

STARMINE 

Smart Contracts for On-Demand Datasets in  
Machine Learning & Artificial Intelligence

whitepaper v0.88

[starmine.ai](https://starmine.ai)

"In the future AI will be diffused into every aspect of the economy." - [Nils J. Nilsson](#), Founding researcher, Artificial Intelligence & Computer Science, Stanford University.



## Introduction:

[Starmine](#) is a robust and highly scalable platform for constructing, trading and exchanging advanced algorithmically generated on-demand datasets for Machine Learning (ML) and Artificial Intelligence (AI) efforts. Datasets remain at the core of most advances in ML and AI. A dataset is typically made up of rows and columns, similar to an organized matrix or spreadsheet. Specifically, columns contain features along with continuously valued attributes or scores. These features and their scored attributes are stored as 'supercolumns' and can be [automatically engineered](#), traded or exchanged by machines without human intervention. This can also be done with full anonymity using Starmine tokens to execute and transact smart contracts where each dataset or each individual feature column in any dataset can be treated as a tradable asset which is transacted and executed via smart contracts on the blockchain.

A commercial example of a Starmine built dataset with supercolumns can be found at Google's data science community, [Kaggle](#). Starmine supercolumns and datasets are currently being built and transacted by a variety of institutions in the Financial and Cryptocurrency markets including hedge funds, data vendors and Data-as-a-Service (DaaS) providers at Starmine's engineering community where Starmine tokens will be used to transact on-demand updated real-time datasets and supercolumns. The Starmine platform is currently being extended to service ML and AI efforts in additional industries such as Life Sciences/Pharmaceutical, Food, Energy, Travel, Automotive, Entertainment, Advertising etc.

## Objective:



Our current commercial grade release customers (numbering over 150) are primarily data vendors and hedge funds who supply processed data to intelligence firms, institutions and companies with machine learning or artificial intelligence efforts. This functionality enables anyone or any machine (via API) to anonymously construct highly customizable featureengineered supercolumns for datasets based on dynamically updated data sources.

The primary objective in extending the commercially released platform is to enable anyone or any machine to anonymously assemble unique and highly valued datasets made up of Starmine supercolumns and 'package' these datasets with their own algorithms, along with monetary value transacted via smart contracts where Starmine tokens are used. Each supercolumn, entire dataset and complete 'package' will have its own value placed on it by its current owner. Once any of these is created, they can be traded, exchanged, bought or sold.

## **Market Opportunity:**

"In the future AI will be diffused into every aspect of the economy." - Nils J. Nilsson, Founding researcher, artificial intelligence & computer science, Stanford University. The adoption of Machine Learning and Artificial Intelligence applications and solutions continues to grow exponentially. The demand for data, including datasets along with the features and columns that make them up, continues to grow. Addressing this market which remains mostly uncultivated, makes up one of the largest opportunities in technology today. Additional industry markets & verticals targeted for growth in applied Machine Learning and Artificial Intelligence:



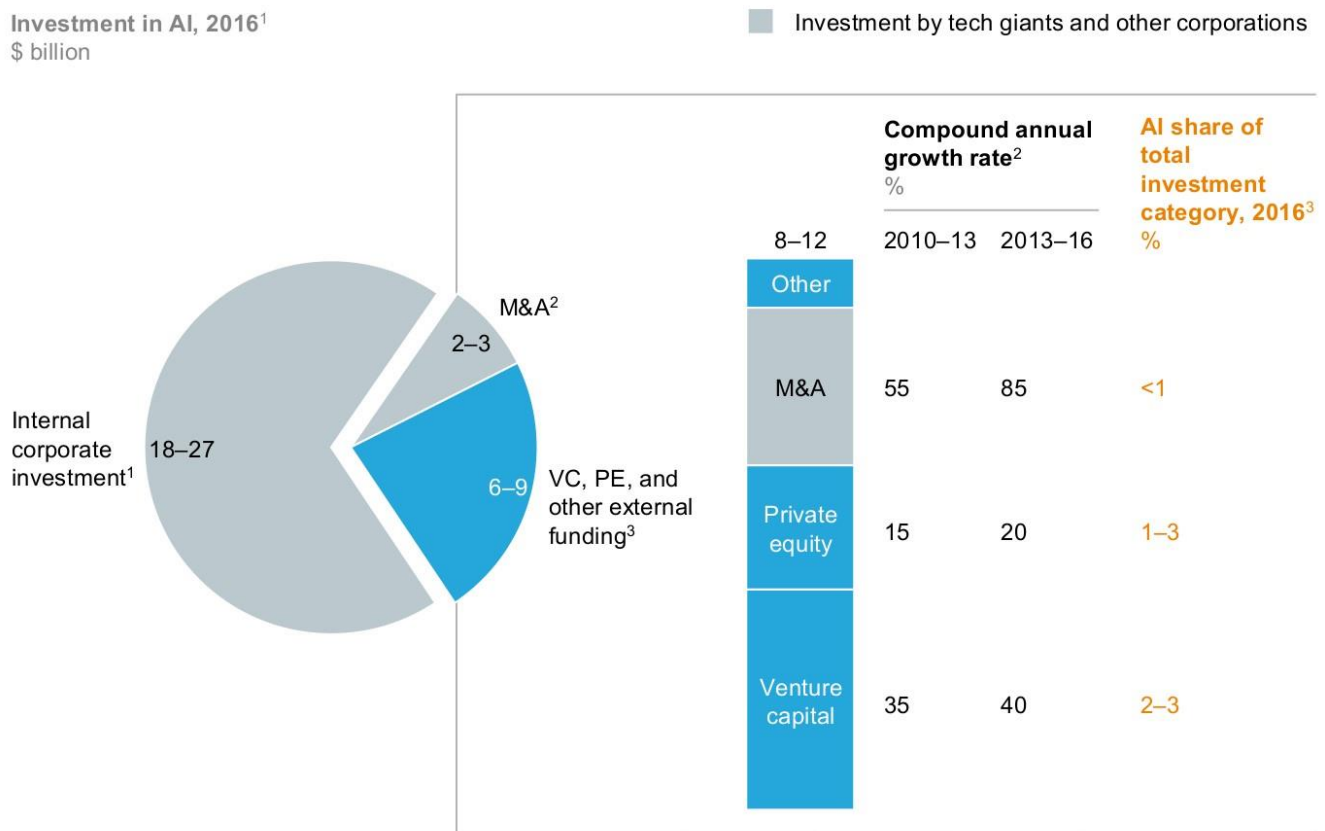
- Food
- Life Sciences/Pharmaceutical
- Movies, TV, Music & Entertainment
- Automotive
- Travel
- Products
- Advertising
- Customer Defined

