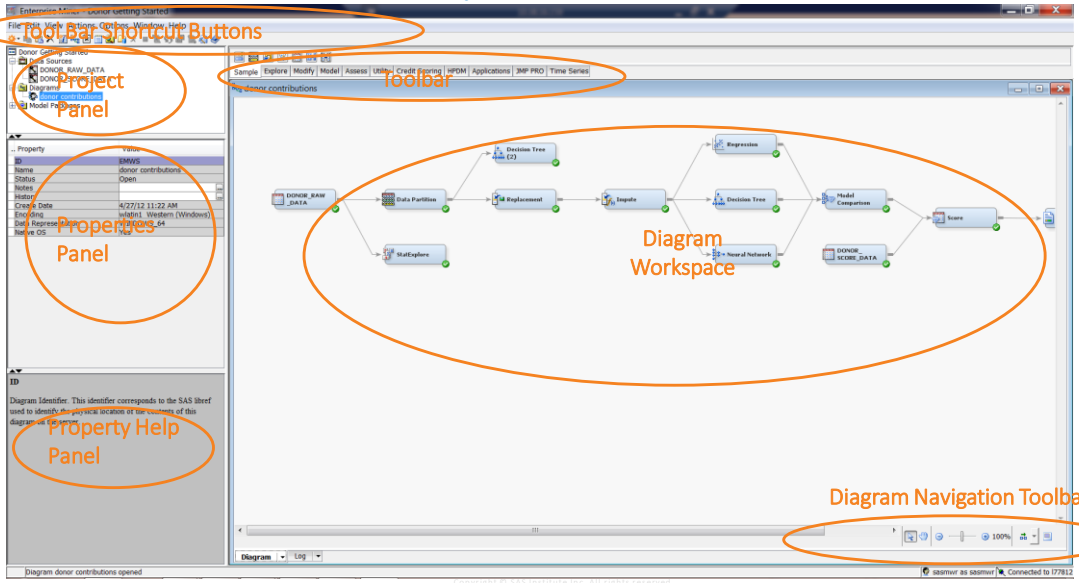
SEMMA

A Glimpse of SAS® Enterprise Miner™



Let's Get Started

SAS Enterprise Miner Interface



Launching SAS Enterprise Miner and Creating a New Project

Let's Get Started

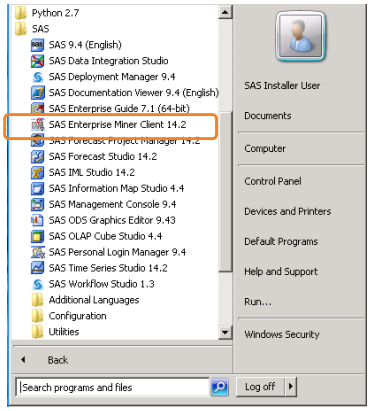
Opening SAS® Enterprise Miner™

Desktop or Local Client

- Desktop Icon
- Start → All Programs→ SAS → Analytics → Enterprise Miner Client
- Pinned to Task Bar or Start Menu

Java Client

- JAVA WebStart

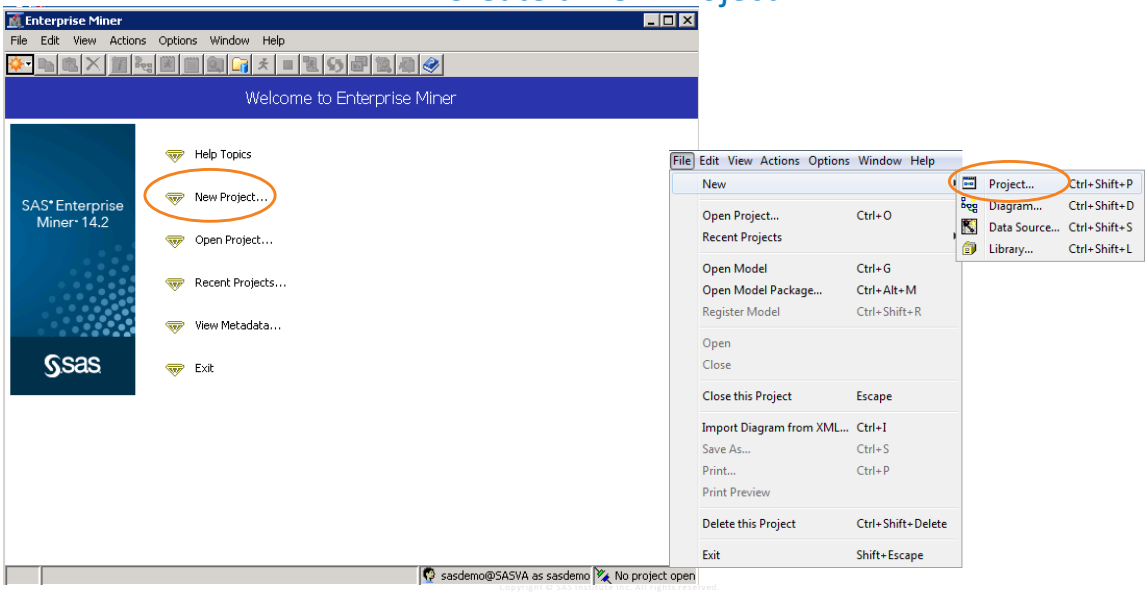


Let's Get Started

Log on to SAS® Enterprise Miner™

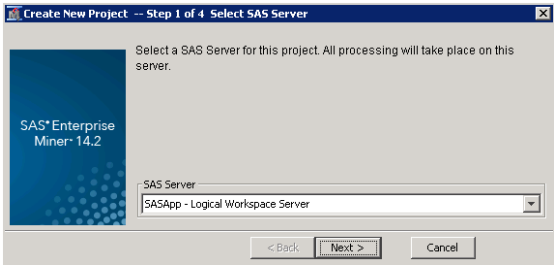
Let's Get Started

Create a New Project

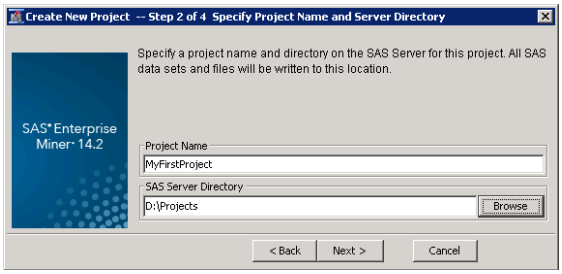


4 Step Wizard to Create a New Project

1. Select the “SAS Server” from the drop down list.

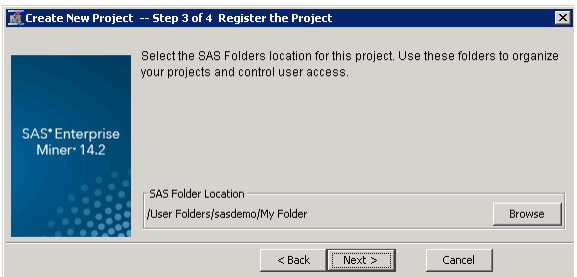


2. Type the name of your Project

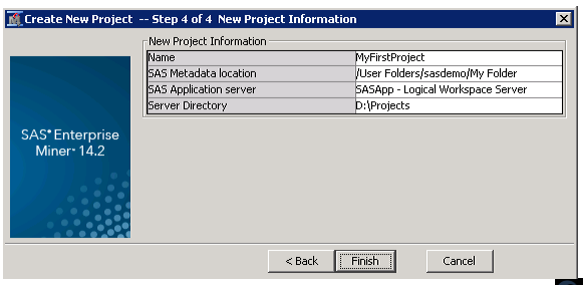


4 Step Wizard to Create a New Project

3. Set the SAS Folder Location to your folder.



4. Check settings. Click Finish



Define a Data Source

Data

Donor_Raw_Data

People likely to donate to a charity

- Y=TARGET_B
- N = 19,372
- Variables = 50 (47 Inputs)

