



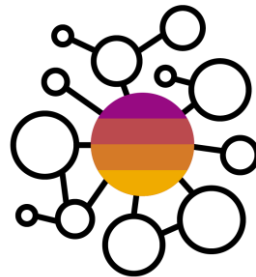
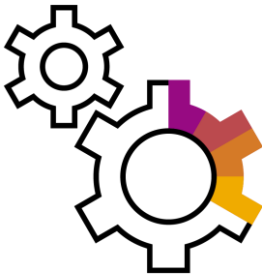
Week 3: Develop, Evaluate, and Deploy Models

Unit 1: Data Description and Data Roles

Data Description and Data Roles

Introduction

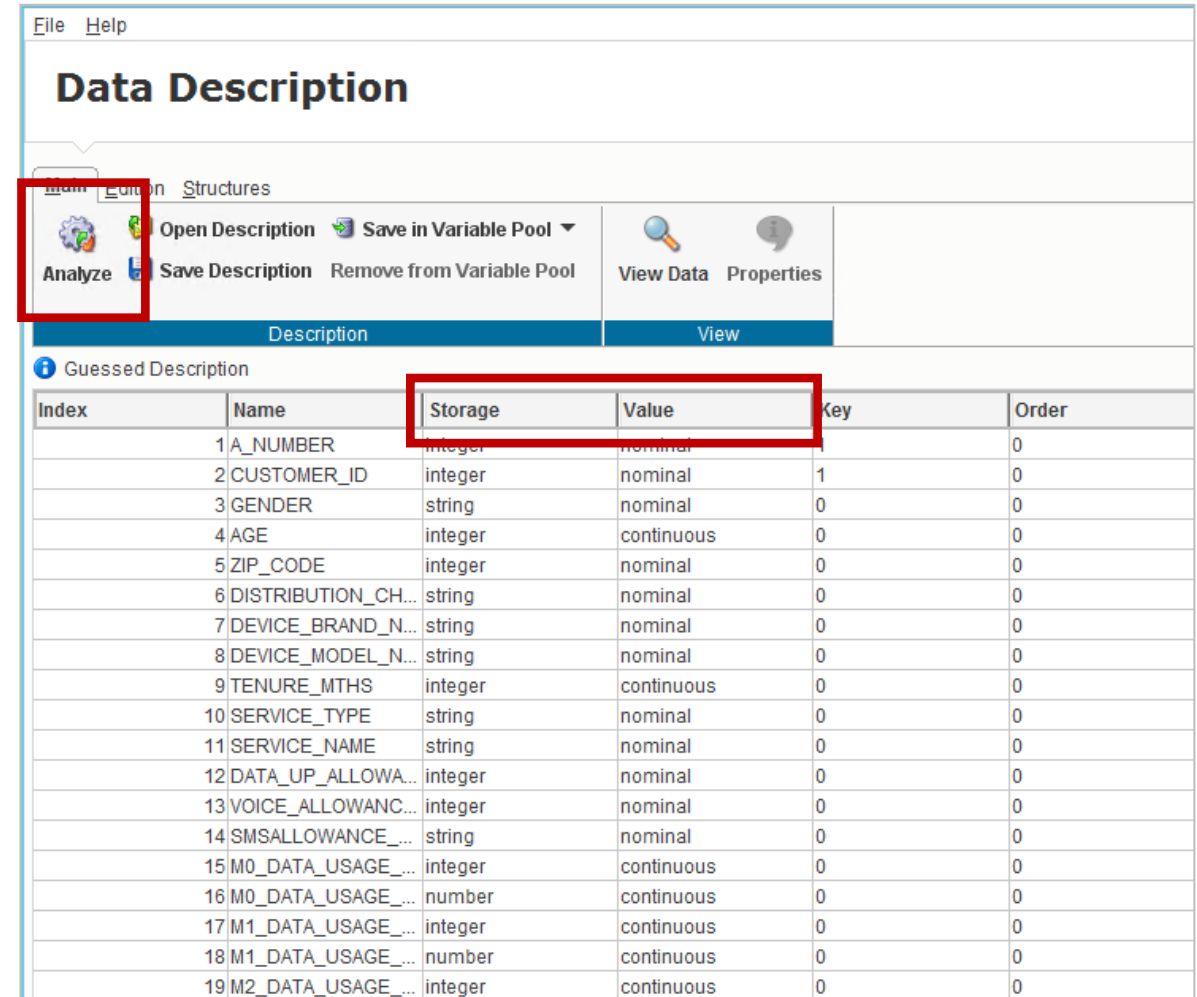
- In SAP Predictive Analytics automated modeling, a variable is defined by three aspects:
 1. The type of variable (continuous, nominal, or ordinal)
 2. The storage format of the variable (for example, date, number, or string)
 3. The role of the variable: Target, Explanatory, or Weight



Data Description and Data Roles

Data description

- To optimize the data encoding, you must check that the data description is correct.
- **Analyze** will scan the first hundred lines of data and provide you with an initial guess of the data file description.
- If the guessed description is not correct, click the **Storage** and/or **Value** columns to edit the variable description.



The screenshot shows the 'Data Description' application window. The 'Analyze' button in the toolbar is highlighted with a red box. Below the toolbar, the 'Gussed Description' table is displayed. The 'Storage' and 'Value' columns in this table are highlighted with a red box.

Index	Name	Storage	Value	Key	Order
1	A_NUMBER	integer	nominal		0
2	CUSTOMER_ID	integer	nominal	1	0
3	GENDER	string	nominal	0	0
4	AGE	integer	continuous	0	0
5	ZIP_CODE	integer	nominal	0	0
6	DISTRIBUTION_CH...	string	nominal	0	0
7	DEVICE_BRAND_N...	string	nominal	0	0
8	DEVICE_MODEL_N...	string	nominal	0	0
9	TENURE_MTHS	integer	continuous	0	0
10	SERVICE_TYPE	string	nominal	0	0
11	SERVICE_NAME	string	nominal	0	0
12	DATA_UP_ALLOWA...	integer	nominal	0	0
13	VOICE_ALLOWANC...	integer	nominal	0	0
14	SMSALLOWANCE_...	string	nominal	0	0
15	M0_DATA_USAGE_...	integer	continuous	0	0
16	M0_DATA_USAGE_...	number	continuous	0	0
17	M1_DATA_USAGE_...	integer	continuous	0	0
18	M1_DATA_USAGE_...	number	continuous	0	0
19	M2_DATA_USAGE_...	integer	continuous	0	0

Data Description and Data Roles

Variable storage



- To describe the data, the application uses six types of storage formats:
 - date
 - datetime
 - number
 - integer
 - string
 - angle

The storage format...	Is used to describe variables when their values correspond to...	For instance...
date	Dates expressed in the following formats: <ul style="list-style-type: none">• YYYY-MM-DD	"2001-11-30"
datetime	Dates and times expressed in the following formats: <ul style="list-style-type: none">• YYYY-MM-DD HH:MM:SS• YYYY/MM/DD HH:MM:SS	<ul style="list-style-type: none">• "2001-11-30 14:08:17"• "1999/04/28 07:21:58"
number	Figures, or numerical values on which operations may be performed	<ul style="list-style-type: none">• The variable "salary", in US dollars: "1000.00", "1593" and "2000.54"•
integer	Figures, or numerical integer values on which operations may be performed	<ul style="list-style-type: none">• The variable "age", in years: "21", "34" and "99"
string	Alphanumeric character strings	<ul style="list-style-type: none">• The variable "family name": "Lake", "Martin" and "Miller"• The variable "occupation": "professor", "engineer" and "translator"• The variable "telephone": "800 555 1234" and "800 555 4321"

Data Description and Data Roles

Data description

- **Key:**
key variable or identifier for the record (0: the variable is not an identifier; 1: primary identifier; 2: secondary identifier...)
- This is very important when working on databases. It enables scoring optimizations in the database.
- **Missing:** string used in the data description file to represent missing values (e.g. 999, ? or #Empty)

Variable Pool ▾					
m Variable Pool		 View Data	 Properties		
		View			
Storage	Value	Key	Order	Missing	Group
integer	nominal	1			
integer	nominal	1			
string	nominal	0			
integer	continuous	0			
integer	nominal	0			
string	nominal	0			
string	nominal	0			
string	nominal	0			
integer	continuous	0			
string	nominal	0			
string	nominal	0			
integer	nominal	0			
integer	nominal	0			
string	nominal	0			
integer	continuous	0			
number	continuous	0			
integer	continuous	0			
number	continuous	0			
integer	continuous	0			

Data Description and Data Roles

Save data descriptions

- Once you are satisfied with the data descriptions, you can save them.
- **Save Description** enables you to save the description:
 - in a file or table
 - in the Variable Pool (metadata repository)

Index	Name	Storage	Value	Key	Order	Missing
1	A_NUMBER	integer	nominal	1	0	
2	CUSTOMER_ID	integer	nominal	1	0	
3	GENDER	string	nominal	0	0	
4	AGE	integer	continuous	0	0	
5	ZIP_CODE	integer	nominal	0	0	
6	DISTRIBUTION_CH...	string	nominal	0	0	
7	DEVICE_BRAND_N...	string	nominal	0	0	
8	DEVICE_MODEL_N...	string	nominal	0	0	
9	TENURE_MTHS	integer	continuous	0	0	
10	SERVICE_TYPE	string	nominal	0	0	
11	SERVICE_NAME	string	nominal	0	0	
12	DATA_UP_ALLOWA...	integer	nominal	0	0	
13	VOICE_ALLOWANC...	integer	nominal	0	0	
14	SMSALLOWANCE_...	string	nominal	0	0	
15	M0_DATA_USAGE_...	integer	continuous	0	0	
16	M0_DATA_USAGE_...	number	continuous	0	0	
17	M1_DATA_USAGE_...	integer	continuous	0	0	
18	M1_DATA_USAGE_...	number	continuous	0	0	
19	M2_DATA_USAGE_...	integer	continuous	0	0	

Data Description and Data Roles

Data roles

- **Explanatory variables:**
Model input variables used to explain the target variable
- **Target variables:**
Variable to be predicted. The engine automatically selects the last variable of the dataset as target. This might need to be changed, of course.
- **Excluded variables:**
Input variables removed from the modeling process

The screenshot shows the 'Selecting Variables' dialog box in SAP. The dialog is titled 'Selecting Variables' and has a menu bar with 'File' and 'Help'. It is divided into several sections:

- Explanatory Variables Selected 44:** A list of variables including GENDER, AGE, ZIP_CODE, DISTRIBUTION_CHANNEL_ID, DEVICE_BRAND_NAME, DEVICE_MODEL_NAME, TENURE_MTHS, SERVICE_TYPE, SERVICE_NAME, DATA_UP_ALLOWANCE_MB, VOICE_ALLOWANCE_MINUTES, SMSALLOWANCE_NUM_MESSAGES, M0_DATA_USAGE_MB, M0_DATA_USAGE_PCT, M1_DATA_USAGE_MB, M1_DATA_USAGE_PCT, M2_DATA_USAGE_MB, M2_DATA_USAGE_PCT, APR_DATA_USAGE_MB, APR_DATA_USAGE_PCT, MAY_DATA_USAGE_MB, MAY_DATA_USAGE_PCT, JUN_DATA_USAGE_MB, JUN_DATA_USAGE_PCT, CHURN_MAY, CHURN_JUN, CNT_0_VOC, CNT_1_VOC, CNT_2_VOC, M_MEAN_VOC_CNT, SUM_0_DURATION_VOC_1, SUM_1_DURATION_VOC_1, and SUM_2_DURATION_VOC_1.
- Target Variables 1:** A list containing the variable TARGET.
- Weight Variable 0:** A text input field with the value 0.
- Excluded Variables 2:** A list containing the variables A_NUMBER and CUSTOMER_ID.

At the bottom of the dialog, there are three buttons: 'Cancel', 'Previous', and 'Next'. There are also checkboxes for 'Alphabetic Sort' next to the variable lists.

Data Description and Data Roles

Summary

- In this unit, you have learnt about the different data storage formats and the data roles.
- You have also seen how to enter these details into SAP Predictive Analytics automated modeling.
- It is very important that you create the correct data description, otherwise it will have an adverse effect on your model.



Thank you.

Contact information:

open@sap.com

© 2017 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See <http://global.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.

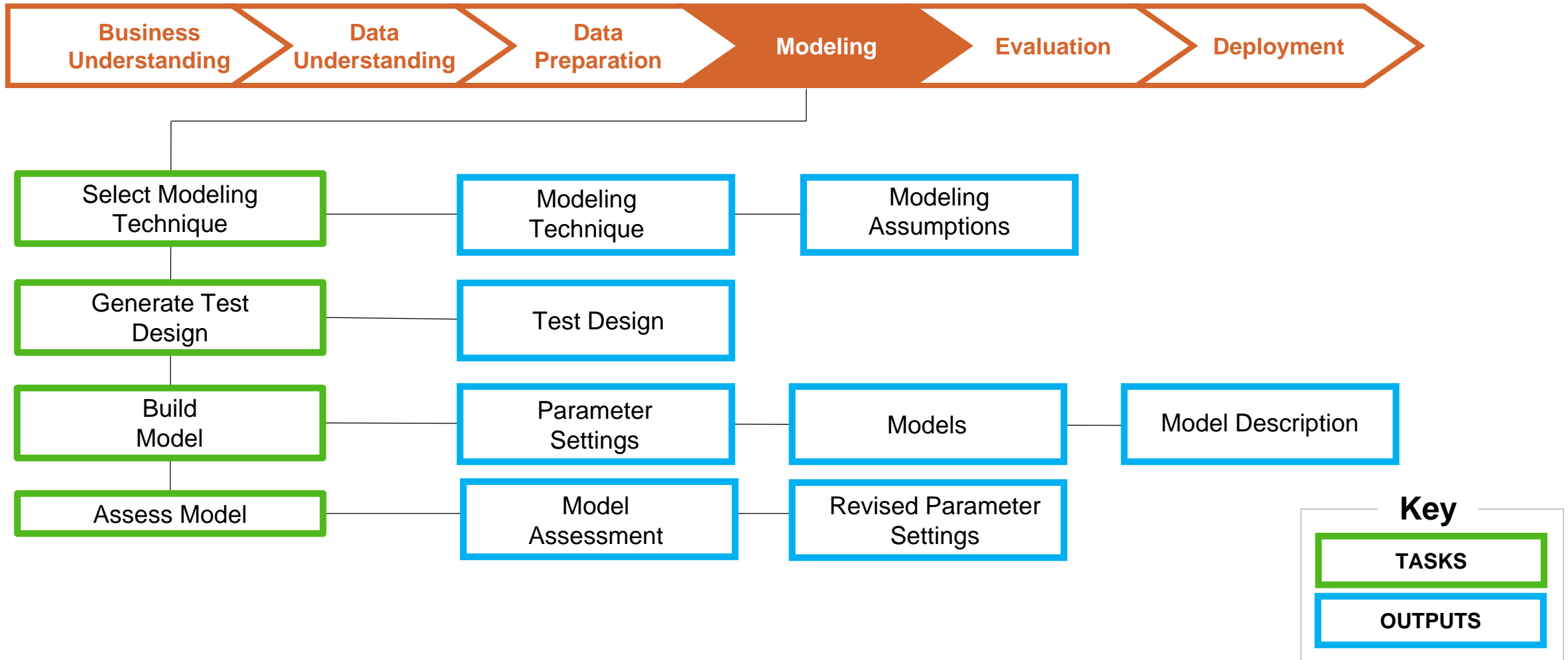


Week 3: Develop, Evaluate, and Deploy Models

Unit 2: Developing an Initial Churn Model

Developing an Initial Churn Model

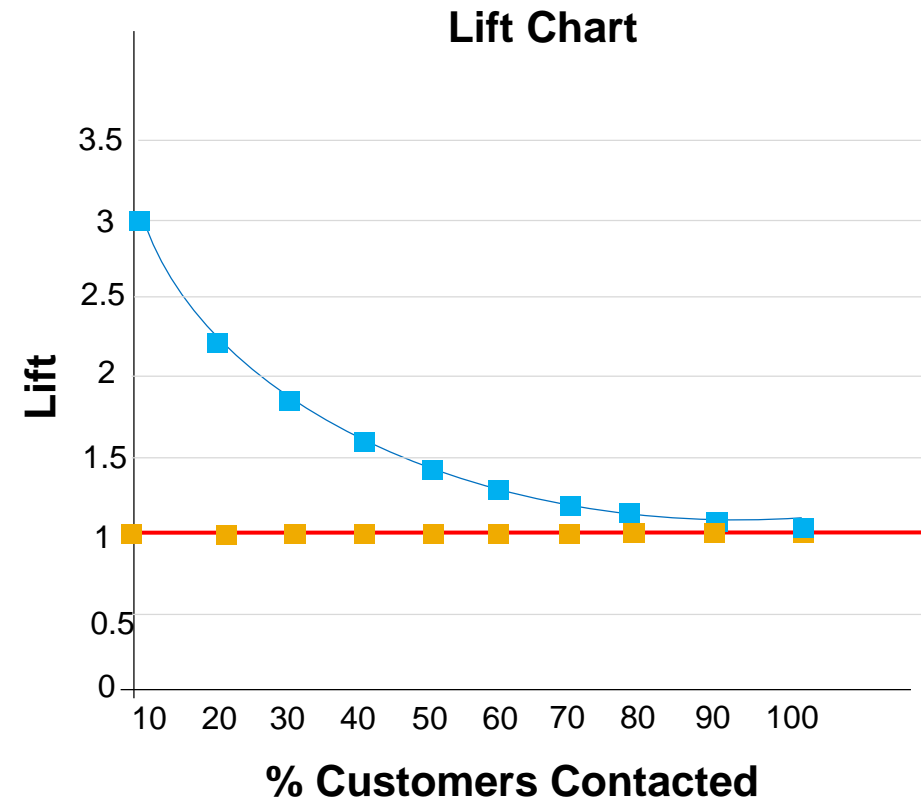
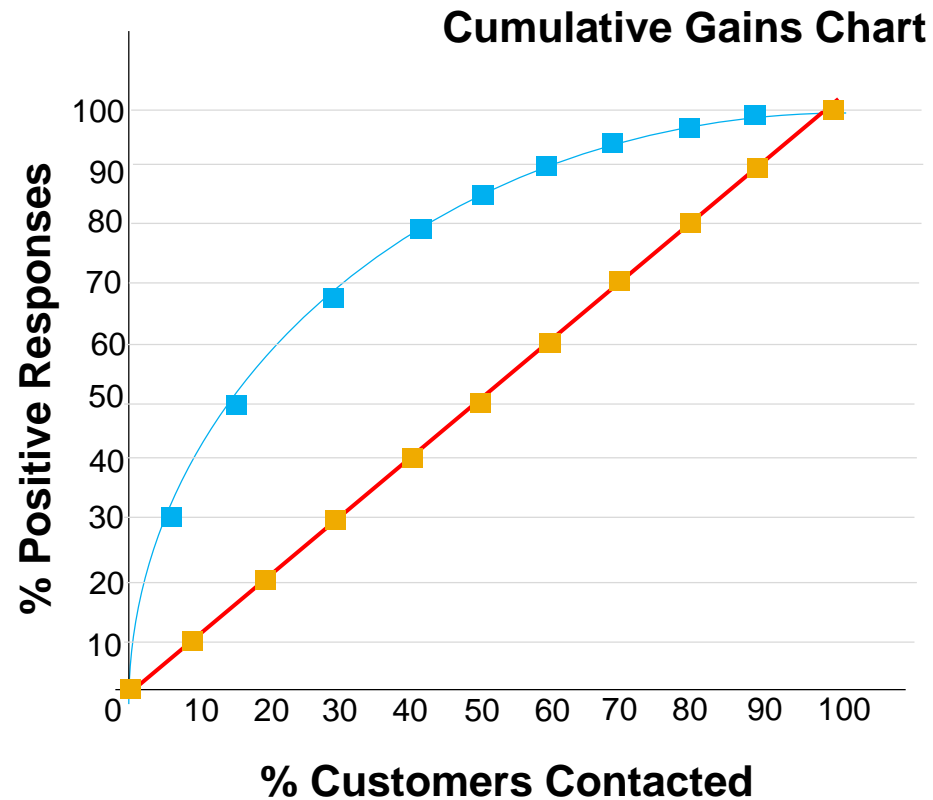
CRISP-DM – Phase 4: Modeling



