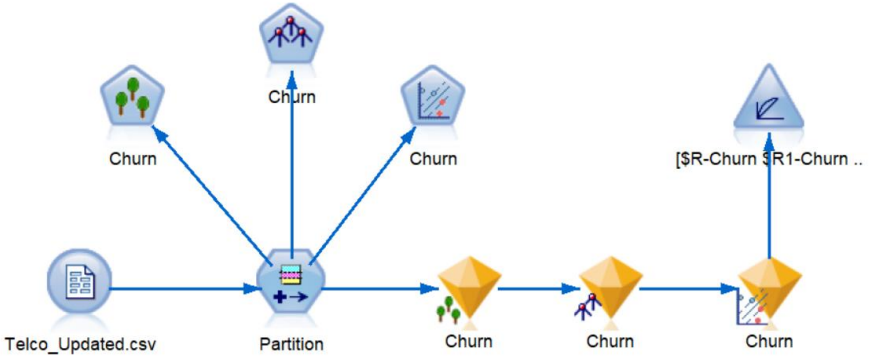
SMARTVISION Europe

Training Webinar



Building and Applying Predictive Models in IBM SPSS Modeler

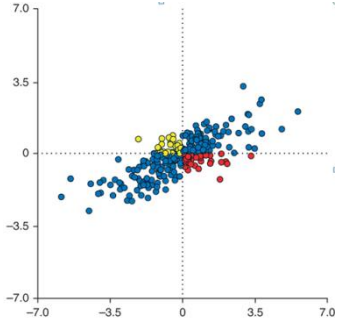
Jarlath Quinn

Contents

- The CRISP-DM process
- Modelling techniques
- Defining model inputs and target fields
- Creating a model interactively using a decision tree algorithm
- Using the Auto-Classifier node
- Assessing model performance
- Using a predictive model to maximise profit
- Applying a model to generate predictions against current data

What do we mean by 'Predictive Analytics'?

- Actually, it's not *always* about prediction
- It's different from Business Intelligence or MI reporting
- However, Predictive Analytics *does* creates important new data
- These data take the form of estimates, probabilities, forecasts, recommendations, propensity scores, classifications or likelihood values
- Which in turn can be incorporated into key operational and/or insight systems