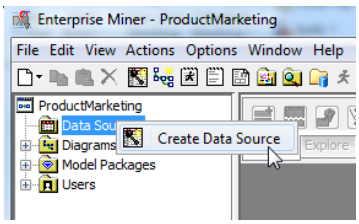
Column Descriptions
Download Data

Alphabetic List of Varia

#	Variable	Type	Len	Form
37	CARD_PROM_12	Num	8	
8	CLUSTER_CODE	Char	2	
3	CONTROL_NUMBER	Char	8	
10	DONOR_GENDER	Char	3	
41	FILE_AVG_GIFT	Num	8	
42	FILE_CARD_GIFT	Num	8	
21	FREQUENCY_STATUS_97NK	Num	8	
9	HOME_OWNER	Char	3	
48	IM_DONOR_AGE	Num	8	2.
46	IM_INCOME_GROUP	Num	8	2.
50	IM_MONTHS_SINCE_LAST_PROM_RESP	Num	8	2.
44	IM_WEALTH_RATING	Num	8	2.
5	IN_HOUSE	Num	8	
36	LAST_GIFT_AMT	Num	8	
32	LIFETIME_AVG_GIFT_AMT	Num	8	
28	LIFETIME_CARD_PROM	Num	8	
30	LIFETIME_GIFT_AMOUNT	Num	8	
31	LIFETIME_GIFT_COUNT	Num	8	
33	LIFETIME_GIFT_RANGE	Num	8	
34	LIFETIME_MAX_GIFT_AMT	Num	8	
35	LIFETIME_MIN_GIFT_AMT	Num	8	
29	LIFETIME_PROM	Num	8	
14	MEDIAN_HOME_VALUE	Num	8	
15	MEDIAN_HOUSEHOLD_INCOME	Num	8	
40	MONTHS_SINCE_FIRST_GIFT	Num	8	
39	MONTHS_SINCE_LAST_GIFT	Num	8	
4	MONTHS_SINCE_ORIGIN	Num	8	
13	MOR_HIT_RATE	Num	8	
47	M_DONOR_AGE	Num	8	
45	M_INCOME_GROUP	Num	8	
49	M_MONTHS_SINCE_LAST_PROM_RESP	Num	8	
43	M_WEALTH_RATING	Num	8	
38	NUMBER_PROM_12	Num	8	
12	OVERLAY_SOURCE	Char	1	
16	PCT_OWNER_OCCUPIED	Num	8	
18	PEP_STAR	Num	8	
17	PER_CAPITA_INCOME	Num	8	
11	PUBLISHED_PHONE	Num	8	
20	RECENCY_STATUS_96NK	Char	5	
25	RECENT_AVG_CARD_GIFT_AMT	Num	8	
23	RECENT_AVG_GIFT_AMT	Num	8	
27	RECENT_CARD_RESPONSE_COUNT	Num	8	
24	RECENT_CARD_RESPONSE_PROP	Num	8	
26	RECENT_RESPONSE_COUNT	Num	8	
22	RECENT_RESPONSE_PROP	Num	8	
19	RECENT_STAR_STATUS	Num	8	
7	SES	Char	4	
1	TARGET_B	Num	8	
2	TARGET_D	Num	8	
6	URBANICITY	Char	4	