Developing an Initial Churn Model

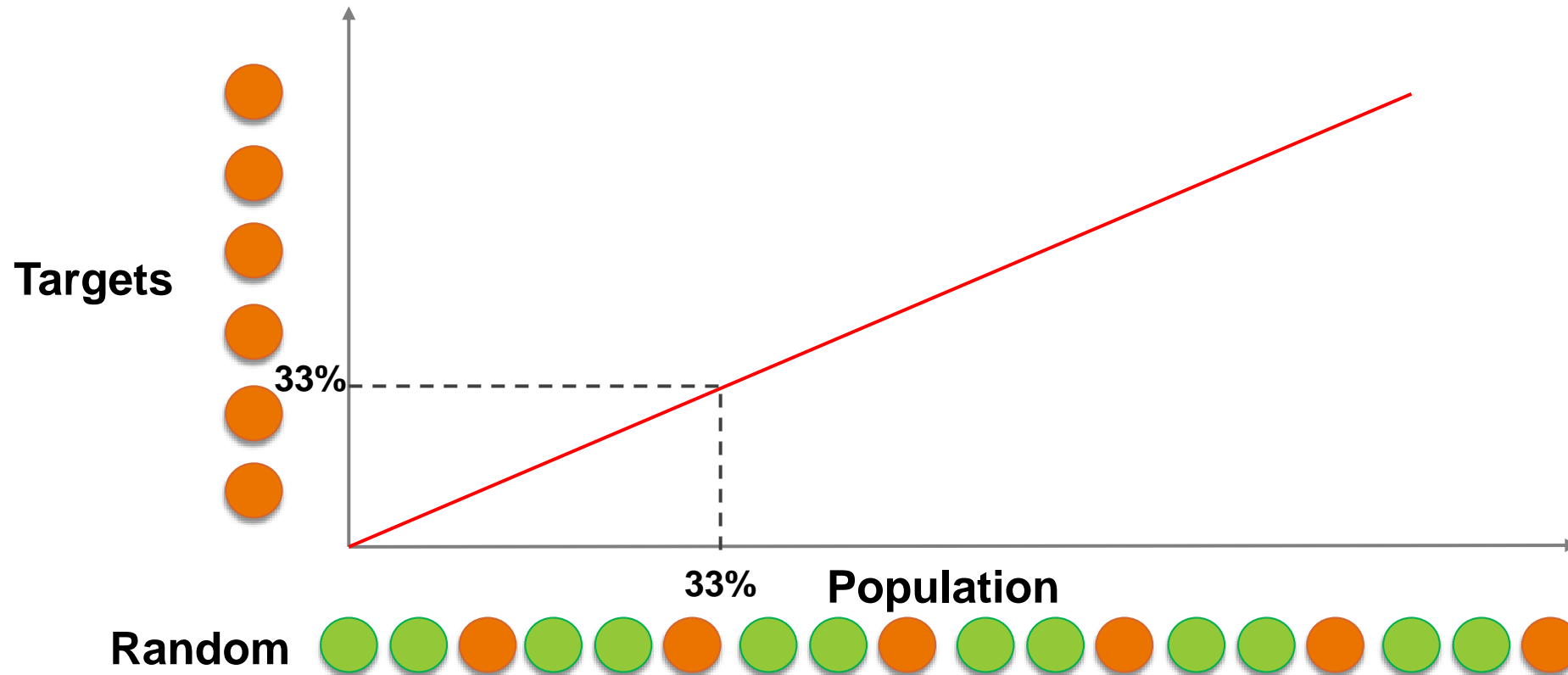
Assessing classification models using the lift and gains charts

—■— Baseline
—■— Lift Curve



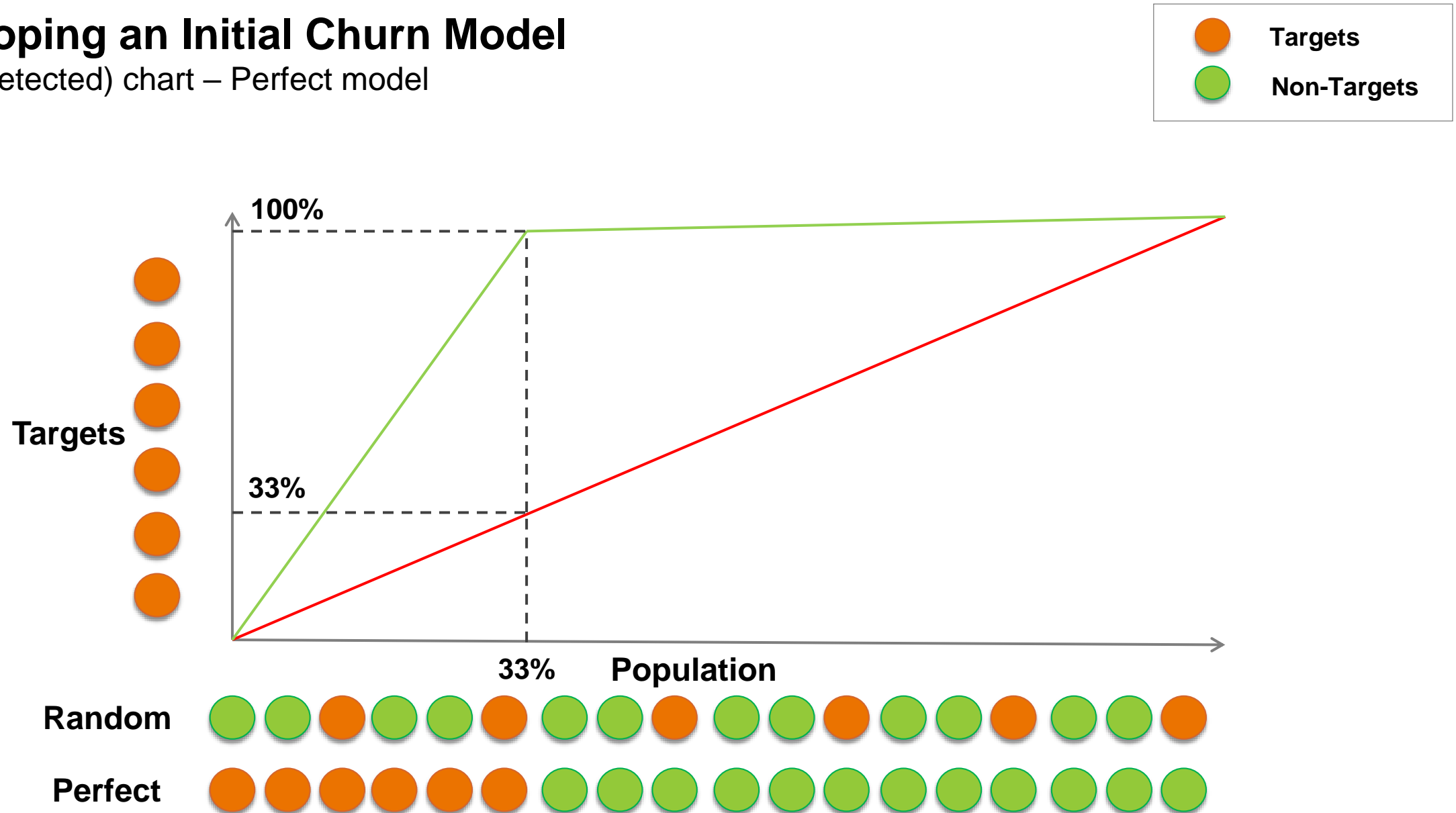
Developing an Initial Churn Model

Gains chart – Random model



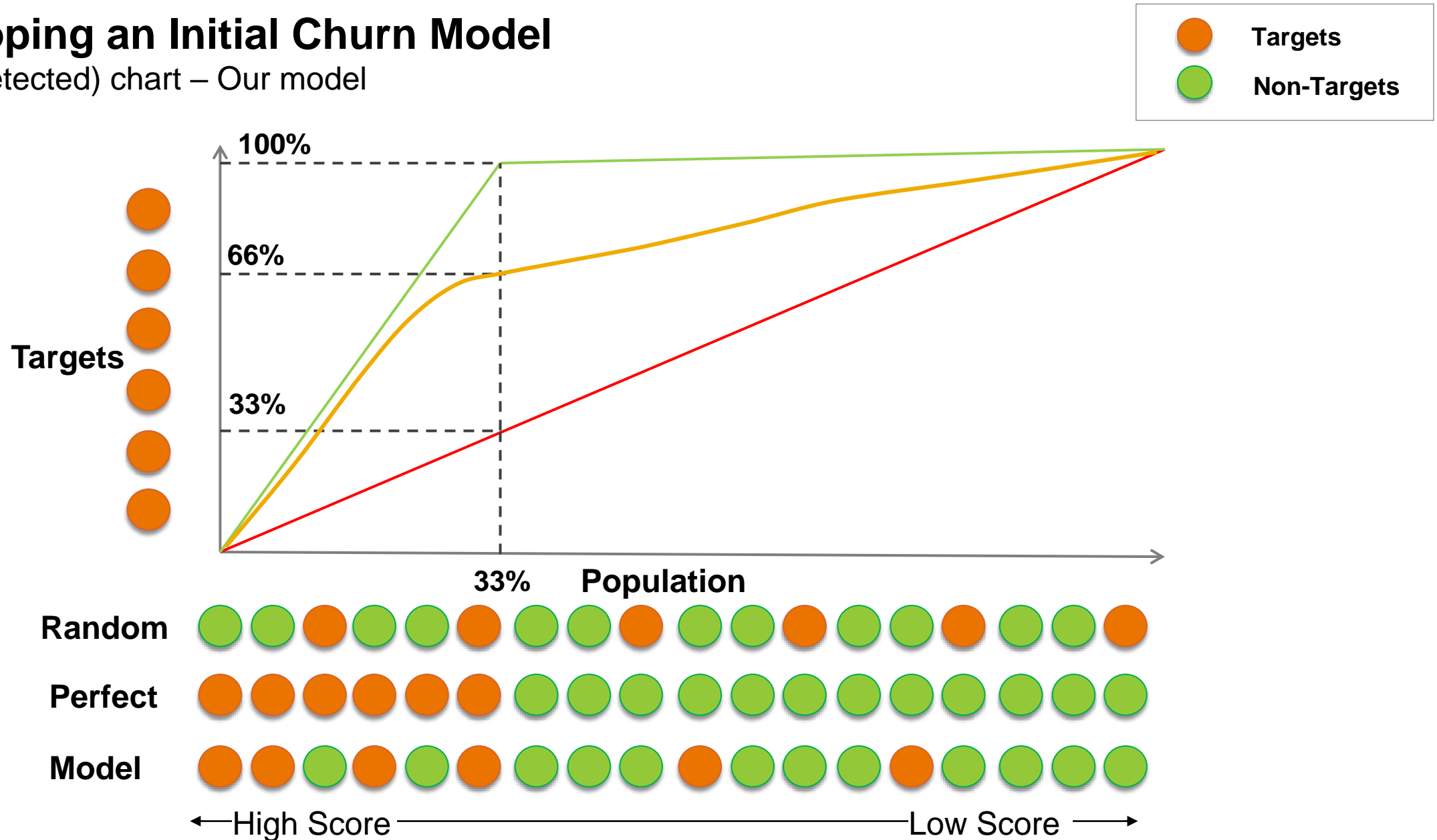
Developing an Initial Churn Model

Gains (detected) chart – Perfect model



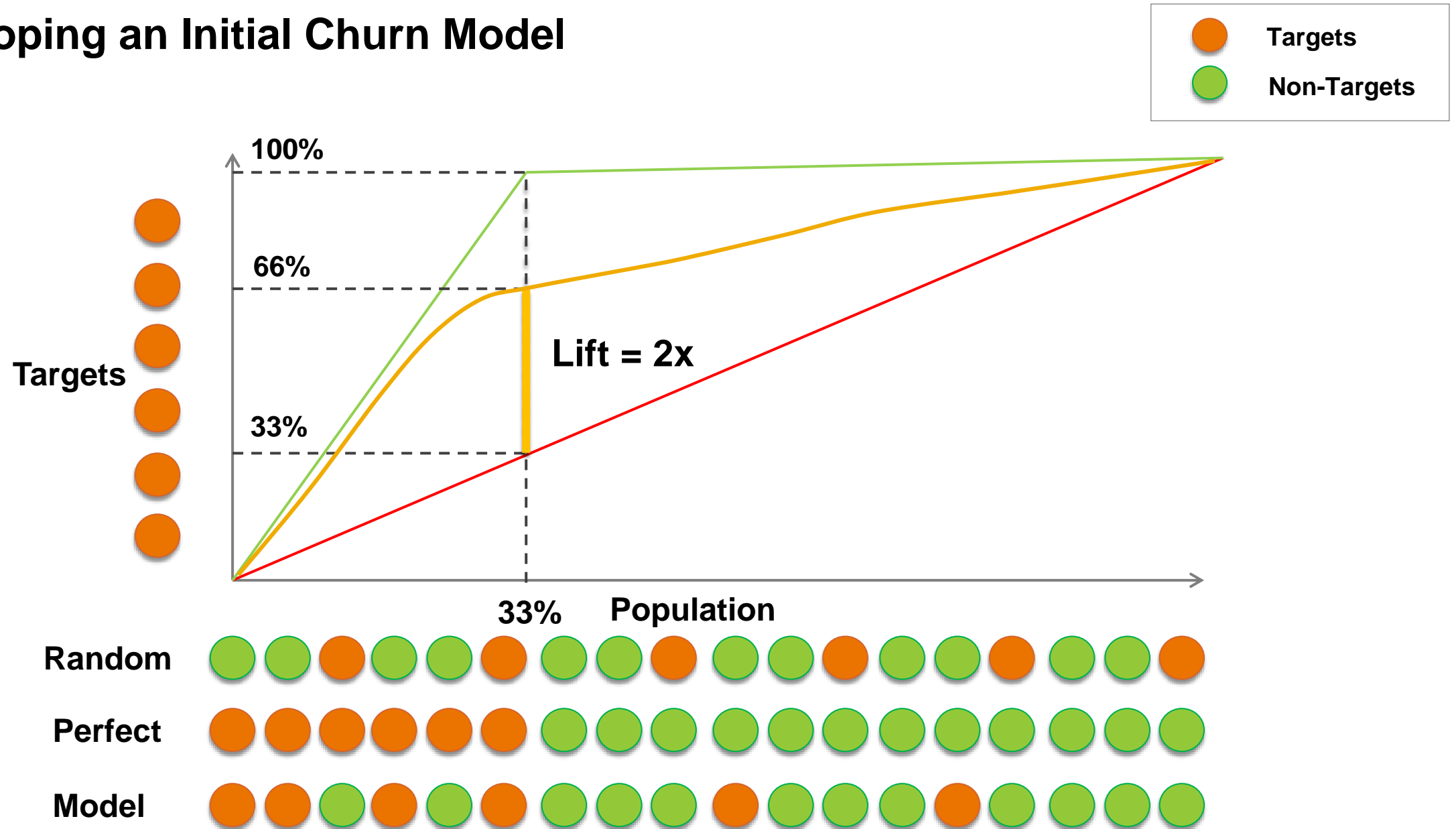
Developing an Initial Churn Model

Gains (detected) chart – Our model



Developing an Initial Churn Model

Lift chart



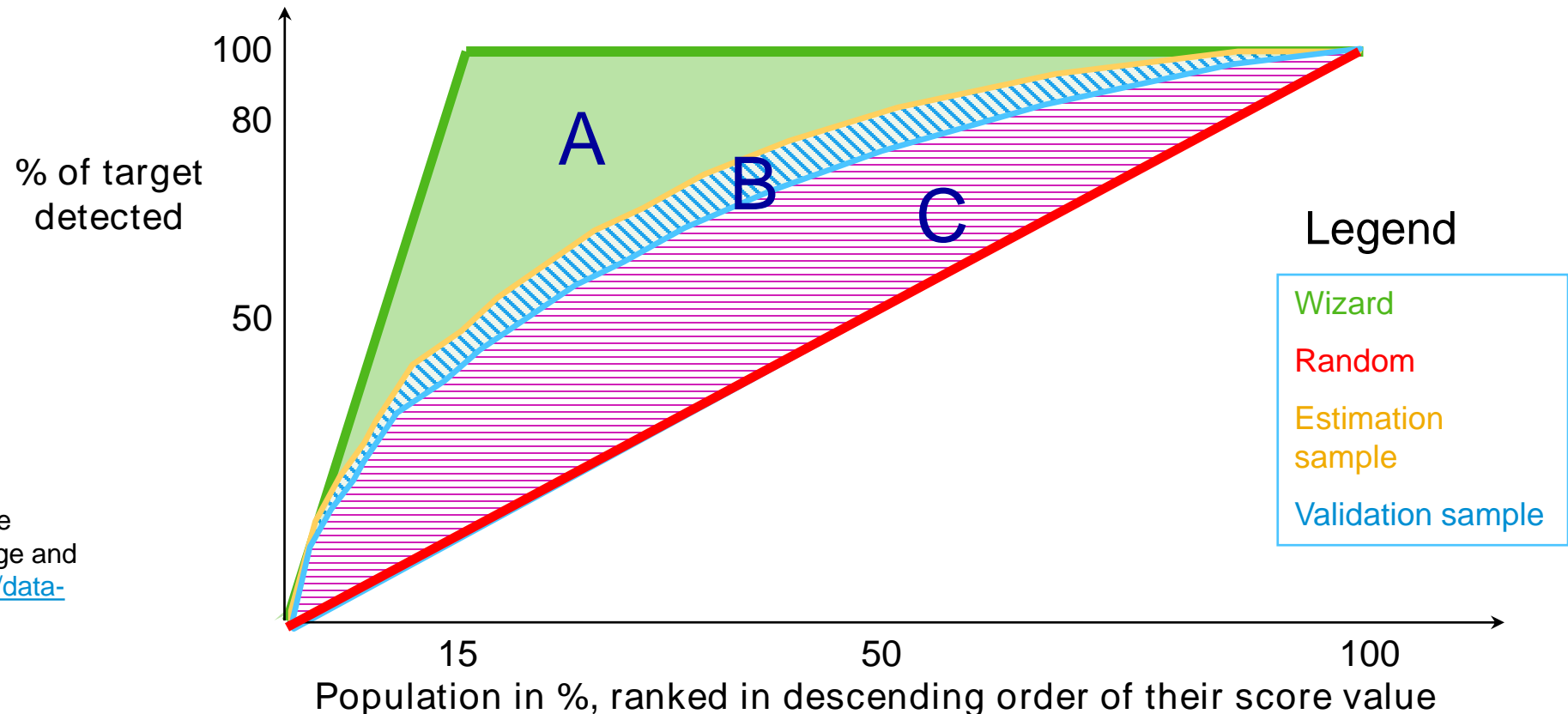
Developing an Initial Churn Model

SAP metrics – Predictive power (KI) and prediction confidence (KR)

Predictive Power (KI)_{Validation} $\approx C/(A+B+C)$

Prediction Confidence (KR) $\approx 1 - B/(A+B+C)$

Predictive Power (KI)_{Estimation} $\approx (B+C)/(A+B+C)$



(To learn about leaker variables, see <https://www.kaggle.com/wiki/Leakage> and <http://machinelearningmastery.com/data-leakage-machine-learning/>)