[McKinsey reports](#) estimate that between \$18 billion and \$27 billion were invested by large corporations in 2016; external investment (from VCs, PE firms, M&A, grants, and seed funding) was around \$8 billion to \$12 billion as shown in the chart below:



## Technology giants dominate investment in AI

Investment in AI, 2016<sup>1</sup>  
\$ billion



1 Estimate of 2016 spend by corporations to develop and deploy AI-based products. Calculated for top 35 high tech and advanced manufacturing companies investing in AI. Estimate is based on the ratio of AI spend to total revenue calculated for a subset of the 35 companies.

2 VC value is an estimate of VC investment in companies primarily focused on AI. PE value is an estimate of PE investment in AI-related companies. M&A value is an estimate of AI deals done by corporations. "Other" refers to grants and seed fund investments. Includes only disclosed data available in databases, and assumes that all registered deals were completed within the year of transaction. Compound annual growth rate values rounded.

3 M&A and PE deals expressed by volume; VC deals expressed by value.

SOURCE: Capital IQ; Pitchbook; Dealogic; S&P; McKinsey Global Institute analysis

To more accurately grasp the huge potential of the market opportunity or the addressable market, an understanding of the data ecosystem that underpins all of Machine Learning and AI efforts today and in the near future. As the McKinsey Global Institute June 2017 report titled "Artificial Intelligence – The Next Digital Frontier?" describes it this way: "Artificial intelligence, as an idea, first appeared soon after humans developed the electronic digital computing that makes it possible. And, like digital



technology, artificial intelligence, or AI, has ridden waves of hype and gloom—with one exception: AI has not yet experienced wide-scale commercial deployment.

That may be changing. Machines powered by AI can today perform many tasks—such as recognizing complex patterns, synthesizing information, drawing conclusions, and forecasting—that not long ago were assumed to require human cognition. And as AI’s capabilities have dramatically expanded, so has its utility in a growing number of fields. At the same time, it is worth remembering that machine learning has limitations. For example, because the systems are trained on specific data sets, they can be susceptible to bias; to avoid this, users must be sure to train them with comprehensive data sets. Nevertheless, we are seeing significant progress...”

“... Data is at the heart of the disruptions occurring across economies and is recognized as an increasingly critical corporate asset. Without data, getting the AI engine started is impossible. Because of this, business leaders should know what data they already have access to and where they can obtain additional data relevant for their company’s future success. Google and Facebook are well-known examples of companies that obtain most of their revenue through insights they extract from the enormous quantities of data their customers generate on a daily basis by using their services.

One important capability will be making data usable that is not available in a relational format or that cannot be analyzed with traditional methodologies. Much data being produced in industry today is “flat data,” without relational structure—in manufacturing, an estimated 90 percent of data is flat. 39



Making this data usable requires new approaches that can efficiently handle large volumes of different types of data—for example, NoSQL and Hadoop technologies.

Firms also need to recognize potentially distinctive types of data that can create a competitive edge in AI-enabled product offerings. Customer sentiment and geo-locational real-time event data are examples of differentiating data for which competition for exclusive access is likely to intensify. Certain data may become valuable only if combined with other data sources in a larger ecosystem. Given the rapidly increasing data output from sensors, machinery, and social networks, organizations face challenges in how to handle such massive streams of information. While some use cases for such data will be very concrete, with clear requirements, other potential uses will be fuzzy or not yet fully defined. Some use cases will require significant time series of data, while others may require real-time data. Companies will need to decide which data to store in their original granularity and which to aggregate or pre-analyze. With increasing data storage capacities in the cloud as well as more powerful “edge” computing capabilities close to sensors, flexibility increases rapidly.”

This kind of data-centric environment places Starmine in front of a unique and powerful opportunity connected to providing on-demand updating algorithmically generated and human curated datasets for any ML or AI effort.