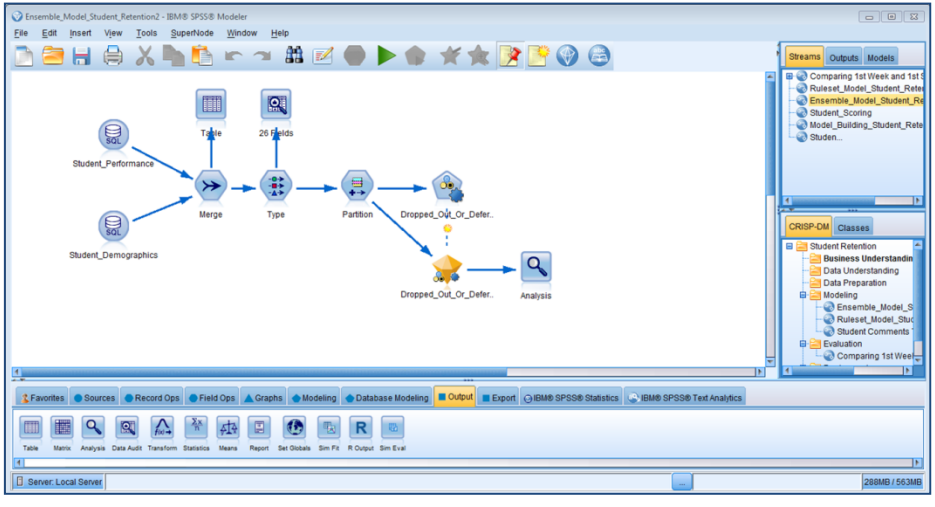
Part A

An Overview of the CRISP-DM Process

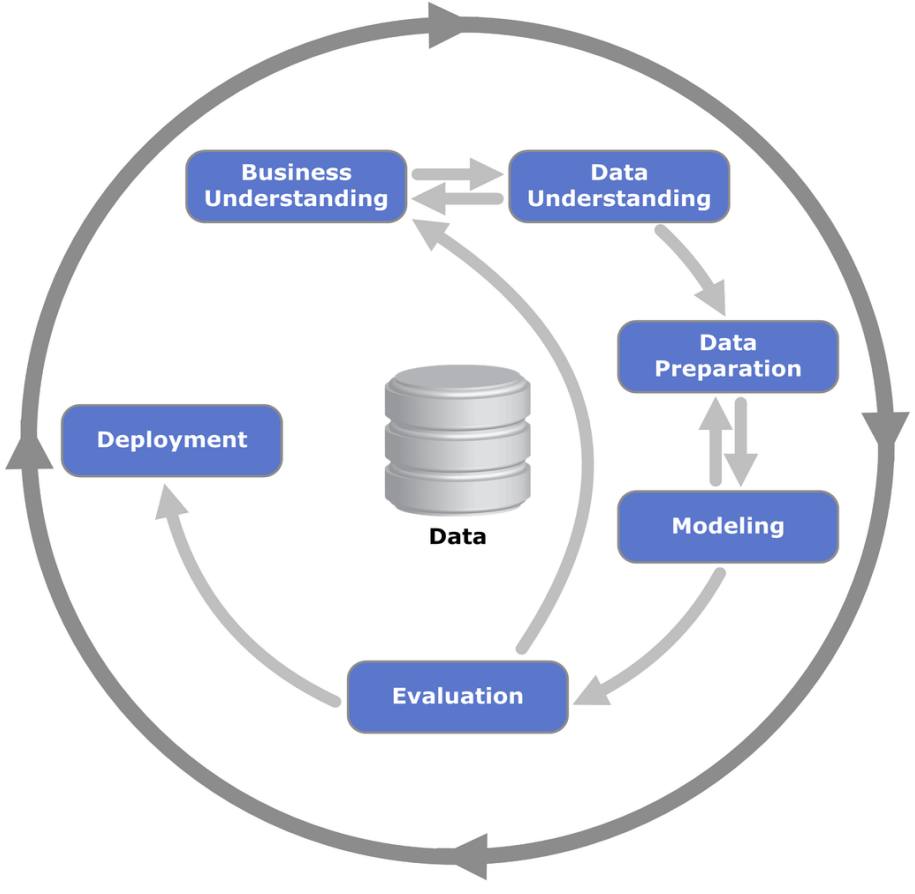
Why IBM SPSS Modeler?

- Powerful, industry-leading analytics platform
- Intuitive, graphical interface used by analytical experts and non-experts alike

- A one-stop shop for:
  - Reading and consolidating data
  - Transforming and manipulating key fields
  - Uncovering patterns and predictive models
  - Integrating results with operational and reporting systems



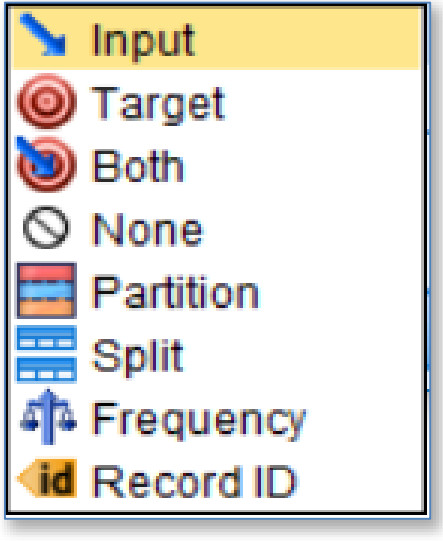
CRISP-DM: Cross Industry Standard Process for Data Mining



https://www.sv-europe.com/crisp-dm-methodology/

Variable Roles

- **Input -** The field will be used as an input (or predictor) for a predictive modelling algorithm.
- **Target -** The target role sets the field to be the outcome that a model attempts to predict. For example, outcomes such as campaign response, credit worthiness, customer churn or fraud are all commonly defined as the Target field within the Type node.
- **Both -** Certain association model algorithms (such as Apriori) allow fields to serve the role of Target or Input. Setting the role to 'Both' will enable this.
- **None -** The field will be ignored by most analytical procedures.
- **Partition -** Partition fields are specially created for the purposes of testing model accuracy. They work by splitting the data into separate samples for training, testing, and (optionally) validation purposes. The model is built using the training partition and then applied to the test (and possibly the validation) sample for see how well it performs. As we will see later, the Partition *node* is used to create fields that do precisely this
- **Split -** The split role only applies to nominal, ordinal or flag fields. It specifies that a model is to be built for each possible value of the field.
- **Frequency -** The frequency role only applies to numeric fields. The role enables a field value to be used as a frequency weighting factor for the record. The Frequency role is only supported by certain algorithms (C&R Tree, CHAID, QUEST and Linear models).
- **Record ID** – This role indicates that the field will be used as the unique record identifier. It is supported by Linear models though most algorithms ignore it.



