Predicting the next market move with A.I.

How Punch built a series of connected models to trade the financial markets.

numin

punch case study

Contents

About Numin

NUMIN IS ONE OF THE WORLD'S MOST INNOVATIVE QUANTITATIVE TRADING FIRMS, APPLYING PREDICTIVE MACHINE LEARNING MODELS, RIGOROUS FEATURE ENGINEERING PRACTICES, INDEPENDENT QUALITY CONTROL METRICS, AND INTERCONNECTED SUBTASK SPECIALIZATION.

ADVANCED QUANTITATIVE RESEARCH

Numin's approach required setting up a machine learning research factory.

Punch helped develop independently measured and monitored quality control metrics for each subtask. We helped organize both teams quants to specialize in a particular subtask, and to become the best at that specific task, while having a holistic view of the entire process.

Numin's approach required setting up a machine learning research factory.

PUNCH SERVICES PROVIDED

Punch provided expertise in machine learning model development, data science feature engineering, data pipelines, and back-end systems.

Machine Learning, Artificial Intelligence Data Science, Architecture, Feature engineering, Hyperparameter Tuning, Quality Control Systems

2

Finding patterns in chaos

NUMIN'S TEAM FACED A PROBLEM – THEY WANTED TO SELF ORGANIZE INTO A RESEARCH FACTORY, AND NEEDED A SECOND TEAM TO FILL IN ANY GAPS AND PROVIDE A SECOND PAIR OF EYES AND EARS.

The challenges facing quant financial researchers is immense. The optionality creates a decisiveness dilemma. The only solution was to have the data guide the process. But this would require immense engineering work to achieve.

1. **Data sourcing.** Financial data can be enormous – there is no shortage of direct, indirect, and alternative data. The challenge for financial researchers is separating signals from noise.
2. **Feature engineering.** Normalization options for financial data can be precarious, losing correlation to the original dataset. New engineered features and original features are pruned so only the essential elements are fed to the model.
3. **Model engineering.** Supervised and unsupervised strategies compete for researchers' attention. Labeling can be spurious or exact. Overfitting can be a death sentence.
4. **Hyperparameter tuning.** Brute force tactics waste computing resources and time.

78,531

NUMBER OF TRADES ON THE S&P 500 INDEX ON WEDNESDAY, SEPTEMBER 30TH, 2009

Millions of trades are executed daily across financial exchanges. Gathering this data helped Punch refine models to create predictions.

