

Financial Markets Machine Learning

M. Temple-Raston, Decision Machine

Introduction

Businesses rely on the fact that people know how to count, and, subject to clear counting rules, the result is independent of the person counting.

Businesses are also subject to chance, probability / contingency. To calculate probability, a set of appropriate states (mutually exclusive and exhaustive) are defined for the problem at hand. Usually the states are obvious: up or not up, etc., so that the question becomes what is the probability that the next measurement will be “up”? The next measurement after that? And so on.

The objective of this paper is forecasting. Pattern recognition is not in scope.

While algorithms can be employed to estimate probability, in many real-world situations the probability can be solved mathematically (closed-form expression) through the Science of Counting formulated as machine learning (ScienceML). This whitepaper reports on the ScienceML web service from Decision Machine. Decision Machine’s Precision Alpha focuses ScienceML on applications in financial markets.

The next section presents the conceptual architecture, shows how to express the Science of Counting as Machine Learning, and plots ScienceML output.

Section three investigates NYSE and NASD closing price data using ScienceML. Strong quantitative evidence is presented to show that assets on the NYSE and NASD are rarely in equilibrium and many symbols display dissipative structures. Dissipative structures are important because they are associated with repeatable behavior in non-equilibrium systems (self-regulating).

Section four looks at forecasting.

Section five discusses contemporary operational challenges to conventional ML that ScienceML addresses. ScienceML is a tool for data scientists and business analysts that focuses their energy on understanding the dynamics of their data, not on model building.

Section six provides a summary and conclusions.

This is multi-year effort by Decision Machine to build E.T. Jaynes’ scientific reasoning “robot”. E.T. Jaynes’ classic book on Probability Theory: The Logic of Science is a natural blueprint for our web services.^[1]

ScienceML

The Science of Counting (SoC) dates to the 19th century and produced at least one well-known practical application: a theory of electrical circuits (generalized to Maxwell's equations later). The connection with electrical circuits suggests that the ScienceML web service should be able to calculate much more than just probability (exactly).

It is useful to have a conceptual architecture for the Science of Counting (Figure 1) circa 1850 as an electrical circuit. The diagram consists of two circuits that express the laws of thermodynamics and interact with each other. The LHS of Figure 1 is the mechanical (non-thermal) circuit driven by the energy, E . The RHS is the thermal circuit driven by heat, $Q = T_R - T$. The temperature of the system is denoted by T , and the temperature of the environment (reservoir), T_R . The heat flow in the RHS circuit is in the direction of hot to cold.

The mechanical resistance to energy, R_E , and the thermal resistance to heat, R_Q , are both in general time-dependent (arrows on the resistors). SoC is not expected to be in statistical or thermal equilibrium. For simplicity of presentation, Figure 1 suppresses the mechanical and thermal strains, ϵ_E and ϵ_Q , that store energy for later release.

Now, let $m = n/N$ be the percentage, given by the number of counted "up" events, n , divided by the total number of events counted, N . The expected value of m , $\langle m \rangle$, is the current in the mechanical circuit below and the total mechanical probability of an up-state. The "needle is moved" when the current, $\langle m \rangle$, increases or decreases.

The expected free energy, $\langle A \rangle$, is the stored energy (potential energy) that can be used to do work on $\langle m \rangle$. When free energy is increasing/decreasing, energy is being stored or released to do work on $\langle m \rangle$.

Energy can be converted into different forms (heat, potential energy and free energy), shared and stored between the mechanical and thermal circuits. A detailed accounting of the energy crossing between the two circuits (the boundary conditions) is required for forecasting, along with a measured value for T_R .

