Vectorspace AI

Real-Time Datasets for Training in Artificial Intelligence (AI) & Machine Learning (ML)

vectorspace.ai

What Do We Do?

1. Vectorspace AI builds real-time on-demand datasets that are used to train the worlds AI & Machine Learning (ML) systems via tiered subscriptions
2. Customer subscriptions are converted to BTC/ETH then to our VXV utility token on the exchange
3. This VXV is placed in a wallet which doubles as a what we call a "VXV wallet-enabled API key"
4. Datasets are distributed using our VXV wallet-enabled API services
5. Datasets are tracked using our VXV Data Provenance Pipeline (DPP) hash

Reddit AMA on r/AskScience
https://www.reddit.com/r/askscience/comments/9k5i8u/askscience_ama_series_were_team_vectorspace_ai/

What Our Customers Are Saying

"Really great work - just read the details. Will you be writing an academic paper based on your findings? "

PJ Hampton, PhD Researcher, NLP

"Hello, I'm pretty curious and amazed by how powerful your summarization of articles is ! I'd like to try it out on a side project so that's why I'm sending you this e-mail :)"

N. Hardy

"Great work with the context summarization! May I request a copy of the code base? I won't use for commercial purposes, or anything like that. Just academic insight."

H. Peterson

"That's actually really cool. Now you've got me looking into these algorithms..."

bcolb

"You guys are working on some really incredible stuff. Can't wait to get my hands on some code especially after looking at those links now wow...this is more of what we need in the world, this will change everything :D Cool... you're based in SF would love to meetup and grab a coffee sometime and tell you what I'm working on."

N. Sherriff

"I do a lot of hobby or experimental projects using NLP, I'd love a chance to use your system. Is this something you think you may release as open source?"

T. Lelelu

"That's.... incredible. Wow."

chrismbarr

"Requesting API Access. Really intrigued by the amazing work your team has done around content summarization. Was wondering what is the process for applying for API access. Is there an open source implementation that we can contribute towards?"

R. Tewari

"I tried it with the 50 Shades of Grey text (btw, impressive text generation work) and I found the results to be just Ok. However, I then tried it with a new text related to an app description and found the result to be quite impressive, both the summarization and the keywords extraction. Well done!"

carlos_argueta

"I'm really impressed by how accurate your algorithm is, and I would like to try it out summarizing some news concerning the same topic (in different languages). My goal is to create an algorithm that can merge some texts speaking about the same subject, so that's why I will read your mathematical explanations on the API page. FYI I tried it on French news articles, and it seems to be working almost as good as english. (it may work better on latin-based languages right?)"

Anon

"Is there somewhere I can use this same software/algorithm to actually summarize other things? I am thoroughly impressed."

primary_action_items

"Amazing work! Is there a repository I can have access to? These are exciting times for aspiring amateur academics like myself. Public knowledge of this calibre is exactly what humanity needs to overcome the debt of our myriad lapses in good faith."

W. Sahatdjian

"I'm comparing this algorithm with the one i wrote for wingztv.com (shameless plug) i am really shocked by how much better these guys algo is. although mine was just kinda was hacked up. it just sees sentences with more words in common and ranks then by it, and then i just grab the top 5. this algo looks a lot more fancy O.o probably why it's a lot better lol."

moridin007

Current Partners & Collaborators

Berkeley Lab

Ernest Orlando Lawrence
Berkeley National Laboratory

U.S. DEPARTMENT OF
ENERGY

Berkeley
UNIVERSITY OF CALIFORNIA

Market Opportunity

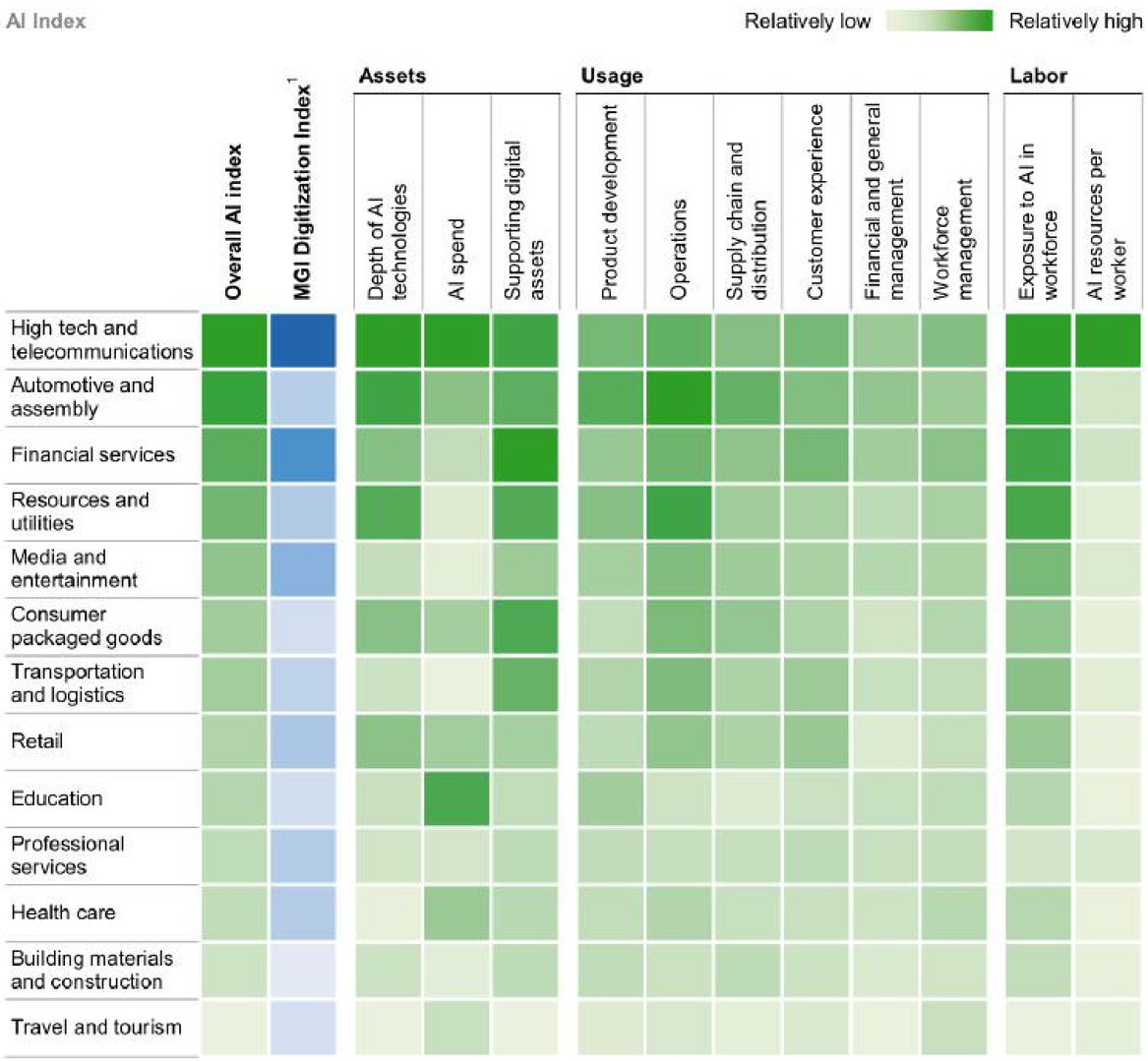
"Ten million dollars for a data set might seem like a lot, but if you can spend $10 million to make a hundred, $10 million suddenly doesn't seem like a lot of money anymore."