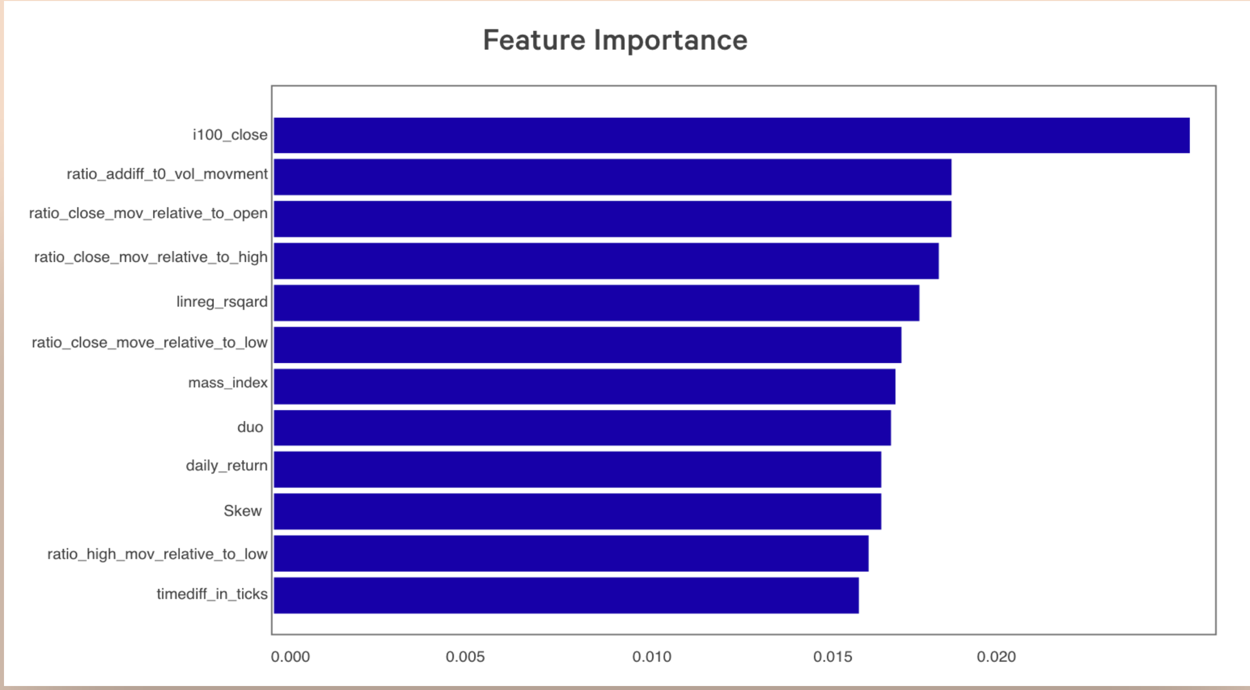
1

3

DATA SOURCING

SEE WHAT DATA IS USEFUL AND NOISE.

This factory is able to evaluate the efficacy of any new data and whether it is additive or subtractive to other data sources and should be admitted or omitted.

