H₂O.ai

# Predictive Model Factory

**Case Study**

CISCO™

**Industry:**

Cisco is a multinational technology company that develops and sells a wide variety of networking equipment including routers to switches. More recently Cisco has also started to sell a variety of analytics and software solutions such as data center analytics and automation software.

**Challenges:**

- Build models to deliver accurate predictions about customer propensity to buy across Cisco's extensive product portfolio.
- Overcome speed and scalability challenges associated with analyzing an exploding amount of information about buying patterns.

**Solution:**

- Deploy H2O's pre-built ready-to-use algorithms.
- Streamline the Cisco prediction factory to a much simpler input controlled using R.
- Leverage H2O's in-memory compute engine to minimize the need for expensive storage resources.
- Incorporate a greater amount of up to date customer buying information.

**Results:**

- A dramatic reduction in processing time from more than a month to two days - despite a dramatically larger dataset.
- The ability to analyze the entire set of customer data, rather than just a small sample - delivering far greater accuracy.
- Immediate incorporation of new buying data with no need to prepare models in advance.
- The power to deploy of advanced machine learning algorithms (including GBM and deep learning) - rather than simple decision trees - for more accurate results.

**Inside Cisco's Predictive Model Factory**
How Cisco's 20-person advanced analytics team modernized data preparation, model training, and score deployment procedures for 60,000 customer intelligence models, across hundreds of millions of observations.

**From an artisanal approach to mass production of predictive models**
Cisco has long-embraced the power of predictive analytics. Every quarter, the networking leader's 20-person advanced analytics team deploys a set of propensity to buy (P2B) models, which predict whether certain companies will buy

certain products within a given timeframe. Marketing and sales teams rely on these predictions to focus on their highest potential revenue opportunities.

What started as a two person effort to develop a predictive model for a couple of Cisco products back in 2007, evolved from an adhoc, "statistician intensive" process to a semi automatic factory, a tight production process that creates, validates and deploys 60 thousand specific predictive models every quarter. Streamlining data preparation steps and repeating common sense methods for training machine learning models have allowed Cisco to predict purchases for very customized combinations of markets, customer segments and sizes, technologies and relationship maturity with Cisco. Since its inception the P2B factory has been associated with

*"We score every single one of the 160 million companies in have in our databases and individually calculate the probability that they will buy each of the most important products and services we sell, and that for very distinct marketing situations involving acquisition, cross-sell and up-sell situations. This information helps teams customize their selling effort, concentrating on the technologies that each of our customers may need at a certain point in time, and clearly stating the value proposition of our products and solutions to them. This strategy requires a high degree of specificity, so we end up creating separate models by country, company size and historical relationship with Cisco."*

– Lou Carvalheira, Advanced Analytics Manager at Cisco

more than $3B in revenue, a tremendous demonstration of the power of data science when combined with excellence in marketing and sales execution.

To guarantee a high level of precision, Cisco creates roughly 60,000 models to calculate P2B scores for multiple types of products. With more than 2,000 variables included in the pool of training data, Cisco relies on historical information (from past purchase behavior, marketing interactions, and macroeconomic indicators) as well as recent data (from purchase behavior, customer satisfaction surveys, and marketing interactions) to predict buying patterns for the upcoming quarter.

**Navigating speed and scalability challenges**
One of the biggest challenges that Cisco's advanced

analytics team faces is the ability to keep up with rapidly changing buying patterns in the technology market.  Smaller companies are beginning to invest in technologies that were once only limited to large organizations. Cisco is constantly introducing new

*"The technology market is so vibrant and dynamic. Patterns of buying behavior change, and new data comes in every time – we are constantly improving the ways we collect data about our customers. We want to bring this intelligence to our models."*

*– Lou Carvalheira, Advanced Analytics Manager at Cisco*

products and services. Even Cisco's customers undergo changes through mergers, acquisitions, growth, and setbacks.

To ensure accurate predictions and to keep models up-to-date, Cisco re-creates all its predictive models from scratch every single quarter – a process that, until recently, took more than four weeks. To avoid even longer processing times, models were trained on relatively small samples that were rarely larger than 100,000 cases.

Operating on a tight schedule, a mistake of any sort inevitably meant long hours for the data mining teams and opportunity cost for the business. As a result, Cisco's advanced analytics team remained limited to standard techniques (decision trees and logistic regression) and was unable to test competing algorithms such as ensembles, grid search, and deep learning.
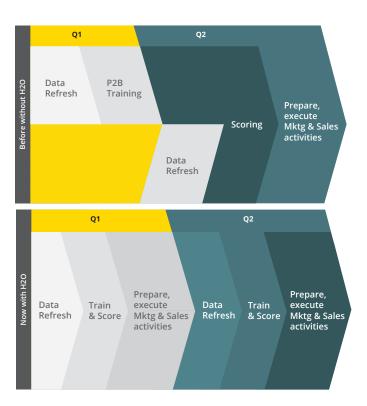
"Decision trees are good. They're quick. They're reliable but not the best in terms of generating predictions." -Lou Carvalheira, advanced analytics manager at Cisco.

