MACHINE LEARNING IN RETAIL
Content

CEO Statement 3
What is Machine Learning 4
A Brief History of Machine Learning 5
Daisy’s Math Factory 6
Daisy’s Theory of Retail™ 7
Implementation 9
CEO Statement

I have always viewed mathematics as a means to an end – not just an academic discipline but as a practical way to solve problems. The issue involved in managing very large databases intrigued me and I was drawn to machine learning science because it focused on applying adaptive learning technologies to solve complex, real world problems. Machine learning is based on algorithms that can adapt and refine themselves based on new input data without human involvement or rules based programming.

I started playing with machine learning in the early 1990’s when I was one of the first 4 people worldwide utilizing IBM’s state of the art data mining software and I wrote an often quoted book on data mining. What was missing I felt, and still feel, was the ability to execute of the promise of the “digital age”. The output of our efforts, at the time, was often technical and did not lead to better business decisions. Even today, analytics is powerful but the full potential has not yet been realized. “Big Data” is part of the solution however “Big Analytics” represented by machine learning approaches is needed to maximize the value of the data.

Machine learning approaches along with the dramatic reduction in the cost of computing power are starting to conquer this challenge. Today, we are using next generational machine learning technologies to fulfill the potential of adaptive insight creation to answer optimization questions for our clients. Even a few years ago, these types of questions would have been too difficult to tackle because the data sets were simply too large or because the interrelationships between the variables were too complex for available analytical approaches. Even more exciting, today’s machine learning approaches are adaptive – the more data the system touches, the better refined the algorithms become and the better the resulting outcomes become.

I am thrilled to have this opportunity. Our distinctive capabilities in machine learning processes combined with our Theory of Retail™ algorithms, cloud based infrastructure, parallel in memory processing engine, data models and our experience working with clients in the retail and insurance space provides us with unique capabilities to help our customers take advantage of the promise of analytics and machine learning.

Twenty five years from now, I believe, the discussion will not be analytics – the discussion will be focused on mining our past decisions to make our machine learning models smarter to improve our overriding policies that govern future decision making processes.

Gary Saarenvirta

CEO
What is Machine Learning?

Machine learning is a branch of computer science that is used to describe both explicitly programmed algorithms for supervised (prediction/classification) and unsupervised (clustering/feature detection) learning. Machine learning is derived from the field of artificial intelligence (AI), which strives to provide computers with the ability to learn without being explicitly programmed.

Machine learning is not one specific scientific domain, but rather, a number of domains that offer many approaches to solving very complex problems. The general focus of most machine learning domains is the development of computer algorithms and programs. Artificial Intelligence and reinforcement learning further focuses on developing algorithms that can teach themselves to iteratively adapt when exposed to new data to continually improve their defined outcomes.

The growing interest in machine learning as a solution for improving various applications derives from the explosion and daunting amount of new data available, the continued reduction in the cost of processing power, the lack of skilled resources in this field and maybe, most importantly, the need to operate more efficiently and effectively in every domain of human output.
As a scientific endeavour, machine learning grew out of the quest for artificial intelligence initiated in the 1950’s. However, machine learning really started to flourish in the 1990’s when it separated from the more academically and symbolical focused study of AI and proceeded to incorporate some of the more practical applications of mathematical optimization approaches and computational statistics.

With Moore’s Law continually driving the cost of computing downward, progress in the field of machine learning has been rapid. Today’s machine learning is very different and far more powerful than machine learning of the past. It is now feasible to analyze all of the data we collect and understand relationships in data sets that we could never have comprehended previously. The machine learning family tree below demonstrates that many machine learning algorithms and methodologies have been around for a long time. However, it is only recently that the ability to apply these complex theories, their more advanced offspring and other mathematical calculations to large datasets has been possible.

Iterative machine learning has become a viable reality.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1952</td>
<td>An IBM scientist named Arthur Samuel uses the game of checkers to create the world’s first learning program that learned from observing winning strategies.</td>
</tr>
<tr>
<td>1957</td>
<td>Frank Rosenblatt designs “the perception”: a type of neural network that mimics the human brain to solve complex problems.</td>
</tr>
<tr>
<td>1967</td>
<td>The first programs able to recognize patterns were designed based on a type of algorithm called the “nearest neighbour”.</td>
</tr>
<tr>
<td>1981</td>
<td>Gerald Dejong introduces explanation based learning (EBL) in a journal article. In EBL, the algorithm discards irrelevant data and forms a general rule to follow.</td>
</tr>
<tr>
<td>1990’s</td>
<td>We begin truly applying machine learning in data mining, adaptive software, text learning and language learning. Reinforcement learning algorithms were also developed during this time period. (This is also when Daisy’s CEO begins getting his hands dirty in machine learning).</td>
</tr>
<tr>
<td>2000’s</td>
<td>The new millenium brings in an explosion of adaptive programming. These advancements are capable of recognizing patterns, learning from experience, abstracting new information from data and optimizing the efficiency and accuracy of its processing and output.</td>
</tr>
</tbody>
</table>
Daisy utilizes a patent pending approach to machine and reinforcement learning that analyses large databases to uncover insightful answers to complex optimization questions.