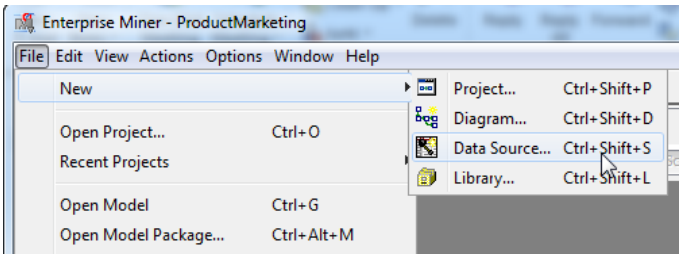
Create Data Source

Right Mouse Click on Data Source and Select Create Data Source



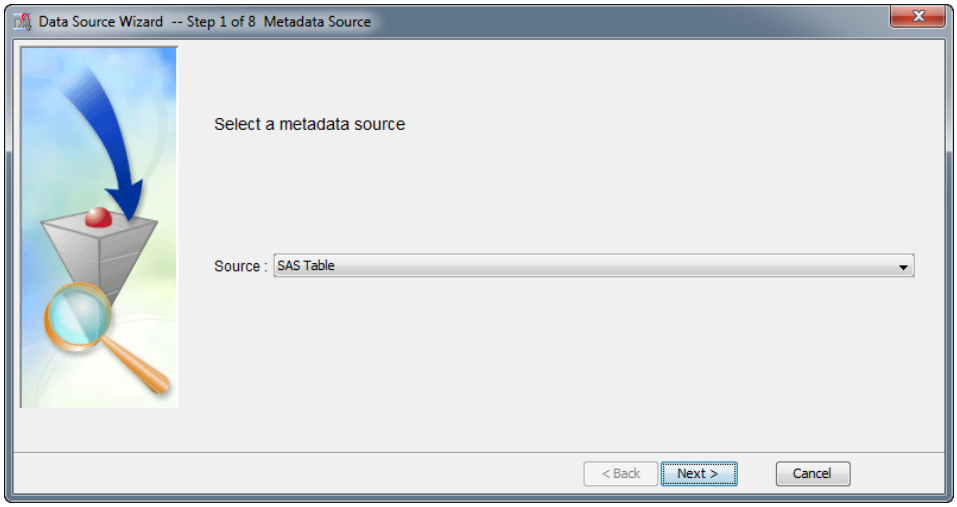
OR

Go to the Menu and select File→ New → Data Source



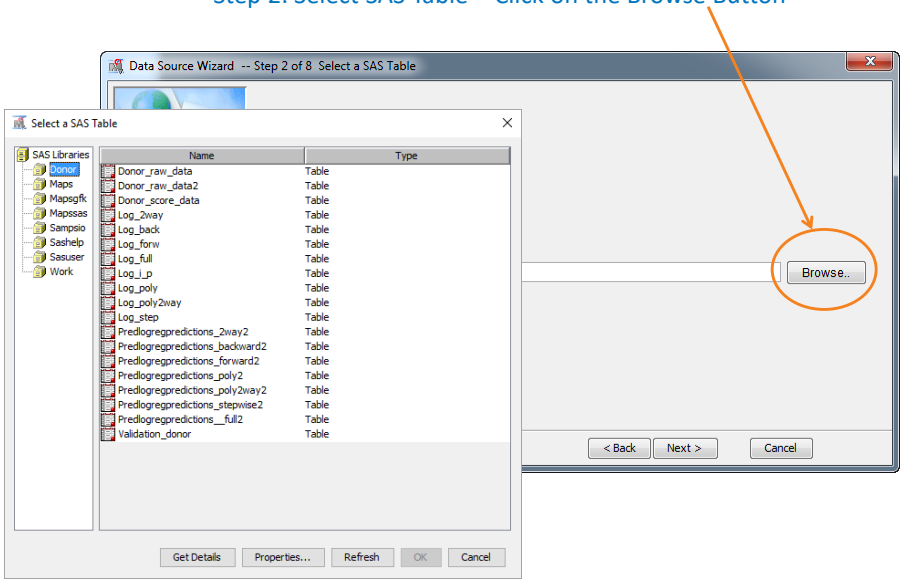
Create Data Source - Wizard

Step 1: Select Source



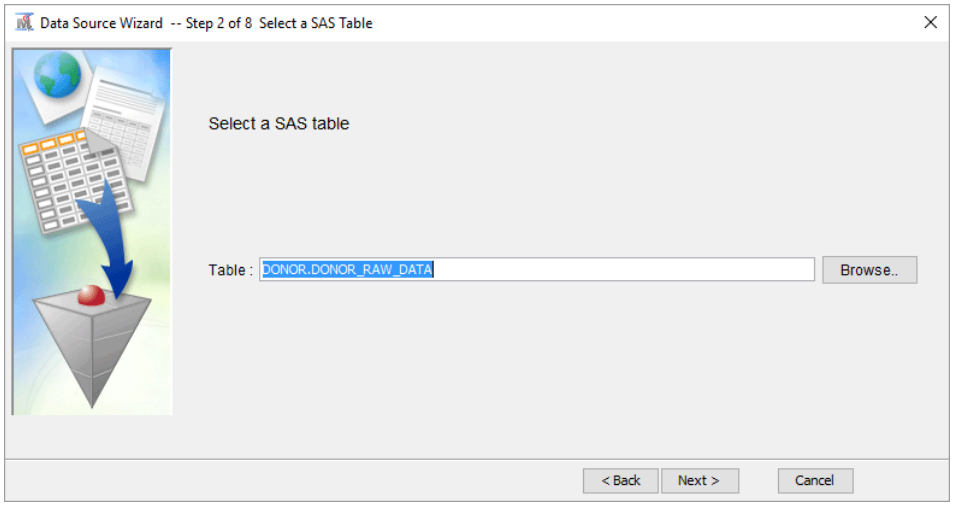
Create Data Source - Wizard

Step 2: Select SAS Table – Click on the Browse Button



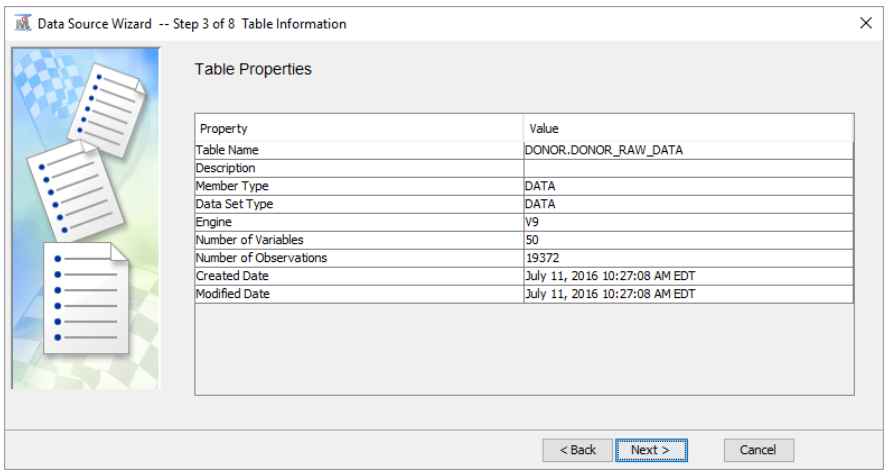
Create Data Source - Wizard

Step 2: Select SAS Table



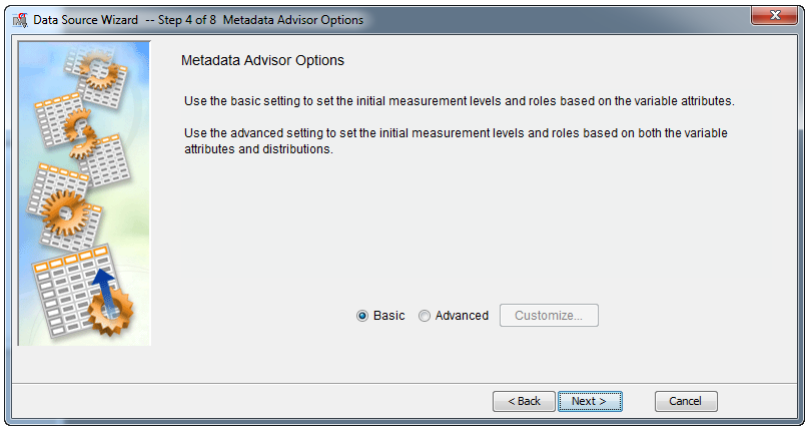
Create Data Source - Wizard

Step 3: Check Properties



Create Data Source - Wizard

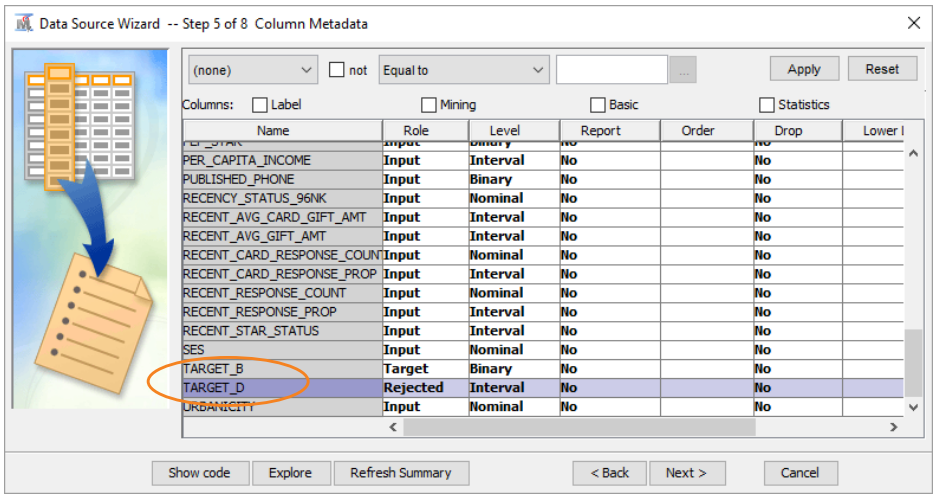
Step 4: Use Defaults or customize how the roles and attributes are assigned



Create Data Source - Wizard

Step 5: Assign Roles and Levels

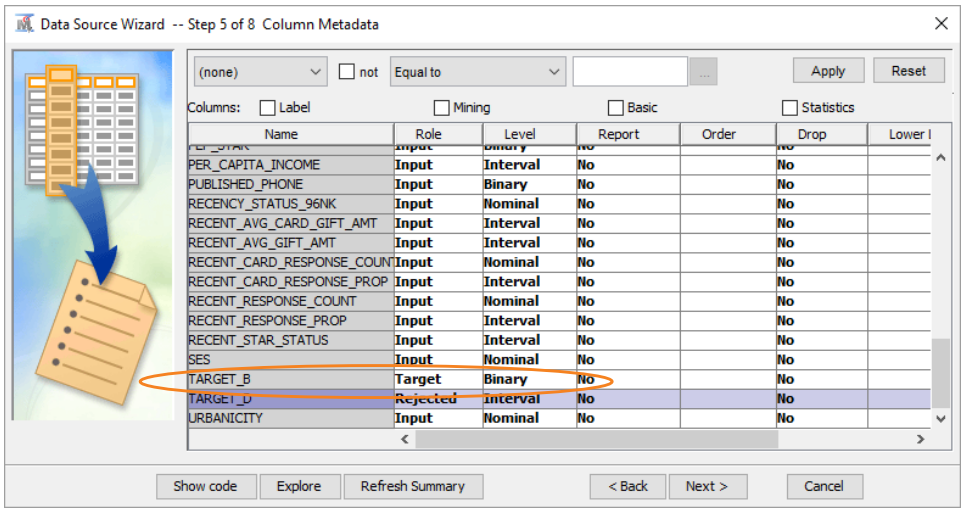
Assign Target_B a role of Target and Target_D a role of Rejected