Developing an Initial Churn Model

Demo





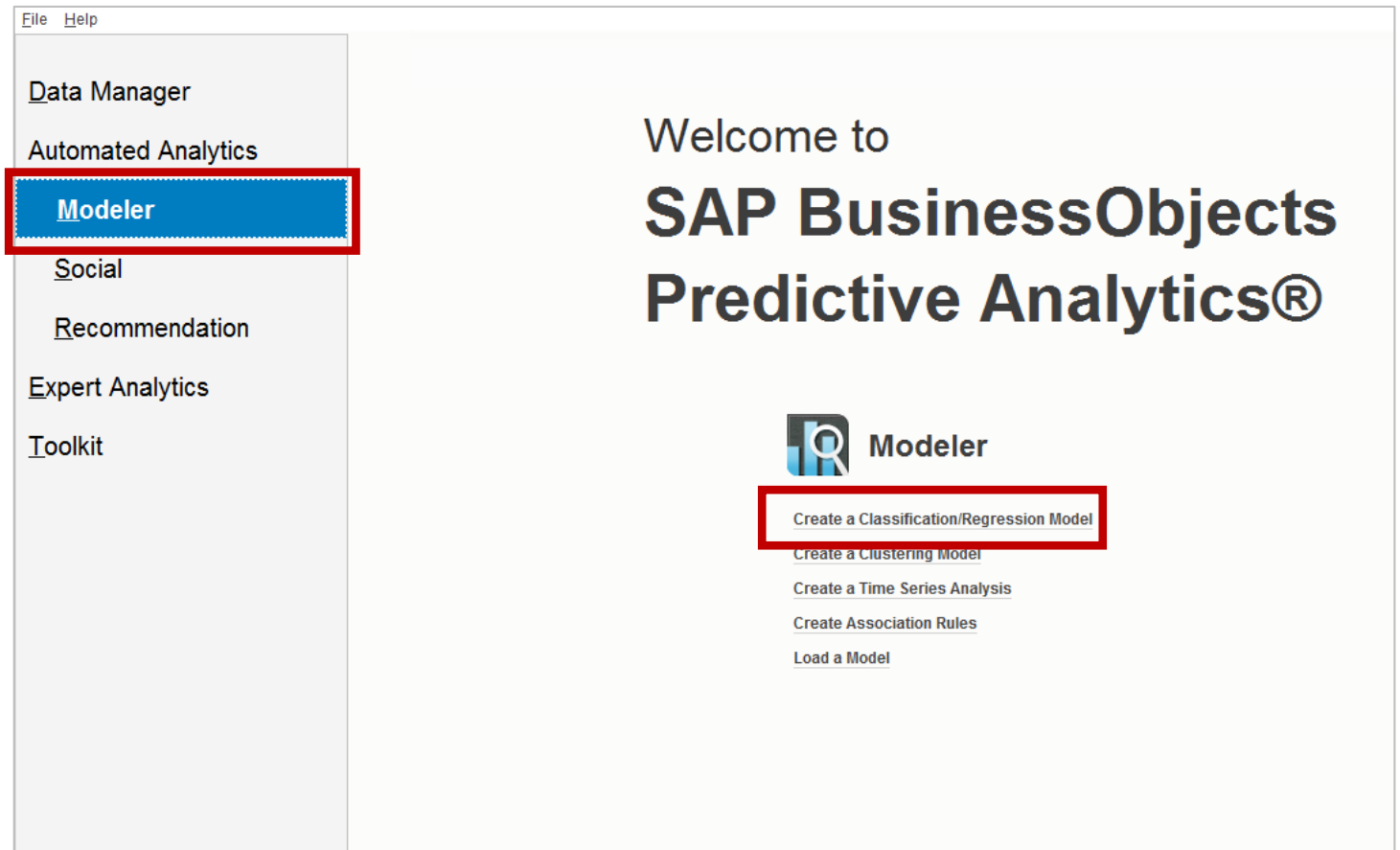
Appendix

Step-by-step guide to building a classification model

Developing an Initial Churn Model

Demo – Step-by-step guide 1

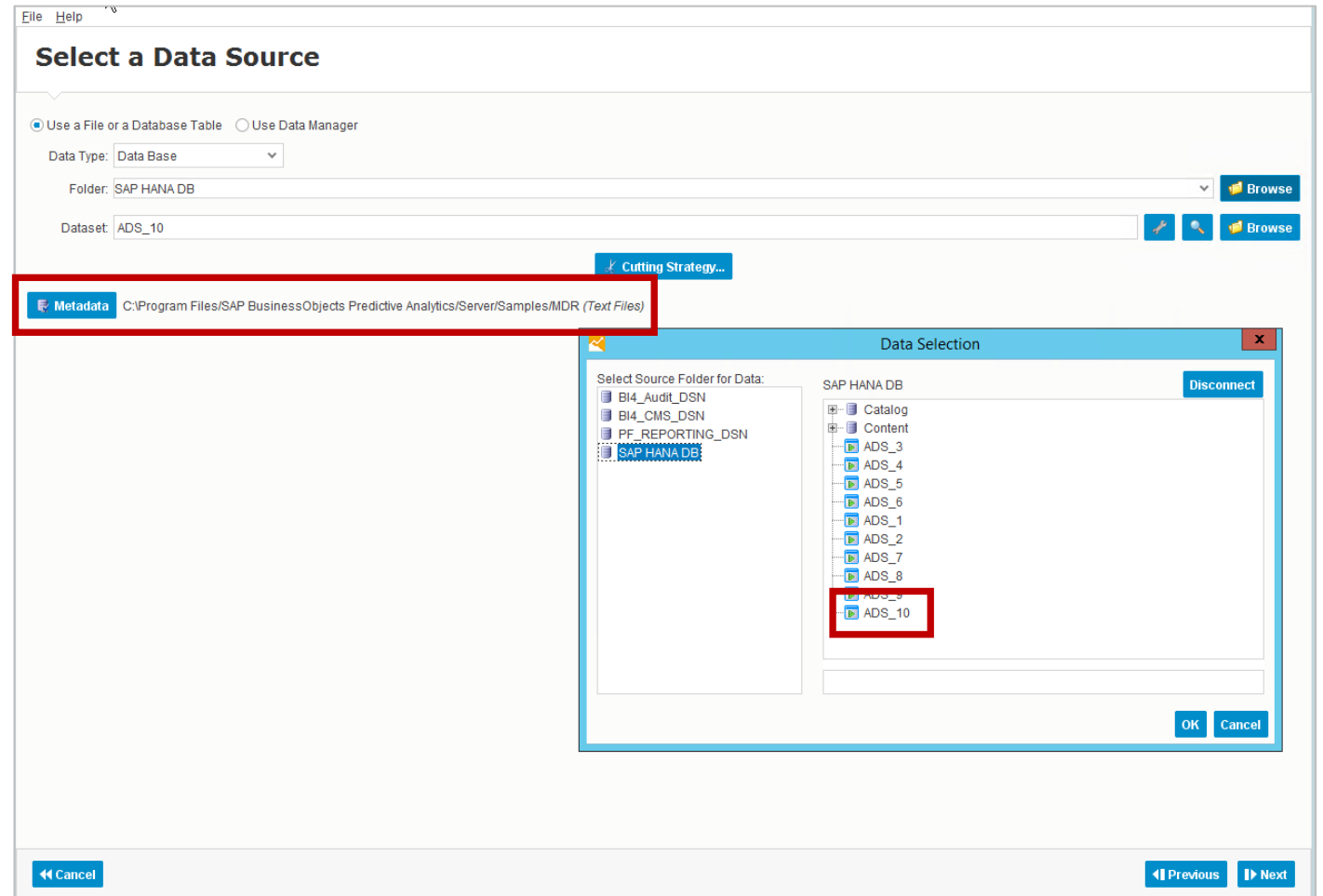
- Open SAP Predictive Analytics desktop
- Select Modeler / Create a Classification/Regression Model
- Please refer to the user guide if you would like to deep-dive into any of the product functionality. See pa32_class-clust_user_en.pdf.



Developing an Initial Churn Model

Demo – Step-by-step guide 2

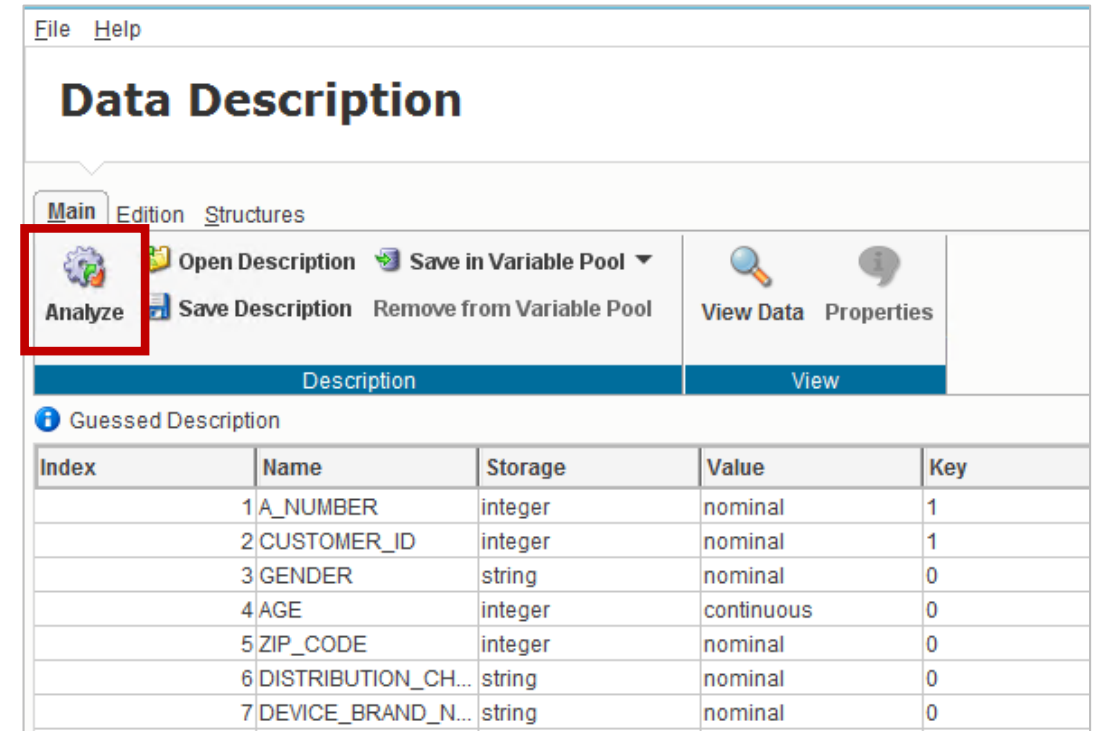
- Data Type = Data Base
- Make sure you are connected to the metadata repository where you saved the data manipulation.
- Use the Browse button to connect to the SAP HANA database using the credential you will be given.
- Select your data manipulation ADS_10. Choose OK.
- Choose Next to go to the next step.



Developing an Initial Churn Model

Demo – Step-by-step guide 3

- Choose *Analyze* to read in the data description you created when you built the data manipulation.
- This should be correct, but there is a check list on the following slide.
- Choose *Next*.



Index	Name	Storage	Value	Key
1	A_NUMBER	integer	nominal	1
2	CUSTOMER_ID	integer	nominal	1
3	GENDER	string	nominal	0
4	AGE	integer	continuous	0
5	ZIP_CODE	integer	nominal	0
6	DISTRIBUTION_CH...	string	nominal	0
7	DEVICE_BRAND_N...	string	nominal	0

Developing an Initial Churn Model

Demo – Data description check

RANK	NAME	STORAGE	VALUETYPE	KEYLEVEL
0	A_NUMBER	integer	nominal	1
1	CUSTOMER_ID	integer	nominal	1
2	GENDER	string	nominal	0
3	AGE	integer	continuous	0
4	ZIP_CODE	integer	nominal	0
5	DISTRIBUTION_CHANNEL_ID	string	nominal	0
6	DEVICE_BRAND_NAME	string	nominal	0
7	DEVICE_MODEL_NAME	string	nominal	0
8	TENURE_MTHS	integer	continuous	0
9	SERVICE_TYPE	string	nominal	0
10	SERVICE_NAME	string	nominal	0
11	DATA_UP_ALLOWANCE_MB	integer	nominal	0
12	VOICE_ALLOWANCE_MINUTES	integer	nominal	0
13	SMSALLOWANCE_NUM_MESSAGES	string	nominal	0
14	M0_DATA_USAGE_MB	integer	continuous	0
15	M0_DATA_USAGE_PCT	number	continuous	0
16	M1_DATA_USAGE_MB	integer	continuous	0
17	M1_DATA_USAGE_PCT	number	continuous	0
18	M2_DATA_USAGE_MB	integer	continuous	0
19	M2_DATA_USAGE_PCT	number	continuous	0
20	APR_DATA_USAGE_MB	integer	continuous	0
21	APR_DATA_USAGE_PCT	number	continuous	0
22	MAY_DATA_USAGE_MB	integer	continuous	0
23	MAY_DATA_USAGE_PCT	number	continuous	0

24	JUN_DATA_USAGE_MB	integer	continuous	0
25	JUN_DATA_USAGE_PCT	number	continuous	0
26	CHURN_MAY	integer	nominal	0
27	CHURN_JUN	integer	nominal	0
28	CNT_0_VOC	integer	continuous	0
29	CNT_1_VOC	integer	continuous	0
30	CNT_2_VOC	integer	continuous	0
31	M_MEAN_VOC_CNT	number	continuous	0
32	SUM_0_DURATION_VOC_1	integer	continuous	0
33	SUM_1_DURATION_VOC_1	integer	continuous	0
34	SUM_2_DURATION_VOC_1	integer	continuous	0
35	M_MEAN_VOC_DUR	number	continuous	0
36	CNT_0_VOC_EV	number	continuous	0
37	CNT_1_VOC_EV	number	continuous	0
38	CNT_2_VOC_EV	number	continuous	0
39	DUR_0_VOC_EV	number	continuous	0
40	DUR_1_VOC_EV	number	continuous	0
41	DUR_2_VOC_EV	number	continuous	0
42	M_MEAN_DATA_USAGE	number	continuous	0
43	DATA_0_EV	number	continuous	0
44	DATA_1_EV	number	continuous	0
45	DATA_2_EV	number	continuous	0
46	TARGET	integer	nominal	0

Developing an Initial Churn Model

Demo – Step-by-step guide 4

- Target Variable = TARGET
- The Excluded Variables are selected because they are either IDs, or they are characteristics that occur after the reference date (end of March). If these variables were included, the model would be perfectly accurate! These are “leakers”.
- Choose Next.

File Help

Selecting Variables

Explanatory Variables Selected 36

- DISTRIBUTION_CHANNEL_ID
- DEVICE_BRAND_NAME
- DEVICE_MODEL_NAME
- TENURE_MTHS
- SERVICE_TYPE
- SERVICE_NAME
- DATA_UP_ALLOWANCE_MB
- VOICE_ALLOWANCE_MINUTES
- SMSALLOWANCE_NUM_MESSAGES
- M0_DATA_USAGE_MB
- M0_DATA_USAGE_PCT
- M1_DATA_USAGE_MB
- M1_DATA_USAGE_PCT
- M2_DATA_USAGE_MB
- M2_DATA_USAGE_PCT
- CNT_0_VOC
- CNT_1_VOC
- CNT_2_VOC
- M_MEAN_VOC_CNT
- SUM_0_DURATION_VOC_1
- SUM_1_DURATION_VOC_1
- SUM_2_DURATION_VOC_1
- M_MEAN_VOC_DUR
- CNT_0_VOC_EV
- CNT_1_VOC_EV
- CNT_2_VOC_EV
- DUR_0_VOC_EV
- DUR_1_VOC_EV
- DUR_2_VOC_EV
- M_MEAN_DATA_USAGE
- DATA_0_EV
- DATA_1_EV
- DATA_2_EV

Target Variables 1

- TARGET

Weight Variable 0

Excluded Variables 10

- A_NUMBER
- CUSTOMER_ID
- APR_DATA_USAGE_MB
- APR_DATA_USAGE_PCT
- MAY_DATA_USAGE_MB
- MAY_DATA_USAGE_PCT
- JUN_DATA_USAGE_MB
- JUN_DATA_USAGE_PCT
- CHURN_MAY
- CHURN_JUN

Alphabetic Sort

Developing an Initial Churn Model

Demo – Step-by-step guide 5

- Enable Auto-selection is enabled by default.
- Auto-selection allows you to automatically reduce the number of variables in the model in relation to certain quality criteria. This selection is done by successive iterations, starting with all of the variables, then removing the worst performing variables.
- Choose Generate to build the model.

The screenshot shows a software interface titled "Summary of Modeling Parameters". It includes a "Model Name" field with the value "TARGET_ADS_10" and an empty "Description" field. Below this, the "Kxen.RobustRegression" section is highlighted in light blue. Inside this section, the following parameters are listed: "Data to be Modeled: SAP HANA DB/ADS_10", "Cutting Strategy: Random without test", "Target Variable: TARGET", and "Weight Variable (Optional): None". At the bottom of the dialog, there are two checkboxes: "Compute Decision Tree:" (unchecked) and "Enable Auto-selection:" (checked). Three buttons are located at the bottom: "Autosave...", "Export KxShell Script...", and "Advanced..." (on the right).

For more information about auto-selection, or to deep-dive into other functionality, please refer to the user guide [pa32_class-clust_user_en.pdf](#)

Developing an Initial Churn Model

Demo – Step-by-step guide 6

- The model overview shows that there are 7445 records in ADS_10.
- The TARGET has 15.4% churners and 84.6% non-churners.
- The auto-selection iteration 2 gave the best model, with a Predictive Power (KI) of 0.7472 and Prediction Confidence (KR) of 0.9795. This model contains 17 variables.
- This model is robust (KR > 0.95) with high predictive power.
- Choose Next.

Model Overview

Overview

Model: TARGET_ADS_10		
	Data Set:	ADS_10
	Initial Number of Variables:	47
	Number of Selected Variables:	36
	Number of Records:	7,445
	Building Date:	2017-10-20 09:21:12
	Learning Time:	10 s
	Engine Name:	Kxen.RobustRegression
	Author:	Administrator

Nominal Targets

TARGET		
	Target Key	1
	0 - Frequency	84.6%
	1 - Frequency	15.4%

Selection Process Selected Iteration

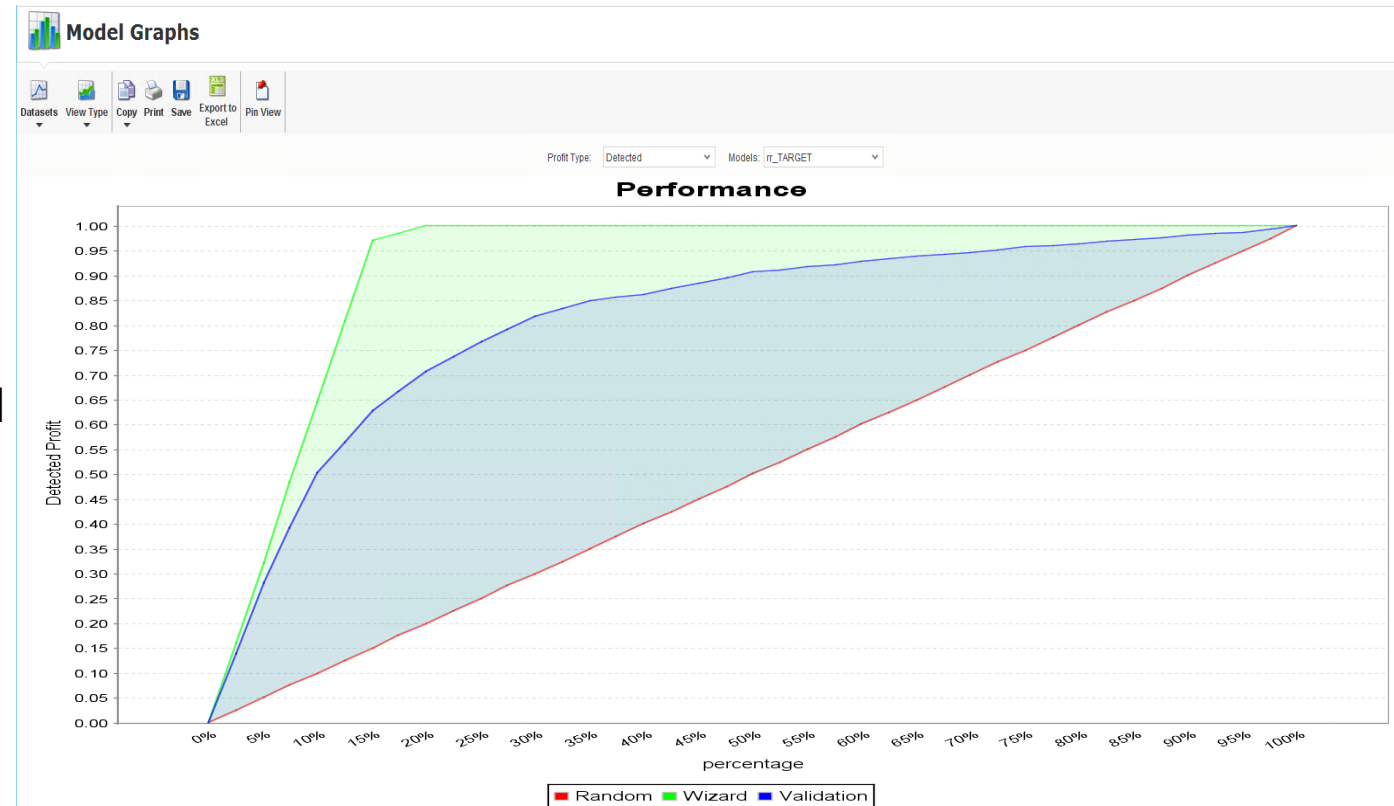
2		
	Predictive Power (KI)	0.7472
	Prediction Confidence (KR)	0.9795
	Nb. Variables Kept	17

Please note that your results might be slightly different, because the random cutting strategy will have selected different records for the estimation and validation data sets.

Developing an Initial Churn Model

Demo – Step-by-step guide 7

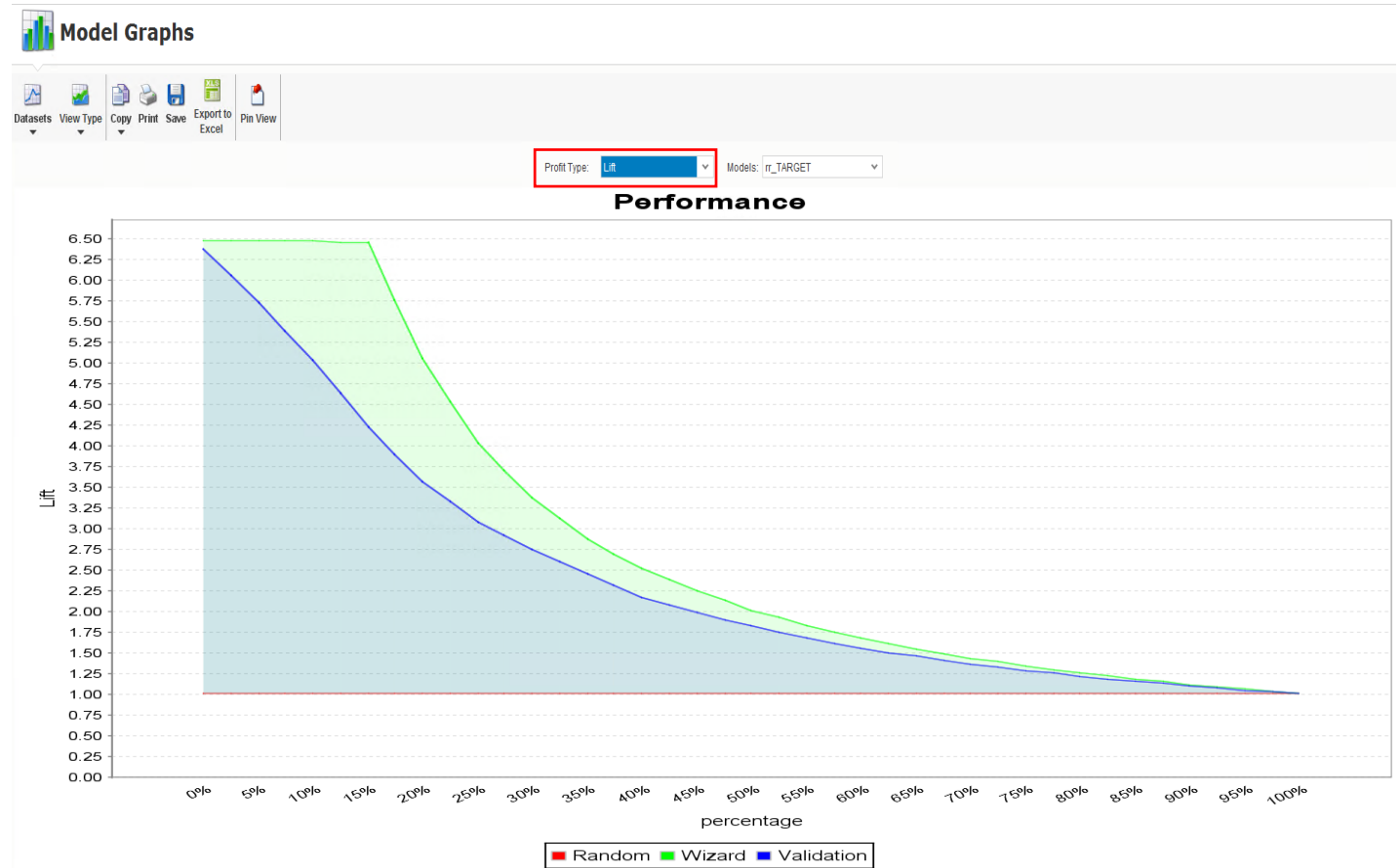
- Select Model Graphs.
- The Detected graph shows the performance of the model (blue curve) relative to the random line (red) and perfect model wizard (green).
- The validation represents how well the model performs on the validation sub-set of data.
- The random represents the worst-case scenario, when only a random selection is taken.
- The wizard represents the best-case scenario, when the model perfectly predicts all of the churners.