Kevin McPartland, researcher at Greenwich Associates
Citation

"Some investment firms are hiring individuals in the emerging role of Head of Data... It may possibly be the next frontier for funds looking for an investment edge" where the new norm is a "continuous information arbitrage."

CME Group
Citation

AI adoption is occurring faster in more digitized sectors and across the value chain



1 The MGI Digitization Index is GDP weighted average of Europe and United States. See Appendix B for full list of metrics and explanation of methodology.

SOURCE: McKinsey Global Institute AI adoption and use survey; *Digital Europe: Pushing the frontier, capturing the benefits*, McKinsey Global Institute, June 2016; *Digital America: A tale of the haves and have-mores*, McKinsey Global Institute, December 2015; McKinsey Global Institute analysis

Technology & Commercially Released Products

"Coming up with features is difficult, time-consuming, requires expert knowledge. 'Applied machine learning' is basically feature engineering."

Andrew Ng, Google Brain Deep Learning founder, former Baidu AI Chief
Citation

"In machine learning and deep learning we can't do anything without data. So the people that create datasets for us to train our models are the (often under-appreciated) heroes."

fast.ai
Citation

"In the future AI will be diffused into every aspect of the economy."

Nils J. Nilsson, Founding researcher, Artificial Intelligence & Computer Science, Stanford University
Citation

Patents, Awards & Papers

Winner R&D100 Award - Lawrence Berkeley National Laboratory

System and method for generating a relationship network - K Franks, CA Myers, RM Podowski - US Patent 7,987,191, 2011

Inter-term relevance analysis for large libraries

Matching and recommending relevant videos and media to individual search engine results

Media discovery and playlist generation

System and method for summarizing search results

Discovering and scoring relationships extracted from human generated lists

Statistical modeling of biomedical corpora: mining the Caenorhabditis Genetic Center Bibliography for genes related to life span - Blei DM, Franks K, Jordan MI, Mian IS.

Vectorspace AI (VXV)

Science & Data Engineering

- VXV Wallet-Enabled API
- NLP/ML/AI & Analytics
  VXV Wallet-Enabled API
- Streaming Real-Time On-Demand Datasets
  (Triangulated Sentiment/Alternative & Proprietary)
- Feature Vector Supercolumns
  (automated feature engineering)
- Data Engineering, Algorithms & Analytics (Advanced NLP/ML/AI)

Business Units

- Smart Baskets
- VXV Life Sciences Fund
- Alternative Data Services
- QuantBot
- On-Demand Datasets
  (Triangulated Sentiment/Alternative)

Customers

Revenue, Profit & Funding:
(Cryptocurrency, Stocks, Assets)

- Subscriptions
- Licensing
- 20% Net
- Revenue Sharing
- Transaction Fees

For:

- AI & Data Operations
- Exchanges
- Trading Desks
- Funds
- Data Vendors
- Partners

VXV Wallet-Enabled API Key

Transaction Network (General)

Science & Data Engineering

VXV Wallet-Enabled API

VXV Wallet Validator

Smart Basket ETF API

QuantBot API

NLP API

On-Demand Datasets API

Business Units

Required VXV per Wallet-Enabled API Key at any given time

42.9 VXV
0x32Be343B94f860124dC4fEe278FDCBD38C102D88

153.1 VXV
0x32Be343B94f860124dC4fEe278FDCBD38C102D88

153.1 VXV
0x32Be343B94f860124dC4fEe278FDCBD38C102D88

153.1 VXV
0x32Be343B94f860124dC4fEe278FDCBD38C102D88

Customers

Financial Product Developers (3)

Exchanges (2)

Current Partners (3)

Data Vendors, ML/AI & Funds (3)

VXV API Key Pricing Algorithm

JSON API Endpoint: GET https://vectorspace.ai/recommended/datasets/api/trending_smart_baskets
Params: vxv_token_addr=0xC2A568489BF6AAC5907fa69f8FD4A9c04323081D

Data Engineering, Algorithms & Analytics (Advanced NLP/ML/AI)

VXV Wallet-Enabled API Keys

Subscription-to-Exchange On-Demand Order Execution

Science & Data Engineering

VXV Smart Contract
(Subscription Handler)

- BTC
- ETH
- USD
- USD to BTC/ETH
- VXV

Python Trade Execution Module

VXV Wallet-Enabled API
Key Generator & Populator

Customer Activation Module
(Tier 1 Services)

VXV API Key
Pricing Algorithm

VXV Wallet-Enabled API Key: 1.000 VXV / 0xC2A568489BF6AAC5907fa69f8FD4A9c04323081D
JSON API Endpoint: GET https://vectorspace.ai/recommended/datasets/api/trending_smart_baskets
Params: vxv_token_addr=0xC2A568489BF6AAC5907fa69f8FD4A9c04323081D

Data Engineering, Algorithms & Analytics (Advanced NLP/ML/AI)

Business Units

Subscriptions

BTC	ETH	USD	VXV

Exchange API

Automated Buy
Order Execution:

VXV

Customers

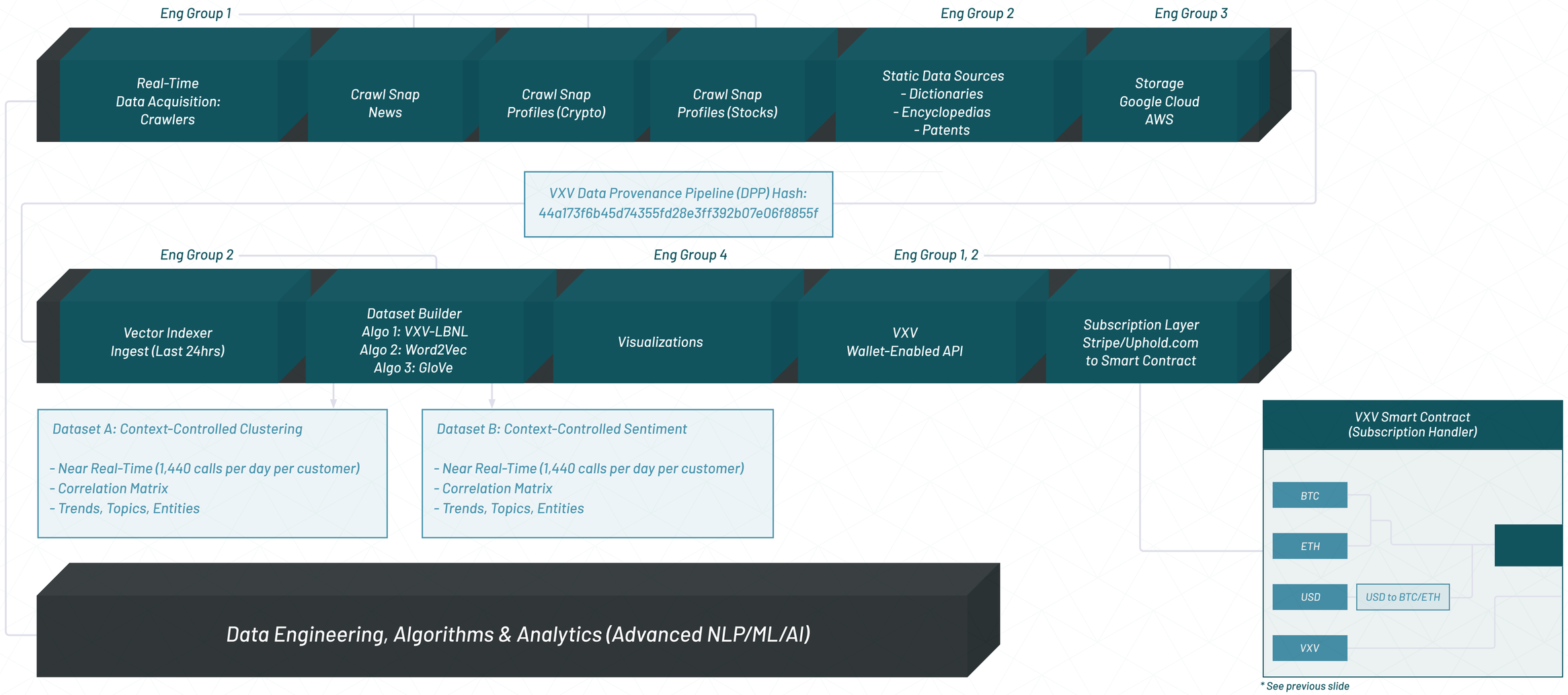
Financial Product Developers
All Industry Sectors: Data/ML/AI

Exchange

Active VXV Customer Pool
(API Calls per Day)

VXV Data Engineering Pipeline

Revenue Architecture Diagram



* See previous slide

VXV Streaming Real-Time On-Demand Datasets

With VXV Wallet-Enabled API Subscriptions

Customers

Streaming Real-Time On-Demand Dataset
(100,000 Datasets per Day)

Subscriptions

BTC	ETH	USD	VXV

Financial Product Developers
All Industry Sectors: Data/ML/AI

VXV Data Provenance Pipeline (DPP) Hash:
44a173f6b45d74355fd28e3ff392b07e06f8855f

Exchange API

Automated Buy
Order Execution:

VXV

Exchange

Static Datasources

- Real-Time Search Trends: Google, Bing, FB, Twitter
- Emerging Technologies
- Bitcoin, Price Relationships
- Chemicals
- Pharmaceuticals
- Context-Controlled Sentiment
- Top Gainers, Top Losers
- Metals
- Foods
- Human Genes
- Real Estate
- Commodities
- Global Geography, Cities
- Botanicals, Phytochemicals, Micronutrients

Dynamic Datasources

Topic
- Data Publishers
- Data Vendors
- Public
- Proprietory
- Ontologies

Real-Time Datasources

Topic
- Reports
- Filings
- News
- Social

Data Engineering, Algorithms & Analytics (Advanced NLP/ML/AI)

Investment & Growth Strategy

Vectorspace AI is raising growth capital in $50k tranches up to $150k in exchange for equity. Use of funds will go towards:

Support for our current customers & customer discovery process

Fund will be used to convert pilot programs with current customers to full-fledged revenue-generating partnerships

Augmenting our Silicon Valley based team of engineers & customer support staff

Build sales & marketing organization to capture business globally

Legal opinions

On-Demand Price Tiers

	Tier 1 – Free Limited	Tier 2 – $0.99 per On-Demand Update	Tier 3 – $1,950.00/Mo + $0.99 per ODU
	Free	**Data Vendor**	**Institutional**
Updates	1 Free On-Demand Update	100 Free On-Demand Updates	10,000 Free On-Demand Updates
Equity Types	NYSE stocks	Bitcoin & Cryptocurrencies, NYSE, Nasdaq, OTC	Any
Data Streams	Featured only	Any	Any
Custom Features		Yes	Yes
Support		Yes	Yes

Revenue Model

	Year 1		Year 2		Year 3	
	Nasdaq/NYSE	Cryptocurrency	Nasdaq/NYSE	Cryptocurrency	Nasdaq/NYSE	Cryptocurrency
Customers	100	75	220	260	380	500
Trading hours per day	6	24	6	24	6	24
Trading days per year (avg.)	252	365	252	365	252	365
On-Demand Updates per minute	9,072,000	39,420,000	19,958,400	136,656,000	34,473,600	262,800,000
Number of Custom Data Sources	5	5	5	5	5	5
Cost per On-Demand Update	$0.99	$0.99	$0.99	$0.99	$0.99	$0.99
Revenue	$44,906,400	$195,129,000	$98,794,080	$676,447,200	$170,644,320	$1,300,860,000
Total Revenue	$240,035,400		$775,241,280.00		$1,471,504,320.00	

Vectorspace CYM Short/Long Baskets

The Approach

It is well-established that companies participate in an economic ecosystem that has parallels to the natural world: sympathetic, symbiotic, and parasitic relationships can all be found among and between corporations. The Efficient Market Hypothesis suggests that all of these potential relationships between companies are known by the entire marketplace at the same time and, thus, they are already factored into company stock prices.