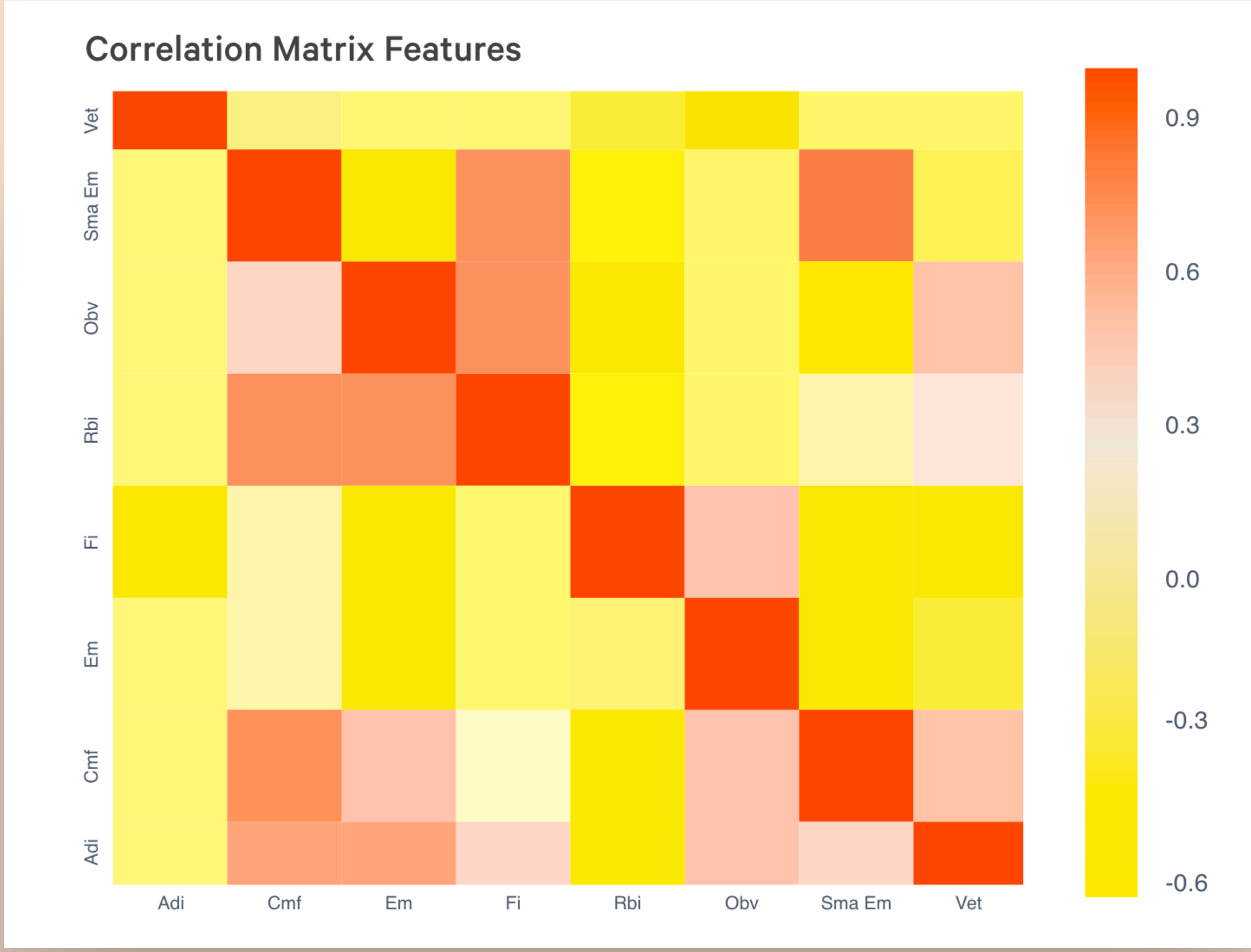


2

FEATURE ENIGINEERING

COMBINING SELECTED DATA WITH ENGINEERED FEATURES TO ACHIEVE A SHARPER EDGE.

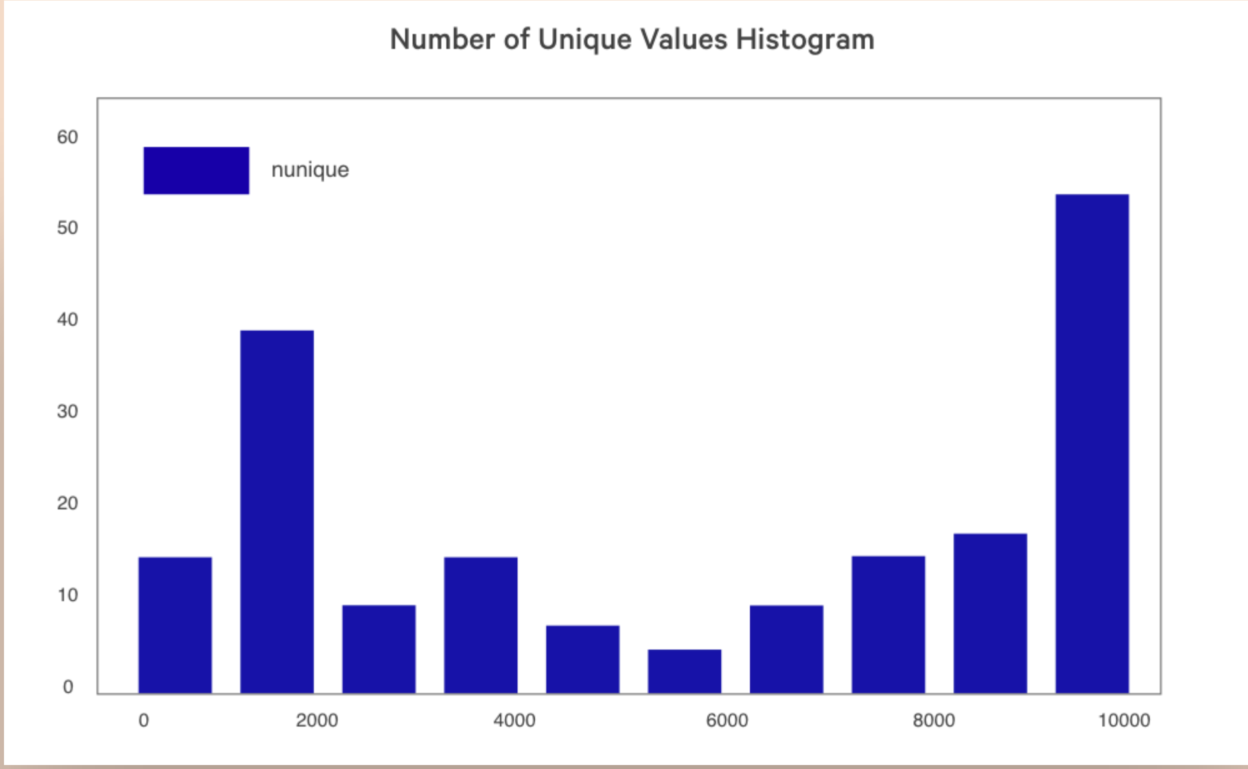
Predictive machine learning helps Numin identify and prioritize potential data features. By applying Granger causality and Pearson correlation testing, with shallow model and statistical analysis, we were able to build a data analysis factory.