Create Data Source - Wizard

Step 5: Assign Roles and Levels *(continued)*

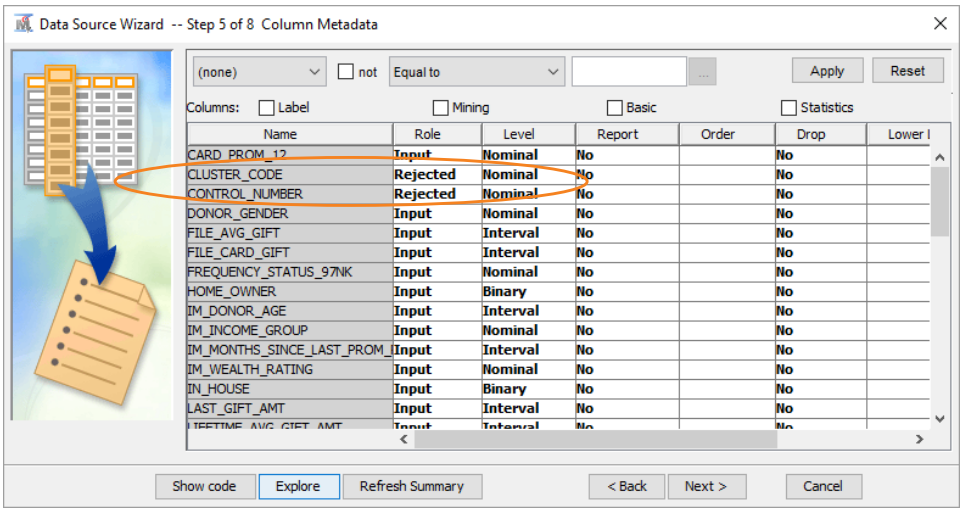
Assign Target_B the Level of Binary



Create Data Source - Wizard

Step 5: Assign Roles and Levels *(continued)*

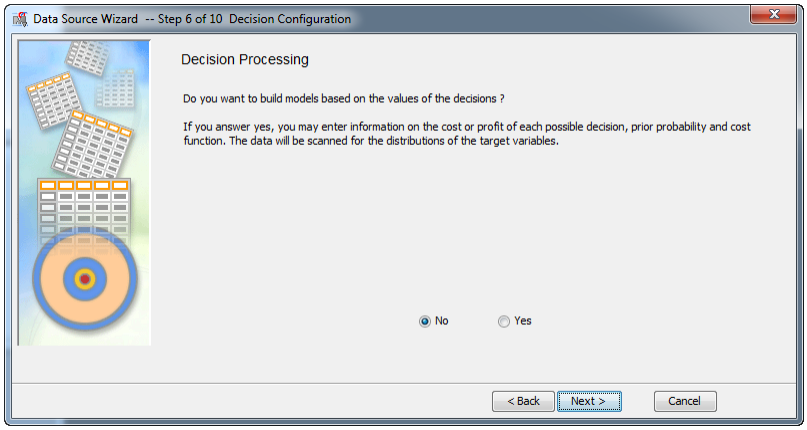
Assign CLUSTER_CODE and CONTROL_NUMBER the Role of Rejected



Create Data Source - Wizard

Step 6: Decision Processing

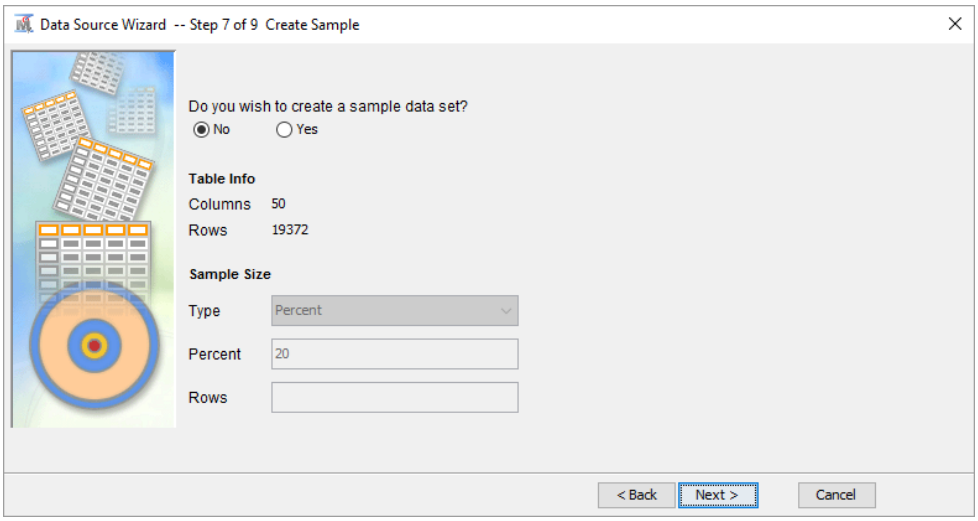
Used for applying prior probabilities, costs and/or profits



Create Data Source - Wizard

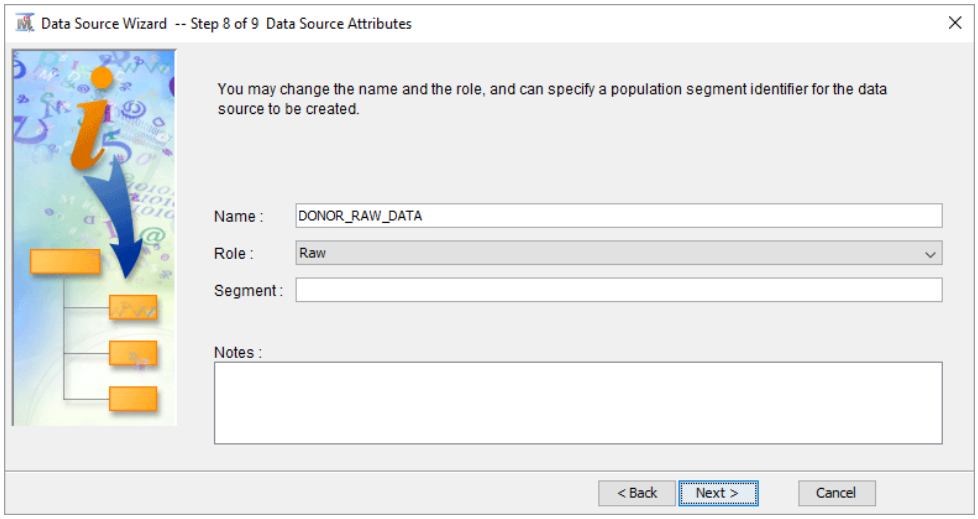
Step 7: Create Sample Data Set

(Useful if data is large; will keep the same sample for graphics)



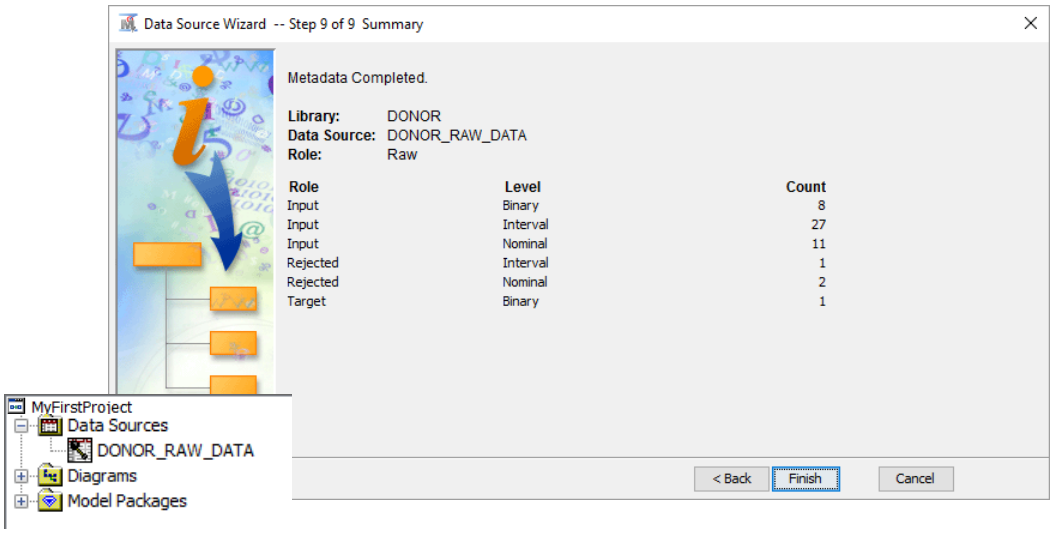
Create Data Source - Wizard

Step 8: Select Role of Data Set & add notes



Create Data Source - Wizard

Step 9: Review Summary of Data Source – Click Finish

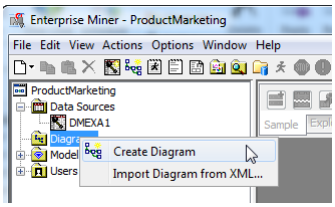


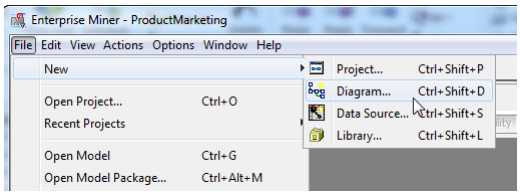
Create a New Diagram

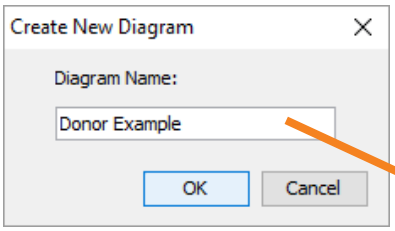
Right Mouse Click on Diagram and Select Create Diagram

