Massive modeling with Predictive Analytics

Or

Why and How to deploy and maintain thousands of models

Hervé KAUFFMANN / Global Predictive Analytics
Legal disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has 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's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP 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. This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This document is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP's willful misconduct or gross negligence.

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, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.
What is Predictive Analytics?

Use Historical DATA to detect PATTERN

Predict by executing PATTERN on current DATA

Control & Maintain the Quality over time

Probabilities
Values
Groups
Rules
The Predictive Process with SAP Predictive Analytics…

AUTOMATED MODE

SEMANTIC layer production is automated thanks to DATA MANAGER

MODEL creation is automated thanks to MODELER

MODEL Management and Consumption are automated thanks to PREDICTIVE FACTORY

All steps of the process are controlled
All this is only possible because Predictive analytics uses RDBMS as DATA SOURCE

DATA MANAGER
- Is where DATASETS are designed
- Pushes code to the RDBMS to produce the DATASET at a given time

MODELER
- Creates the model handling all technicalities
- Push code to the RDBMS to produce the results

PREDICTIVE FACTORY
- Schedules model Control & Maintenance
- Schedules batch production of results

DATA is not moved from RDBMS, results are produced in RDBMS
An Expert Mode to go Further

Automated mode has been designed for productivity:

- Automate Analytical data set generation
- Speed up models generation
- Industrialize models execution
- Automate models maintenance
- Generate documentation

Beyond the Automated Mode, it is possible to use the more traditional Expert mode approach:

- Custom R-Code
- Predictive Analytics Libraries implemented in HANA
Massive Modeling needs AUTOMATION

SAP answer: Automated Mode is based on SRM (Structured Risk Minimization)

SRM merges three steps into one

1. **Variable selection**
   - Uses all available variables

2. **Automated data preparation**
   - Variable encoding – Nominal, ordinal, and continuous
   - Target optimized binning – creating meaningful value groups using SRM principles
   - Missing value handling
   - Outlier detection
   - Event aggregation – by time, value, or sequence in Data Manager

3. **Automated model testing** on holdout sample
Adding explanatory variables has potentially high benefits:

- Does not cause over-fitting
- More variables can only add to the model quality
- Random variables do not harm quality or reliability
- Highly correlated explanatory variables do not harm the modelling process (multicolinearity)
- Efficient scaling with additional variables

Free of distribution assumptions:

- Normal distributions are not necessary
- Skewed distributions do not harm quality
- Resistant to outliers

Indicates robustness of the model

- Prediction Confidence metric indicates if there are insufficient training examples to develop a robust model
Deploy Where and When it Matters

Producing the results should take into account how they will be used

BATCH
- Offline campaigns
- Churn detection
- Customer knowledge
- ...

ON THE FLY / REAL TIME when it makes sense
- Credit card fraud detection
- Online recommendations
- Call Centers
- Quality control
- ...

Models are transformed into EXECUTABLE code in DATABASE (SQL) or in 3rd party environments (C++, JAVA…) to be executed when and where it matters
Why building many models is important

Exploration models
- Analyze data through statistical prism
- Control data consistency
- Validate an « a priori » business question

All divisions of the company need answers
- Sales
- Marketing
- Finances
- Human Resources
- Logistic
- Pricing
- Fraud
- Production
- Strategy

In most cases answering one question requires more than one model
- Heterogeneous population studied
- Multiple type of targets to be addressed
How to build many models: Massive Modeling

In most cases, building one global model has no sense:

- Stores are different
- Borders are different
- Products are different
- Customers are different

For each question you need to define homogenous populations to analyze:

- Split the data in different populations thanks to a physical segmentation (eg: 1 store = 1 model)
- Split the data in different populations thanks to a business rule segmentation (eg: re-group customers depending on their purchase rate)
- Split the data in different populations thanks to a statistical segmentation (eg: clustering / supervised clustering)
SAP Predictive Analytics answer: Segmented Modeling