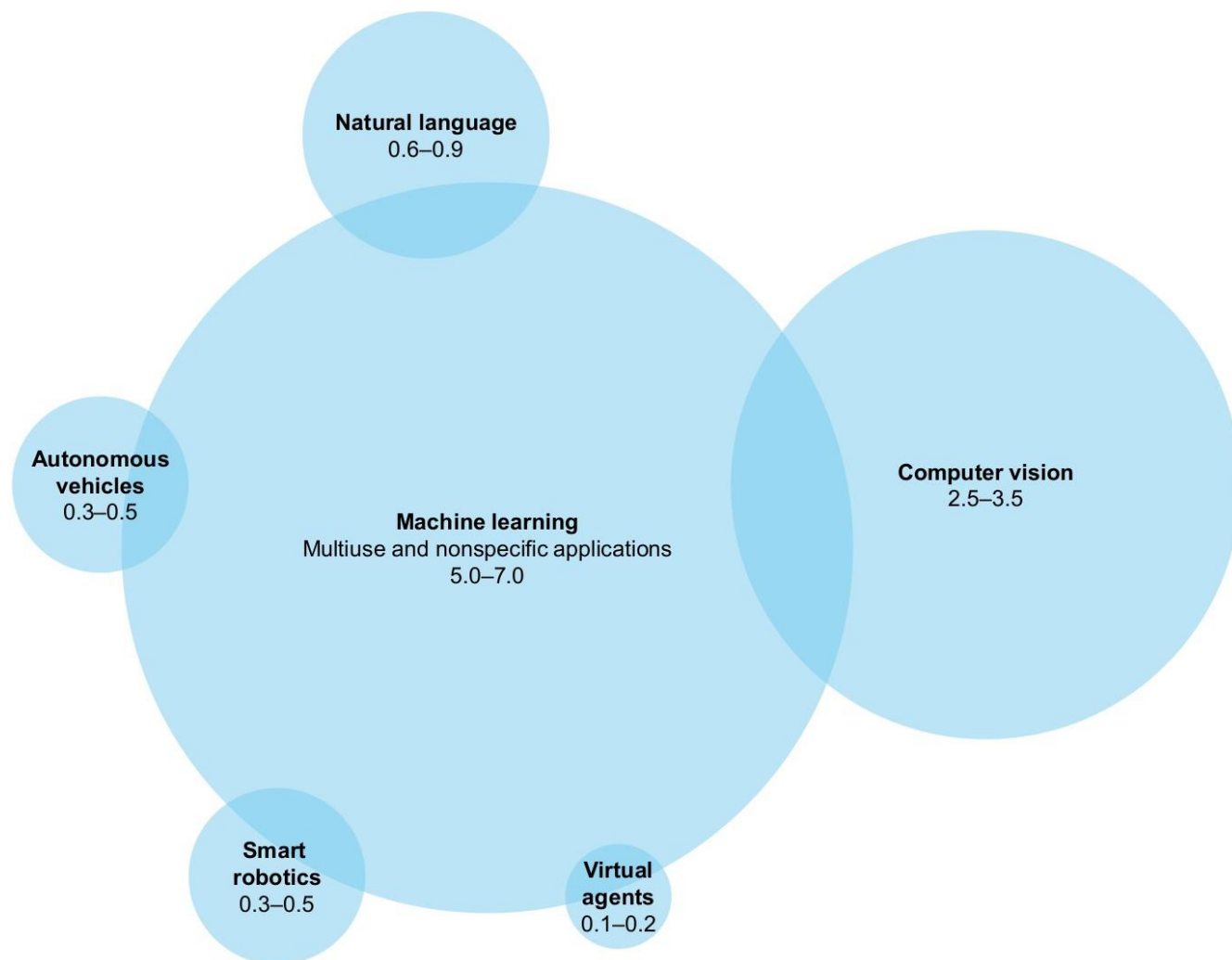
As the market for ML and AI applications increases the market for algorithmically generated datasets also increases. Here are a few data points describing growth in this sector:



**Machine learning received the most investment, although boundaries between technologies are not clear-cut**

External investment in AI-focused companies by technology category, 2016<sup>1</sup>

\$ billion



<sup>1</sup> Estimates consist of annual VC investment in AI-focused companies, PE investment in AI-related companies, and M&A by corporations. Includes only disclosed data available in databases, and assumes that all registered deals were completed within the year of transaction.

SOURCE: Capital IQ; Pitchbook; Dealogic; McKinsey Global Institute analysis














































































































































































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

AI Index

Relatively low  Relatively high

	Overall AI index	MGI Digitization Index <sup>1</sup>	Assets			Usage						Labor	
			Depth of AI technologies	AI spend	Supporting digital assets	Product development	Operations	Supply chain and distribution	Customer experience	Financial and general management	Workforce management	Exposure to AI in workforce	AI resources per worker
High tech and telecommunications													
Automotive and assembly													
Financial services													
Resources and utilities													
Media and entertainment													
Consumer packaged goods													
Transportation and logistics													
Retail													
Education													
Professional services													
Health care													
Building materials and construction													
Travel and tourism													

<sup>1</sup> 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.