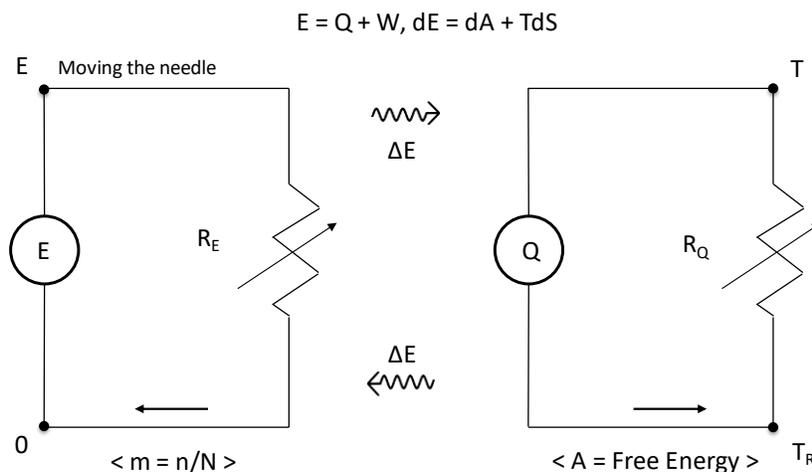


Figure 1. Conceptual architecture for the Science of Counting circa 1850. Implements the laws of thermodynamics for the Science of Counting (SoC).

A single useful takeaway from the conceptual architecture in Figure 1: there are strict algebraic relationships (equalities) between all the various circuit quantities presented above.

Since the relationships in the conceptual architecture are strictly algebraic, a user-supplied time-series viewed as the current in Figure 1 generates a set of scientific measurements, including notably probability. Scientific Machine Learning takes the time-series from an environment to produce a standard set of scientific measurements that can be applied to any vertical or specific interest.

The proof is by construction. ScienceML explicitly constructed closed-form expressions for the scientific measurements presented in Table A, that drive the dynamics of $\langle m \rangle$.

ScienceML	Comments
Probability (non-thermal)	Probability used in conventional Risk Analysis.
Probability (thermal)	When time-series is not (near-) statistical, thermal effects must be included. Thermal probability can be significantly larger than non-thermal probability. ^[2]
Energy	The energy measured from equilibrium. Equals zero at equilibrium.
Power	Energy per unit time. Also equals zero at equilibrium.
Mechanical Resistance	Force opposed to changing current, $\langle m \rangle$.
Mechanical Strain	Stored Energy. Emits or absorbs energy.
Noise	Energy stored as a deformation or lost to heat.
Temperature	Temperature is well-defined and equals $e/4$ at equilibrium.
Free Energy	Energy available to do work. For stocks, the energy available for price movements. Not all forms of energy can move prices.

Thermal Resistance	Force opposed to changing free energy, $\langle A \rangle$.
Thermal Strain	Stored Energy. Emits or absorbs heat.

Table A. Scientific measurements ScienceML uses to derive the dynamics of the system.

As a sample, Figure 2 (below) plots the probabilities and the energy. Six months of closing price data for stock symbol ZIM (bottom plot) is processed by ScienceML and plotted. Note that out of six months of data, only the last three months of processed data are plotted. The middle plot is the energy; the energy is zero at statistical equilibrium. The top plot presents two different types of probability: non-thermal (dotted) and thermal (solid). Observe that when the thermal probability dominates the non-thermal probability, the thermal probability drives the price movements. Compare in Figure 2.