Developing an Initial Churn Model

Demo – Step-by-step guide 8

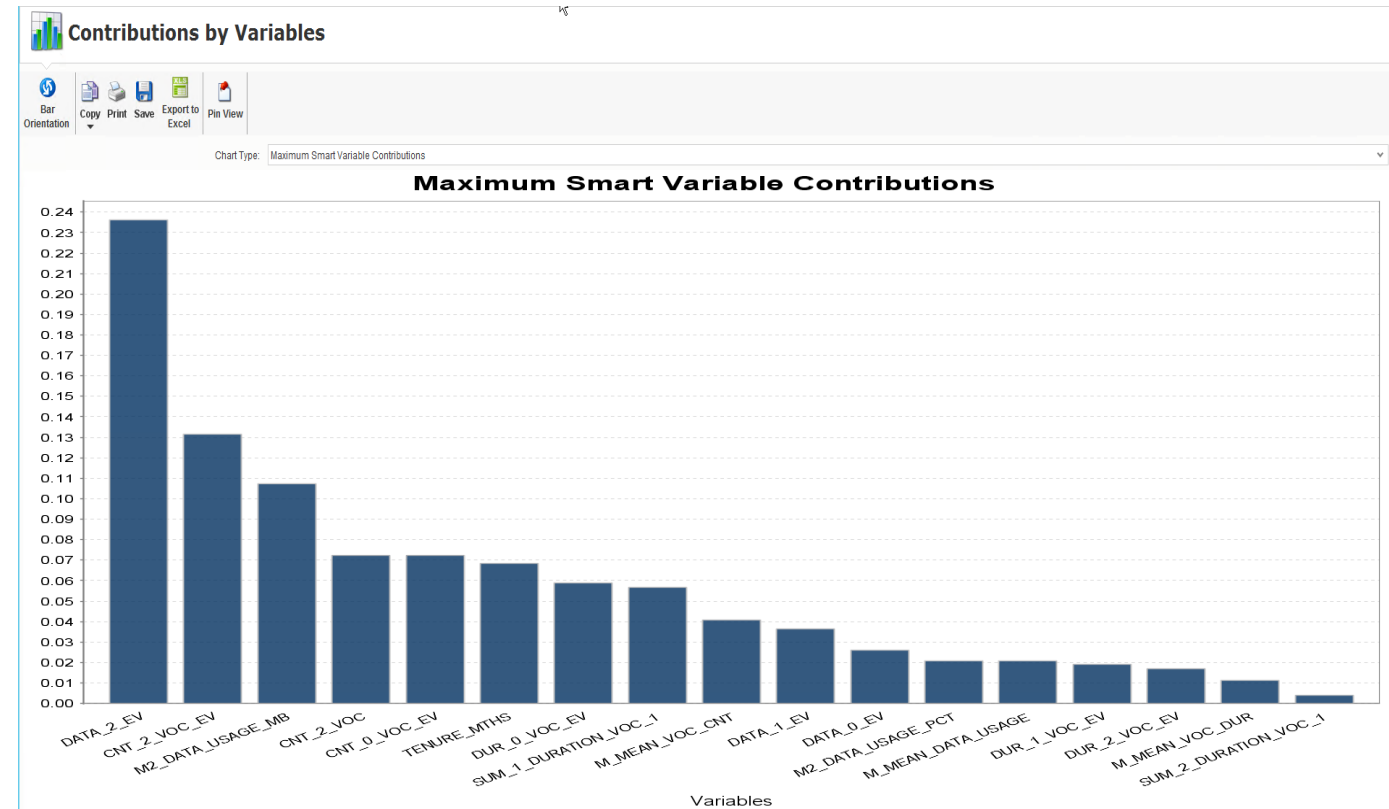
- Select Profit Type = Lift
- This shows the model performance relative to a constant random line, where Lift = 1.
- The “lift” of a model is the ratio of model accuracy divided by the accuracy of a baseline measure, usually the expected performance of a random guess (the red line shown here, where Lift = 1).
- To see how well the model is performing, you can click the model validation performance curve (blue line).
- Choose Previous.



Developing an Initial Churn Model

Demo – Step-by-step guide 9

- Examine the contributions of the variables.
- Select Contributions by Variables.
- This shows the relative importance of each variable in the model.
- In my model, the data usage and voice call count in the month prior to the reference date (DATA_2_EV and CNT_2_VOC_EV) are important contributors in the model.
- You can drill into each variable by double-clicking the bar.
- Double-click TENURE_MTHS.



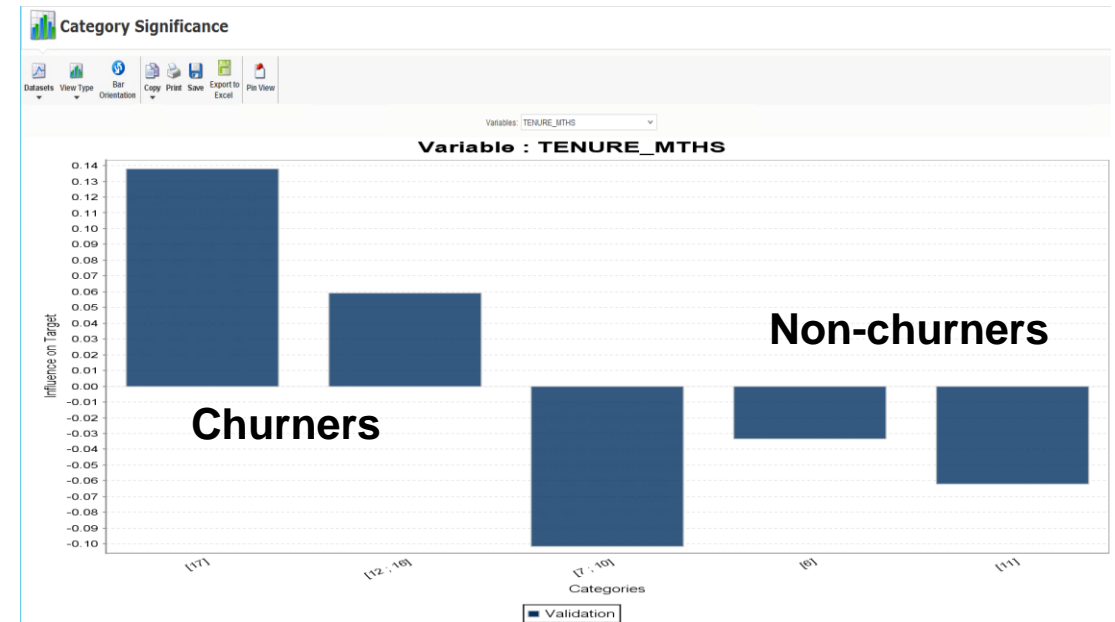
Developing an Initial Churn Model

Demo – Step-by-step guide 10

- This shows the relative significance of the different categories of a given variable with respect to the target variable.
- The importance of a category depends on both its difference to the target category mean and the number of represented cases.
- High importance can result from:
 - a high discrepancy between the category and the mean of the target category of the target variable
 - or a minor discrepancy combined with a large number of records in the category
 - or a combination of both.
- Choose *Previous*.

The height of the bar shows the “profit” from that category. The positive bars correspond to categories which have more than the mean number from the target category (i.e. churners) and the negative bars correspond to categories which have less than the mean number from the target category (i.e. non-churners).

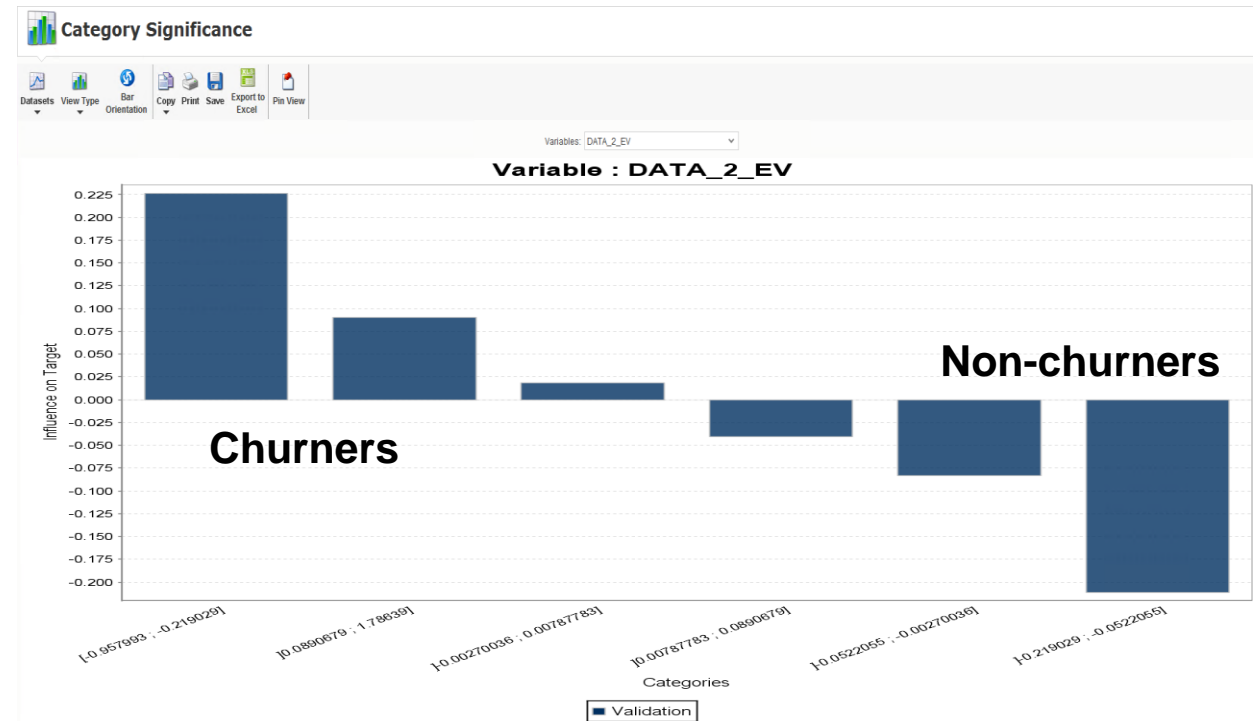
Categories sharing the same effect on the target variable are grouped. They appear as follows [Category_a;Category_b;Category_c].



Developing an Initial Churn Model

Demo – Step-by-step guide 10

- Double-click DATA_2_EV.
- If you remember, $DATA_2_EV = (M2_DATA_USAGE_MB - M_MEAN_DATA_USAGE) / M_MEAN_DATA_USAGE$
- This shows that customers who are most likely to churn (on the left side of the graph) have a large reduction in data usage in the month prior to the reference date (M2), compared to the mean over the 3-month period.
- These are clearly customers who are cutting back on using data on this service.



Intervals are represented by brackets []. Closed square brackets [] mean an *inclusive* interval - that is, the numbers inside the brackets are included. Reverse or open square brackets] [mean an *exclusive* interval, that is, the numbers inside the brackets are excluded from the interval. For example:

$[4, 9] = 4 \leq x \leq 9$ or all numbers from 4 to 9.

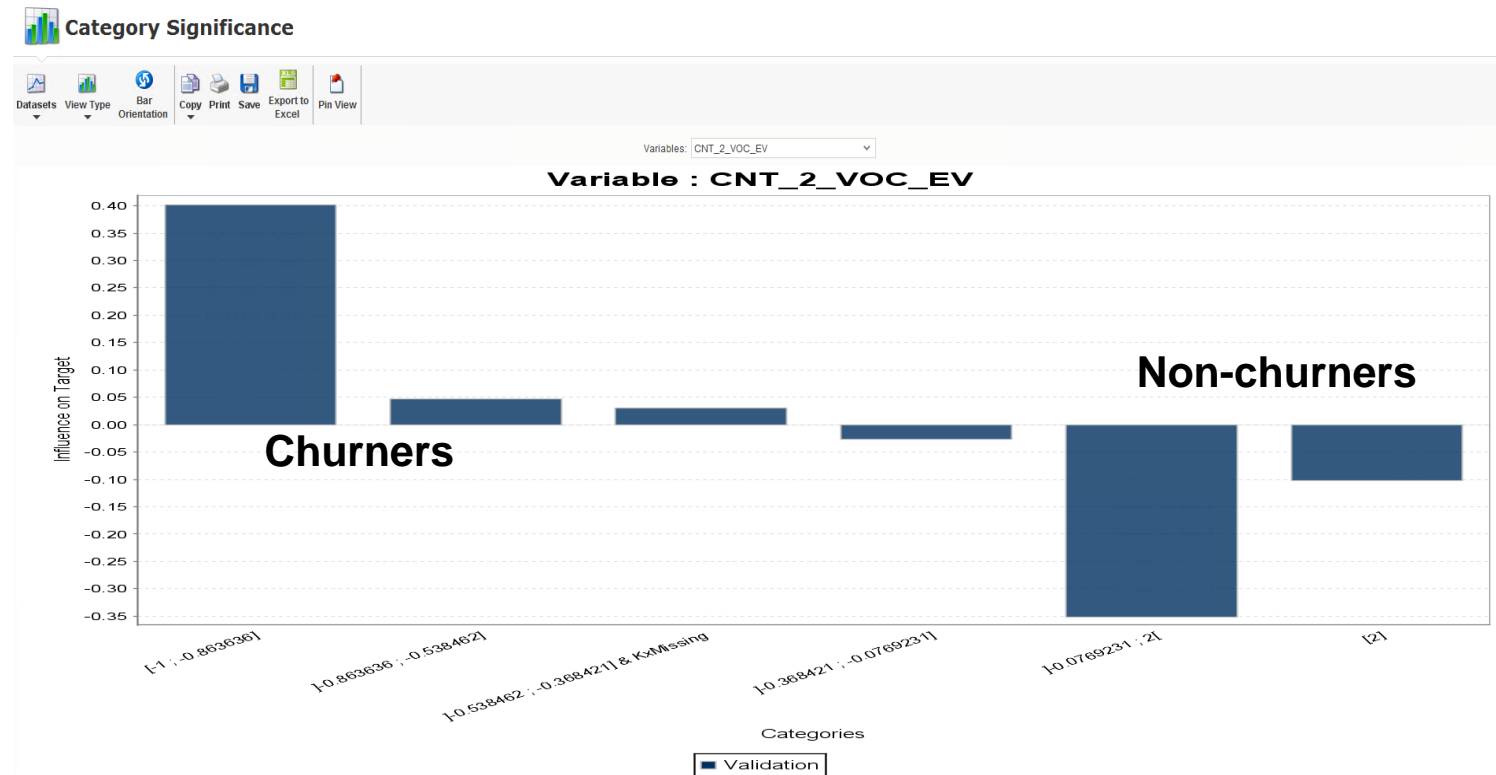
$]4, 9[= 4 < x < 9$ or all numbers between 4 and 9.

$[4, 9[= 4 \leq x < 9$ or all numbers between 4 and 9 including 4.

Developing an Initial Churn Model

Demo – Step-by-step guide 10

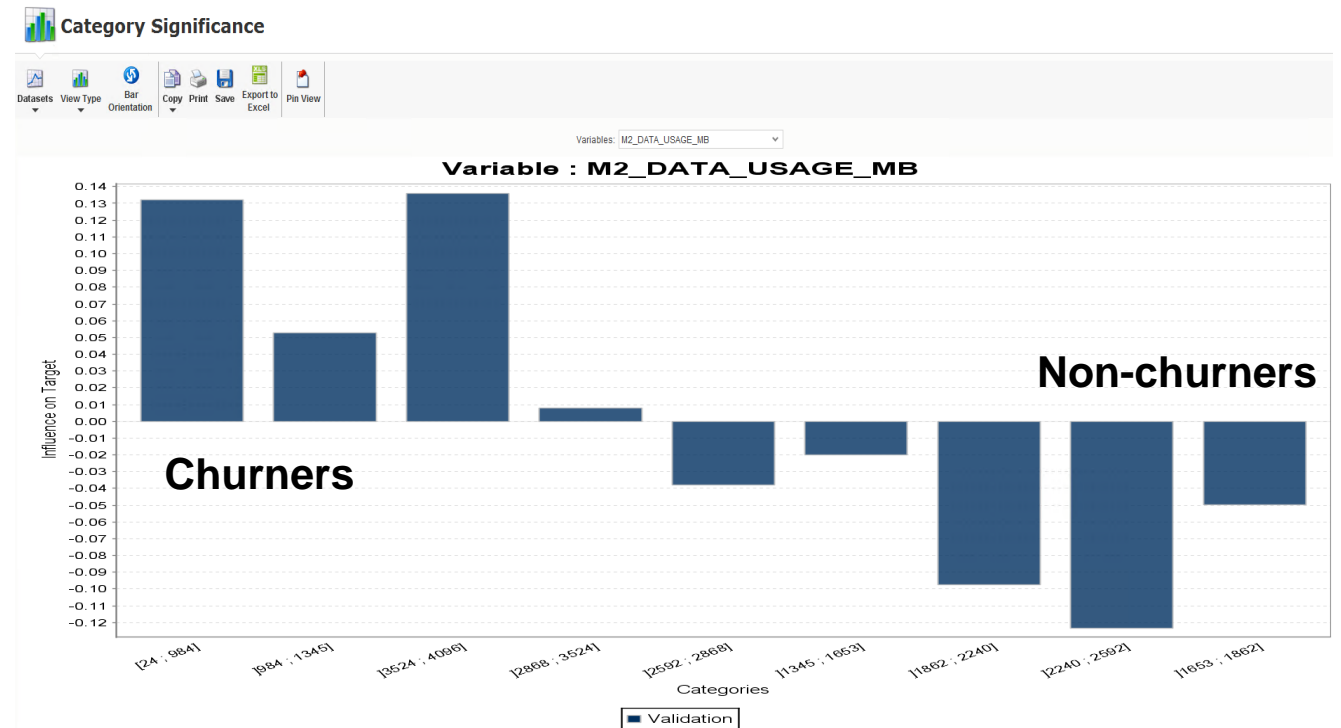
- Double click CNT_2_VOC_EV.
- If you remember, $\text{CNT_2_VOC_EV} = (\text{CNT_2_VOC_M_MEAN_VOC_CNT}) / \text{M_MEAN_VOC_CNT}$
- This shows that customers who are most likely to churn (on the left side of the graph) have a large reduction in voice call usage in the month prior to the reference date (M2), compared to the mean over the 3-month period.
- These are clearly customers who are cutting back on voice calls on this service.



Developing an Initial Churn Model

Demo – Step-by-step guide 10

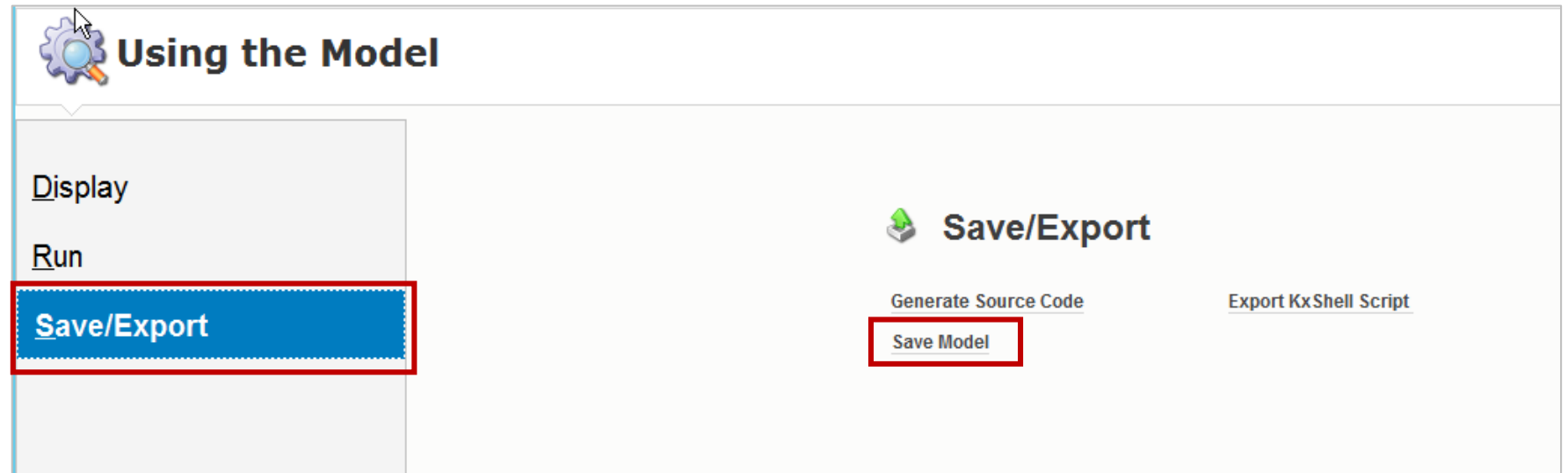
- Double-click M2_DATA_USAGE_MB.
- This variable reflects the data usage in the month previous to the reference date.
- This shows that customers who are most likely to churn (on the left side of the graph) have a low data usage, in the range 24-984Mb.
- This supports what we saw in the previous graph for DATA_2_EV.
- Examine the other graphs so you can start to build up a picture of churning behavior.
- Choose *Previous*.
- Choose *Previous* again.