IDENTIFYING UNIQUENESS IN THE DATA

RUNNING SAMPLE DATA THROUGH MULTIPLE FORMS OF ANALYSIS YIELDS NEW DISCOVERIES

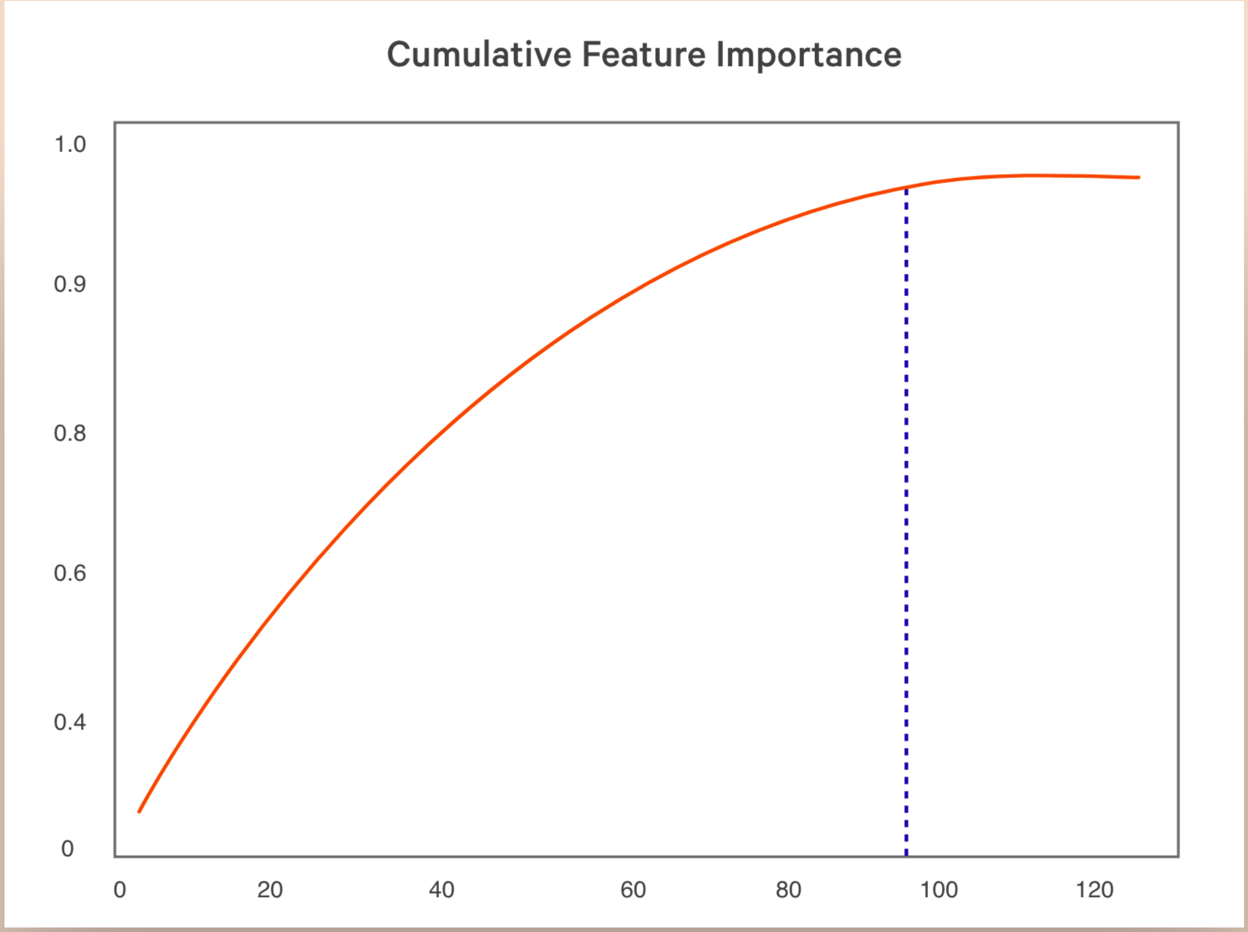
Testing synthesization of many data structures we discovered all machine learning models fail at synthesizing certain data types (such as ratios and ratios of differences). We corrected for synthesization weakness by engineering new features to help our models learn from hidden relationships within the data.