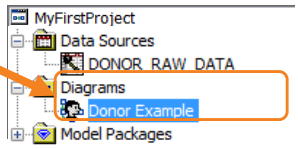
OR

Go to the Menu and select File→ New → Diagram





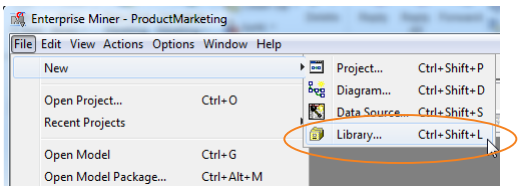


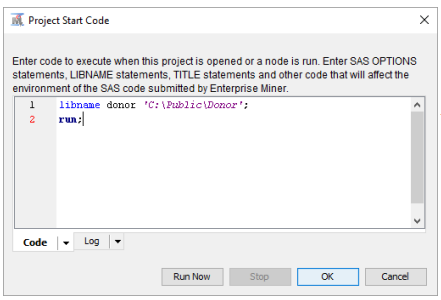


To define a LIBNAME

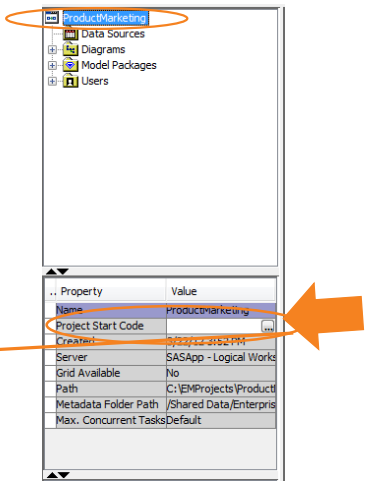
To assign a LIBNAME through a Wizard Click on File→New→Library

OR





To assign by code use Project Start Code Property. Highlight the project name.



Click on the ellipse next to Project Start Code

Process Flow Demo

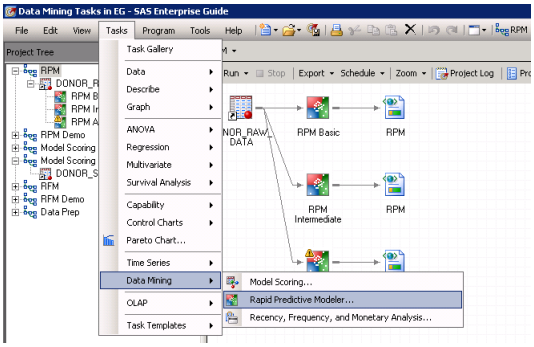
Starts at 10:13

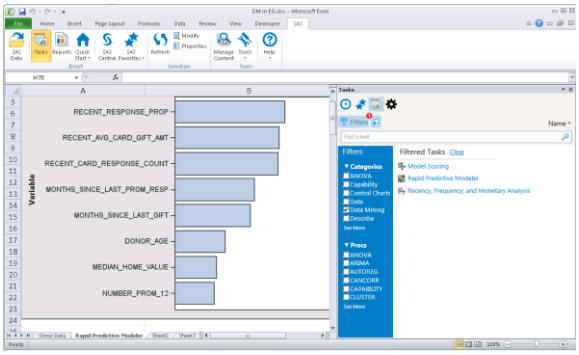
SAS® Rapid Predictive Modeler

SAS Rapid Predictive Modeler

Packaging

- Included in the SAS Enterprise Miner bundle at no additional charge
- Delivered as a customized task in SAS Enterprise Guide, SAS Add-in for MS Office (MS Excel only) and SAS Studio

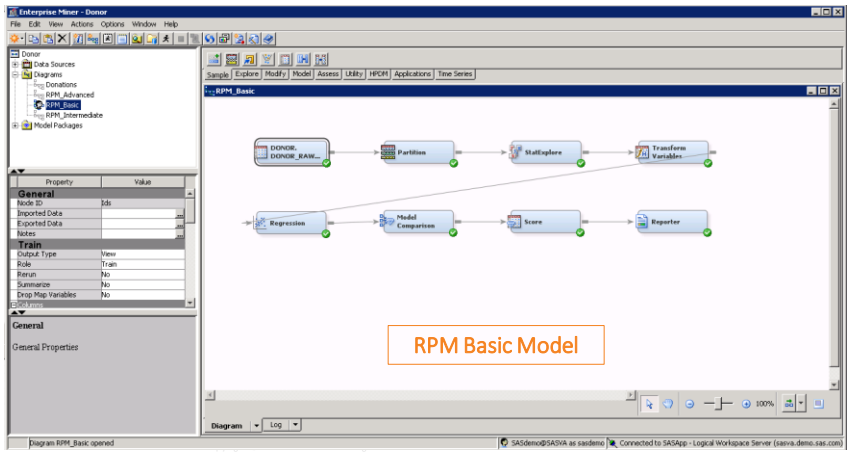




SAS Rapid Predictive Modeler

Key Capabilities

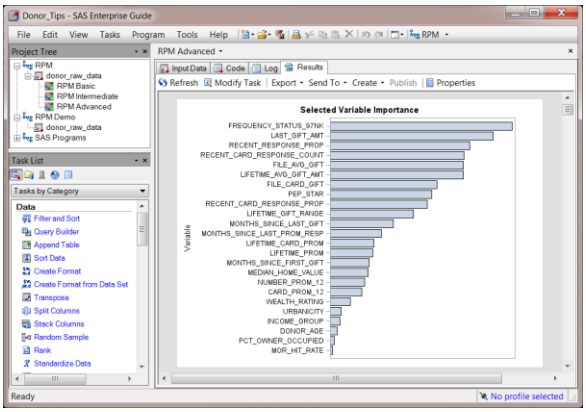
- Chose from prebuilt Enterprise Miner model flows that use a broad range of classical and modern modeling techniques.
- Analytic experts can further customize and improve SAS RPM developed models using SAS Enterprise Miner.

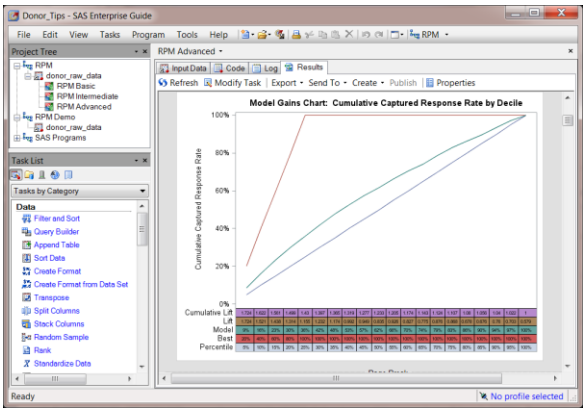


SAS Rapid Predictive Modeler

Key Capabilities

- Analytic results are presented as a simple to understand reports including a scorecard, lift charts, and listing of key variables.

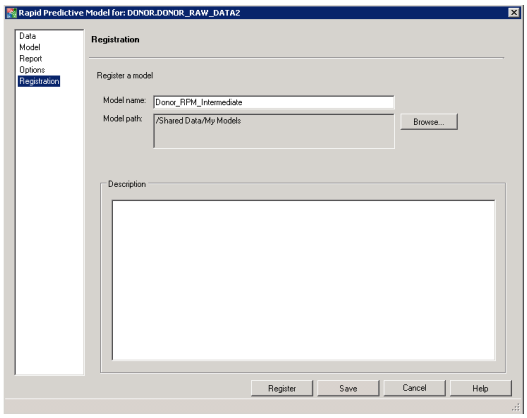


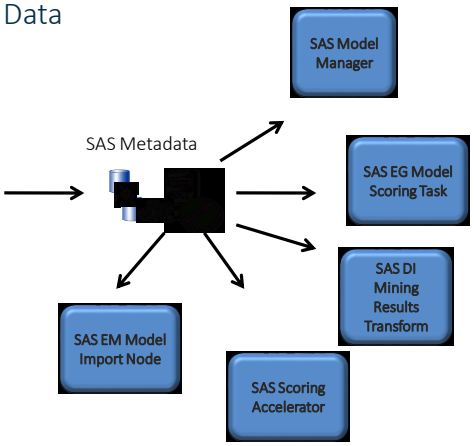


SAS Rapid Predictive Modeler

Key capabilities

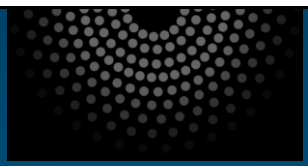
- Models can be registered in SAS metadata for direct use in other products such as SAS® Enterprise Guide, SAS® Data Integration Studio, and SAS® Model Manager.