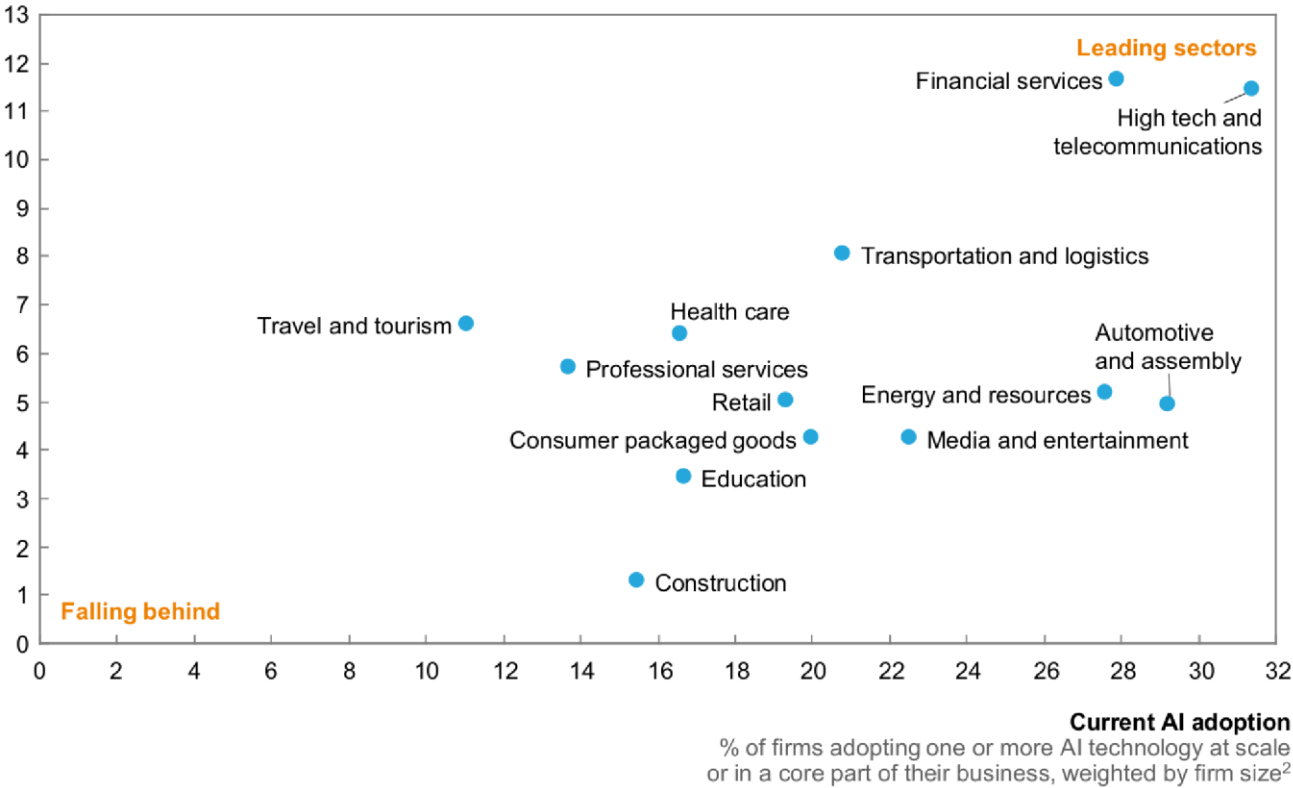
SOURCE: McKinsey Global Institute AI adoption and use survey; *Digital Europe: Pushing the frontier, capturing the benefits*, McKinsey Global Institute, June 2016; *Digital America: A tale of the haves and have-mores*, McKinsey Global Institute, December 2015; McKinsey Global Institute analysis



Sectors leading in AI adoption today also intend to grow their investment the most

Future AI demand trajectory<sup>1</sup>

Average estimated % change in AI spending, next 3 years, weighted by firm size<sup>2</sup>