In reality, few investors have the proper tools to uncover these hidden corporate relationships and accurately project the directional knock-on impact of a change in one company in such a relationship.

The CYM System utilizes cutting-edge artificial intelligence technology based on patented Vectorspace techniques to identify and understand hidden relationships between companies to enable investors to get ahead of the broader impacts of breaking news. Strategies can be tailored based on customer needs.

The Story & Team

The CYM Team is lead by Kasian Franks, who has been working with artificial intelligence for more than 15 years. While at Lawrence Berkeley National Laboratory, he led a team of researchers in developing NLP/NLU algorithms to extract hidden connections between human genes involved in extending human lifespan under the direction of Mina Bissell, director of the Life Sciences division. NLP/NLU based pattern recognition and matching technologies were necessary to interpret this genetic data and map it to medical health data and protein functionality. The team then developed the first NLP/NLU system to uncover hidden relationships between proteins, disease states, and biological functions in the area of space biosciences for the purpose of LET radiation repair and protection during extended human space travel.

The heart of Kasian's NLP/NLU technology was a system designed to understand and read scientific abstracts, medical journals and updating news while uncovering hidden relationships among the biological components studied. In 2008, the team adapted this technology to the investment world and created CYM.

The Technology

The CYM System draws in a massive amount of triangulated and alternative data. The System then uses proprietary, patented algorithms and software to detect hidden relationships within a given context.

Based on a global event occurring, the CYM system then automatically generates short baskets of equities by leveraging context controlled NLP/NLU. The result is an ability to spot information arbitrage anomalies based on global events.

System Testing

The CYM System has been tested by outside auditors. Historical performance is based on the analysis of human language surrounding public companies located in public and private data pools.

Vectorspace CYM Short/Long Basket Returns

S&P 500 Stocks (ETFs not included)

	Hidden Relationship Returns	S&P Returns	Spread
2008	3.69%	-36.81%	40.50%
2009	46.84%	26.37%	20.47%
2010	28.29%	15.06%	13.23%
2011	15.29%	1.89%	13.40%
2012	22.08%	15.99%	6.09%
2013	27.24%	32.31%	-5.07%
2014	14.66%	13.46%	1.20%
2015-04/16	6.79%	-2.93%	9.72%

CYM Data 2008-2015

cym: 2.1.1.1 Date: 20150416

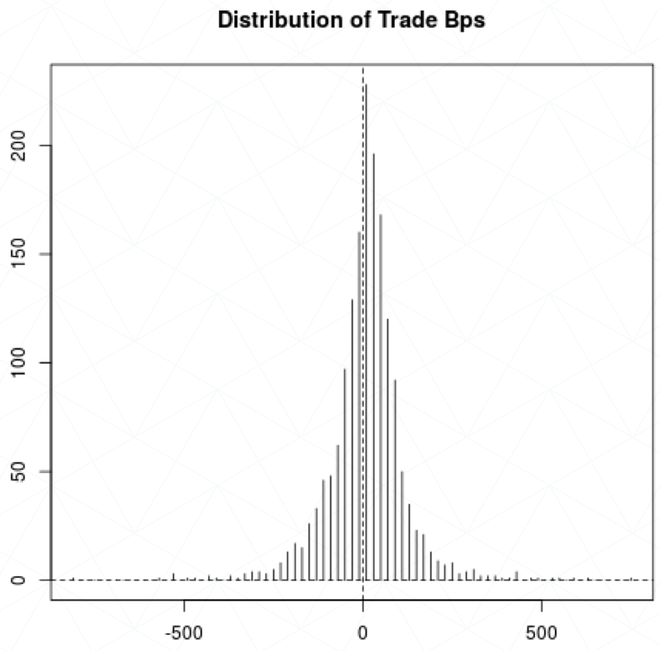
Bps	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	52645	120106	43.83%	3.87	0.84	-210.97	-3.36	-1.72	-0.45	0.08	0.14	0.67	2.09	3.82	186.92
Short:	0	0	0.00%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Position:	52645	120106	43.83%	3.87	0.84	-210.97	-3.36	-1.72	-0.45	0.08	0.14	0.67	2.09	3.82	186.92
Day:	685	1687	40.60%	110.51	6.48	-809.94	-160.75	-107.23	-35.03	14.32	9.77	58.59	111.01	164.05	742.97
Tdays:	685	1687	40.60%	110.51	6.48	-809.94	-160.75	-107.23	-35.03	14.32	9.77	58.59	111.01	164.05	742.97
Week:	135	381	35.43%	218.56	-0.71	-1107.76	-292.65	-175.62	-39.67	42.25	43.28	146.89	245.14	325.37	1165.27
Month:	25	88	28.41%	397.92	155.48	-1192.51	-365.19	-277.43	-31.32	166.34	187.36	384.29	592.96	803.76	1739.93
Quarter:	5	30	16.67%	627.32	155.48	-493.67	-481.30	-210.73	157.51	521.03	549.60	775.25	1435.90	1872.24	2442.76
Year:	0	8	0.00%	1291.19	679.26	369.60	369.60	369.60	1465.77	1868.40	2060.99	2828.98	4683.93	4683.93	4683.93

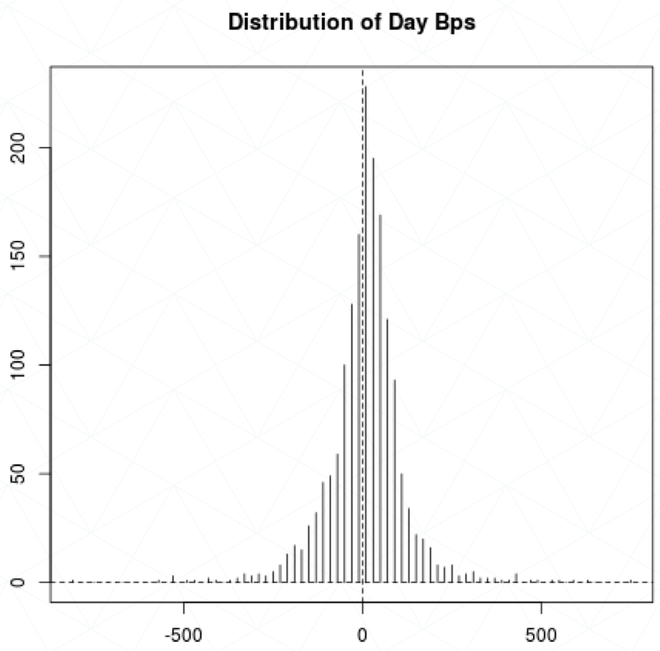
Hold-time	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	52645	120106	43.83%	8	380	351	355	359	368	377	373	380	380	380	380
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	52645	120106	43.83%	8	380	351	355	359	368	377	373	380	380	380	380