SAS Rapid Predictive Modeler Demo

10 Minute Break



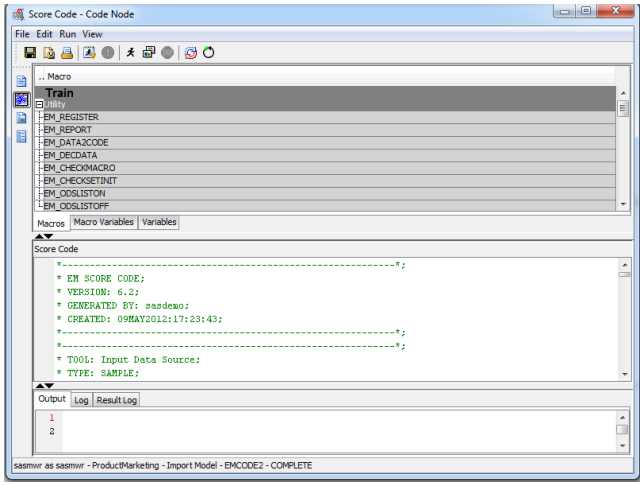
Champion Challenger Models

Integrating new algorithms, previous models and models created outside SAS Enterprise Miner

Integrating Outside Models
SAS Code Node

Enables you to incorporate new or existing SAS code into process flow diagrams that were developed using SAS Enterprise Miner.

- Train
- Score
- Report

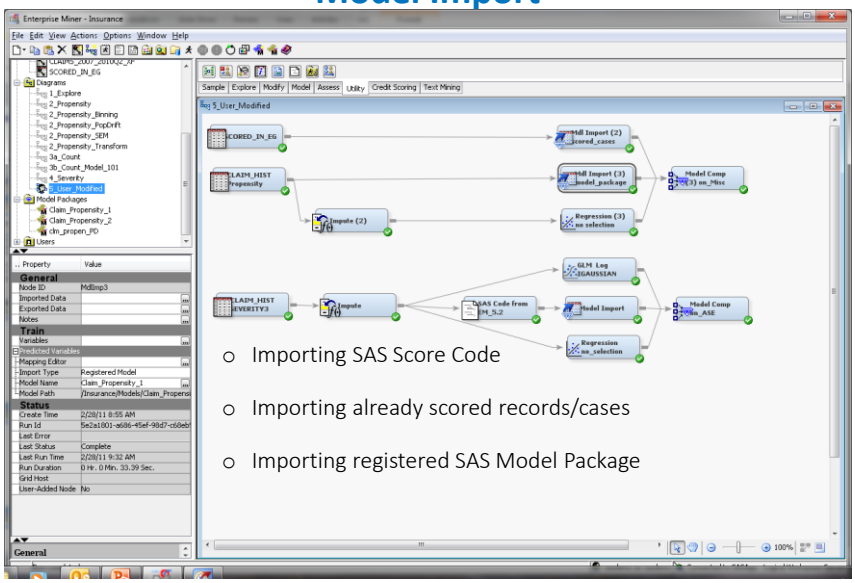


Integrating Outside Models
Model Import

The Model Import node enables you to import models into the Enterprise Miner environment that were not created by Enterprise Miner. Models that were created by using SAS PROC LOGISTIC, for example, can now be run, assessed, and modified in Enterprise Miner.

Integrating Outside Models

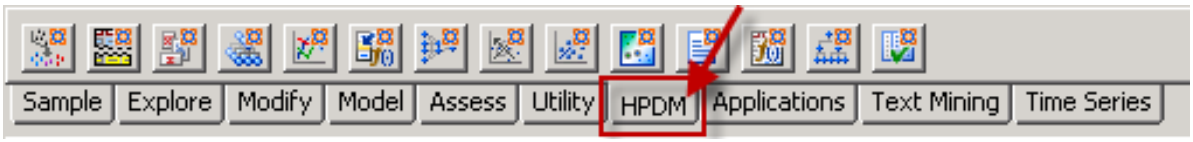
Model Import



- Importing SAS Score Code
- Importing already scored records/cases
- Importing registered SAS Model Package

Modeling Algorithms

High Performance Nodes



HP BN Classifier
HP Cluster
HP Data Partition
HP Explore
HP Forest
HP GLM
HP Impute
HP Neural
HP Principal Components
HP Regression
HP SVM
HP Text Miner
HP Transform
HP Tree
HP Variable Selection

New in starting in SAS® Enterprise Miner™ 12.3

Champion Challenger Models Demo

Model Import

Open Source

SAS and Open Source

R integration
SAS® Enterprise Miner

- SAS Enterprise Miner Open Source Integration node
  - Enables the **execution of R code** within an Enterprise Miner flow
    - Facilitates **multitasking** in R
    - Generates **text and graphical output** from R
    - Integrates both **supervised and unsupervised** learning tasks
  - Transfers data, metadata, and results automatically between Enterprise Miner and R
  - Uses SAS/IML under the covers

Open source integration node
Modes of operation

- Training Mode
  - Supervised
  - Unsupervised
- Output Mode
  - PMML: Creates SAS Data step score code
  - Merge: Merge inputs with predictions
  - None: Troubleshooting R code, output graphs, simulations etc

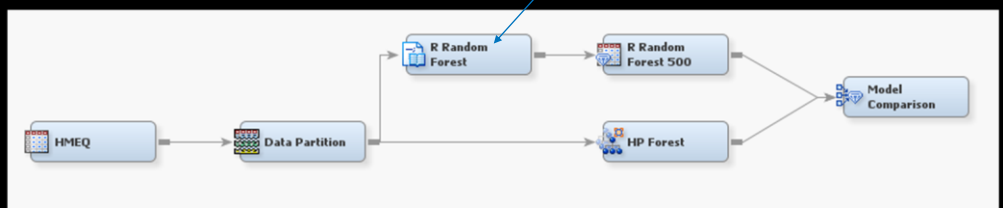
Property	Value
General	
Node ID	EMOPEN
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Code Editor	
Language	R
Training Mode	Supervised
Output Mode	PMML

SAS and Open Source

SAS Enterprise Miner

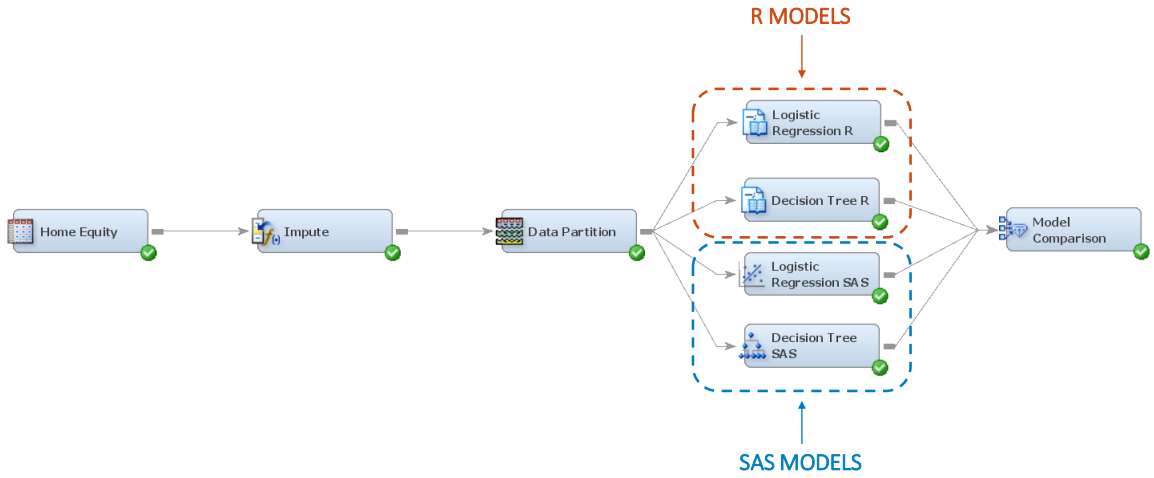
- Integrate R code inside Enterprise Miner
  - Includes R models in model assessment
  - SAS and R ensemble models

```
library(randomForest)

&EMR_MODEL <- randomForest(&EMR_CLASS_TARGET ~  &EMR_CLASS_INPUT + &EMR_NUM_INPUT, ntree= 500, mtry= 5, data= &EMR_IMPORT_DATA, importance= TRUE)

&EMR_EXPORT_TRAIN <- predict(&EMR_MODEL, &EMR_IMPORT_DATA, type="prob")


&EMR_EXPORT_VALIDATE <- predict(&EMR_MODEL, &EMR_IMPORT_VALIDATE, type="prob")
&EMR_EXPORT_TEST <- predict(&EMR_MODEL, &EMR_IMPORT_TEST, type="prob")

&EMR_EXPORT_TRAIN[1:10,]

png("EMR_forestMeePlot.png")
plot(&EMR_MODEL, main= 'randomForest MSE Plot')
dev.off()

write.table(round(importance(&EMR_MODEL),2), file= "EMR_forestImportance.csv", sep= ",", row.names= TRUE, col.names= TRUE)

print(&EMR_MODEL)
  round(importance(&EMR_MODEL),2)
```