Developing an Initial Churn Model

Demo – Step-by-step guide 11

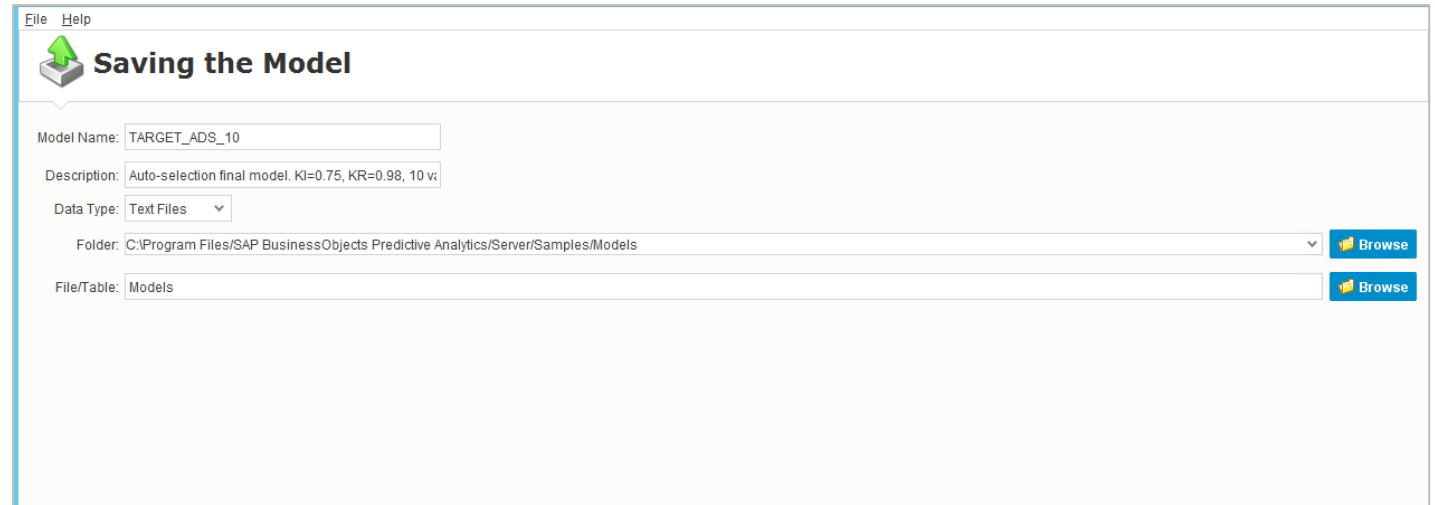
- Save the model.
- Select *Save/Export* and *Save Model*.



Developing an Initial Churn Model

Demo – Step-by-step guide 11

- Enter a description so that you will recognize the model in future. This could include the KI, KR, and number of variables in the model.
- Save the model as a text file into the folder C:\Program Files\SAP BusinessObjects Predictive Analytics/Server/Samples/Models
- If necessary, create a “Models” folder in the Samples folder.
- The File/Table = Models
- There is a version control, so all of the models can be written here, with the same file name, and they will be automatically saved as a new version.
- Choose Save.



The screenshot shows a software window titled "Saving the Model" with a menu bar containing "File" and "Help". The window contains the following fields and controls:

- Model Name:** A text input field containing "TARGET_ADS_10".
- Description:** A text input field containing "Auto-selection final model. KI=0.75, KR=0.98, 10 v".
- Data Type:** A dropdown menu currently set to "Text Files".
- Folder:** A text input field containing "C:\Program Files\SAP BusinessObjects Predictive Analytics/Server/Samples/Models", followed by a dropdown arrow and a "Browse" button.
- File/Table:** A text input field containing "Models", followed by a "Browse" button.

Developing an Initial Churn Model

Summary

- The modeling phase, which is Phase 4 of the CRISP-DM process, includes a model assessment task. In this unit, you have learnt how to assess the gains and lift charts that are created when you build a classification model.
- You have also seen a demonstration showing you how to use the SAP Predictive Analytics software to automatically build a classification model and assess the model.
- There are step-by-step instructions in the appendix in this deck, so you can refer to it when you build the model.



Thank you.

Contact information:

open@sap.com

© 2017 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See <http://global.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.

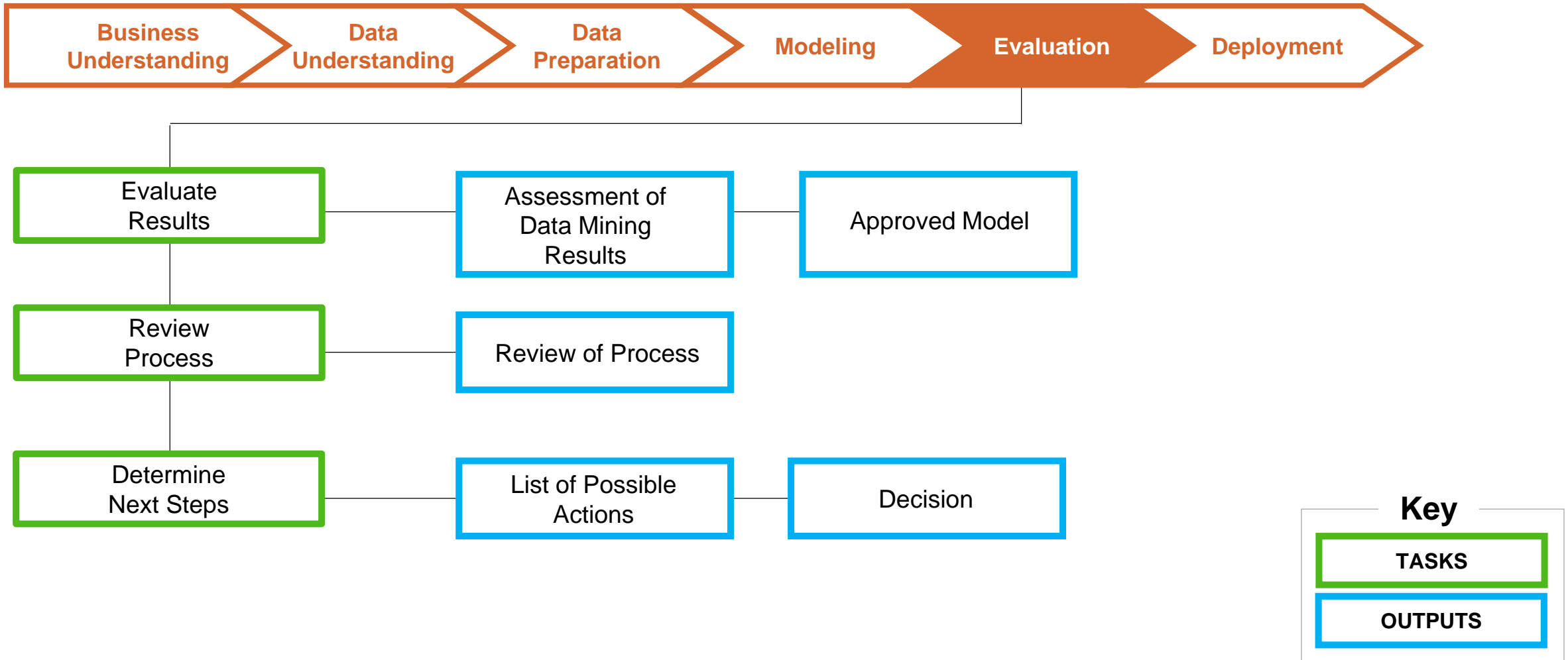


Week 3: Develop, Evaluate, and Deploy Models

Unit 3: Evaluating the Initial Churn Model

Evaluating the Initial Churn Model

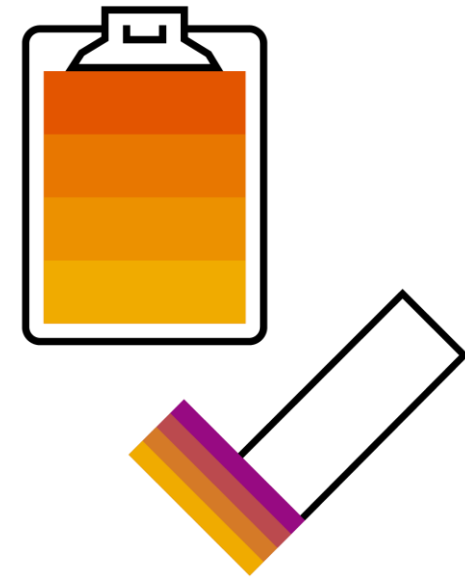
CRISP-DM – Phase 5: Evaluation



Evaluating the Initial Churn Model

Evaluate results

- Task
 - Assess the degree to which the model meets the business objectives.
 - Test the model(s) in the business environment if time and budget constraints permit.
- Output – Assessment of the model with respect to business success criteria



For more information on control groups, see:

<http://www.latentview.com/blog/measure-campaign-effectiveness/>

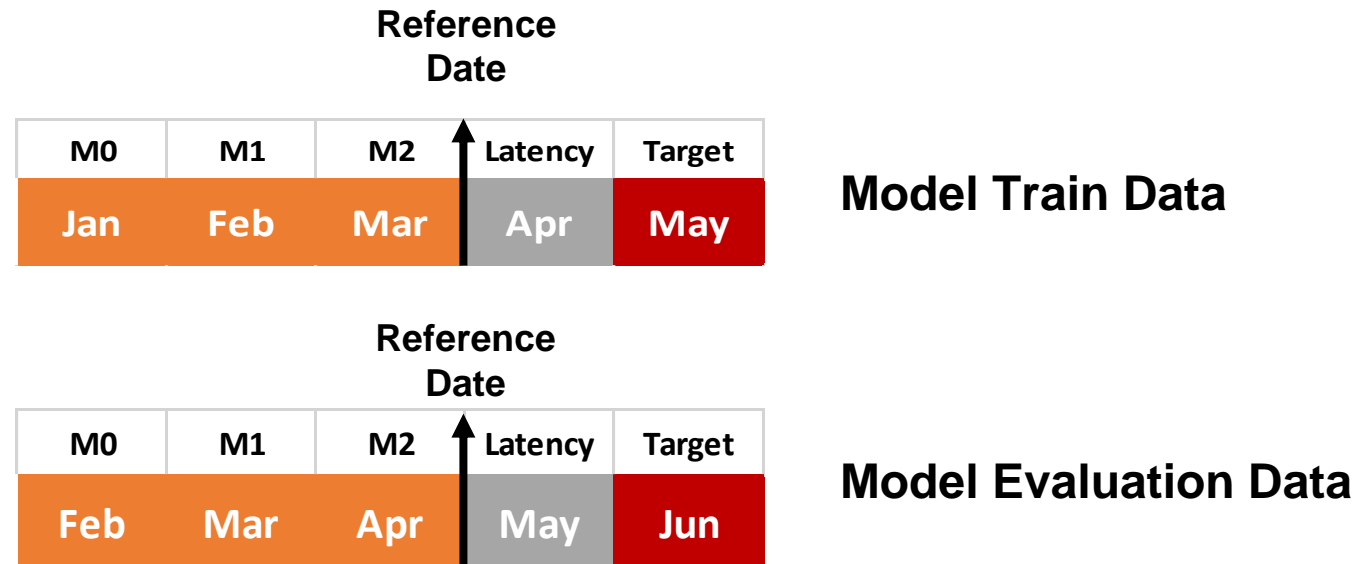
<http://20bits.com/article/statistical-analysis-and-ab-testing>

<https://blog.kissmetrics.com/how-ab-testing-works/>

Evaluating the Initial Churn Model

Evaluate results using June churn data

- The telco provided data for churn in May and June. Therefore, one way you could try to evaluate the model is to apply it onto the usage data in Feb to April, and compare the predicted output to the actual June data. You can also compare the distributions of the usage data in the training and evaluation periods.
- First, the Analytical Data Set needs to be updated. Essentially, the data set we used to train the model needs to be moved forward 1 month:



Evaluating the Initial Churn Model

Deviation analysis

- The software compares the distribution of the variables in the evaluation data to the distribution of the variables in the data used to train the model.
- In our example, the deviation test finds a lot of deviations in the data frequencies and in the model output. Any deviation probability > 0.95 is significant.
- This might be due to:
 - the quality of the data. The customer must check that the data has been supplied correctly
 - a change in the general usage characteristics. Maybe there is a “seasonal” type variation in Jan to Mar that does not match Feb to Apr.
- This analysis will help to pinpoint the changes in usage.
- The customer needs to be consulted and the data checked.

Analyze Deviations Debriefing

Stop

View Type

Copy

Print

Save

Current Report

All Reports

Data Set:

"EVALUATION"

Source:

SAP HANA DB/Catalog/TRIALSCHEMA

Number of Records:

6,297

Number of Variables Showing Deviation:

17

Deviant Variables

Variable	Deviation Probability
M0_DATA_USAGE_MB	1
M0_DATA_USAGE_PCT	1
M1_DATA_USAGE_MB	1
M1_DATA_USAGE_PCT	1
CNT_0_VOC	0.98552
CNT_1_VOC	1
SUM_0_DURATION_VOC_1	0.999197
SUM_1_DURATION_VOC_1	1
CNT_0_VOC_EV	1
CNT_1_VOC_EV	1
CNT_2_VOC_EV	1
DUR_0_VOC_EV	1
DUR_1_VOC_EV	1
DUR_2_VOC_EV	1
DATA_0_EV	1
DATA_1_EV	1
DATA_2_EV	1
rr_TARGET	1
err_rr_TARGET	1

Performance on Control Data Set


Target: TARGET






rr_TARGET	KI	0.4923
	KR	0.7432

Evaluating the Initial Churn Model

Deviation analysis update

- When the customer checked the data, they found that there were data supply problems that were affecting some of the frequency counts in April, and the June target distribution.
- The data was refreshed and the Evaluation data set was updated.
- There are now fewer data deviations and no model deviations, and the model performance (KI and KR) metrics are acceptable with KR > 0.95.

**Analyze Deviations Debriefing**



StopView TypeCopyPrintSave

Current ReportAll Reports

Data Set used for Deviation Control

Data Set:"EVALUATION_2"

Source:SAP HANA DB/Catalog/TRIAL.SCHEMA

Number of Records:6,297

Number of Variables Showing Deviation:11

Deviant Variables

Variable	Deviation Probability
CNT_0_VOC	1
CNT_1_VOC	1
M_MEAN_VOC_CNT	0.982428
SUM_0_DURATION_VOC_1	1
SUM_1_DURATION_VOC_1	0.999881
CNT_0_VOC_EV	1
CNT_1_VOC_EV	0.999978
CNT_2_VOC_EV	1
DUR_0_VOC_EV	0.999999
DUR_1_VOC_EV	0.999855
DUR_2_VOC_EV	0.999999

Performance on Control Data Set

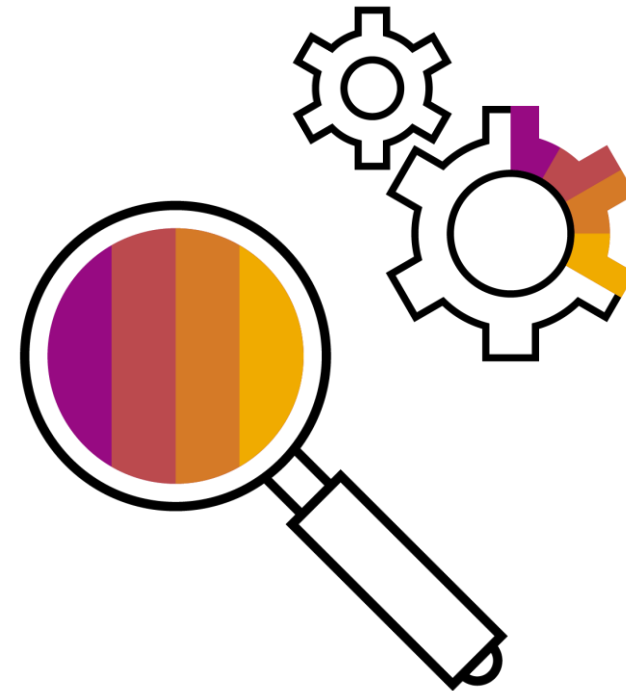
Target: TARGET

rr_TARGET	
	KI0.7469
	KR0.9842

Evaluating the Initial Churn Model

Review process

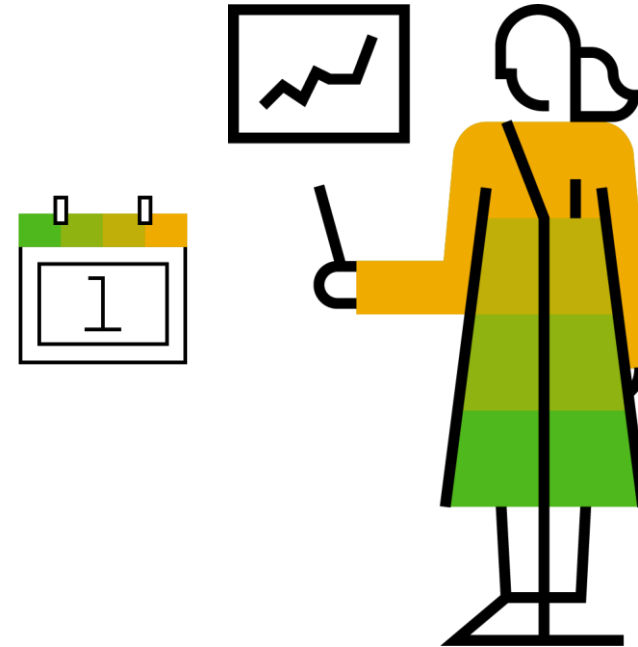
- Task
 - Conduct a more thorough review of the data mining engagement.
 - Identify any quality assurance issues.



Evaluating the Initial Churn Model

Determine next steps

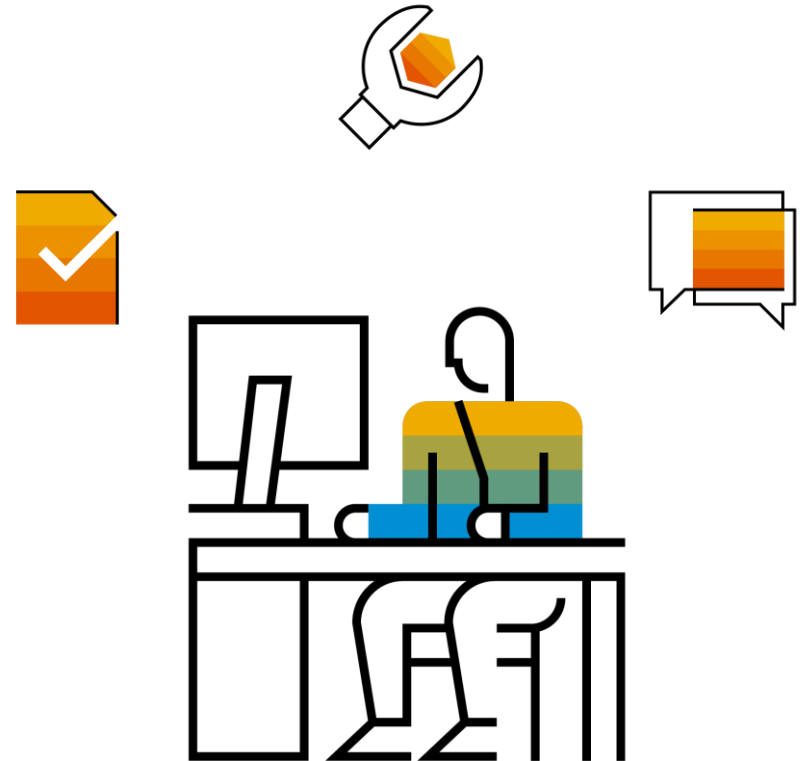
- Task
 - Assess how to proceed with the project.
- Output – List of Possible Actions
 - List the potential further actions.
- Output – Decision
 - Describe the decision on how to proceed.



Evaluating the Initial Churn Model

Determine next steps

- The next steps:
 - The model meets the business success criteria, but needs to be monitored closely.
 - The company will start to use this initial churn model immediately, and its performance will be monitored and assessed further when it is deployed.
 - A social link analysis and segmentation are to be developed.



