STATS4TRADE and BigML

Unlock the Power of Machine-Learning for Investors

BigML Case Study

MACHINE LEARNING MADE BEAUTIFULLY SIMPLE

"We believe that managers of actively-traded equity funds are well aware of the coming "Uber-ization" of their industry; our commitment is to help these fund-managers better serve their clients with modern machine-learning approaches that yield better-than-index returns with less volatility and lower costs."

Jean-Marc Guillard Ph.D.,
Founder and CEO of STATS4TRADE

KEY ACHIEVEMENTS

- Better-than-index portfolio performance with minimized volatility
- Drastically faster iterative experimentation and time-to-market
- Lower operational costs

1. BUSINESS CONTEXT

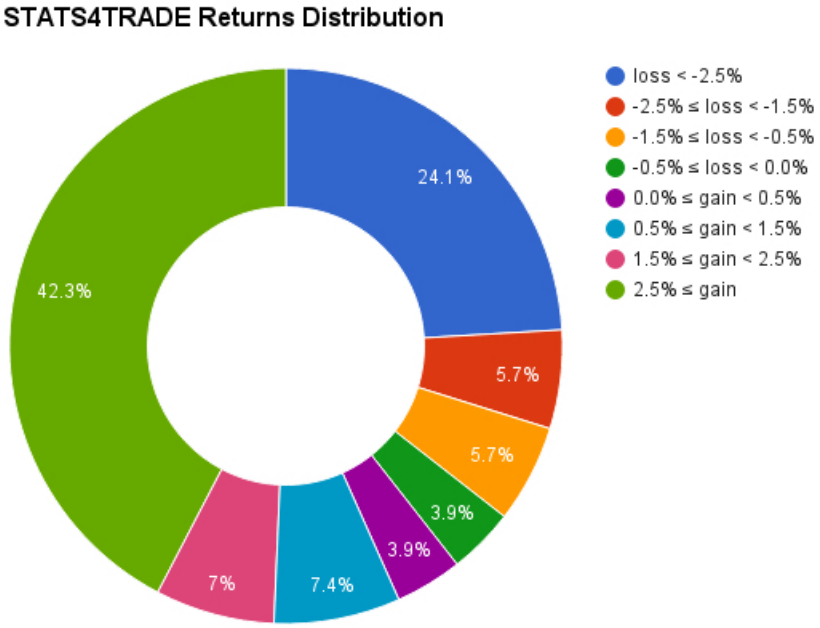
The financial services industry is faced with drastic change in the coming years and actively-managed equity funds are not immune. Investors are rightfully questioning high fees in the face of continued poor performance compared to passive funds with much lower fees. Similar to the disruptive changes now occurring in the transport industry, active-fund managers must contemplate an "Uber-ization" of their business model with software driving innovation to provide investors promised returns at lower cost.

STATS4TRADE and BigML are uniquely positioned to help active funds navigate this coming change. With the aid of machine-learning and cloud-computing technologies, we offer fund managers a new approach for selecting stocks and making buy/sell decisions – a software-driven approach that not only yields consistently better-than-index performance but also minimizes volatility and decreases operational costs while protecting capital.

2. SOLUTION

STATS4TRADE effectively combines the benefits of two fundamental albeit different investment strategies – namely the theoretical higher returns provided by active strategies with more frequent trades backed by in-depth analysis (e.g. active funds) combined with the low operational costs and diversification provided by traditional buy-and-hold strategies (e.g. passive index funds).

Of course none of this is magic. Instead we rely on consistently evaluating and increasing the statistical probability of making a favorable buy or sell decision for any given trade. As proof just consider the following chart that shows the distribution of individual returns per trade based on our approach between January 2014 and October 2016 for all stocks in the CAC 40. Clearly the relative frequency of positive returns is higher than negative returns: 60.3% positive versus 39.7% negative. This result is fundamental to the ability to yield better-than-benchmark returns with machine-learning technologies from BigML.



The key to our approach is the ability to identify subtle, complex statistical patterns in equity prices within a particular market like the CAC 40 or Dow Jones 30. From these patterns we can forecast the likelihood of certain events occurring – such as a price increase for a certain stock in the coming days or weeks. However markets are inherently complex with much random content. Therefore a machine-learning approach with a statistical model is employed that replicates market behavior in terms of statistical parameters over time.

3. IMPLEMENTATION

The process for creating a statistical model is shown in the schematic. At the heart of the process is a machine-learning algorithm from BigML that loads training data and then trains, evaluates and tunes a model to reveal statistically significant patterns.

The first step is "data wrangling". For STATS4TRADE this step involves collecting historical price-data over several decades for all stocks in a particular market like the CAC 40 or Dow Jones. The price-data are then "pre-conditioned" by converting the raw daily prices into more complex datasets. In turn these datasets are assigned a range of descriptive parameters intended for later manipulation by BigML's machine-learning tools.

After "data wrangling" the pre-conditioned data are loaded into BigML's environment. During the loading process BigML converts the data into a useful internal format with features such univariate statistics and histograms in its "Dashboard" graphical-user interface.

Next the model is trained to find relationships among statistical parameters in the underlying data over different timespans. The training process can be quite complex mathematically; however BigML's platform abstracts this complexity away from the end-user while yielding statistically valid results. Following this approach we experimented with many models using BigML's Dashboard.