MEASURING FEATURE CONTRIBUTION

RUNNING GRADIENT BOOSTING TESTS ALLOW FOR SOME PRELIMINARY FEATURE TRIMMING

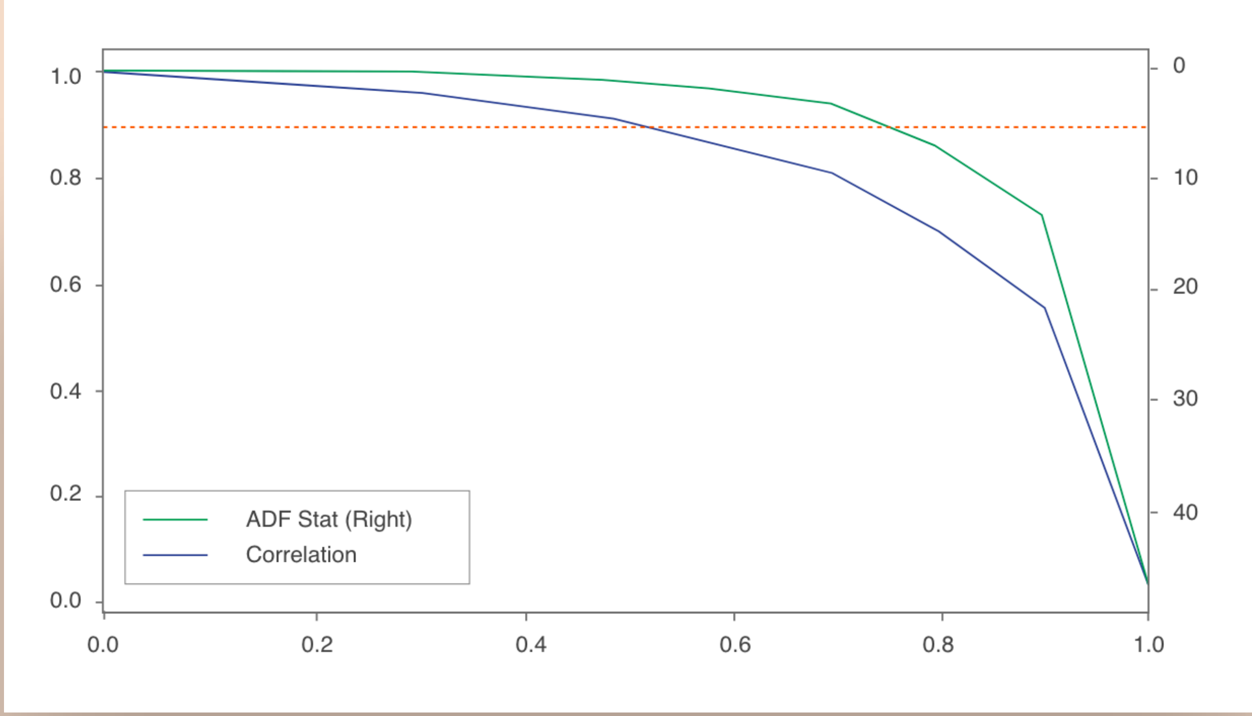
Feature importance was evaluated on an individual level and in aggregate. 70% of the features achieved 99% of the feature importance scoring.



PRESERVING MEMORY IN NORMALIZATION

PASSING THE AUTOMATED DICKEY-FULLER TEST WITH HIGH CORRELATION BETWEEN DATA SETS

We used partial differentiation to achieve data stationarity while preserving a 99.5% correlation to the original data set on a per-feature basis.