Evaluating the Initial Churn Model

Demo





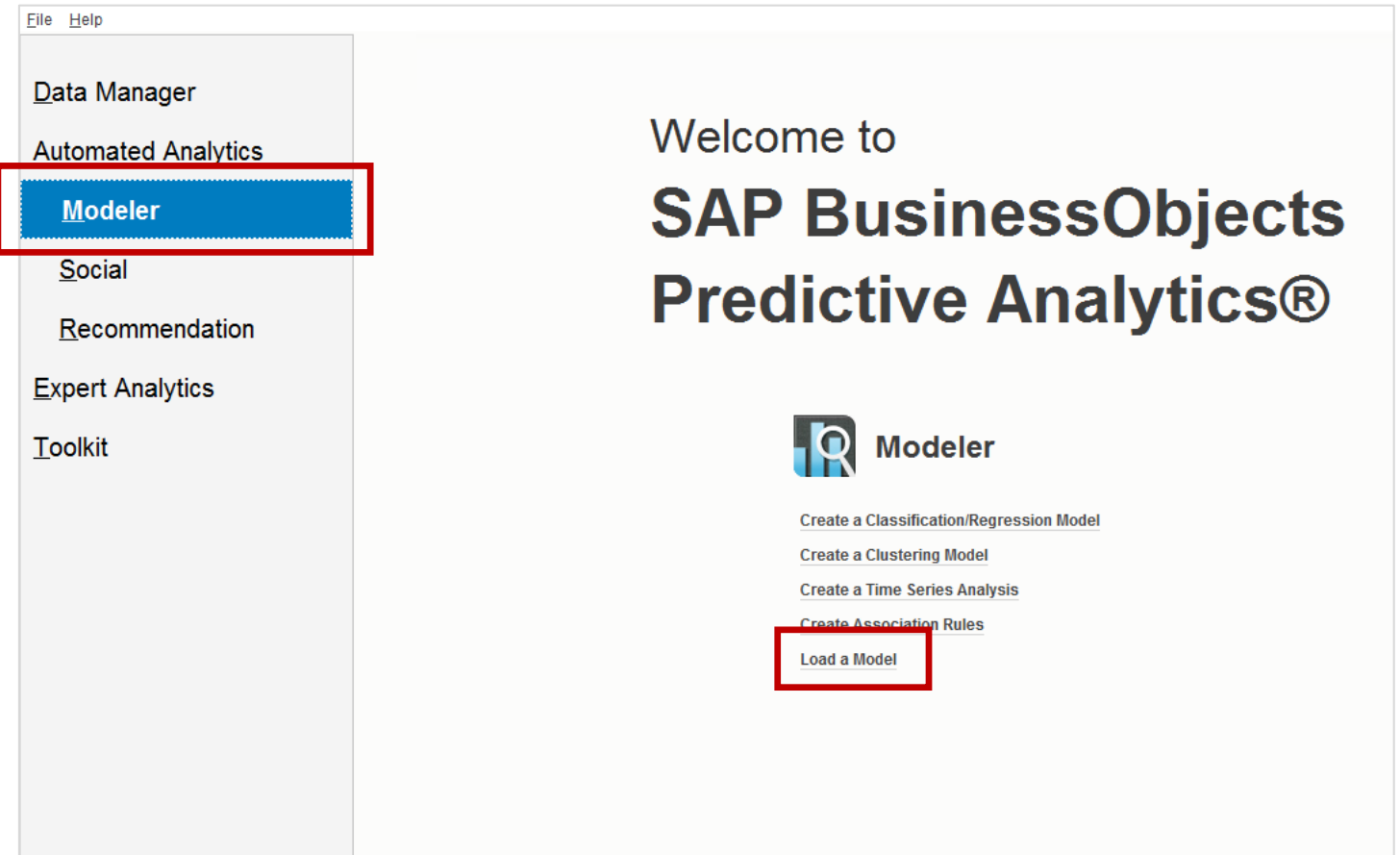
Appendix

Step-by-step guide to testing data for deviations

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 1

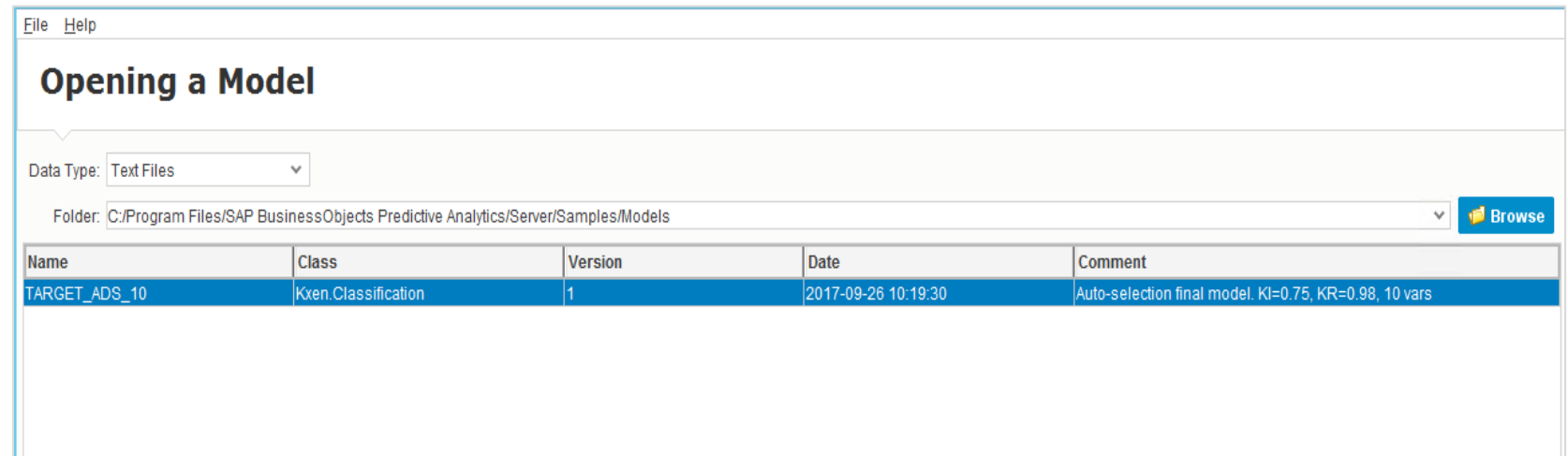
- Select Modeler / Load a Model
- Please refer to the user guide if you would like to deep-dive into any of the product functionality. See pa32_class-clust_user_en.pdf.



Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 2

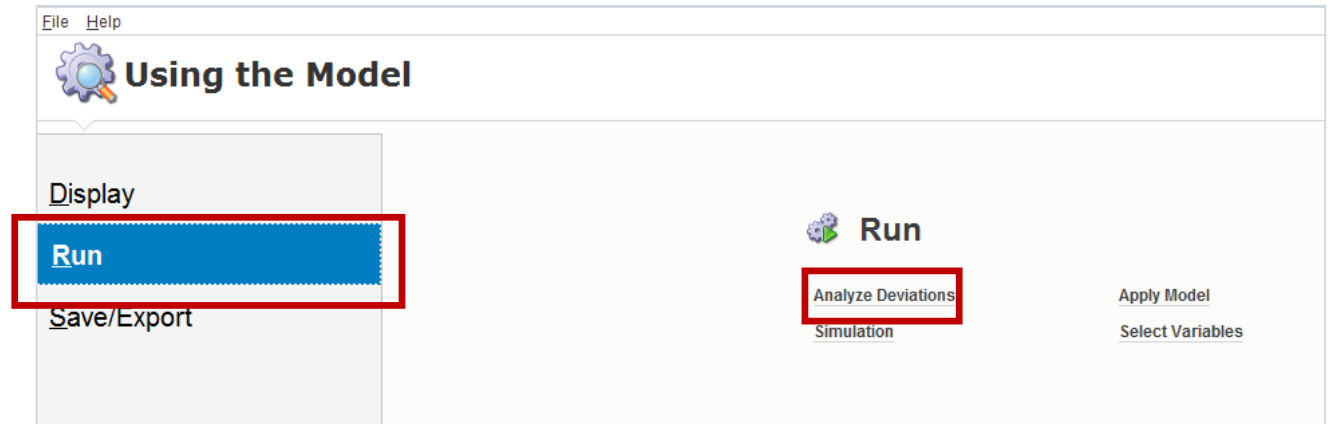
- Locate the model you saved in C:/Program Files/SAP BusinessObjects Predictive Analytics/Server/Samples/Models
- Choose Open.
- This will open the model at the Using the Model interface.



Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 3

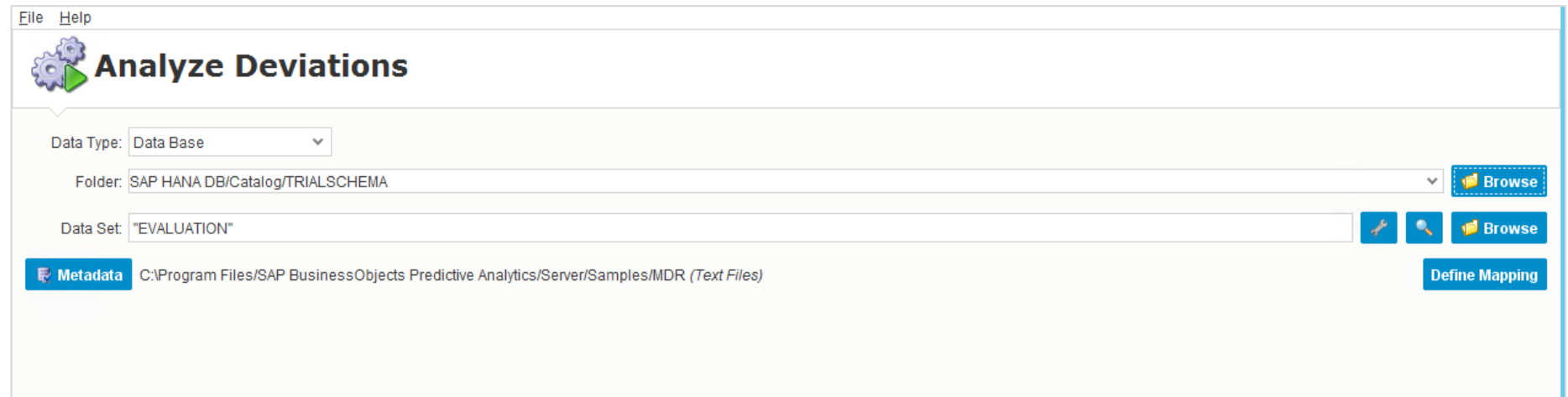
- Select Run / Analyze Deviations



Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 4

- Data Type = Data Base
- Choose Browse to find the database and data. You may need to enter your credentials again.
- Folder = SAP HANA DB
- Go to Catalog / TRIAL SCHEMA
- The evaluation data, with the reference date set at end of April 2016, is called EVALUATION



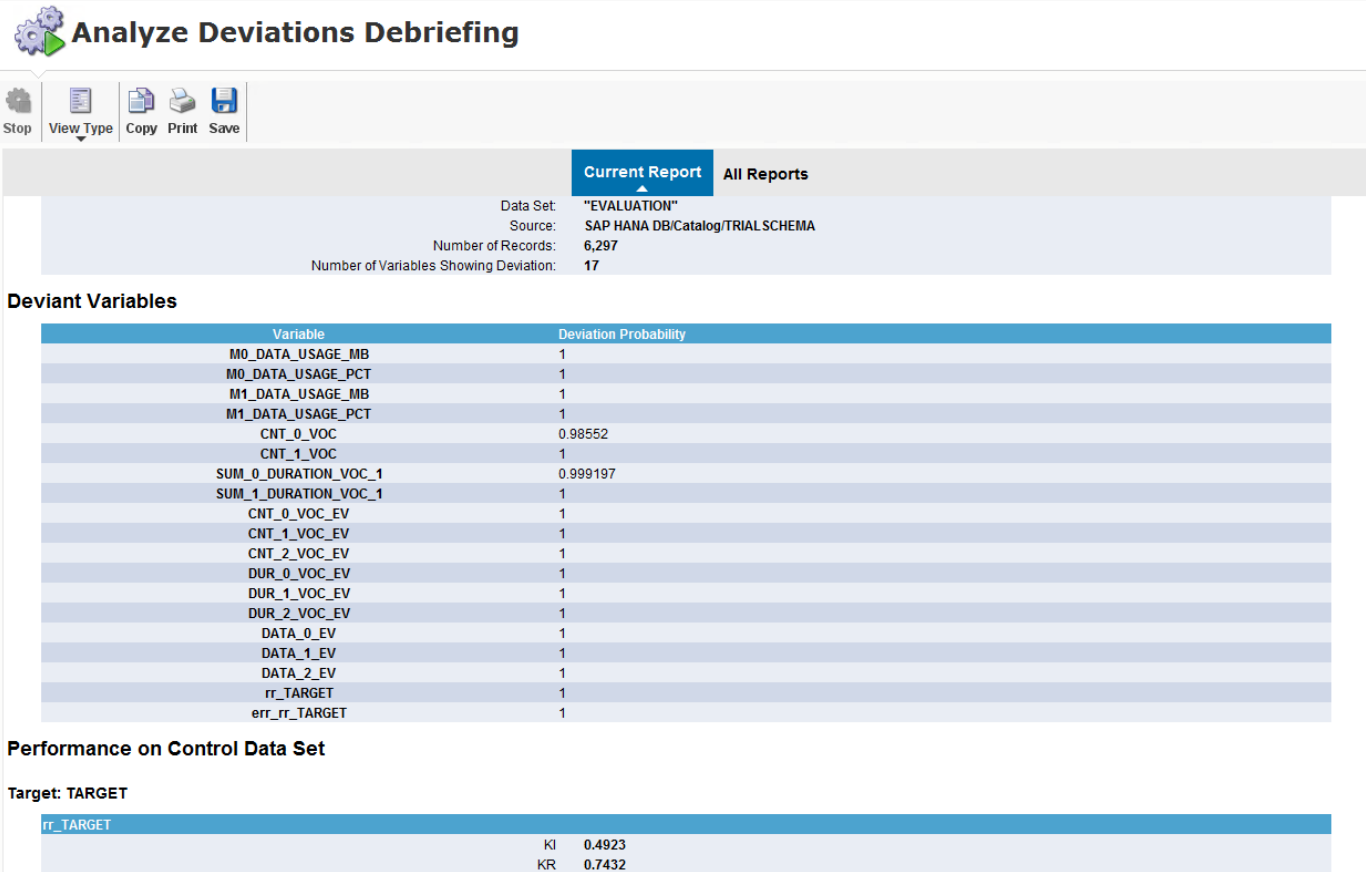
The screenshot shows the 'Analyze Deviations' window with the following configuration:

- Data Type:** Data Base (selected in a dropdown menu)
- Folder:** SAP HANA DB/Catalog/TRIALSCHEMA (selected in a dropdown menu, with a 'Browse' button to the right)
- Data Set:** "EVALUATION" (entered in a text field, with a 'Browse' button to the right)
- Metadata:** C:\Program Files\SAP BusinessObjects Predictive Analytics/Server/Samples/MDR (Text Files) (displayed next to a 'Metadata' icon)
- Buttons:** 'Define Mapping' (bottom right), and icons for 'Tools', 'Search', and 'Refresh' (bottom right)

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 5

- The software compares the distribution of the variables in the evaluation data to the distribution of the variables in the data used to train the model.
- It finds there are 6297 records in the new data set, and that there are 17 variables that have a strong probability of deviation.
- Also, the model target rr_TARGET is deviating.



Analyze Deviations Debriefing

Stop View Type Copy Print Save

Current Report All Reports

Data Set: "EVALUATION"
Source: SAP HANA DB/Catalog/TRIALSCHEMA
Number of Records: 6,297
Number of Variables Showing Deviation: 17

Deviant Variables

Variable	Deviation Probability
M0_DATA_USAGE_MB	1
M0_DATA_USAGE_PCT	1
M1_DATA_USAGE_MB	1
M1_DATA_USAGE_PCT	1
CNT_0_VOC	0.98552
CNT_1_VOC	1
SUM_0_DURATION_VOC_1	0.999197
SUM_1_DURATION_VOC_1	1
CNT_0_VOC_EV	1
CNT_1_VOC_EV	1
CNT_2_VOC_EV	1
DUR_0_VOC_EV	1
DUR_1_VOC_EV	1
DUR_2_VOC_EV	1
DATA_0_EV	1
DATA_1_EV	1
DATA_2_EV	1
rr_TARGET	1
err_rr_TARGET	1

Performance on Control Data Set

Target: TARGET

rr_TARGET	KI	0.4923
	KR	0.7432

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 6

- The output also gives the KI and KR indicators obtained by the model on the control data set.
- If the KI and/or KR of the model on the control data set are significantly lower than for the original model, it means that the relation between the variables and the target variable has changed. As a consequence, the model should be rebuilt on the new data.
- If the KI and KR are not much different, it means that the relation between the input variables and the target behavior has not changed. However, differences in the distributions are still possible.

Analyze Deviations Debriefing

Stop View Type Copy Print Save

Current Report All Reports

Data Set: "EVALUATION"
Source: SAP HANA DB/Catalog/TRIALSCHEMA
Number of Records: 6,297
Number of Variables Showing Deviation: 17

Deviant Variables

Variable	Deviation Probability
M0_DATA_USAGE_MB	1
M0_DATA_USAGE_PCT	1
M1_DATA_USAGE_MB	1
M1_DATA_USAGE_PCT	1
CNT_0_VOC	0.98552
CNT_1_VOC	1
SUM_0_DURATION_VOC_1	0.999197
SUM_1_DURATION_VOC_1	1
CNT_0_VOC_EV	1
CNT_1_VOC_EV	1
CNT_2_VOC_EV	1
DUR_0_VOC_EV	1
DUR_1_VOC_EV	1
DUR_2_VOC_EV	1
DATA_0_EV	1
DATA_1_EV	1
DATA_2_EV	1
rr_TARGET	1
err_rr_TARGET	1

Performance on Control Data Set

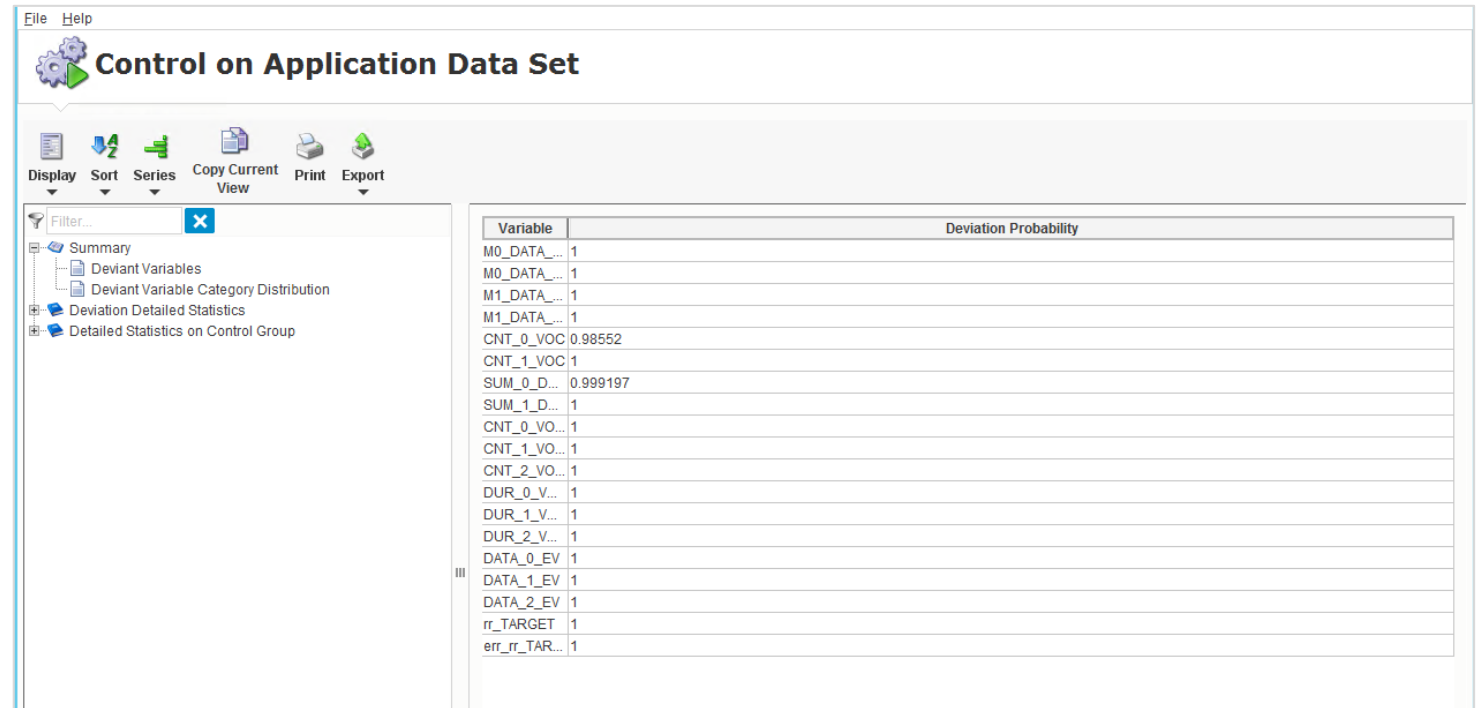
Target: TARGET

rr_TARGET	KI	KR
	0.4923	0.7432

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 7

- The Control on Application Data Set presents you with the following options:
 - Summary
 - Deviation Detailed Statistics
 - Detailed Statistics on Control Group
- Examine these reports



Variable	Deviation Probability
M0_DATA_...	1
M0_DATA_...	1
M1_DATA_...	1
M1_DATA_...	1
CNT_0_VOC	0.98552
CNT_1_VOC	1
SUM_0_D...	0.999197
SUM_1_D...	1
CNT_0_VO...	1
CNT_1_VO...	1
CNT_2_VO...	1
DUR_0_V...	1
DUR_1_V...	1
DUR_2_V...	1
DATA_0_EV	1
DATA_1_EV	1
DATA_2_EV	1
rr_TARGET	1
err_rr_TAR...	1


These tables provide the probabilities of deviation of each variable distribution, be it by variable, variable category or group of categories.






A probability over 0.95 indicates that the variable or category global distribution is significantly different in the application dataset than in the original dataset.