Then the tedious yet necessary process of evaluating and tuning the model begins. For STATS4TRADE these two steps entailed iteratively testing the model results – in the form of forecast probabilities – against actual historical prices over different timespans for both accuracy and consistency. Because we want to minimize any error caused by "training-data in, training-data out", it was critical that we test against price-data not used to train the model. This ensures that the underlying statistical model is truly vetted with reasonable consistency across different timespans.

During the evalulation and tuning process we focus on model precision and variance metrics to ensure a balance between forecast accuracy and consistency. For example we test with different timespans and baskets of stocks while observing how prediction confidence-levels varied. Only by exhaustively evaulating prediction results for different configurations over time are we able to iteratively build robust models that consistently provide accurate and consistent forecast results.

Overall and when asked about his end-to-end implementation experience with the BigML platform, Jean-Marc Guillard reports the following.

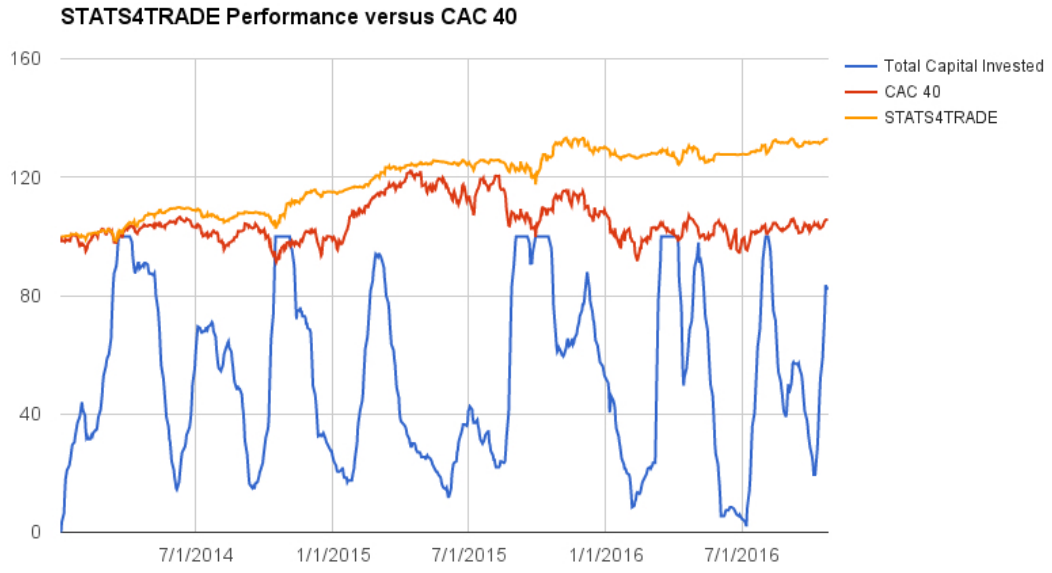
"STATS4TRADE uses a mix of the BigML Dashboard and APIs. The Dashboard iseasy to master and made rapid testing of different features and various ensemble configuration-parameters relatively pain-free. Once we determine the features and parameters that improve model performance, we implement the optimal model as part of our investment strategy by using the REST API. In our experience BigML is superior to other solutions that we have tested because of the ability to very quickly experiment with many different machine-learning workflows without spending much time programming. The built-in visualizations greatly help in interpreting each model as well as communicating learning within the team before we begin implementation phases in earnest. We have indeed saved several months of development time by starting first with BigML."

4. RESULTS

Consistently beating widely-accepted benchmarks like diversified stock indices is no small feat for an active fund-manager. However investors rightfully demand it – else why not just invest in a low-fee passive fund that simply tracks an index? Moreover most investors are risk-averse; hence active fund-managers must also minimize volatility relative to benchmark indices.

STATS4TRADE is no different than an active-fund manager. Therefore our machine-learning approaches must also yield higher returns with less volatility than benchmark indices. One ideal index for benchmarking active management is the challenging French CAC 40, which has demonstrated little sustained growth coupled with relatively high volatility since the 2008 financial crisis.

And how have we performed against the CAC 40? Well consider the following chart that shows the performance of STATS4TRADE's machine-learning approach for all stocks in the CAC 40 (yellow curve) versus the index itself (red curve) from 2014 through October 2016. The results are quite clear! Over the two-year window the machine-learning approach earns a return of approximately 31% with visibly more consistency (e.g. less volatility) while the index yields only 5% with significantly greater fluctuations.



One of the reasons that we outperform the CAC 40 is because we are not always fully invested in the market. The percent invested varies between 0-100% (blue curve) thus delivering a hedging effect. This limits investor exposure to potential downturns versus a traditional buy-and-hold strategy and is a direct consequence of the ability to statistically forecast price movements with machine-learning approaches.

Note that all of this was achieved automatically by machine-learning-enabled software without need for constant human intervention after models are created. Thus actively-managed funds can save significant operational costs like the salaries of expensive stock analysts. In turn these savings can passed on to investors in the form of lower fees approaching those of passively-managed funds.

5. FUTURE OPPORTUNITIES

STATS4TRADE is encouraged by the results that machine-learning offers and is planning to extend its offerings in cooperation with BigML. In the near-term additional equity markets beyond the initial CAC 40 and Dow Jones are planned as well as increasing the present forecast timespan from twenty days to twenty weeks. Longer term plans include investigating more complex trading strategies as well as currency, bond and even commodity markets.

In continuing its development efforts STATS4TRADE intends to deepen its partnership with BigML by utilizing new and more advanced features of BigML's platform. For example, WhizzML, BigML's Domain Specific Programming Language that helps automate Machine Learning workflows.

bigml®