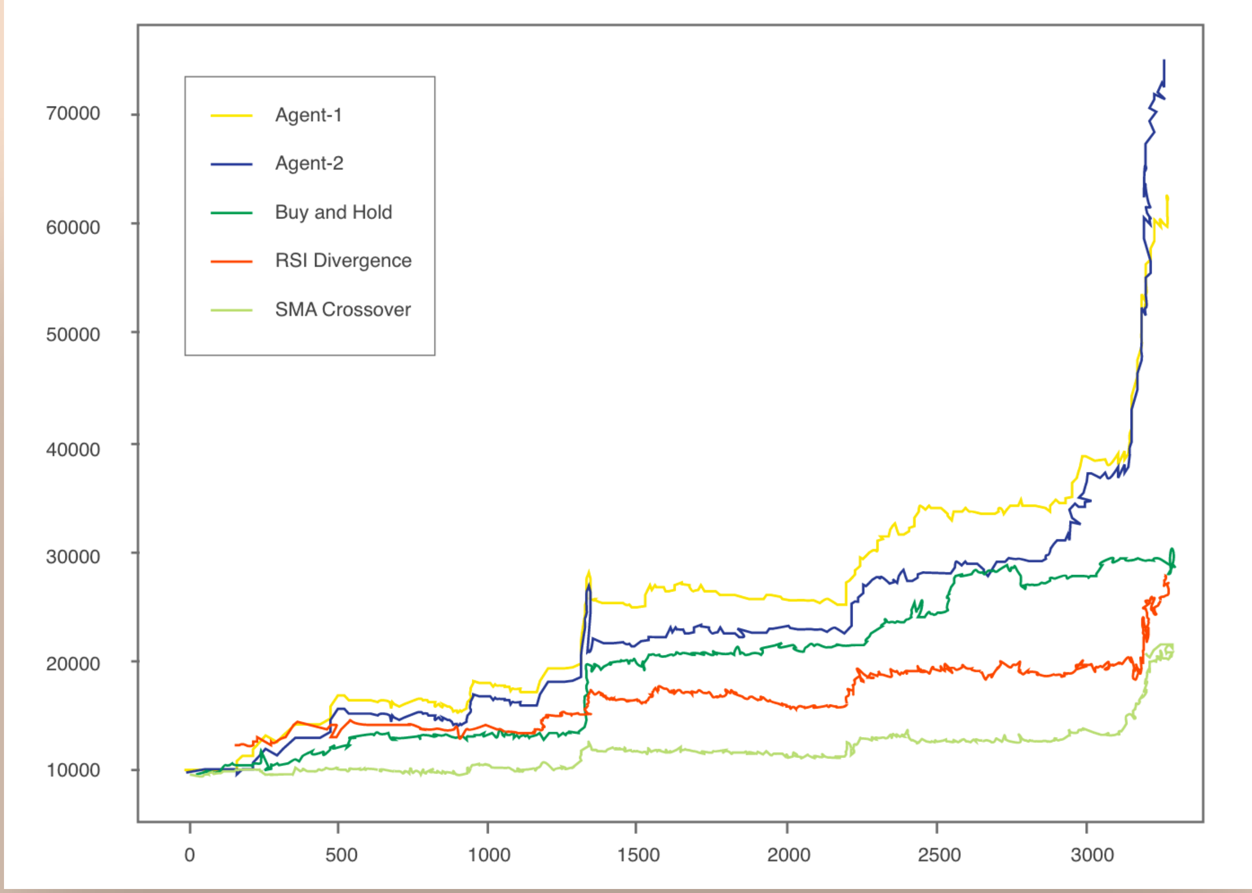


3

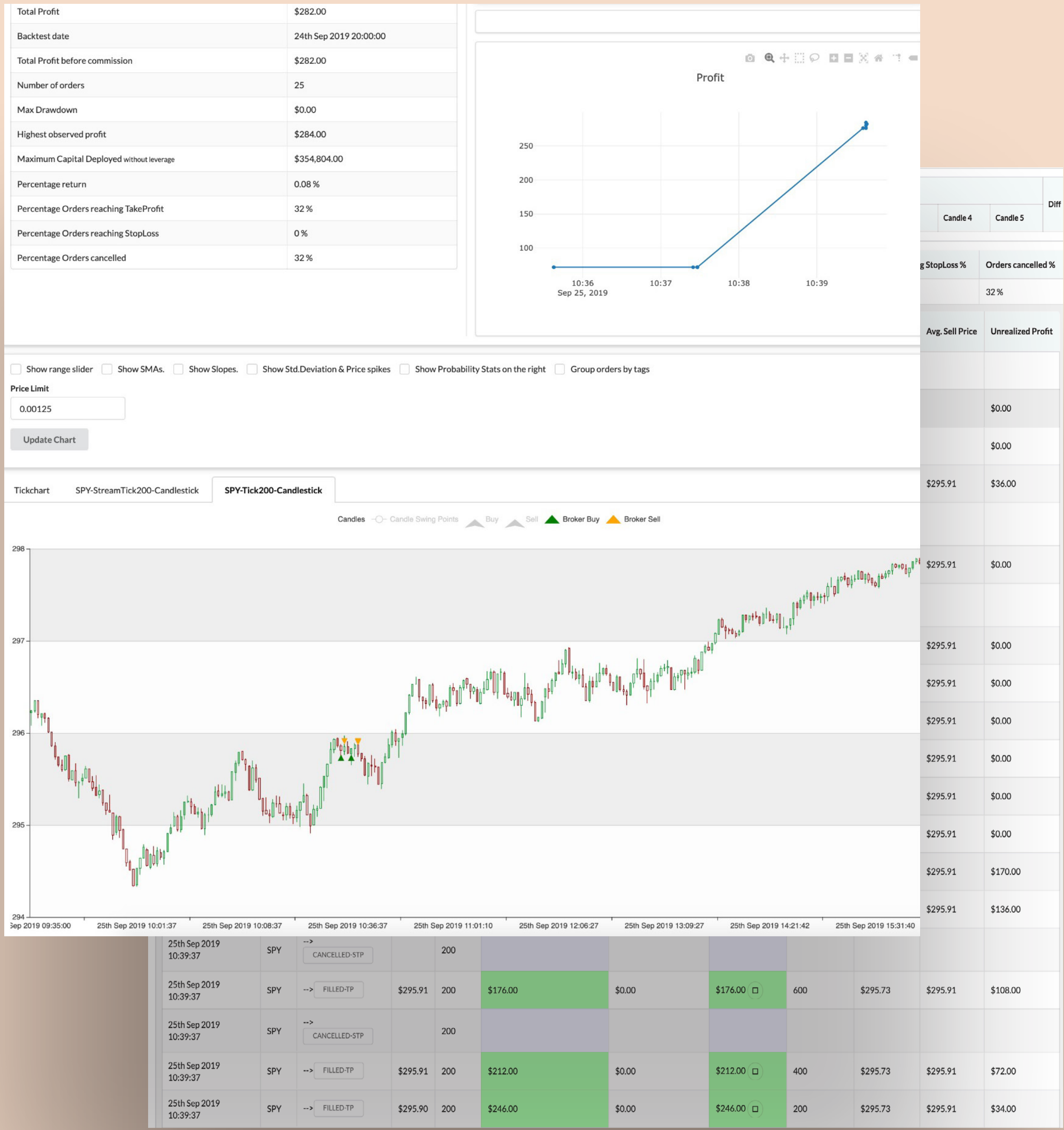
MODEL ENGINEERING

A SYSTEM THAT LEARNS TO TRADE

Benchmarks against our model were set using 3 popular strategies: 1. simple moving average crossover, 2. buy and hold, and 3. relative strength index (RSI) overbought/oversold conditions.



We turned the model's black box into a white box by building a trading visualization system to monitor the trades being placed in real-time. We added financial metrics to analyze the model's results such as the coefficient of variation.



The models began to learn to identify reversal areas and to avoid tight ranges.