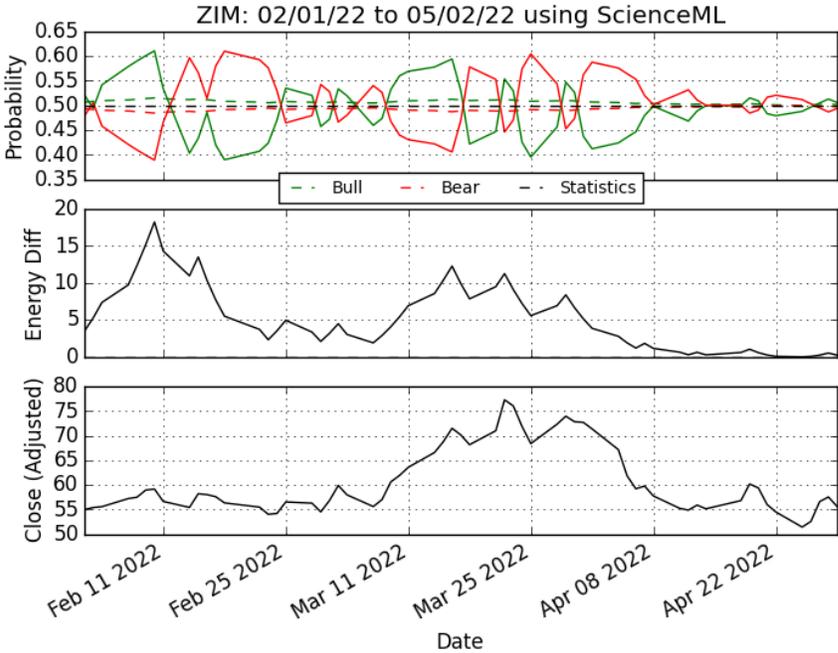


Figure 2. Market closing prices (input, bottom plot) for ZIM. Energy measured from equilibrium (middle plot). Both non-thermal (green/red dotted) and thermal probability (green/red solid) are in top plot.

Dissipative Structures^[3]

In the first week of January 2022, Precision Alpha analyzed all symbols on the NYSE and NASD using ScienceML. The analysis focused on symbols that had non-thermal probability greater than 0.51 or less than 0.49 for an entire year. In each exchange, roughly one third of all symbols were found to satisfy that (arbitrary) threshold. And of those symbols, all were found to have evidence of dissipative structures, both bullishly and bearishly. In January 2022 at least one third of the stocks on NYSE and NASD displayed strong thermal behavior.

Figure 3 plots the temperature and the free energy for the same closing price time-series data used in Figure 2 for ZIM. The system temperature and free energy are well-defined. The free energy is the energy available to do price movement work. When free energy is falling, free energy is doing price movement work. Note that neither the temperature nor the free energy is constant over the time duration. Heat flows in, and heat flows out.

This is characteristic of a thermodynamically open system. From the data, ZIM has operated out of, and often far from, statistical or thermodynamic equilibrium (steady state) in an environment with which it exchanges energy and information. The temperature of the external trading environment is denoted by T_R .

A thermodynamically open system is said to be a dissipative structure when it has a dynamical regime that is in some sense reproducible or periodic. That behavior is present in Figure 3.

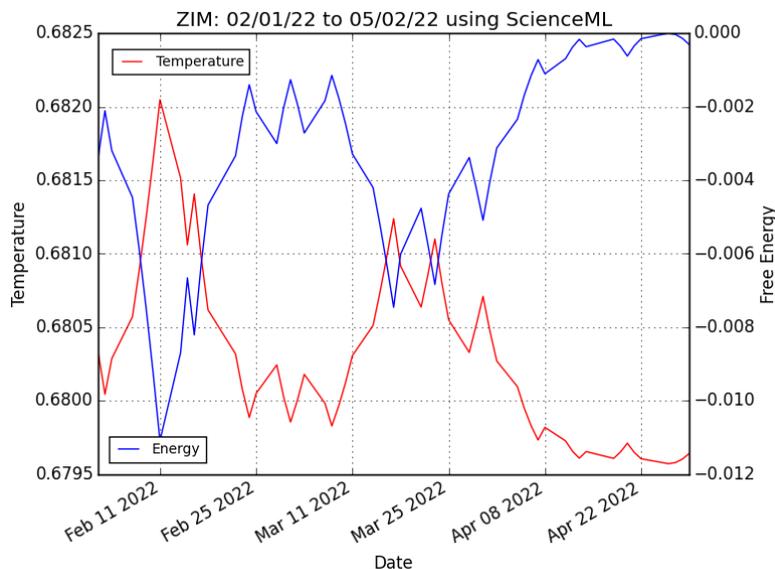


Figure 3. Plots of temperature and free energy over three months using six months of closing prices. Provides evidence that asset prices for ZIM have been far from either statistical or thermal equilibrium.

The temperature of the system, T , and the temperature of the external trading environment (T_R) are generally different. However, the two temperatures are the same when the system is in thermal equilibrium. In the Figure above, there is an interval in March when the temperature and the free

energy of ZIM are constant; this is thermal equilibrium with the trading environment. Thermal equilibrium provides a measurement of the temperature of the trading environment, in this case, $T_R = 0.6809$, approximately.

Decision Machine's analysis suggests that asset prices that behave as dissipative structures are common on the NYSE and NASD.