Part B

Building a Predictive Model

Types of Model

- Modeler supports three modes of analytics
  - Predict
    - Category membership
    - Numeric Value
  - Segmentation
    - Cluster (group) things together based on their characteristics/attributes
  - Association
    - Find events that occur together

Predictive Modelling Algorithms

Statistical

Rule Induction/Trees

Machine Learning

Other Modelling Algorithms

Association

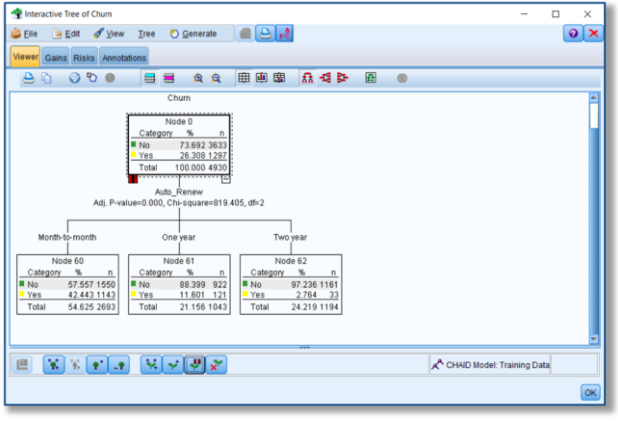
Clustering and Anomaly Detection

Time Series

Automatic

SMARTVISION Europe

CHAID



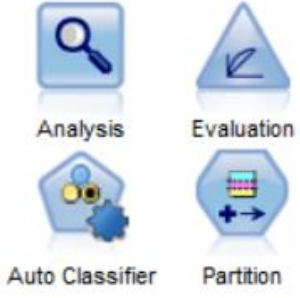
Demo: Using the CHAID node to build an Interactive Decision Tree model

Part C

Assessing and Selecting Predictive Models

Assessing Model Performance

- Accuracy
- Interpretability
- Stability
- Coherence
- Simplicity
- Performance



What is your success criteria?

Scenario 1:

- *Currently around 100K customers cancel their contracts each month.*
- *Previous tests have shown that sending offers to a random group of 100K current customers approaching contract end dates each month, reduces the churn amount by 7K customers (7%).*
- *We would like to send 50K offers to customers with a high risk of churning with view to reducing churn by 14K (14%).*
- *The model should therefore identify the 50K current customers approaching their contract end date who have the highest likelihood of churning with a view to retaining at least 14K of them.*

What is your success criteria?

Scenario 2:

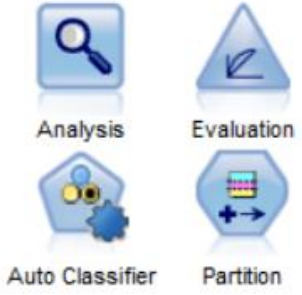
- *Currently we have a monthly outbound email campaign that targets 400K customers approaching their contract end.*
- *Tests indicate that this retains around 15K customers who would otherwise have churned.*
- *However, the total cost of the offers is very expensive as we suspect many customers with a low likelihood of churning also redeem them.*
- *We would like to reduce the outbound mailing to 100K customers and still retain the 15K customers who would otherwise have churned.*

What is your success criteria?

Scenario 3:

- *Currently we have a monthly outbound email campaign that targets 60K key customers approaching their contract end.*
- *The campaign offers the next month's standard call costs for free if they extend their contract by two months.*
- *However, the campaign makes a net loss of $130K each month as many customers churn after the additional two months anyway.*
- *We would like to make a net profit of $70K per month by targeting only those with a low likelihood of churning after accepting the offer.*

SMARTVISION Europe



Demo: Methods of Assessing and Choosing Models

Part D

Scoring Models

SMARTVISION Europe

Demo: Scoring Models and Exporting Predictions

Working with Smart Vision Europe Ltd.

- **Sourcing Software**
  - You can buy your analytical software from us often with discounts
  - Assist with selection, pilot, implementation & support of analytical tools
  - http://www.sv-europe.com/buy-spss-online/
- **Training and Consulting Services**
  - Guided consulting & training to develop in house skills
  - Delivery of classroom training courses / side by side training support
  - Identification & recruitment of analytical skills into your organisation
- **Advice and Support**
  - offer ‘no strings attached’ technical and business advice relating to analytical activities
  - Technical support

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you