The process, at the time, was akin to an assembly line and provided Cisco's advanced analytics team very limited opportunity to explore the data – and business challenges – as deeply as the team would have liked.

"The reality is that we just don't have enough people on the data mining team." -Lou Carvalheira, advanced analytics manager at Cisco.

**Modernizing the factory**
Cisco's advanced analytics team is always looking for ways to improve and build upon its operations. In 2014, Cisco began using H2O's open source in-memory distributed machine learning platform. Cisco uses

**Before without H2O**

| Q1 | | Q2 | |
|---|---|---|---|
| Data Refresh | P2B Training | | Prepare, execute Mktg & Sales activities |
| | Data Refresh | Scoring | |

**Now with H2O**

| Q1 | | | Q2 | | |
|---|---|---|---|---|---|
| Data Refresh | Train & Score | Prepare, execute Mktg & Sales activities | Data Refresh | Train & Score | Prepare, execute Mktg & Sales activities |

H2O to make better and faster predictions through sophisticated, ready-to-use algorithms and processing power to analyze bigger datasets, more variables, and more models. Carvalheira and his team used H2O to streamline their process flow, which is now controlled using R:

**H20 Process Flow - controlled with R**

1. Define environment and main parameters
2. Read training and scoring files
   - Reserve subset or training for validation
3. Define list of predictors and target variables
4. Train first stage for each target product (what is the

*"We use H2O where it shines in terms of model training and scoring."*

*– Lou Carvalheira, Advanced Analytics Manager at Cisco*

probability of purchase?)
- Train GLM and evaluate model against validation
- Train a couple of Random Forests using different architectures (fewer, deeper trees vs more, shallower trees) and evaluate model against validation
- If product has traditionally been hard to predict then train a GBM and evaluate model against validation

- Use best model (AUC) to score. If more than one has good result, use ensemble to compose probability of purchase

5. Train second stage for each target product (how much will be purchased?)

   - Train a GLM and evaluate results on validation set
   - Train a GBM and evaluate results on validation set
   - Choose the best model to score and predict purchase value

6. Save intermediate results and treat the next target product (step 4)

7. Save final score files and clean things up

Without H2O, models needed to be prepared in advance to avoid any delays in scoring with new buying patterns. The team spent so much time preparing models that there was limited time left for using the scores in resulting sales and marketing activities. With H2O, new buying patterns are incorporated into models immediately. Scores are published sooner, leaving more time for campaign planning and execution.

What once took longer than a month now takes Cisco two days – a 15x increase in speed. With faster production times, Cisco's analytics team is empowered to evolve its models.

### Key achievements

Cisco relies heavily on advanced analytics to drive marketing and sales efforts, and the P2B Factory is a fundamental part of the company's vision. After seven years, however, existing procedures needed to expand, predict new products and services, and become more accurate – all in less time.

H2O made that possible allowed that improvement to happen with its powerful in-memory distributed computing algorithms, great support team, and cost-effective solution. Cisco's advanced analytics team is now in a position to accomplish the following:

- Create more models for more products – with more historical data.
- Train its tens of thousands of models using data from tens of millions, rather than hundreds of thousands of cases from its database of 160 million customers – at a processing time of 1/15th the previous production time.
- Experiment with deep learning and GBM algorithms to tackle complex prediction challenges – and establish a process to constantly test new models for higher precision.
- Accomplish more with fewer technical resources – Cisco's SAS environment and shared by many groups, resulting in a busy queue.

On a 4-node cluster with 24 cores and 128GB memory each, Cisco can now run its P2B factory to train models for the same amount of products, scoring all 160 million companies in a couple of days. H2O has significantly improved Cisco model factory's speed and throughput, enabling sales and marketing to predict customer behavior sooner and more accurately, and leading to more efficient forecasting and campaign planning.

### About H2O.ai

At H2O.ai we see a world where all software will incorporate AI, and we're focused on bringing AI to business through software. H2O.ai is the maker behind H2O, the leading open source machine learning platform for smarter applications and data products. H2O operationalizes data science by developing and deploying algorithms and models for R, Python and the Sparkling Water API for Spark. Some of H2O's mission critical applications include predictive maintenance, operational intelligence, security, fraud, auditing, churn, credit scoring, user based insurance, predicting sepsis, ICU monitoring and more in over 5,000 organizations. H2O is brewing a grassroots culture of data transformation in its customer communities. Customers include Capital One, Progressive Insurance, Zurich North America, Transamerica, Comcast, Nielsen Catalina Solutions, Neustar, Macy's, Walgreens, Kaiser Permanente and Aetna.