The trading environment temperature, T_R , is essential to accurate forecasting. This is taken up in the next section.

Forecasting

To summarize what we have learned from the previous two sections that could improve forecasting, we note that:

- Exact probability should lead to improved forecasting over one that has only estimated probability,
- Dissipative structures have elevated thermal probability that can dominate the system dynamics and should improve the forecast,
- The temperature of the environment, T_R , in which the system operates should improve a forecast over one that is unaware or discards the information.

Empirical testing suggests that the thermal probability is largest when T_R is set to the correct value, and therefore it is a critical value for forecasting. The previous section gave an example on how to measure T_R .

With a value for T_R , the equations of motions for the Science of Counting can also be derived for all the quantities in Table A. The exact forecasted probability based on learning can be put to immediate use: the exact probability is fed to a state machine to produce an output state. Now add the output state to the historical record and repeat the process. From this an individual forecast is produced for any finite horizon. An ensemble forecast is created by generating multiple individual forecasts.

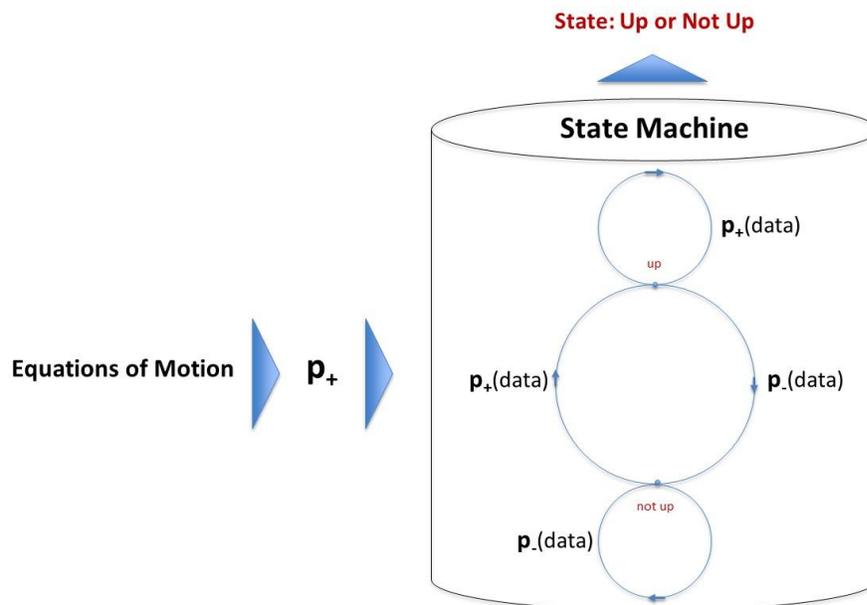


Figure 4. The equations of motion are used to generate an exact value for the probability based on the historical data (learning). The exact probability is then fed to the state machine to produce an output state. The output state is added to the historical data. Repeat.

In Figure 5 below, ScienceML produces an individual forecast for the same six months of ZIM closing prices as in previous sections. We plot the last 30 days and then the individual forecast to a horizon of 30 days (total of 60). The dotted vertical line is where the individual forecast begins. The state machine

with the probability computed by ScienceML appears to behave consistently as we move from past to future.

Precision Alpha will be field testing and optimizing ScienceML forecasting over the course of the next six months. ScienceML for forecasting will be generally available in Fall 2022.