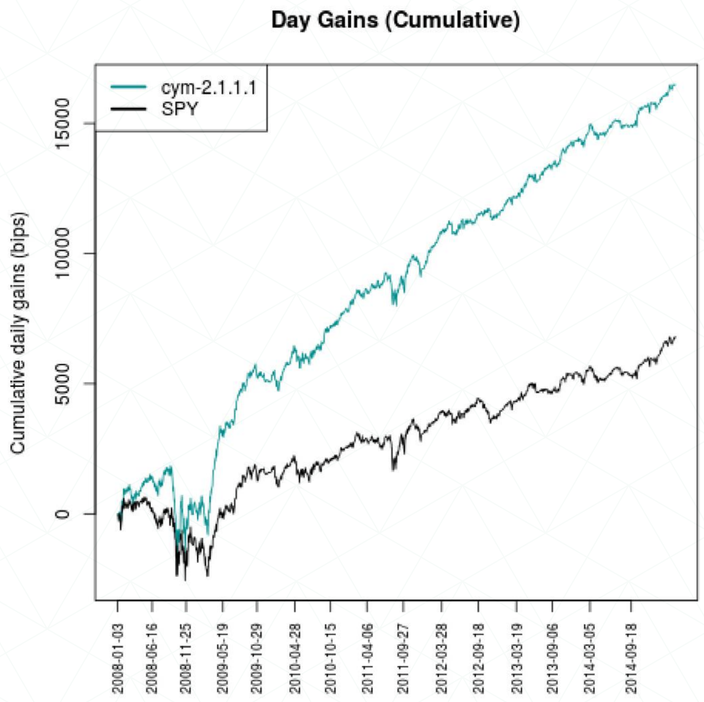
Trade eff	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	52645	120106	43.83%	33.38	33.65	-0.00	2.50	5.86	18.52	47.06	49.61	80.33	100.00	100.00	100.00
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	52645	120106	43.83%	33.38	33.65	-0.00	2.50	5.86	18.52	47.06	49.61	80.33	100.00	100.00	100.00
Day:	9767	24080	40.56%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Total Positions: 120106 Days: 1687 Traded Days: 1687 Positions/Day: 71.20 Positions/TradedDay: 71.20
Gain: 16487.92 Annualized: 2462.93 MaxDD: 28.24 MaxDDtime: 137 CurrDD: 0.03 CurrDDtime: 8
MaxDDBps: 3346.91 CurrDDBps: 6.99 MaxCnegetiveDays: 7 AvgCnegativeDays: 2 Negativedayscount: 161
Wbpp: 1.49 Lbpp: -1.58 Bpp: 0.14 BppL:0.31
Wbpd: 70.10 Lbpd: -78.47 Bpd: 9.77 BpdL:24.07 Wbps/Lbps: -1.3067
RR: 5.84 Sharp: 1.40 Beta: 0.8610 Gini: 0.0672 Skew: -0.1047 Kurt: 7.6224
AR3: 0.24 AR4: 0.32 AR7: 0.51 AR8: 0.59
Day- MarketGain: 6799.45 ExcessGain: 9688.47 AvgExcessGain: 5.74 MedExcessGain: 5.34 AccExcessGain: 55.78%
TDays- MarketGain: 6799.45 ExcessGain: 9688.47 AvgExcessGain: 5.74 MedExcessGain: 5.34 AccExcessGain: 55.78%
%Pos: 54.66% areaVar5: -17.80% areaVar10: -26.61% areaPor10: 27.55% areaPor5: 18.57%
Acc: 56.17% PrecLong: 56.17% PrecShort: 0.00% RecUp: 100.00% RecDown: 0.00%
F-Up: 71.93 F-Down: 0.00

CYM Data 2008-2015 (cont.)







CYM Data 2010-2015

cym: 2.1.1.1 Date: 20150416

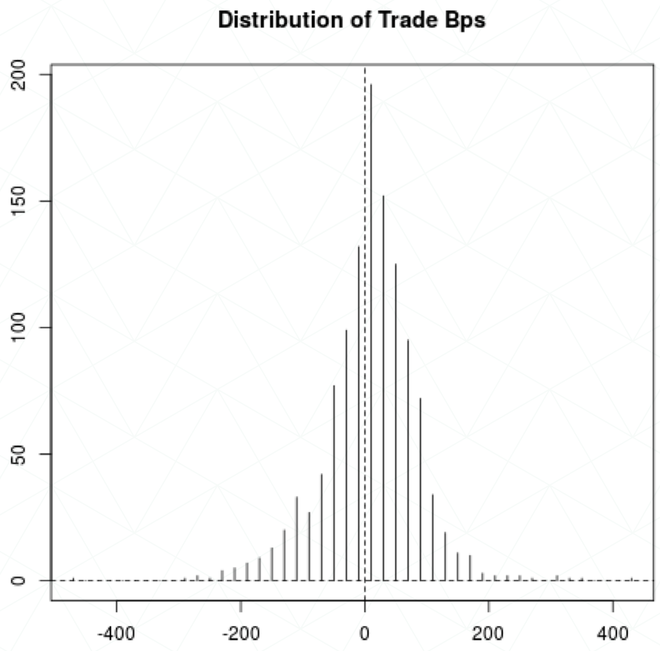
Bps	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	21474	50971	42.13%	4.54	0.84	-142.09	-4.68	-2.41	-0.50	0.09	0.23	0.91	3.05	5.37	116.60
Short:	0	0	0.00%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Position:	21474	50971	42.13%	4.54	0.84	-142.09	-4.68	-2.41	-0.50	0.09	0.23	0.91	3.05	5.37	116.60
Day:	472	1202	39.27%	76.06	6.48	-462.77	-120.38	-81.28	-28.50	13.14	9.51	52.40	90.68	114.16	432.11
Tdays:	472	1202	39.27%	76.06	6.48	-462.77	-120.38	-81.28	-28.50	13.14	9.51	52.40	90.68	114.16	432.11
Week:	88	276	31.88%	144.18	-0.71	-697.63	-199.25	-133.85	-26.47	45.86	41.43	133.29	208.58	263.77	447.53
Month:	16	64	25.00%	275.32	155.48	-466.62	-277.43	-161.91	2.99	166.34	178.66	354.18	480.96	625.00	1059.48
Quarter:	2	22	9.09%	421.14	155.48	-481.30	-145.53	93.52	251.35	521.03	519.75	775.25	981.22	1164.99	1435.90
Year:	0	6	0.00%	758.51	679.26	679.26	679.26	679.26	1465.77	1868.40	1905.73	2723.58	2828.98	2828.98	2828.98