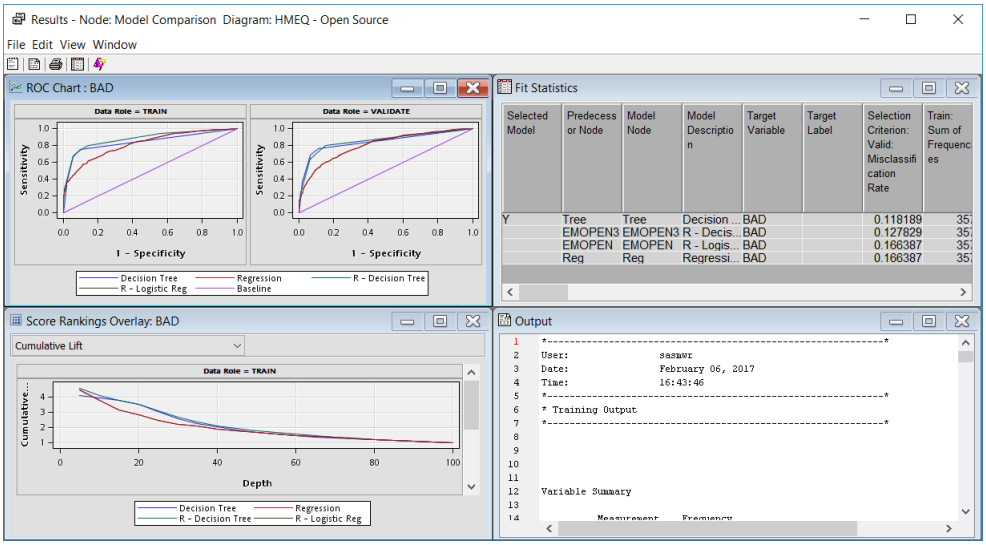
SAS and Open Source

Discovery – R in Enterprise Miner



SAS and Open Source

SAS and R Model Comparison



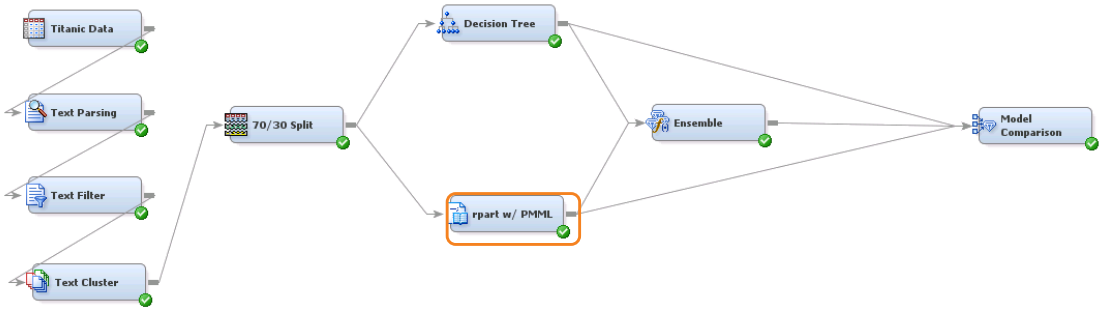
SAS and Open Source

SAS and R Model Comparison

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error	Train: Root Average Squared Error	Train: Divisor for ASE	Train: Total Degrees of Freedom	Valid: Sum of Frequencies
Y	Tree	Tree	Decision ...	BAD		0.118189	3574	0.114997	0.935823	655.6948	0.091731	0.302872	7148	3574	2386
	EMOPEN3	EMOPEN3	R - Decis...	BAD		0.127829	3574	0.113318	0.964939	632.8894	0.088541	0.297558	7148	.	2386
	EMOPEN	EMOPEN	R - Logis...	BAD		0.166387	3574	0.148573	0.999547	821.2193	0.114888	0.338951	7148	.	2385
	Reg	Reg	Regressi...	BAD		0.166387	3574	0.148293	0.999547	821.2158	0.114887	0.338951	7148	3574	2386

Using R in SAS Enterprise Miner
PMML output mode



Using R in SAS Enterprise Miner
PMML output mode

```
library(rpart)

&EMR_MODEL <- rpart(&EMR_CLASS_TARGET ~ &EMR_CLASS_INPUT +
&EMR_NUM_INPUT, data= &EMR_IMPORT_DATA, method= "class")
```

Using R in SAS Enterprise Miner

PMML output mode

```
library(rpart)

&EMR_MODEL <- rpart(&EMR_CLASS_TARGET ~ &EMR_CLASS_INPUT +
&EMR_NUM_INPUT, data= &EMR_IMPORT_DATA, method= "class")
```

SAS Enterprise Miner data source

Using R in SAS Enterprise Miner

PMML output mode

SAS Enterprise Miner metadata

```
library(rpart)

&EMR_MODEL <- rpart(&EMR_CLASS_TARGET ~ &EMR_CLASS_INPUT +
&EMR_NUM_INPUT, data= &EMR_IMPORT_DATA, method= "class")
```

SAS Enterprise Miner data source

Using R in SAS Enterprise Miner

PMML output mode

SAS Enterprise Miner metadata

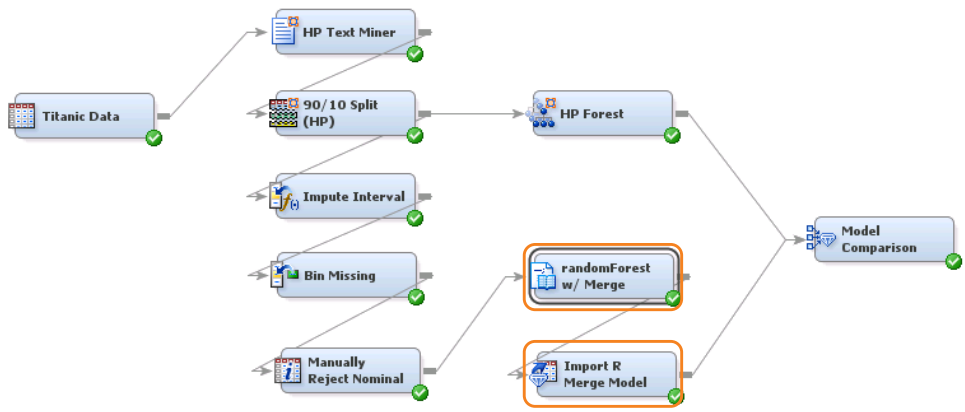
```
library(rpart)

&EMR_MODEL <- rpart(&EMR_CLASS_TARGET ~ &EMR_CLASS_INPUT +
&EMR_NUM_INPUT, data= &EMR_IMPORT_DATA, method= "class")
```

R object translated to SAS DATA step code using PMML

SAS Enterprise Miner data source

Using R in SAS Enterprise Miner

Merge output mode



Using R in SAS Enterprise Miner
Merge output mode

```
library(randomForest)

&EMR_MODEL <- randomForest(&EMR_CLASS_TARGET ~ &EMR_CLASS_INPUT + &EMR_NUM_INPUT, ntree= 250, mtry= 5, maxnodes= 50, data= &EMR_IMPORT_DATA,
importance= TRUE)

&EMR_EXPORT_TRAIN <- predict(&EMR_MODEL, &EMR_IMPORT_DATA, type="prob")
&EMR_EXPORT_VALIDATE <- predict(&EMR_MODEL, &EMR_IMPORT_VALIDATE, type="prob")

&EMR_EXPORT_TRAIN[1:10,]
```

Best practice

User must explicitly create exported variables to be merged with Enterprise Miner data sources