"The compartment alization of the quants on both teams have ***owned their domain****, creating a* ***research factory*** *of* ***continuous improvement*** *that would have been difficult to achieve otherwise."*

ANUP

Director of Engineering
Numin

"We were able to increase ***F-1 Scores*** *substantially by focusing more energy on* ***feature inputs*** *through* ***subtask specialization*** *with Punch."*

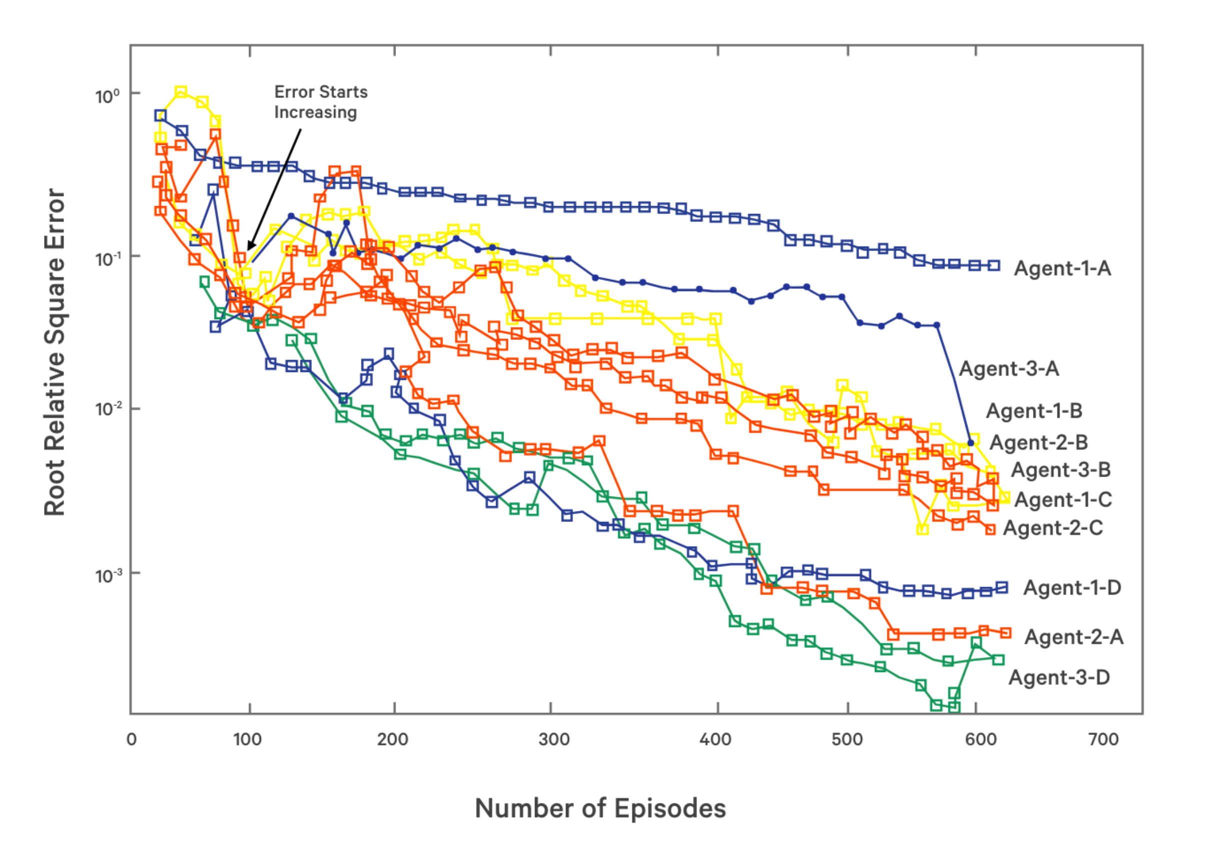
KENT

Senior Architect
Numin

4

HYPERPARAMETER TUNING

Increasing the learning data set resulted in slower training times. Deploying distributed asynchronous algorithm driven hyperparameter optimization allowed for parallel training.



Results

PUNCH PRODUCED A STREAMLINED ASSEMBLY LINE FOR QUANTITATIVE STRATEGY DEVELOPMENT.

AI PROJECT STATISTICS

	Items
Terabytes of data	**8**
Length of project	**2 years**
Tech stack	Machine Learning and Google Cloud
Teams involved	San Francisco & Lahore

Legal

punch