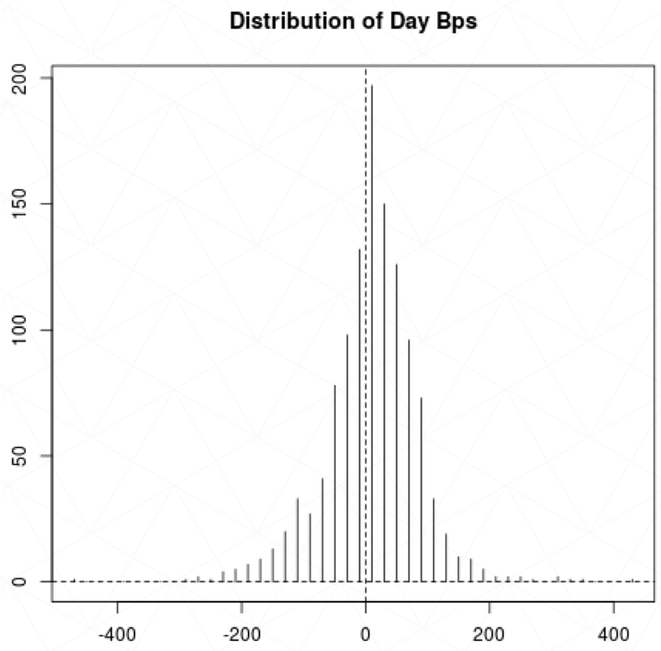
Hold-time	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	21474	50971	42.13%	8	380	351	354	358	368	378	373	380	380	380	380
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	21474	50971	42.13%	8	380	351	354	358	368	378	373	380	380	380	380

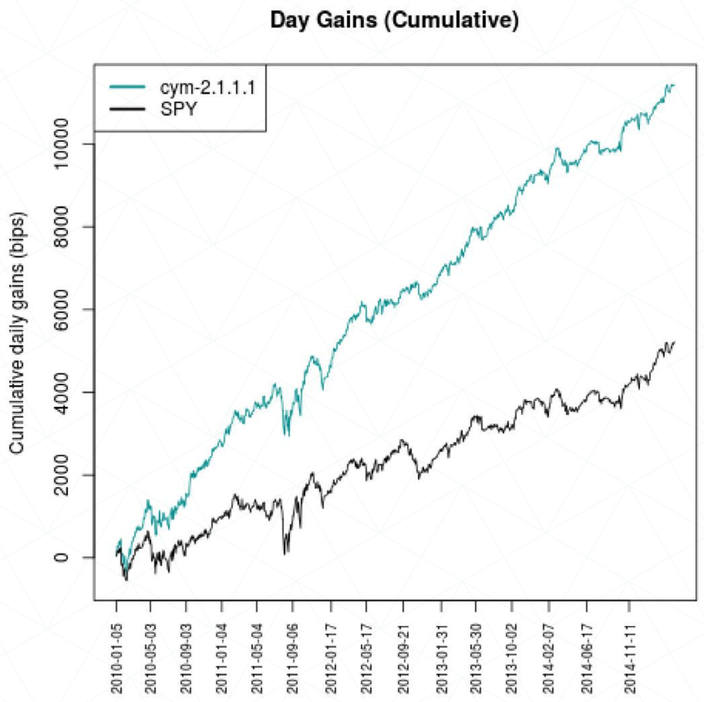
Trade eff	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	21474	50971	42.13%	32.21	33.65	-0.00	2.90	6.35	18.27	45.30	48.00	76.67	96.40	100.00	100.00
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	21474	50971	42.13%	32.21	33.65	-0.00	2.90	6.35	18.27	45.30	48.00	76.67	96.40	100.00	100.00
Day:	3860	10223	37.76%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Total Positions: 50971 Days: 1202 Traded Days: 1202 Positions/Day: 42.41 Positions/TradedDay: 42.41
Gain: 11434.39 Annualized: 2397.23 MaxDD: 8.98 MaxDDtime: 74 CurrDD: 0.03 CurrDDtime: 8
MaxDDBps: 1276.55 CurrDDBps: 6.99 MaxCnegetiveDays: 7 AvgCnegativeDays: 2 Negativedayscount: 111
Wbpp: 1.92 Lbpp: -2.08 Bpp: 0.22 BppL:0.53
Wbpd: 53.64 Lbpd: -58.73 Bpd: 9.51 BpdL:24.23 Wbps/Lbps: -1.4125
RR: 12.73 Sharp: 1.98 Beta: 0.8811 Gini: 0.0852 Skew: -0.2539 Kurt: 3.6339
AR3: 0.28 AR4: 0.16 AR7: 0.52 AR8: 0.41
Day- MarketGain: 5218.70 ExcessGain: 6215.69 AvgExcessGain: 5.17 MedExcessGain: 5.27 AccExcessGain: 56.66%
TDays- MarketGain: 5218.70 ExcessGain: 6215.69 AvgExcessGain: 5.17 MedExcessGain: 5.27 AccExcessGain: 56.66%
%Pos: 55.89% areaVar5: -15.96% areaVar10: -25.03% areaPor10: 24.62% areaPor5: 15.58%
Acc: 57.87% PrecLong: 57.87% PrecShort: 0.00% RecUp: 100.00% RecDown: 0.00%
F-Up: 73.31 F-Down: 0.00

CYM Data 2010-2015 (cont.)