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 8

- Test the model on the refreshed data
EVALUATION_2

 **Analyze Deviations Debriefing**

 Stop  View Type  Copy  Print  Save

Current Report All Reports

Data Set used for Deviation Control

Data Set:	"EVALUATION_2"
Source:	SAP HANA DB/Catalog/TRIAL SCHEMA
Number of Records:	6,297
Number of Variables Showing Deviation:	11

Deviant Variables

Variable	Deviation Probability
CNT_0_VOC	1
CNT_1_VOC	1
M_MEAN_VOC_CNT	0.982428
SUM_0_DURATION_VOC_1	1
SUM_1_DURATION_VOC_1	0.999881
CNT_0_VOC_EV	1
CNT_1_VOC_EV	0.999978
CNT_2_VOC_EV	1
DUR_0_VOC_EV	0.999999
DUR_1_VOC_EV	0.999855
DUR_2_VOC_EV	0.999999

Performance on Control Data Set

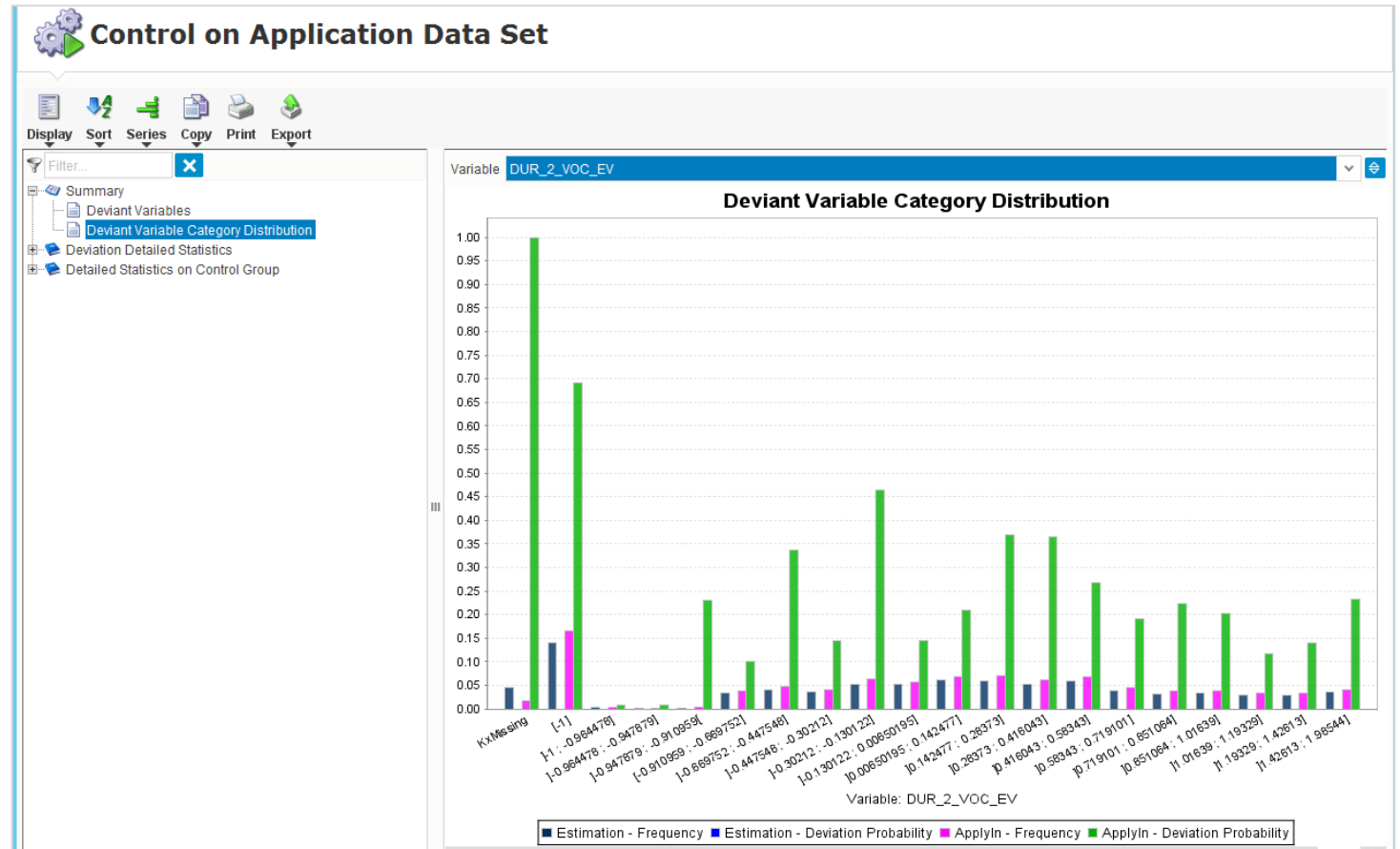
Target: TARGET

rr_TARGET		
KI	0.7469	
KR	0.9842	

Evaluating the Initial Churn Model

Demonstration – Step-by-step guide 9

- The detailed evaluation points to the frequency counts that have deviations in the lower and missing values.
- For example, for DUR_2_VOC_EV, the high probability deviations are in the missing values.
- Examine the other reports.
- Importantly, the TARGET and score (rr_TARGET) do not have high probability of deviation.



Evaluating the Initial Churn Model

Summary

- Phase 5 of the CRISP-DM process is the evaluation of the model.
- In this unit, you have seen how to evaluate a predictive model using a deviation analysis.
- Deviations can occur for a number of reasons, including changes in the quality of the data, changes in the general usage characteristics, or maybe due to a “seasonal” type variation in the data.
- You have seen how to use the Deviation Analysis in SAP Predictive Analytics.



Thank you.

Contact information:

open@sap.com

© 2017 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See <http://global.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.



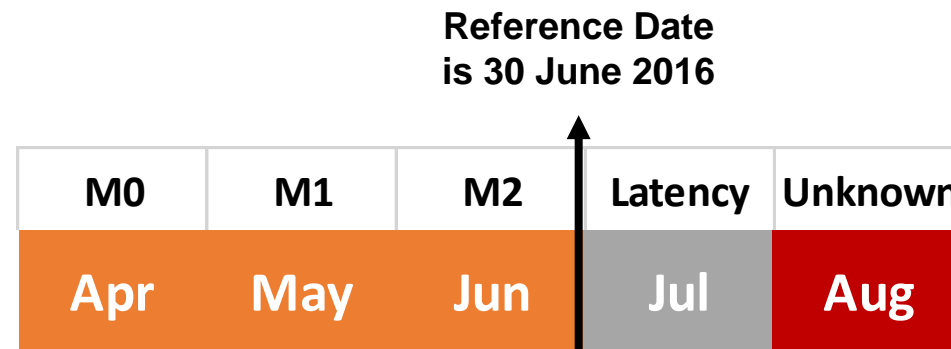
Week 3: Develop, Evaluate, and Deploy Models

Unit 4: Deploying the Initial Model Using SAP Predictive Analytics

Deploying the Initial Model Using SAP Predictive Analytics

Introduction

- When you apply a model onto a new data set, the data you use will usually represent more up-to-date data than the data you trained the model on.
- The Apply Data Set must contain all of the same explanatory variables that are in the churn model, but updated so that M0 refers to April, M1 refers to May, and M2 refers to June data.
- The model will then predict the probability of an account churning in August.



Deploying the Initial Model Using SAP Predictive Analytics

Demo





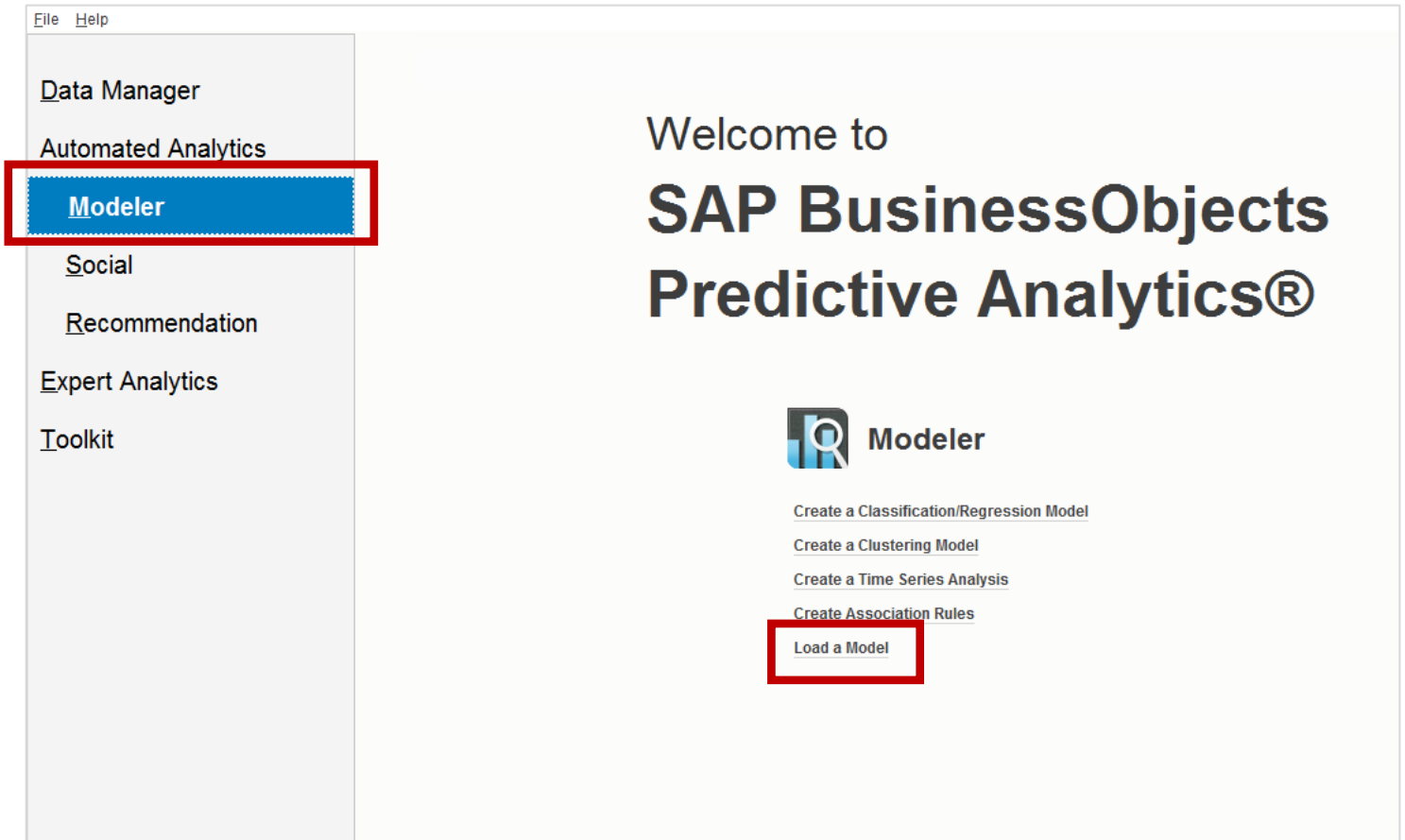
Appendix

Step-by-step guide to applying the model

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 1

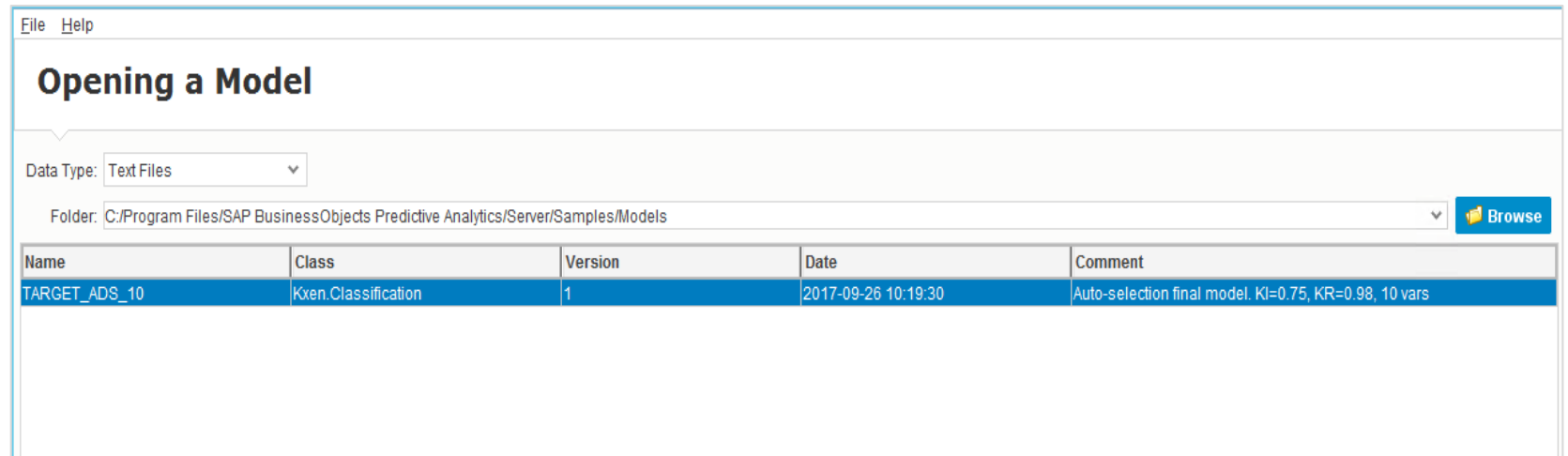
- Select Modeler / Load a Model



Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 2

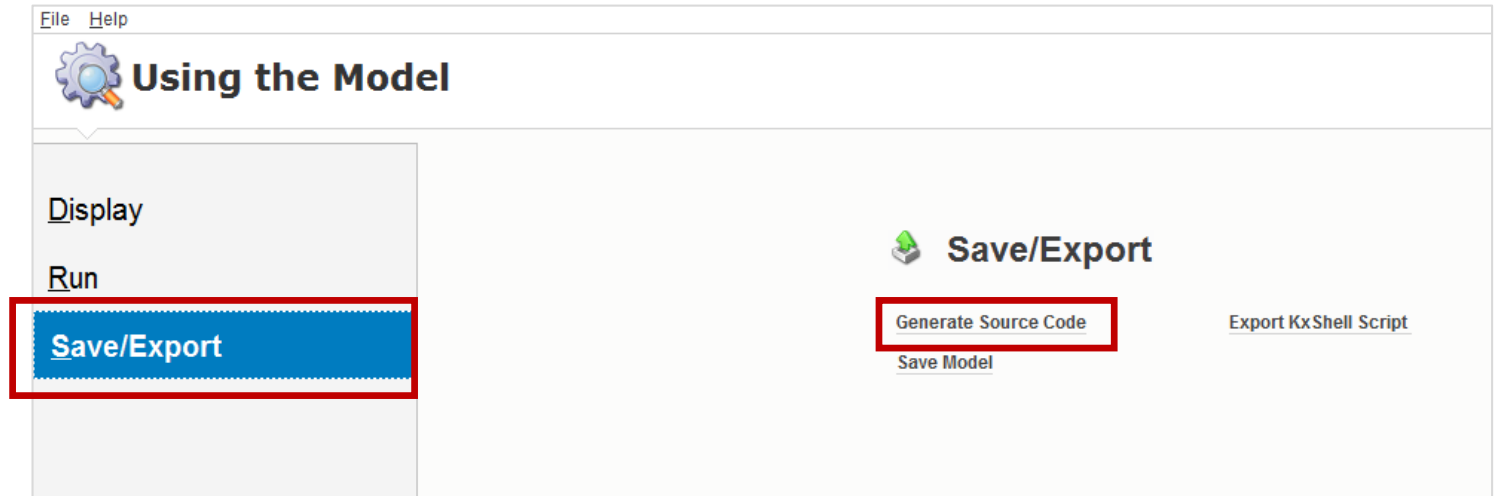
- Locate the model you saved in
C:/Program Files/SAP BusinessObjects Predictive Analytics/Server/Samples/Models
- Choose Open.
- This will open the model at the Using the Model interface.



Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 3

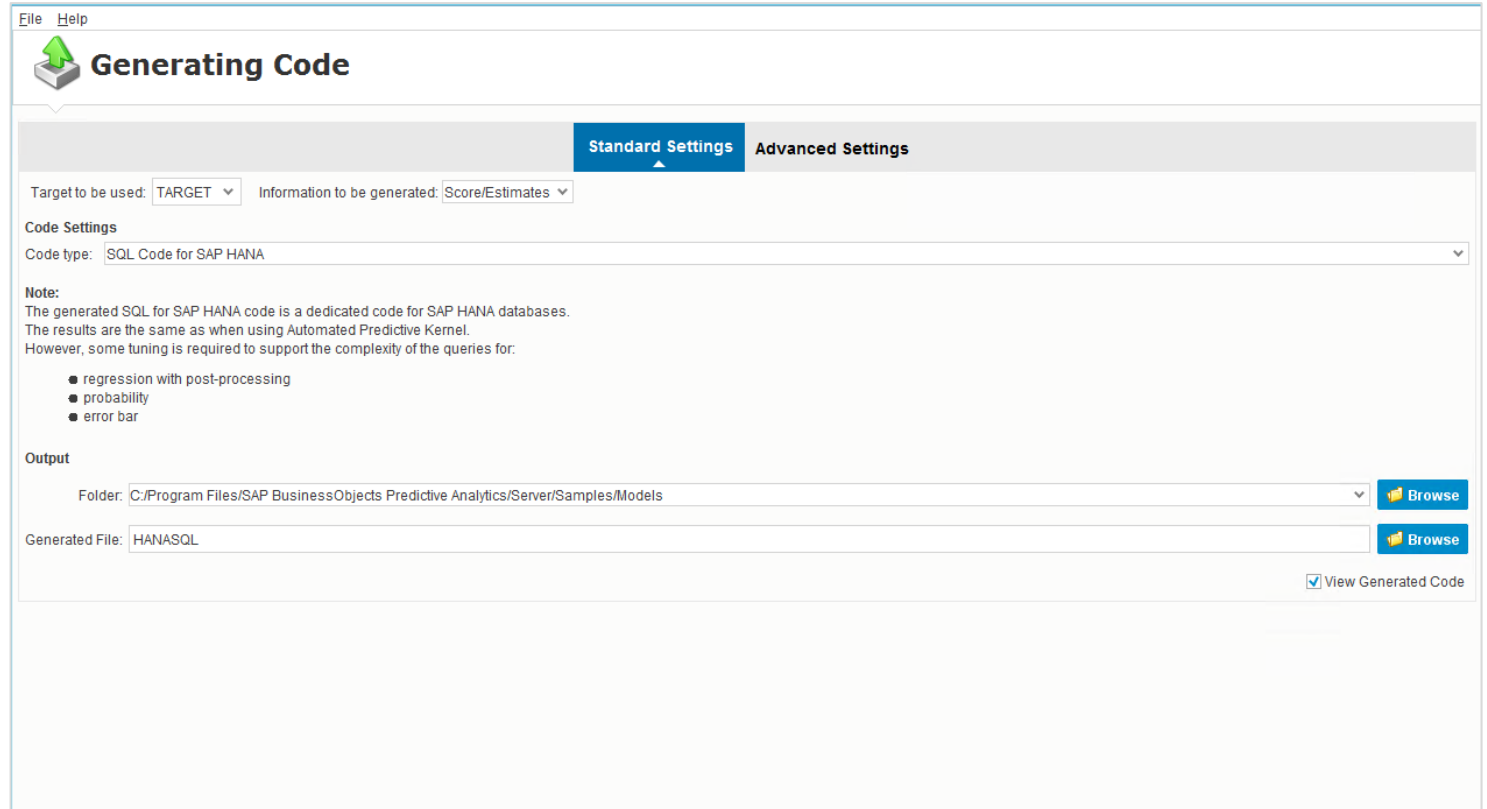
- You can view the SAP HANA SQL Code that will be deployed in the database to produce the model scores.
- Select:
Save/Export/Generate Source Code



Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 4

- Code Type – SQL Code for SAP HANA
- Folder = C:/Program Files/SAP BusinessObjects Predictive Analytics/Server/Samples/Models
- Generated File = HANASQL
- Select View Generated Code
- Choose Generate



The screenshot shows the 'Generating Code' dialog box in SAP Predictive Analytics. The 'Standard Settings' tab is active. The 'Target to be used' is set to 'TARGET' and 'Information to be generated' is set to 'Score/Estimates'. Under 'Code Settings', the 'Code type' is 'SQL Code for SAP HANA'. A note states: 'The generated SQL for SAP HANA code is a dedicated code for SAP HANA databases. The results are the same as when using Automated Predictive Kernel. However, some tuning is required to support the complexity of the queries for: regression with post-processing, probability, error bar'. In the 'Output' section, the 'Folder' is 'C:/Program Files/SAP BusinessObjects Predictive Analytics/Server/Samples/Models' and the 'Generated File' is 'HANASQL'. Both fields have 'Browse' buttons. A 'View Generated Code' checkbox is checked at the bottom right.

File Help

Generating Code

Standard Settings Advanced Settings

Target to be used: TARGET Information to be generated: Score/Estimates

Code Settings

Code type: SQL Code for SAP HANA

Note:
The generated SQL for SAP HANA code is a dedicated code for SAP HANA databases.
The results are the same as when using Automated Predictive Kernel.
However, some tuning is required to support the complexity of the queries for:

- regression with post-processing
- probability
- error bar

Output

Folder: C:/Program Files/SAP BusinessObjects Predictive Analytics/Server/Samples/Models Browse

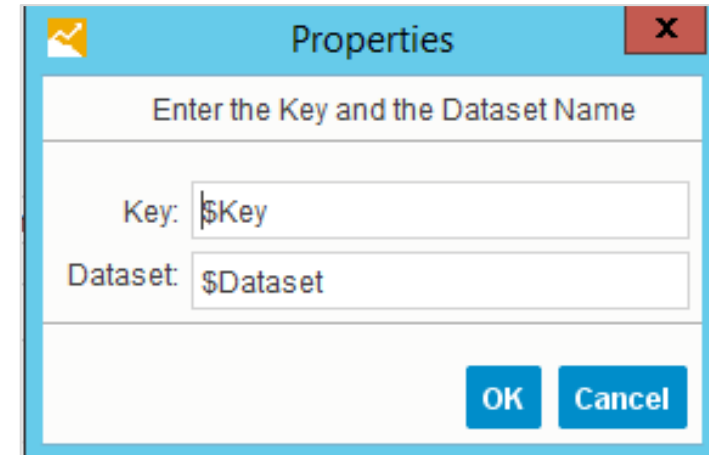
Generated File: HANASQL Browse

☒ View Generated Code

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 5

- Use the default values for the Key and Dataset names when you are just going to view the code.
- Choose *OK*.