Open Source Integration
Other Benefits of Open Source in Enterprise miner

Multi-Threaded Processing of Workflows
- Use Open Source node in various flows simultaneously
- Enterprise Miner handles multi-threaded execution

Scoring
- Create supported models in R that can converted into scoring code for operational deployment (i.e. in-database)

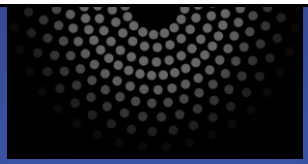
Collaboration
- Many users can access the same EM diagram

General Use
- Use any of R's packages within Enterprise Miner (EM), leveraging all EM functionality in prior nodes (i.e. data prep, prep-processing)

Open Source Integration

- DATA STEP COMPONENT
- ESSENTIALLY A SAS LANGUAGE API INTO JAVA:
  - Instantiates Java classes and calls Java methods with arguments from the DATA step (Supported since 9.1.3)

Tip: How to execute a Python script in SAS® Enterprise Miner™



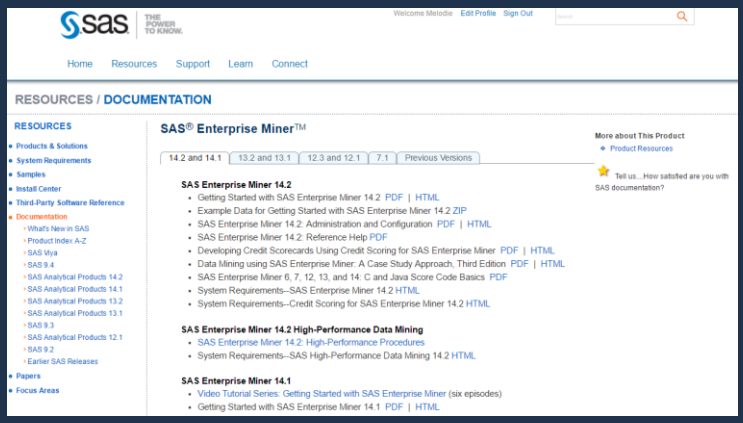
Open Source Demo

SAS and Open Source

Resources

SAS® Enterprise Miner™
Getting Started Documentation

- Using same data from "Getting Started with SAS® Enterprise Miner™" documentation
- Both the data and the documentation are available on support.sas.com http://support.sas.com/documentation/onlinedoc/miner/



****Tab and Scroll to find your version*

Resources

- SAS Enterprise Miner Learning Page Brand New!!
  - SAS Enterprise Miner Technical Support Web Site
  - SAS Enterprise Miner Training
  - Data Mining and Machine Learning Community Join Today
- E-Learning Class for SAS Rapid Predictive Modeler (RPM)
  - Rapid Predictive Modeling for Business Analysts
- YouTube SAS Rapid Predictive Modeler Videos
  - A Modeling Sampler, Part 1
  - A Modeling Sampler, Part 2
  - A Modeling Sampler, Part 3
  - A Modeling Sampler, Part 4
  - Creating Data Miner Models Using SAS Studio and the Rapid Predictive Modeler Task

GitHub

Examples

SAS Enterprise Miner Process Flow Diagrams

- GitHub Examples - https://github.com/sassoftware/dm-flow
- Video – Learn by Example https://www.youtube.com/watch?v=oSLrkvQH7iU

Video Resources
Modeling Techniques

- Getting Started with SAS Enterprise Miner Tutorial Videos
  - https://www.youtube.com/playlist?list=PLVBcK_IpFVi-xzvJiOlf33UvVbRoLRu0z
- Incremental Response Modeling
  - https://www.youtube.com/watch?v=zabWaSS_BDI
- The New HP GLM Node
  - https://www.youtube.com/watch?v=88qWDc1pGUU
- Making Machine Learning Manageable
  - https://www.youtube.com/watch?v=88qWDc1pGUU
- Random Forest and Support Vector Machines
  - https://www.youtube.com/watch?v=EOxwpnbFqIU
- Deep Learning in SAS Enterprise Miner
  - https://www.youtube.com/watch?v=HOEqvyyuPrk

SAS and Open Source
Video Demos / Information

Empowering the SAS Enterprise Miner user

Video: *Using R in SAS Enterprise Miner*
https://www.youtube.com/watch?v=TbXo0xQCqDw

Blogs: *Spectral Clustering in SAS® Enterprise Miner™ Using Open Source Integration Node*
https://communities.sas.com/docs/DOC-8011

Blogs: *How to execute a Python script in SAS® Enterprise Miner™*
https://communities.sas.com/docs/DOC-10832

Article: *The Open Source Integration node installation cheat sheet*
https://communities.sas.com/docs/DOC-9988

Usage Notes:
http://support.sas.com/dsearch?Find=Search&ct=&qt=open+source&col=supppprd&nh=25&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0

Resources
SAS Courses

- Predictive Modeling Using SAS High-Performance Analytics Procedures
- Predictive Modeling Using Logistic Regression
- Applied Analytics Using SAS Enterprise Miner
- Data Mining: Principles and Best Practices
- SAS Enterprise Miner High-Performance Data Mining Nodes
- Data Mining Techniques: Theory and Practice

For a complete list of courses, please see
https://support.sas.com/edu/courses.html?ctry=us

Learning More

Decision Trees for Analytics Using SAS® Enterprise Miner™

By: Barry de Ville and Padraic Neville

ISBN: 978-1-61290-315-6

Copyright Date: July 2013

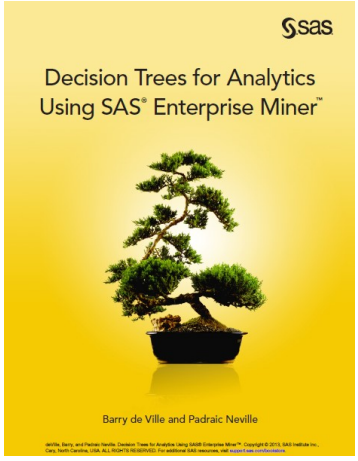
SAS Bookstore: https://www.sas.com/store/prodBK_63319_en.html

Table of Contents [PDF]

Free Chapter [PDF]

Example Code and Data

Available on Amazon



Learning More

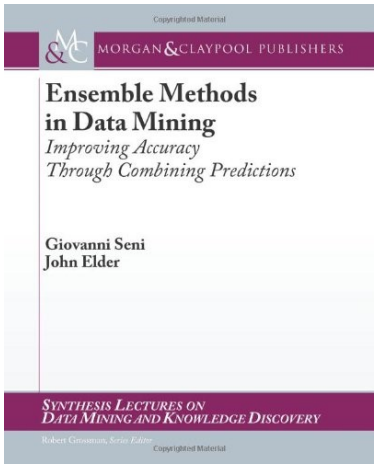
Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions

By: Giovanni Seni & John Elder

ISBN-10: 1608452840

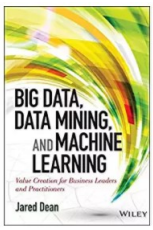
Publisher: Morgan and Claypool Publishers (February 24, 2010)

Available on Amazon



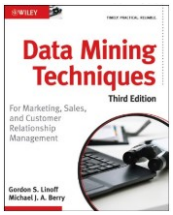
Resources

Suggested Reading



Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners
By Jared Dean

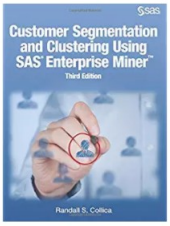
Available on Amazon



Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management
by Gordon S. Linoff and Michael J. A. Berry

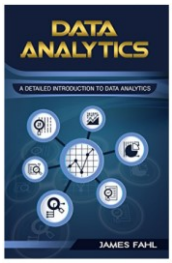
Available on Amazon

Resources
Suggested Reading



Customer Segmentation and Clustering Using SAS Enterprise Miner, Third Edition
By Randall S. Collica

Available on Amazon



Data Analytics: A Practical Guide To Data Analytics For Business, Beginner To Expert(Data Analytics, Prescriptive Analytics, Statistics, Big Data, Intelligence, ... Master Data, Data Science, Data Mining)
by James Fahl

Available on Amazon

Questions?

Thank you for your time and attention!

sas.com

CUSTOMER LOYALTY TEAM • Support You Can Count On