CYM Data 2012-2015

cym: 2.1.1.1 Date: 20150416

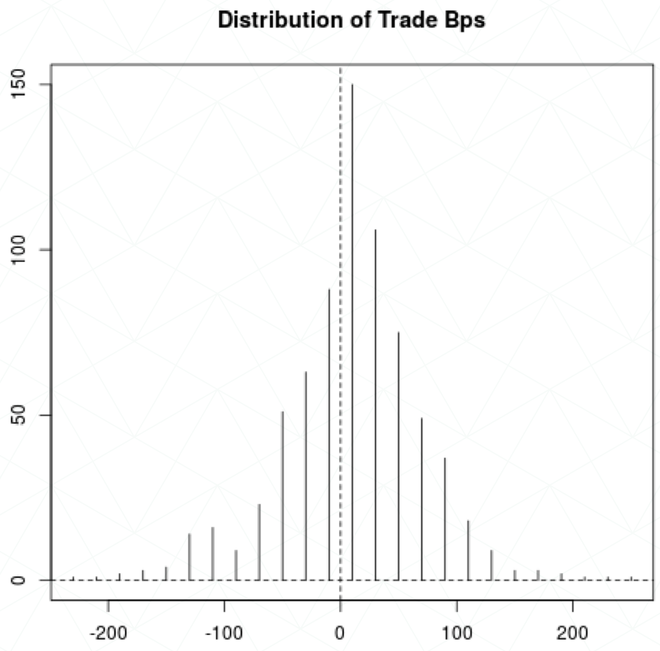
Bps	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	8779	20416	43.00%	5.31	0.84	-142.09	-5.87	-3.32	-0.91	0.20	0.36	1.56	4.08	6.74	113.88
Short:	0	0	0.00%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Position:	8779	20416	43.00%	5.31	0.84	-142.09	-5.87	-3.32	-0.91	0.20	0.36	1.56	4.08	6.74	113.88
Day:	274	730	37.53%	60.30	6.48	-236.78	-105.60	-58.97	-21.12	10.96	9.69	43.93	81.66	101.67	245.59
Tdays:	274	730	37.53%	60.30	6.48	-236.78	-105.60	-58.97	-21.12	10.96	9.69	43.93	81.66	101.67	245.59
Week:	52	172	30.23%	114.57	-0.71	-394.36	-171.53	-98.25	-17.07	39.42	41.14	118.36	173.71	225.28	351.51
Month:	10	40	25.00%	234.05	155.48	-365.19	-175.76	-157.86	2.99	161.38	176.91	353.01	501.99	625.00	636.29
Quarter:	1	14	7.14%	390.39	155.48	-145.53	-145.53	93.52	251.35	477.56	505.45	715.78	981.22	1435.90	1435.90
Year:	0	4	0.00%	771.87	679.26	679.26	679.26	679.26	1465.77	1836.72	1769.07	2723.58	2723.58	2723.58	2723.58

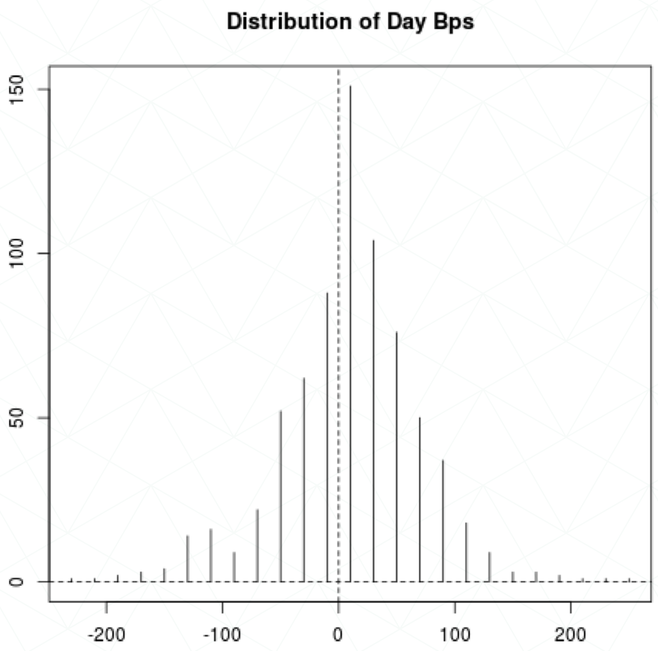
Hold-time	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	8779	20416	43.00%	8	380	351	354	358	366	377	372	380	380	380	380
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	8779	20416	43.00%	8	380	351	354	358	366	377	372	380	380	380	380

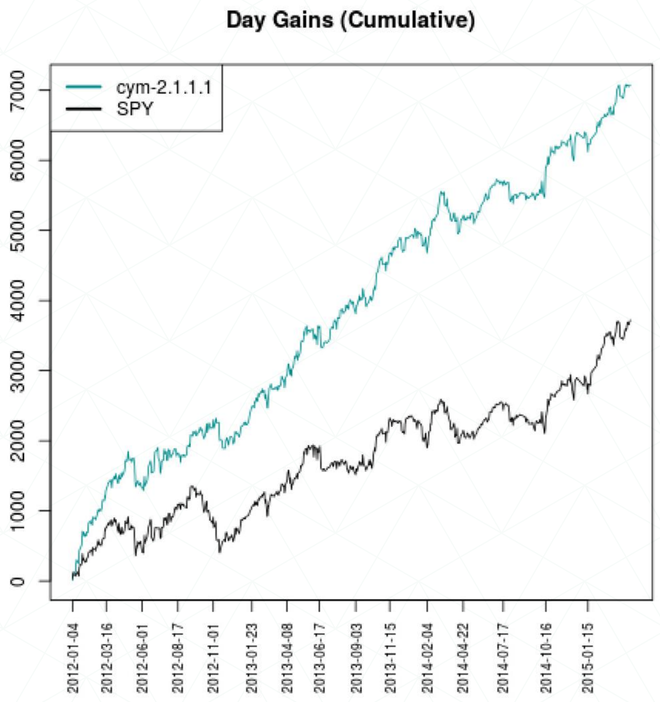
Trade eff	Loss	Total	Lossy%	STD	Last	Min	Var5	Var10	1stQ	Med	Avg	3rdQ	Por10	Por5	Max
Long:	8779	20416	43.00%	31.58	33.65	-0.00	3.57	7.35	19.32	46.21	48.39	76.39	94.59	100.00	100.00
0:	0	0	0.00%	0	0	0	0	0	0	0	0	0	0	0	0
Position:	8779	20416	43.00%	31.58	33.65	-0.00	3.57	7.35	19.32	46.21	48.39	76.39	94.59	100.00	100.00
Day:	1591	4097	38.83%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Total Positions: 20416 Days: 730 Traded Days: 730 Positions/Day: 27.97 Positions/TradedDay: 27.97
Gain: 7076.28 Annualized: 2442.77 MaxDD: 4.70 MaxDDtime: 63 CurrDD: 0.04 CurrDDtime: 8
MaxDDBps: 601.68 CurrDDBps: 6.99 MaxCnegetiveDays: 7 AvgCnegativeDays: 2 Negativedayscount: 64
Wbpp: 2.64 Lbpp: -2.67 Bpp: 0.35 BppL:0.81
Wbpd: 44.09 Lbpd: -47.55 Bpd: 9.69 BpdL:25.83 Wbps/Lbps: -1.5431
RR: 15.05 Sharp: 2.54 Beta: 0.8525 Gini: 0.1324 Skew: -0.2869 Kurt: 1.5604
AR3: 0.29 AR4: 0.16 AR7: 0.52 AR8: 0.38
Day- MarketGain: 3719.94 ExcessGain: 3356.34 AvgExcessGain: 4.60 MedExcessGain: 5.13 AccExcessGain: 56.44%
TDays- MarketGain: 3719.94 ExcessGain: 3356.34 AvgExcessGain: 4.60 MedExcessGain: 5.13 AccExcessGain: 56.44%
%Pos: 56.76% areaVar5: -15.40% areaVar10: -24.00% areaPor10: 25.03% areaPor5: 15.10%
Acc: 57.00% PrecLong: 57.00% PrecShort: 0.00% RecUp: 100.00% RecDown: 0.00%
F-Up: 72.61 F-Down: 0.00

