



# PATTERN COMPUTER®

**PATTERN DISCOVERY:  
PARTNERING WITH DATA  
SCIENTISTS**

**Ty Carlson  
CTO**

Pattern Computer Inc.  
38 Yew Lane  
Friday Harbor, WA 98250



Copyright © 2021 Pattern Computer Inc.

All Rights Reserved

*References to specific products may be registered trademarks of their respective companies.*

---

*As a Data Scientist, partnering with Pattern Computer may be the fastest, most cost-efficient manner to discover patterns in your high dimensional datasets leading to rapid production of results needing your subject matter expertise, insights and interpretation of the results to provide meaningful patterns to your internal customers.*

## Introduction

The challenge in the age of ‘big data’ is finding the meaning and insights in these large datasets. Companies and business leaders have been archiving terabytes and petabytes of information in somewhat of a cryogenic fashion waiting for the day when someone will finally unlock the patterns contained within the data and discover the insights hidden within. Talented statisticians and data scientists have used many methods to gain insights, but often they are finding information based on prior knowledge – in some sense finding what they already know – and updating that information with the current results over the last year. What it is not allowing for, however, is for the data itself to tell its own story – what is the most significant data in the dataset, and how one set of features relate to other features in the data? What features within the datasets are the most responsible elements for a specific outcome, or set of outcomes? When the data tells the story, it reveals the full extent of what is within the data, and how the data is related to that which is known, and unknown. It can confirm what the data scientist already knows and may also provide the hints that experienced data scientists and subject matter experts use to develop new insights into related factors which heretofore were unrecognized players in the outcome.

When neural networks made their resurgence in the realm of machine learning methods, many felt that the neural networks provided the machine learning answers that people have been looking for. Typically, neural networks are first trained by thousands of examples of different things, such as images of a fish, dog, and cat. Then when the neural net is given an image, it can accurately recognize the three different things. However, without prior training, it can’t recognize a cow. Neural networks are very useful for pattern recognition, but they cannot perform pattern discovery. Another limitation of neural networks is that, presently, neural networks cannot inform the user regarding the decisions made within the neural network. Therefore, the factors which influence the decision as to whether the image is that of a cat, dog or fish are not known<sup>1</sup>.

The Pattern Discovery Engine™ ingests the customer datasets and identifies relationships in the datasets in an unsupervised mode, or in supervised mode, identifies the key sets of features most responsible for specific outcomes. For example, the set of gene expressions most directly associated with poor survivability in breast cancer. Another example might be the specific test sensor output(s) and range(s) and operational behavior most directly associated with the later failure of a part in service prior to its expected end-of-life.

## The importance of Data Scientists and Subject Matter Experts

Given the result set from the Pattern Discovery Engine, the Data Scientists and Subject Matter Experts work together to understand the story behind the results. “Under what conditions would these genes

---

<sup>1</sup> Pattern Computer may have further information in this area if you are interested. Please contact them if this area is of interest to you.

---

express themselves in the following manner?” or in the case of the part that failed within its operational service lifetime, “Given the test sensor acceptance data, and the operational information – we have discovered a pattern in the test results and specific operational behavior which is associated with an early failure in service.” Now the question is what do the results *mean*? The Data Scientist works with the Subject Matter Expert (SME) to understand the story the data is telling them. The SME will typically posit specific questions of the “What If” variety and the Data Scientists show the related information. Often the SME will immediately recognize at least a significant portion of the story upon first review. We hear “Oh, this is interesting,” and then the SME will relate that these conditions occur in the following scenario, and while it had been suspected that Feature B and D had something to do with the failures, it had not yet been considered that both would play a combined role in the failure state. In the case of medical genomics, the bioinformatician and researcher will look at the gene combination and recognize a biological pathway, where these genes interact as part of the human system. With these potential pathways identified, potential scenarios whereby the indicated gene expression pattern would occur are posited and reviewed carefully.

Having the information related to the significant features in the datasets allow data scientists and SMEs to collaborate – this may involve bringing together additional information to provide further detailed insight into the pattern discovery process now that they know that a given system or sub-system is involved. This “zooming-out” process of focusing on the wider datasets to provide a more complete scenario understanding is atypical of the data science process of reducing the size of the features and the scope of the datasets in order to make the problem space computationally solvable. Similarly, the data scientist and SMEs may identify secondary and tertiary scenarios by looking at the additional ranked items in the result sets – identifying that the specific failed state that they are analyzing has more than one, and potentially sets of different scenarios where failed states occur and appear in clusters to the pattern discovery engine. Together, the complete story related to the different failure scenarios can be mapped out and ranked, for example, with respect to scope of impact and ease of solution.

The work of the Pattern Discovery Engine, Data Scientists and Subject Matter Experts is a collaborative effort. Data for ingestion into the discovery engine provides more accurate results when the data provided is clean and does not have a lot of missing values. While the discovery engine can still discover the patterns, the cleaner the input signal, the clearer and stronger the results. Data Scientists know how to prepare the data while not compromising its integrity – which is often domain specific. For example, often in biological data, one does not create any number which has not been directly measured. However, in other domains, a missing value is simply filled in with a zero. This data preparation and domain familiarity are critical for maintaining the data integrity necessary for credible results. SMEs know their specific domain information. More than that – they have the history of how their company evolved those processes or systems to be able to interpret unified datasets with two different results – coincident with the deployment of the new systems or processes – things that the data sees, but without the SME there, it might take days or weeks to tease apart that subtle change in a system.

## Mathematical Model and Visualization

The Pattern Discovery Engine provides a mathematical model of the relationship between the features of the discovered pattern toward a specific result and report the accuracy of the model as well. In this

---

manner, the specific contributions of the various features and the nature of their impact on the result can be known and modeled. Using Pattern Computer's Dimensional Navigator™ data scientists and SMEs can have an immersive experience within the discovered model, visualizing up to 8 dimensions of the data and seeing the structure and the relationship of the discovered pattern to the data. This visualization experience provides the ability to see the clusters of data within the dataset and can provide new insights – based on human intuition enabled by the visual experience.

Together, the Data Scientists, Subject Matter Experts and the Pattern Discovery Engine create a powerful solution to the challenges of big data, allowing new discoveries to be gained from existing datasets where all the insights were thought to have been known. Data Scientists know how to correctly handle and curate the data, as well as how to correctly join the different datasets in the process of 'zooming-out' to see the larger scale of implications from the discovered patterns. SMEs can quickly identify the scenarios under which the result sets and related conditions given by the pattern discovery engine occur – the story and understanding quickly comes to life. Systems are optimized, new drug therapies can be realized, manufacturing processes improved, and costs and lives can be saved.

Welcome to the world of Pattern Discovery!