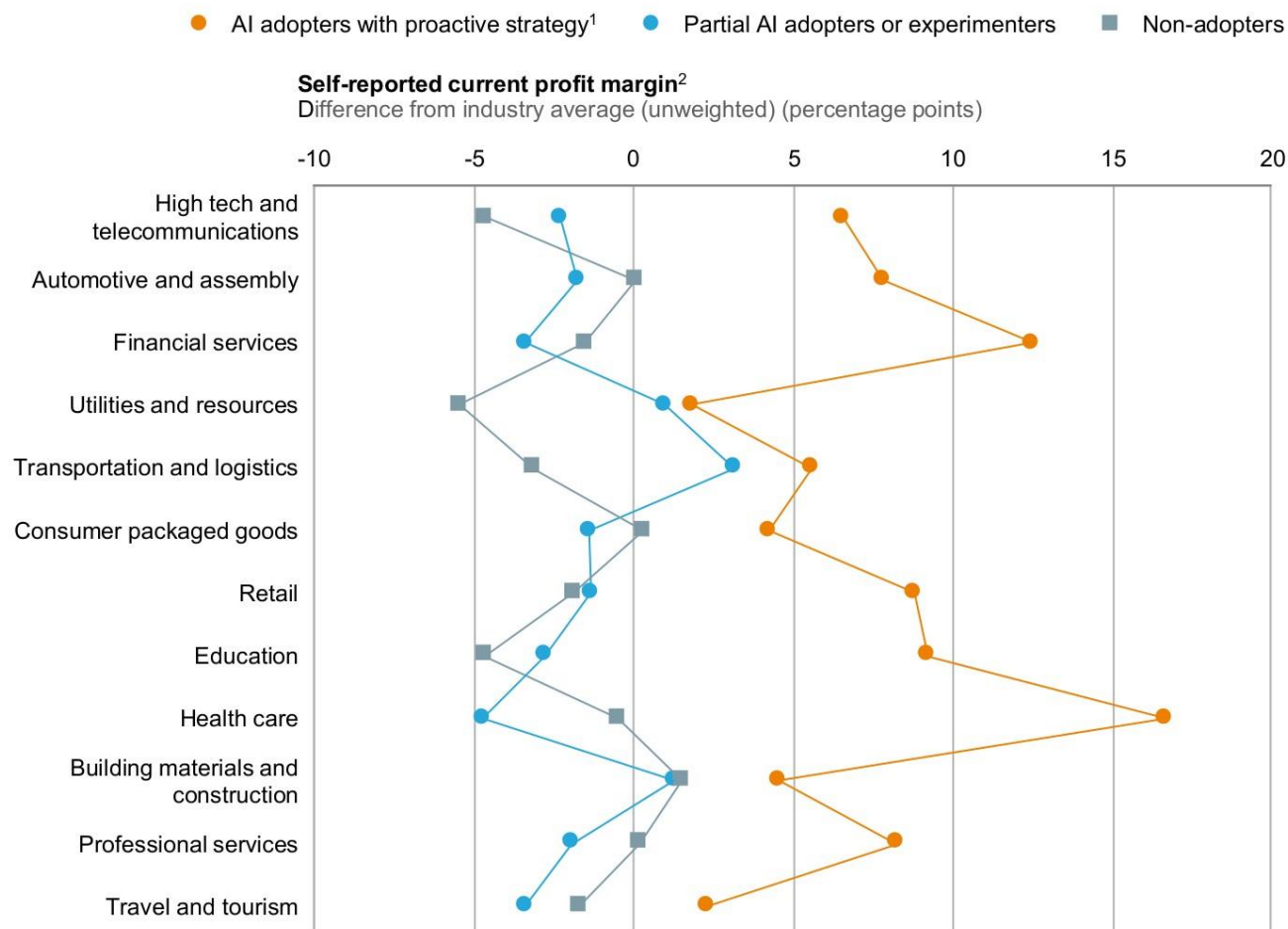


1 Based on the midpoint of the range selected by the survey respondent.  
2 Results are weighted by firm size. See Appendix B for an explanation of the weighting methodology.

SOURCE: McKinsey Global Institute AI adoption and use survey; McKinsey Global Institute analysis



## AI adopters with a proactive strategy have significantly higher profit margins



- 1 Firms that are big data and cloud services users and report their strategic posture toward AI to be: "Disrupting our industry using AI technology is at the core of our strategy," "We have changed our longer-term corporate strategy to address the AI threat or opportunity disruption," or "We have developed a coordinated plan to respond to the AI threat or opportunity but have not changed our longer-term corporate strategy."
- 2 Operating profit margin for selected sectors as a share of turnover, for continuing operations and before exceptional items.

SOURCE: McKinsey Global Institute AI adoption and use survey; McKinsey Global Institute analysis



## Artificial intelligence can create value across the value chain in four ways

		Applicability			
		Project	Produce	Promote	Provide
		Enlightened R&D, real-time forecasting, and smart sourcing	Operations with higher productivity, lower cost, and better efficiency	Products and services at the right price, with the right message, and to the right targets	Enriched, tailored, and convenient user experience
Applicable technologies	Retail	Anticipate demand trends, while optimizing and automating supplier negotiation and contracting	Automate warehouse and store operations; optimize merchandising, product assortment, and microspace	Optimize pricing, personalize promotions, and tailor website displays in real time	Personalize tips and suggestions, offer immediate assistance with virtual agents, automate in-store checkout, and complete last-mile delivery by drones
	Electric utilities	Enhance demand and supply prediction, assess reliability of integrated generation assets, and automate demand-side response	Optimize preventive maintenance, improve electricity production yield, reduce energy waste, and prevent electricity theft	Optimize pricing with time-of-day and dynamic tariffing; match producers and consumers in real time	Automate supplier selection, provide consumption insights, automate customer service with virtual agents, and tailor usage to consumer's preferences
	Manufacturing	Improve product design yield and efficiency, automate supplier assessment, and anticipate parts requirements	Improve processes by the task, automate assembly lines, reduce errors, limit product rework, and reduce material delivery time	Predict sales of maintenance services, optimize pricing, and refine sales-leads prioritization	Optimize flight planning and route and fleet allocation; enhance maintenance engineer and pilot training
	Health care	Predict disease, identify high-risk patient groups, and launch prevention therapies	Automate and optimize hospital operations; automate diagnostic tests and make them faster and more accurate	Predict cost more accurately, focus on patients' risk reduction	Adapt therapies and drug formulations to patients, use virtual agents to help patients navigate their hospital journey
	Education	Anticipate job market demand, identify new drivers of performance to assess students, and help graduates highlight their strengths	Automate teachers' routine tasks, identify early disengagement signs, and optimize group formation for learning objectives		Personalize learning, shift from stop-and-test model to continuous learning cadenced by virtual coaches and tutors, and build student self-awareness