The image shows a 'Properties' dialog box with a light blue header bar containing a small icon and a close button (X). The main area has a title 'Enter the Key and the Dataset Name'. Below this, there are two input fields: 'Key:' with the value '\$Key' and 'Dataset:' with the value '\$Dataset'. At the bottom right, there are two buttons: 'OK' and 'Cancel'.

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 6

- This is the SAP HANA SQL that has been generated, and that will be deployed in the database to calculate the model scores.
- Choose *Next*.

 **View Source Code HANASQL.sql**

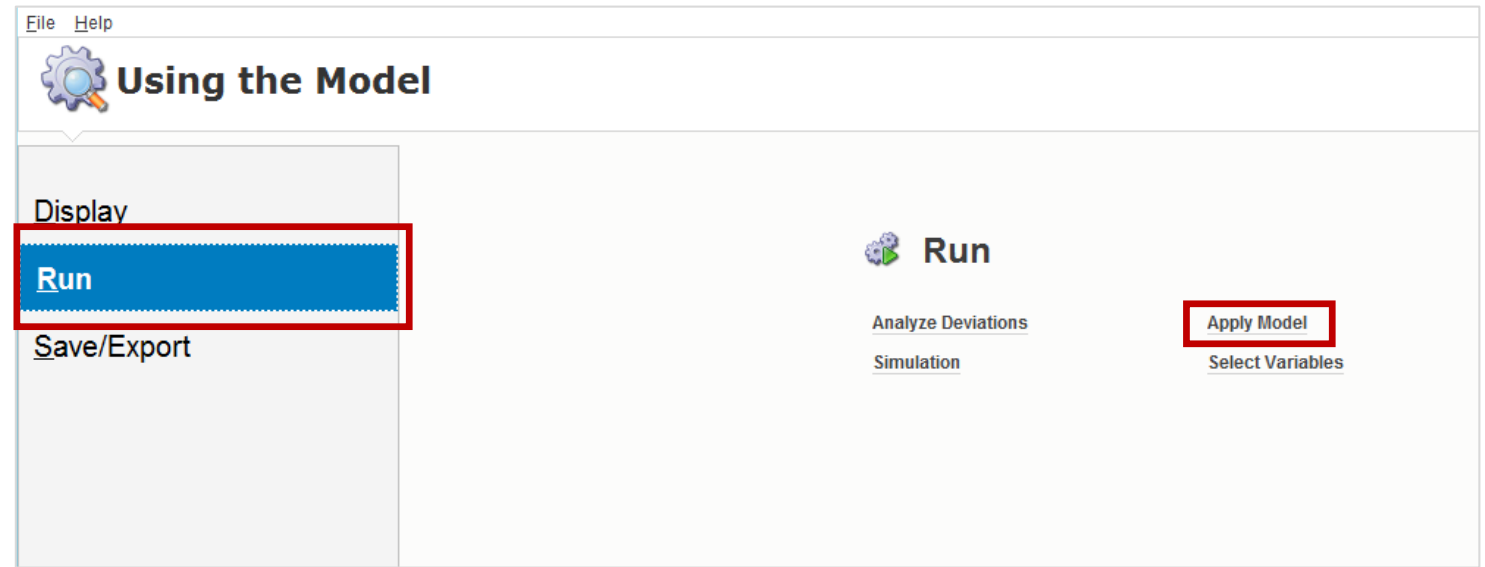
Copy Print Save

```
-- Automated Analytics - Registered Version 9.2.0.8 - Copyright 2017 SAP SE or an SAP affiliate company. All rights reserved. - Model built in 9.2.0.8 - Model Name is TARGET_ADS_10 - Model Version is 1
SELECT $key,
(
  ( CAST( (CASE
    WHEN ( "TENURE_MTHS" IS NULL ) THEN -7.069544926594e-2
    WHEN "TENURE_MTHS" <= 6.0e0 THEN 1.226758389888e-2
    WHEN "TENURE_MTHS" <= 7.0e0 THEN (-6.221855052189e-4*"TENURE_MTHS"+1.600069693012e-2)
    WHEN "TENURE_MTHS" <= 8.0e0 THEN 1.102321288829e-2
    WHEN "TENURE_MTHS" <= 9.0e0 THEN 7.782164183374e-3
    WHEN "TENURE_MTHS" <= 1.1e1 THEN (-3.820015406248e-2*"TENURE_MTHS"+3.495062454213e-1)
    WHEN "TENURE_MTHS" <= 1.4e1 THEN (9.189871991548e-3*"TENURE_MTHS"-1.269055614763e-1)
    WHEN "TENURE_MTHS" <= 1.6e1 THEN (7.848666461582e-2*"TENURE_MTHS"-1.138316465451e0)
    WHEN "TENURE_MTHS" <= 1.7e1 THEN (-3.292497340401e-3*"TENURE_MTHS"+1.701501258486e-1)
    ELSE 1.141776710617e-1
  END) AS DOUBLE)+
  CAST( (CASE
    WHEN ( "M2_DATA_USAGE_MB" IS NULL ) THEN -1.428912857686e-2
    WHEN "M2_DATA_USAGE_MB" <= 2.4e1 THEN 2.074839008267e-2
    WHEN "M2_DATA_USAGE_MB" <= 9.83e2 THEN (4.655927543005e-5*"M2_DATA_USAGE_MB"+1.963096747681e-2)
    WHEN "M2_DATA_USAGE_MB" <= 9.84e2 THEN 4.493632953604e-2
    WHEN "M2_DATA_USAGE_MB" <= 1.246e3 THEN (-3.148501262073e-5*"M2_DATA_USAGE_MB"+7.591758195401e-2)
    WHEN "M2_DATA_USAGE_MB" <= 1.345e3 THEN (-5.121703274155e-4*"M2_DATA_USAGE_MB"+6.752294179068e-1)
    WHEN "M2_DATA_USAGE_MB" <= 1.446e3 THEN (-4.920884870023e-4*"M2_DATA_USAGE_MB"+6.48219342546e-1)
    WHEN "M2_DATA_USAGE_MB" <= 1.653e3 THEN (-8.50187517538e-5*"M2_DATA_USAGE_MB"+5.92401722412e-2)
    WHEN "M2_DATA_USAGE_MB" <= 1.684e3 THEN (-8.088882202204e-4*"M2_DATA_USAGE_MB"+1.255796403614e0)
    WHEN "M2_DATA_USAGE_MB" <= 1.757e3 THEN (-1.165835622737e-4*"M2_DATA_USAGE_MB"+8.946768103369e-2)
    WHEN "M2_DATA_USAGE_MB" <= 1.809e3 THEN (2.728149556378e-4*"M2_DATA_USAGE_MB"-5.948237963965e-1)
    WHEN "M2_DATA_USAGE_MB" <= 1.862e3 THEN (2.728149556378e-4*"M2_DATA_USAGE_MB"-5.948237963965e-1)
    WHEN "M2_DATA_USAGE_MB" <= 1.919e3 THEN (2.522629590705e-4*"M2_DATA_USAGE_MB"-5.565559787853e-1)
    WHEN "M2_DATA_USAGE_MB" <= 2.063e3 THEN (7.938756016091e-5*"M2_DATA_USAGE_MB"-2.24708121376e-1)
    WHEN "M2_DATA_USAGE_MB" <= 2.151e3 THEN (-7.119498232286e-6*"M2_DATA_USAGE_MB"-4.618275921956e-2)
    WHEN "M2_DATA_USAGE_MB" <= 2.24e3 THEN (-7.119498232285e-6*"M2_DATA_USAGE_MB"-4.618275921962e-2)
    WHEN "M2_DATA_USAGE_MB" <= 2.329e3 THEN (-6.561829240495e-6*"M2_DATA_USAGE_MB"-4.743193776129e-2)
    WHEN "M2_DATA_USAGE_MB" <= 2.418e3 THEN (-6.561829240495e-6*"M2_DATA_USAGE_MB"-4.743193776135e-2)
    WHEN "M2_DATA_USAGE_MB" <= 2.559e3 THEN (7.995088523802e-5*"M2_DATA_USAGE_MB"-2.566319263138e-1)
    WHEN "M2_DATA_USAGE_MB" <= 2.592e3 THEN (6.948441788656e-4*"M2_DATA_USAGE_MB"-1.830661058132e0)
    WHEN "M2_DATA_USAGE_MB" <= 2.675e3 THEN (2.94919575665e-4*"M2_DATA_USAGE_MB"-7.940564866342e-1)
    WHEN "M2_DATA_USAGE_MB" <= 2.868e3 THEN (9.400476327475e-5*"M2_DATA_USAGE_MB"-2.564198461068e-1)
    WHEN "M2_DATA_USAGE_MB" <= 3.334e3 THEN (6.608614947386e-5*"M2_DATA_USAGE_MB"-1.763492617228e-1)
    WHEN "M2_DATA_USAGE_MB" <= 3.524e3 THEN (1.651601084299e-4*"M2_DATA_USAGE_MB"-5.067262864729e-1)
    WHEN "M2_DATA_USAGE_MB" <= 4.096e3 THEN (4.655927543005e-5*"M2_DATA_USAGE_MB"-6.631994013435e-2)
  )
)
```

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 7

- Select Run / Apply Model



Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 8

- Application Dataset
 - Data Type = Data Base
 - Folder = SAP HANA DB/Catalog/TRIALSCHEMA
 - Data = “CHURN_MODEL_APPLY_DATA”
- Generation Options
 - Generate = Predicted Value Only (these are scores)
 - Mode = Apply
- Results Generated by the Model will be written to:
 - Data Type = Data Base
 - Folder = SAP HANA DB
 - Data = “TRIALSCHEMA”.“OUTPUT_SCORES”
- Deselect *Add Score Deviation*
- Deselect *Use Direct Apply in the Database*
- Choose *Apply*

Applying the Model

Application Dataset

Data Type: Data Base

Folder: SAP HANA DB/Catalog/TRIALSCHEMA

Data: CHURN_MODEL_APPLY_DATA

Metadata C:\Program Files\SAP BusinessObjects Predictive Analytics\Server\Samples\MDR (Text Files)

Generation Options

Generate: Predicted Value Only

Mode: Apply

☐ Add Score Deviation

Results Generated by the Model

Data Type: Data Base

Folder: SAP HANA DB

Data: TRIALSCHEMA"."OUTPUT_SCORES"


☐ Use Direct Apply in the Database






Please note that the privileges to enable the Direct Apply in Database have not been granted in this training system. The Score Deviation, which is created when you apply the model, is only available when you use the direct apply in the database. Therefore, these options are deselected.


Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 9


- Choose View Output

 **Applying the Model**

 Stop  View Type  Copy  Print  Save

 Your model was applied successfully

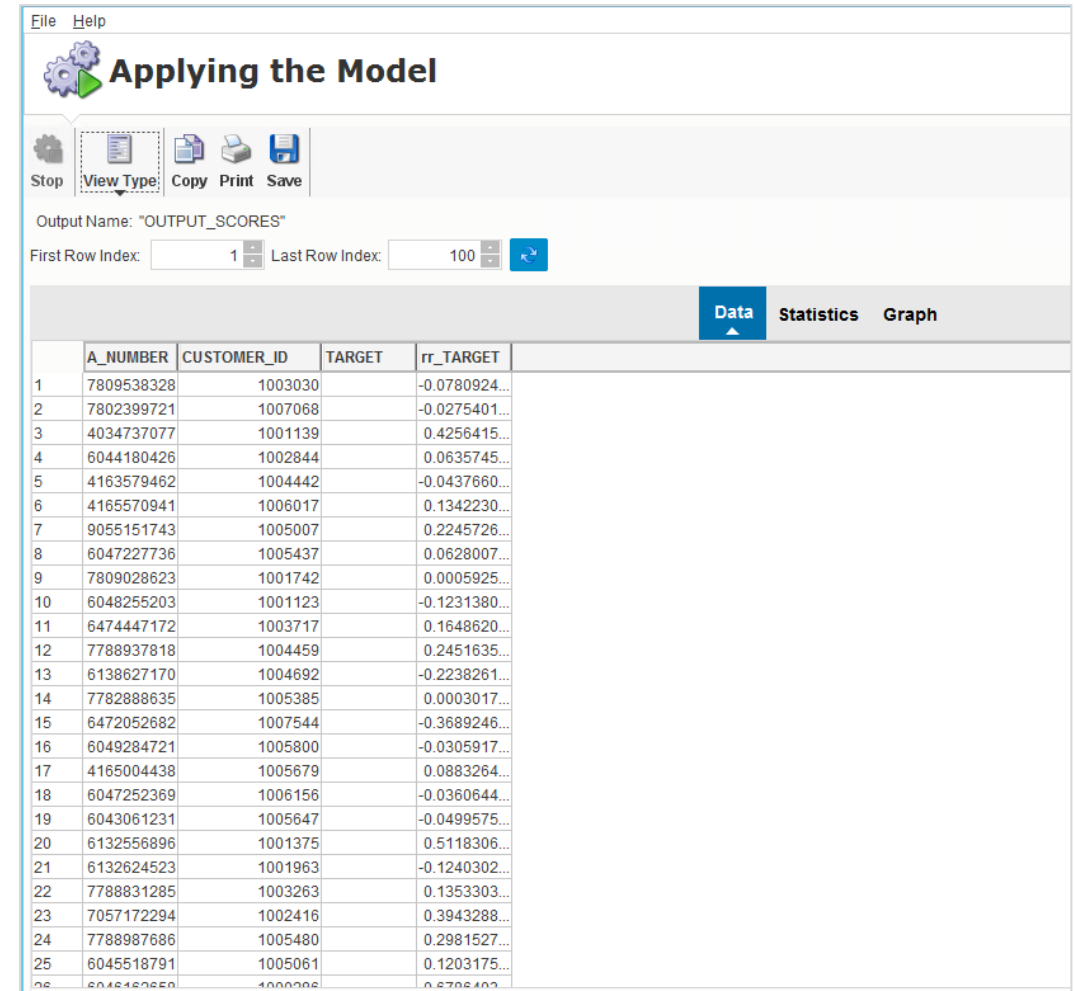
Summary	
Model:	TARGET_ADS_10
Input:	SAP HANA DB/Catalog/TRIALSCHEMA/"CHURN_MODEL_APPLY_DATA" (Kxen.ODBCStore)
Output:	SAP HANA DB/"TRIALSCHEMA"."OUTPUT_SCORES" (Kxen.ODBCStore)


View Output

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 10

- You are viewing the top 100 rows only. This can be increased by using the Last Row Index and refreshing the data.
- Each row represents a separate A_NUMBER.
- TARGET is blank, because there are no actual values for TARGET in August. The model is predicting these values.
- rr_TARGET is the score that the model has calculated. It has negative and positive values. The higher the score, the more likely the customer is a churner. The lower the score, the less likely the customer is to be a churner.
- You can click the header in the table for rr_TARGET, and this will order the score. You can then select the A_NUMBERS with the highest scores, as these will most likely be the churners.



File Help

Applying the Model

Stop View Type Copy Print Save

Output Name: "OUTPUT_SCORES"

First Row Index: 1 Last Row Index: 100

Data Statistics Graph

	A_NUMBER	CUSTOMER_ID	TARGET	rr_TARGET
1	7809538328	1003030		-0.0780924...
2	7802399721	1007068		-0.0275401...
3	4034737077	1001139		0.4256415...
4	6044180426	1002844		0.0635745...
5	4163579462	1004442		-0.0437660...
6	4165570941	1006017		0.1342230...
7	9055151743	1005007		0.2245726...
8	6047227736	1005437		0.0628007...
9	7809028623	1001742		0.0005925...
10	6048255203	1001123		-0.1231380...
11	6474447172	1003717		0.1648620...
12	7788937818	1004459		0.2451635...
13	6138627170	1004692		-0.2238261...
14	7782888635	1005385		0.0003017...
15	6472052682	1007544		-0.3689246...
16	6049284721	1005800		-0.0305917...
17	4165004438	1005679		0.0883264...
18	6047252369	1006156		-0.0360644...
19	6043061231	1005647		-0.0499575...
20	6132556896	1001375		0.5118306...
21	6132624523	1001963		-0.1240302...
22	7788831285	1003263		0.1353303...
23	7057172294	1002416		0.3943288...
24	7788987686	1005480		0.2981527...
25	6045518791	1005061		0.1203175...

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 11

- Application Dataset
 - Data Type = Data Base
 - Folder = SAP HANA DB/Catalog/TRIALSCHEMA
 - Data = “CHURN_MODEL_APPLY_DATA”
- Generation Options
 - Generate = Probability & Error Bars (these are probabilities)
 - Mode = Apply
- Results Generated by the Model will be written to:
 - Data Type = Data Base
 - Folder = SAP HANA DB
 - Data = “TRIALSCHEMA”.“OUTPUT_PROBA”
- Deselect *Add Score Deviation*
- Deselect *Use Direct Apply in the Database*
- Choose *Apply*

Applying the Model

Application Dataset

Data Type: Data Base

Folder: SAP HANA DB/Catalog/TRIALSCHEMA

Data: CHURN_MODEL_APPLY_DATA

Metadata C:\Program Files\SAP BusinessObjects Predictive Analytics\Server\Samples\MDR (Text Files)

Generation Options

Generate: Probability & Error Bars

Mode: Apply

☐ Add Score Deviation

Results Generated by the Model

Data Type: Data Base

Folder: SAP HANA DB

Data: TRIALSCHEMA"."OUTPUT_PROBA"


☐ Use Direct Apply in the Database






Please note that the privileges to enable the Direct Apply in Database have not been granted in this training system. The Score Deviation, which is created when you apply the model, is only available when you use the direct apply in the database. Therefore, these options are deselected.


Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 12


- Choose View Output

 **Applying the Model**

 Your model was applied successfully

Summary	
Model:	TARGET_ADS_10
Input:	SAP HANA DB/Catalog/TRIALSCHEMA/"CHURN_MODEL_APPLY_DATA" (Kxen.ODBCStore)
Output:	SAP HANA DB/"TRIALSCHEMA"."OUTPUT_PROBA" (Kxen.ODBCStore)


View Output

Deploying the Initial Model Using SAP Predictive Analytics

Demonstration – Step-by-step guide 13

- You will see two extra columns.
 - proba_rr_TARGET - This is the prediction probability that the observation belongs to the target category of the target variable, from 0 to 1. There are no negative values.
 - bar_rr_TARGET – the prediction range, or maximum error.
- The higher the probability, the more likely the customer is a churner. The lower the probability, the less likely the customer is to be a churner.
- You can click the header in the table for proba_rr_TARGET, and this will order the score. You can then select only the A_NUMBERS with the highest probability to churn.

File Help

Applying the Model

Stop View Type Copy Print Save

Output Name: "OUTPUT_PROBA"

First Row Index: 1 Last Row Index: 100

	A_NUMBER	CUSTOMER_ID	TARGET	rr_TARGET	proba_rr_TARGET	bar_rr_TARGET
1	7809538328	1003030		-0.078092448...	0.034674841849679	0.5351689374714
2	7802399721	1007068		-0.027540102...	0.039224373350674	0.645360961186
3	4034737077	1001139		0.4256415795...	0.551786604936495	1.4564123856
4	6044180426	1002844		0.0635745208...	0.050879797425063	0.7985692442311
5	4163579462	1004442		-0.043766090...	0.037624030664131	0.7079702293889
6	4165570941	1006017		0.1342230187...	0.124299261071695	1.11327045098
7	9055151743	1005007		0.2245726692...	0.139268127783759	1.244160126618
8	6047227736	1005437		0.0628007944...	0.0508021808093	0.7985692442311
9	7809028623	1001742		0.0005925076...	0.042829314802061	0.6535491435685
10	6048255203	1001123		-0.123138026...	0.028722908009669	0.603747212722
11	6474447172	1003717		0.1648620535...	0.128968873843629	1.11327045098
12	7788937818	1004459		0.2451635662...	0.170331491889083	1.244160126618
13	6138627170	1004692		-0.223826178...	0.013894408224266	0.4881546094626
14	7782888635	1005385		0.0003017485...	0.042776577406031	0.6535491435685
15	6472052682	1007544		-0.368924643...	0.013888888888889	0.4881546094626
16	6049284721	1005800		-0.030591702...	0.038923399046318	0.7079702293889
17	4165004438	1005679		0.0883264766...	0.086364789506198	1.076495193089
18	6047252369	1006156		-0.036064484...	0.03838362744723	0.7079702293889
19	6043061231	1005647		-0.049957508...	0.037013381124768	0.7079702293889
20	6132556896	1001375		0.5118306864...	0.718003592370937	1.4564123856
21	6132624523	1001963		-0.124030294...	0.028547918781414	0.603747212722
22	7788831285	1003263		0.1353303044...	0.124468019498718	1.11327045098
23	7057172294	1002416		0.3943288444...	0.501664173778161	1.498547824127
24	7788987686	1005480		0.2981527427...	0.222735779001911	1.391140671566
25	6045518791	1005061		0.1203175002...	0.122179958446801	1.076495193089

Deploying the Initial Model Using SAP Predictive Analytics

Summary

- You have now learnt how to apply a predictive model.
- You also now know the difference between the model scores and probabilities that are produced in SAP Predictive Analytics.
 - The score has negative and positive values. The higher the score, the more likely the customer is to be a churner. The lower the score, the less likely the customer is to be a churner.
 - The probability varies from 0 to 1. There are no negative values. It is the probability that the observation belongs to the target category of the target variable.



Thank you.

Contact information:

open@sap.com

© 2017 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See <http://global.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.