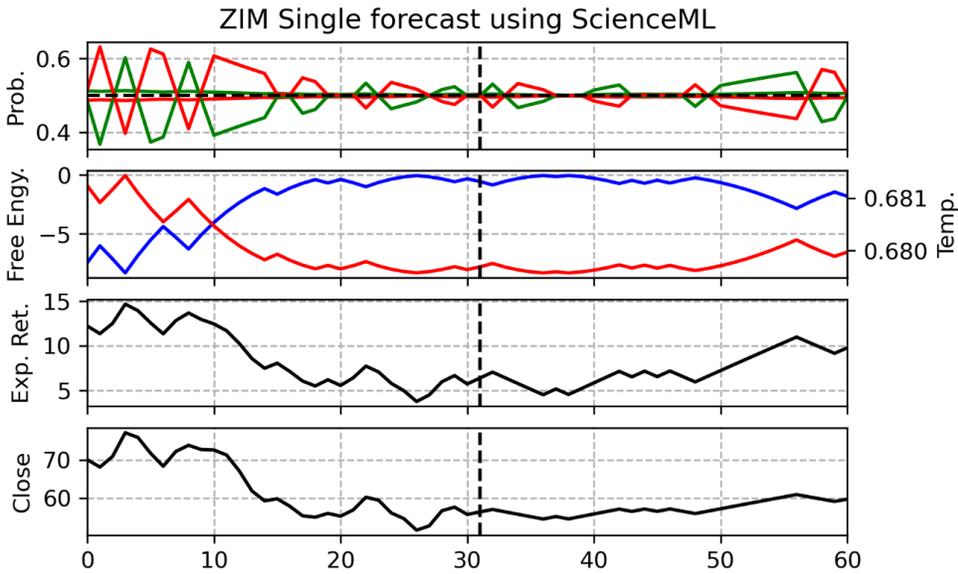


Figure 5. Individual forecast for both thermal and non-thermal probabilities, temperature and free energy. The Expected Return is based on the start date.

Operational Benefits

The consistency and transparency of counting on which business relies, is lost in conventional formulations of machine learning.

ScienceML, however, returns transparency to its original usage: it's just counting/math. Anyone else that does the math should and must get the same results.

With the transparency of math, other problems with conventional machine learning vanish for ScienceML. Notably, in ScienceML there are:

- No model parameters to estimate,
- No model biases to be concerned about,
- No iterative model development processes to design, implement and accelerate,

simply because there is no model! Table 2 summarizes the key benefits realized through ScienceML.

There is still bias to defend against, however, but the bias is entirely a “data bias”. Data biases are in the time-series data itself and depend on what and how the time-series data was acquired. In short, the tasks that remain for the data scientist and analyst using ScienceML resembles the world of the experimental scientist, rather than the theoretician (model builder).

Benefit	Conventional ML	ScienceML
Output	Estimates	Unique and exact answers, no model bias
Time	Iterative Model Development	No model creation. Real-time learning and processing. Machine-assisted human learning.
Ease of Use	Data Scientists required	Self-service for both data scientists and business analysts.
Cost	Expensive.	Web service. Inexpensive.

Table B. Decision Machine’s ScienceML vs. Conventional Machine Learning.

Conclusion

This paper presents an extraordinarily rich, model-free Machine Learning web service, ScienceML, that implements the Science of Counting.

ScienceML:

- Scientifically decomposes any time-series into a set of (mostly) intuitive scientific measurements: probability, energy, power, resistance, temperature, etc., (Monitoring)
- Exact Probability is used with a State Machine to produce rigorous individual and ensemble forecasts for scientific measurements. (Forecasting)

When ScienceML is applied to financial markets, which might be expected to be a challenging environment, impressive market structure is uncovered. There are several notable observations about financial markets:

- Financial markets are thermodynamically open, and exchange energy with the “environment”. The “environment” would include non-trading activities that affect price movement, like earnings announcements, Federal Reserve policy, the news cycle, and so on, all together yielding an ambient temperature, T_R . The ambient temperature is generally different than the financial market temperature, T .
- Financial markets (at least NYSE and NASD) are rich with dissipative structures that are self-regulating and far from equilibrium.
- Forecasting requires that the ambient temperature, T_R , be determined through user measurement. At thermal equilibrium $T = T_R$. Thermal equilibrium has distinctive behavior that can be identified in ScienceML output to measure the ambient temperature, T_R .
- Assets that display dissipative structures (repeatable behavior) should be easier to forecast.

The latest version of the ScienceML web service for monitoring is current generally available. ScienceML for forecasting will be generally available in Fall 2022.

Bibliography

- [1] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press (2003).
[2] G.E. Crooks, Phys. Rev. E 60, 2721-2726.
[3] D. Kondepudi, I. Prigogine, Modern Thermodynamics: From Heat Engines to Dissipative Structures, Wiley (2015).