1 Machine learning for multiuse and nonspecific applications.

SOURCE: McKinsey Global Institute analysis





## AI can help capture significant gains across the value chain

### Examples of AI-related business impact from current use cases

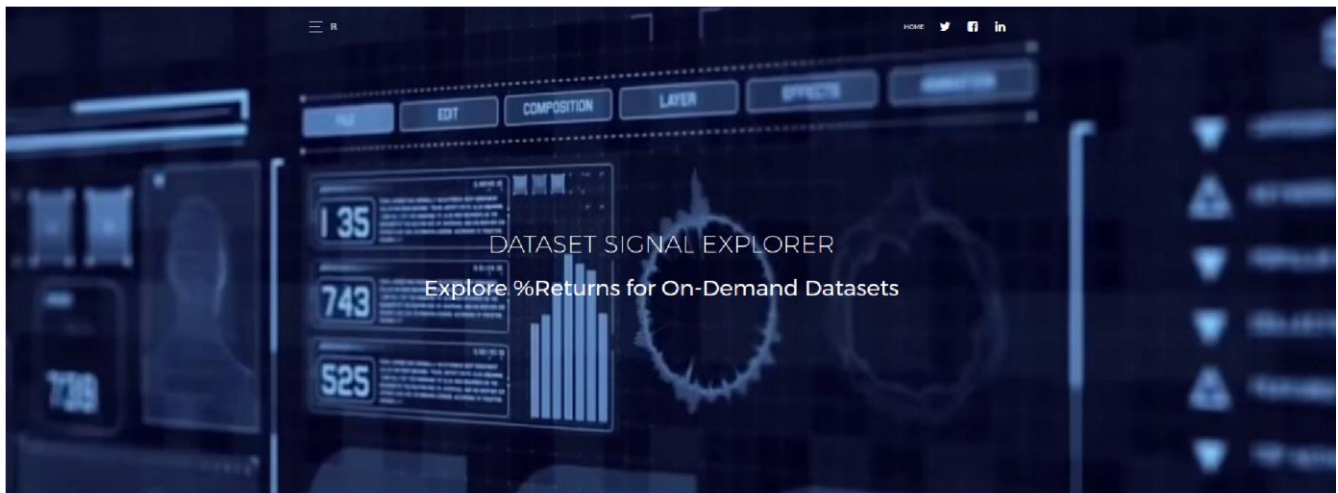
	Project	Produce	Promote	Provide
	Accurate demand forecasting, smart sourcing, and enlightened R&D	Higher productivity and minimized maintenance and repairs	Products and services at the right price, with the right message, to the right targets	Enriched, tailored, and convenient user experience
Retail	<ul style="list-style-type: none"> <li>▪ <b>1–2% EBIT<sup>1</sup> improvement</b> using machine learning to anticipate fruit and vegetable sales</li> <li>▪ <b>20% stock reduction</b> using deep learning to predict e-commerce purchases</li> <li>▪ <b>2 million</b> fewer product returns per year</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>30% reduction of stocking time</b> using autonomous vehicles in warehouses</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>50% improvement of assortment efficiency</b></li> <li>▪ <b>4–6% sales increase</b> using geospatial modeling to improve micromarket attractiveness</li> <li>▪ <b>30% online sales increase</b> by using dynamic pricing and personalization</li> </ul>	
Electric utilities	<ul style="list-style-type: none"> <li>▪ Objective to cut <b>10% in national electricity usage</b> by using deep learning to predict power demand and supply</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>20% energy production increase</b> using machine learning and smart sensors to optimize assets' yield</li> <li>▪ <b>10–20% EBIT improvement</b> by using machine learning to enhance predictive maintenance, automate fault prediction, and increase capital productivity</li> </ul>		<ul style="list-style-type: none"> <li>▪ <b>\$10–\$30 savings on monthly bills</b> by using machine learning to automatically switch electricity supply deals</li> </ul>
Manufacturing	<ul style="list-style-type: none"> <li>▪ <b>10% yield improvement</b> for integrated-circuit products using AI to improve R&amp;D process</li> <li>▪ <b>39% IT staff reduction</b> by using AI to fully automate procurement processes</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>30% increase of material delivery time</b> using machine learning to determine timing of goods' transfer</li> <li>▪ <b>3–5% production yield improvement</b></li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>13% EBIT improvement</b> by using machine learning to predict sources of servicing revenues and optimize sales efforts</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>12% fuel savings</b> for manufacturers' customers, airlines, by using machine learning to optimize flight routes</li> </ul>
Health care	<ul style="list-style-type: none"> <li>▪ <b>\$300 billion possible savings</b> in the United States using machine learning tools for population health forecasting</li> <li>▪ <b>£3.3 billion possible savings</b> in the United Kingdom using AI to provide preventive care and reduce nonelective hospital admissions</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>30–50% productivity improvement</b> for nurses supported by AI tools</li> <li>▪ Up to <b>2% GDP savings</b> for operational efficiencies in developed countries</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>5–9% health expenditure reduction</b> by using machine learning to tailor treatments and keep patients engaged</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>\$2 trillion–\$10 trillion savings</b> globally by tailoring drugs and treatments</li> <li>▪ <b>0.2–1.3 additional years</b> of average life expectancy</li> </ul>
Education		<ul style="list-style-type: none"> <li>▪ Virtual teaching assistants can <b>answer 40% of students' routine questions</b></li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>1% increase in enrollment</b> by using a virtual assistant to follow up with applicants</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>85% match with human grading</b>, using machine learning and predictive modelling</li> </ul>