Create and manage a set of models sharing the same settings as a single model

Define the right granularity
- Products, Products per Stores, Type of Products per Store…
- Stores, Stores per Town, Stores per Revenue…
- Customers, Customers per Stores, Customers per baskets value

Build the attribute for the chosen granularity
- If necessary

The number of models will also depend on the type of analysis
- Classification
- Regression
- Forecasting

Tell predictive Analytics on which attribute build automatically the models
- 1.000 stores = 1.000 models
- 100.000 couples product/store = 100.000 models
Cox Communications: Supercharging Customer Relationships with SAP® Predictive Analytics

Company
Cox Communications Inc.

Headquarters
Atlanta, Georgia

Industry
Telecommunications

Products and Services
Cable entertainment and broadband services

Employees
50,000

Revenue
US$15 billion

Web Site
http://ww2.cox.com

Challenge
• Cox needed to increase sales and retain customers in multiple diverse regions around the country. Marketing tactics in one region were not as effective in others.

Key Influencers
• Building models the traditional way took too long to meet the needs for each region. Cox needed to understand the propensity of individuals to buy new services in different parts of the country.

How SAP PA Addressed the Problem
• It took the Cox data scientists an average of 4 weeks to develop one model per region for a total of 24. Automation allowed them to develop models in a day or two. Productivity increased 24x. They were able to generate > 600 models with the same number of folks. These additional targeted models facilitated campaigns tailored to individual products in each market segment.

Results of implementing PA
• The specificity of the campaigns allowed increased upsell revenues and churn reductions.
  • Additional product sales to existing customers increased 14%
  • The rate of customers leaving was reduced by 28%

“With SAP Predictive Analytics we’ve increased the products sold per household by 14%.”
Parimala Narasimha, Director of Marketing Science, Cox Communications Inc.

14%
More products per customer household

28%
Reduction in customer churn rate

80%
Reduction in model creation time

24x
Greater throughput for central analysts (from 24 to 600 predictive models)
US Large Retailer: Data Driven Company

SAP Predictive Analytics

Users by Division

260k models daily in production. At POS level

Data analysis used for every decisions
Eldorado: Boosting Sales Forecast Accuracy with the SAP® Predictive Analytics Solution

Company
Eldorado LLC

Headquarters
Moscow, Russia

Industry
Retail

Products and Services
Consumer electronics and domestic appliances

Employees
15,000

Revenue
€2.4 billion (2012)

Web Site
www.eldorado.ru

Objectives
• Analyze data stored in the SAP® 360 Customer solution from over 1.5 million point-of-sale transactions for more than 420 product groups and sales of over 8,000 products each month
• Improve forecast precision to boost sales and reduce inventory costs

Why SAP
• Trusted technology partner with a proven record of delivering success across the industry
• Ability to further leverage real-time access to large volumes of data already available with the SAP NetWeaver® Business Warehouse application powered by the SAP HANA® platform and the SAP Planning for Retail application
• Ease of use, precision of predictive models, and innovative automated tools available with the SAP InfiniteInsight® solution

Future plans
• Migrate further SAP applications to SAP HANA, leveraging the full potential of in-memory computing technology
• Continue to expand and evolve the business using world-class IT systems and innovation
• Other use cases: pricing and promotion analysis, store clustering, store location selection, marketing mix

"SAP InfiniteInsight has given us a scalable approach to create accurate forecasts across our business."

Elena Zhukova, Head of Analytics, Eldorado LLC

Fast
Building approximately 500 predictive models a month, a task impossible with traditional modeling techniques that required weeks or months to build a single model

Flexible
Creating forecasts for assortment planning, shelf replenishment, pricing and promotion analysis, store clustering, store location selection, and sales and purchasing planning

Accurate
Achieving up to 82% accuracy in sales forecasts, a 10% improvement over prior forecasting techniques