To be technical, we use the some of the most advanced methods in machine learning and computer science available. Some of our applied research interests include: association rule mining with progressive border sampling, regularized L1 regression, kernel based principal component analysis, Condorcet and other similarity based clustering methods, constrained non-linear optimization using genetic algorithms, non-linear constrained optimization using multi-swarm/particle swarm optimization with penalty functions, combinatorial optimization using multi-swarm/ particle swarm optimization, Monte Carlo gradient descent feature detection, back propagating feed-forward neural networks, radial basis function regression, regression tree feature detection, L1 regularized kernel regression with radial basis function kernels and auto-encoding deep belief nets for feature detection.

Yes, we are very much into math.

Our algorithms mimic biological processes and continuously adapt to improve themselves providing more accurate and effective results. They are also predictive: their outcomes lead to specific actionable recommendations. Further, the algorithms can be customized to specific customer parameters or constraints to shape specific outcomes. For example, store department floor areas could be a constraint used to optimized store layout to maximize sales. Or a retailer's desire to promote a certain event (Halloween, BBQ season, Black Friday) may be used as a constraint to focus output recommendations to a certain category of products within a certain time period.

We executed our processes on Daisy's massively parallel in-memory computational platform using Message Passing Interface and Daisy's proprietary master-slave architecture to parallelize computations and accelerate performance that delivers near real-time results.

To ensure security and scalability, we self-manage our own large private cloud. Our architecture includes a parallel computing layer, a relational data warehouse layer and big data layer to store years of point of sale and Internet of Things (IoT) data.
Daisy’s Theory of Retail™

The theory of evolution. The theory of general relativity. Newton’s laws of motion. There are many theories that work to help us connect the relationships between variables to better understand our world.

Industry specific enterprise data warehouses (EDW’s) introduced the concept of data models to represent important elements to track and measure performance. Advanced analytics introduced more data models to help identify which derived and available data elements were required to predict and optimize very specific outcomes. Machine learning draws from EDW and advanced analytic technologies and further advances this progression by adding the ability to capture any observable data element and allow the system, over time, to learn how to best incorporate all the information available to optimize a wide range of outcomes across far more elements.

Daisy’s machine learning environment intensely studies consumer purchasing behavior found in a client’s POS data and other relevant data inputs. Then, Daisy’s Theory of Retail™ model solves practical marketing optimization questions using a unique mathematical approach to measure the impact product marketing decisions have on other key “observables” such as cannibalization, promotional cadence, attached product sales, price and seasonality. Cyclically, results are returned into the system to further optimize future recommendations (see our Theory of Retail™ whitepaper for a full explanation).

What should I promote this week? At what price? What should be in our flyer?

What products drives the most associated product sales? When can I repeat a promotion?

Which SKUs should I delist? What store layout drives the most engagement and incremental sales?

**OBSERVATIONS**

Variables that describe the properties/state of the system that are observed.

**ACTIONS**

Variables that describe the actions taken.

**OUTCOMES**

The outcomes should be a form of reward, in this case an increase in sales for achieving certain system states.
Implementation

POS data
Daisy starts with a minimum of 2+ years worth of transactional data to ensure as many patterns and trends are uncovered.

Theory of Retail™
Daisy's Theory of Retail™ attempts to make sense of the transactional data being used and delivers actionable insights to improve engagement and sales.

Machine Learning Algorithms
Daisy’s algorithms mimic biological processes and continuously adapt to improve themselves providing more accurate and effective results over time.
After having analysed 100% of a client’s POS data using true machine learning, Daisy delivers weekly product, pricing and inventory recommendations that drive incremental sales growth.

**Optimization Recommendations**

Big data requires a lot of processing power. Daisy uses an in-memory parallel computing platform to crunch the data.
“Big Data is high volume, high velocity and high variety information assets that demand cost effective, innovative forms of information processing for enhanced insight and decision making.”

— Jeff Roster, Research VP, Retail & Wholesale for Gartner