CYM Data 2012-2015 (cont.)







Trade File Format

One file (csv) per symbol, 18 columns per file

No header row, 1 row per trade

Column #	Field	Details
1	Trade reference number	Count of the trade for a given date Reset to 1 for the first trade of each date Suffix "Ref_" string to trade counter
2	Date	YYYYMMDD format
3	Position type	Long / Short
4	Profit (in $)	Profit per position per share in $s
5	Symbol	
6	Entry price	Assume top-ask for buy and top-bid for sell
7	Exit price	Assume top-ask for buy and top-bid for sell
8	Position size	Number of shares / contracts traded
9	Entry value	Column 6 x Column 8
10	Exit value	Column 7 x Column 8
11	Profit (in bips)	Profit per position in basis points
12	Entry time	Time in HH:MM format – in Eastern time (US)
13	Exit time	Time in HH:MM format – in Eastern time (US)
14	Hold Time	Time from entry to exit (in minutes)
15	Minimum profit (in bips)	Minimum bps seen during the life of the trades Measured from time of entry - granularity of atleast once a minute – can be more granular (Once every order-book change / execute)
16	Maximum profit (in bips)	Maximum bps seen during the life of the trades Measured from time of entry - granularity of atleast once a minute – can be more granular (Once every order-book change / execute)
17	Minimum profit time	Time in HH:MM format – in Eastern time (US)
18	Maximum profit time	Time in HH:MM format – in Eastern time (US)

Daily File Format

One file (csv) per symbol, 10 columns per file

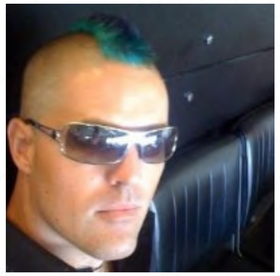
No header row, 1 row per trading day – if there were no trades on a trading day, populate all fields (except Date) with "0"

Column #	Field	Details
1	Date	YYYYMMDD format
2	Positive positions	Count of positions >=0
3	Total positions	Count of positions for the day
4	Long positions	Count of long positions for the day
5	Short positions	Count of short positions for the day
6	Daily bps	Cumulative bps for all trades for the symbol for the date
7	Positive longs	Count of long positions >=0
8	Negative longs	Count of long positions < 0
9	Positive shorts	Count of short positions >=0
10	Negative shorts	Count of short positions < 0

Team

Kasian Franks

A 25 year Silicon Valley veteran and pioneer in digital content streaming much before Netflix and Amazon entered the space. Franks started as a software engineer working for companies such as Genentech, Sun, Oracle, Cisco, Motorola and Morningstar. In 2005, as a genomic research scientist at Lawrence Berkeley National Laboratory, he was the lead inventor of new vector space representations of hidden relationship networks in data along with pattern recognition systems aiming to mimic portions of human cognition. While at the Lab, he co-authored a paper with Michael I. Jordan (machine learning maestro and doctoral advisor to Andrew Ng) titled "Statistical modeling of biomedical corpora: mining the Caenorhabditis Genetic Center Bibliography for genes related to life span - Blei DM1, Franks K, Jordan MI, Mian IS.". Following this, he co-founded SeeqPod in partnership with Lawrence Berkley National Laboratory of U.S . Department of Energy that was then headed by Steven Chu, Energy Secretary in President Obama's first term and winner of Nobel Prize in Physics (1997). SeeqPod was a consumer-facing streaming data search/discovery/recommendation platform originally powering Spotify and others while attracting 50 million monthly active users and 250 million monthly search and recommendation queries. In 2008, his team won the R&D100 award. The company was acquired in 2009. He continues to spend his time mentoring startup founders and hedge funds on Machine Learning, Natural Language Processing (NLP), Artificial Intelligence and data science strategies.



Mike Muldoon

Mike's first program was an ad-lib game, which he wrote in 5th grade on a TRS-80 owned by the school's computer club. He has since established a track record of leading large projects from concept to delivery, and brings over 20 years of experience to Starmine.ai. As employee #1 at SeeqPod, he took the product from whiteboard to 50M monthly active users, delivering an architecture that deployed hundreds of servers across seven different data centers globally pushing 1.6Gb/s of traffic.

Caleb Pate

Caleb is currently working in Data Science, AI & Machine Learning with a focus on feature engineering and cryptocurrencies while continuing to define, explore and solve problems related to recommendation systems. As a member of the founding team at SeeqPod, he built the core Music Recommendation & Curation strategy. He played in a band with an international following and ran an independent music label and continues to create new musical worlds as a Producer, Musician and DJ.

Team Continued: https://vectorspace.ai/index.html#Team

vectorspace.ai