1 Earnings before interest and taxes.

SOURCE: McKinsey Global Institute analysis

## Products:





## Generate a Dataset, Explore Signals and %Returns

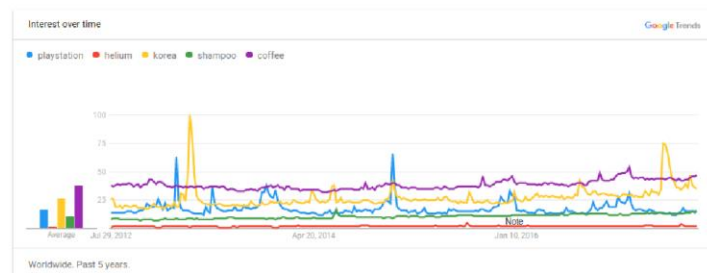
Enter 1 to 5 concepts, keywords or [Google Search Trends](#) to generate a dataset  
example: [Playstation](#), [Helium](#), [Korea](#), [Shampoo](#), [Coffee](#)

1 2 3 4 5

Returns for Playstation, Helium, Korea, Shampoo, & Coffee in comparison to the S&P 500: (interactive)



Google Search Trends for Playstation, Helium, Korea, Shampoo, & Coffee



We have focused the current commercial grade Starmine platform with on the financial industry.

Within the financial domain, products are powered by public and private data sources including global



search trends, sentiment, social trends, scientific literature, public filings and contracts, patents and APIs combined with human curation. ***Some of these sources are currently described in the 2<sup>nd</sup> option in one of our customizable interfaces as seen in the figure below and via commercially available interfaces located <http://starmine.ai> :***

#### Option 1. Download a Dataset

- US Equities: Known & hidden relationships to Elements in the Periodic Table(CSV)
- Bitcoin & other cryptocurrencies: Known & hidden relationships to Global Trending Searches (processing...)
- US Equities: Known & hidden relationships to Global Trending Searches (processing...)

#### Option 2. Build a Dataset

##### Step 1.

- ☐ Real-time Search Trends: Google, Bing, Facebook, 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

##### Step 2.

Your address, to notify on completion

Build Request

#### Option 3. Create a Dataset

##### Step 1.

Equity Type: Nasdaq, NYSE & OTCBB stocks

Data Stream: Nasdaq, NYSE & OTCBB • Genomics & Molecular Biology

##### Step 2.

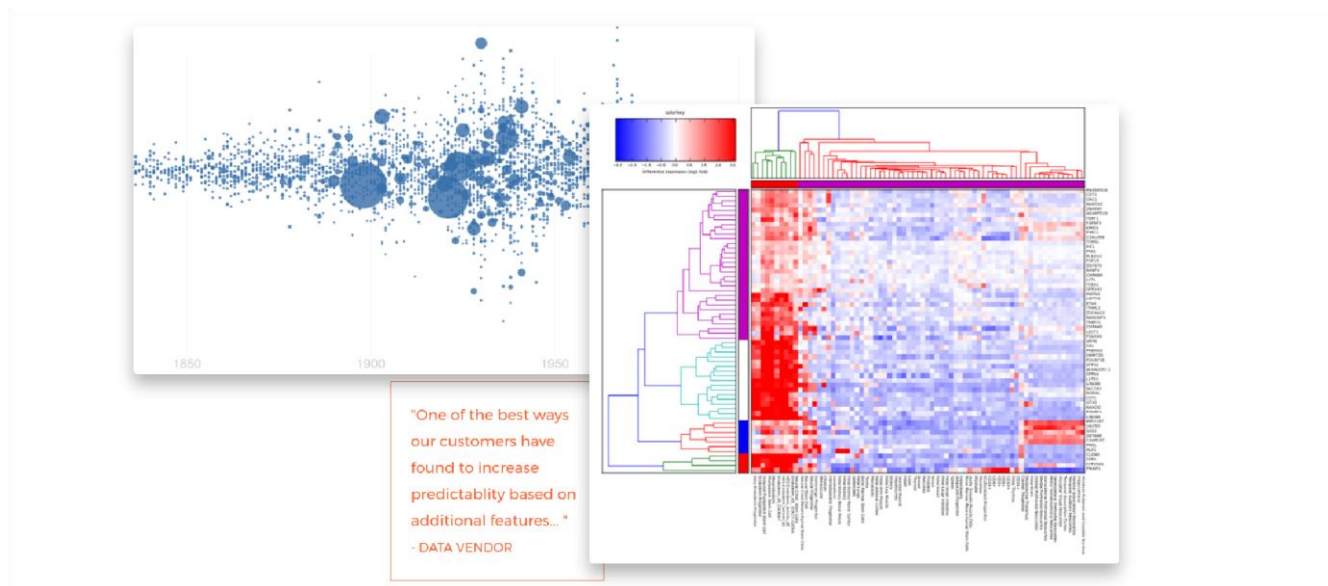
Enter 1 to 5 features, concepts or keywords:

example: Playstation, Helium, Korea, Shampoo, Coffee

1.	2.	3.	4.	5.
----	----	----	----	----

Create Request



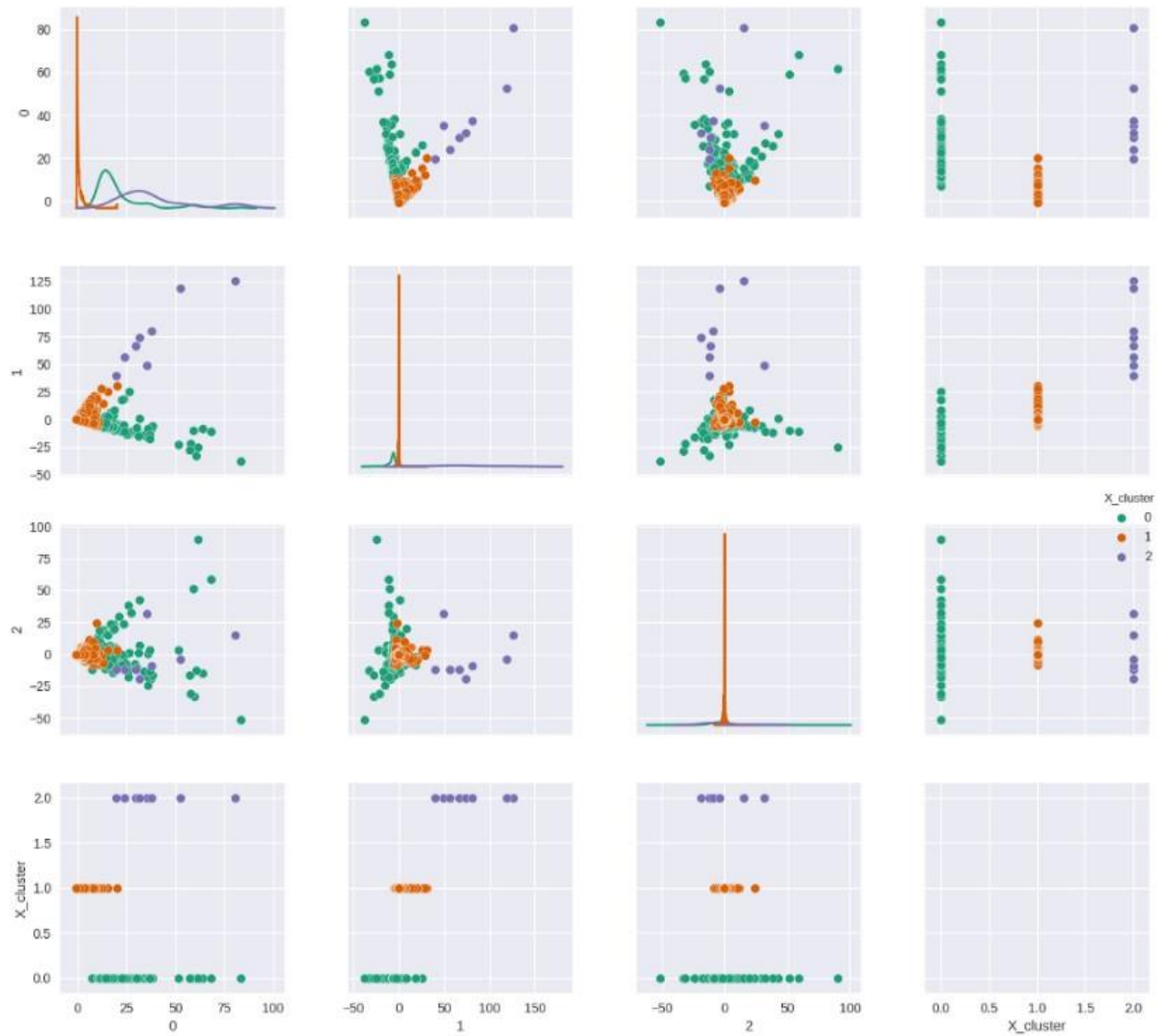


The primary objective in extending the commercially released platform is to enable anyone or any machine to anonymously assemble unique and highly valued datasets made up of Starmine supercolumns and ‘package’ these datasets with their own algorithms, along with monetary value via where Starmine tokens are used. Each supercolumn, entire dataset and complete ‘package’ will have its own value placed on it by its current owner. Once any of these is created, they can be traded, exchanged, bought or sold.

The value or worth of one of these packages can be best described in an article titled “[Profiting from Python & Machine Learning in the Financial Markets](#)” where supercolumns were built for a dataset that resulted in unique clusters of equities resulting in gains beyond overall market returns as shown here:







Resulting clusters of stocks.





Clusters converted to baskets of stocks with historical returns.

Next phases include commercial grade roll-out to a variety of industry verticals.

Customer use cases currently include: Unsupervised Machine Learning, Artificial Intelligence, additional signal boosting, advanced real-time clustering or visualization for Stocks (NYSE, Nasdaq, OTC), Cryptocurrencies (BTC, ETH, LTC), ETFs & Options based on user-defined supercolumns & feature labels.

## Technology:

"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)



Adding feature attributes via columns placed to a dataset can result in boosting the overall signal found in the data. To achieve this goal, a data acquisition process occurs followed by the processing of that data to establish mathematical vectors that form a relationship network which can be used to automatically engage in large-scale feature engineering.

For automated feature engineering and supercolumn construction for datasets, Starmine uses an ensemble of algorithmic approaches in the areas of Matrix Factorization, Reinforcement Learning (RL), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), advanced Natural Language Processing (NLP) and Probabilistic Graphical Models (PBMs). The finished product from this pipeline results in one or more supercolumns that can be combined to form a unique dataset or dataset augmentation for generating context-controlled clusters, acute relationships or boosting overall signal in ML and AI operations.

In general, a dataset can be represented by a number of blocks or smart contracts on the blockchain as each of these would be

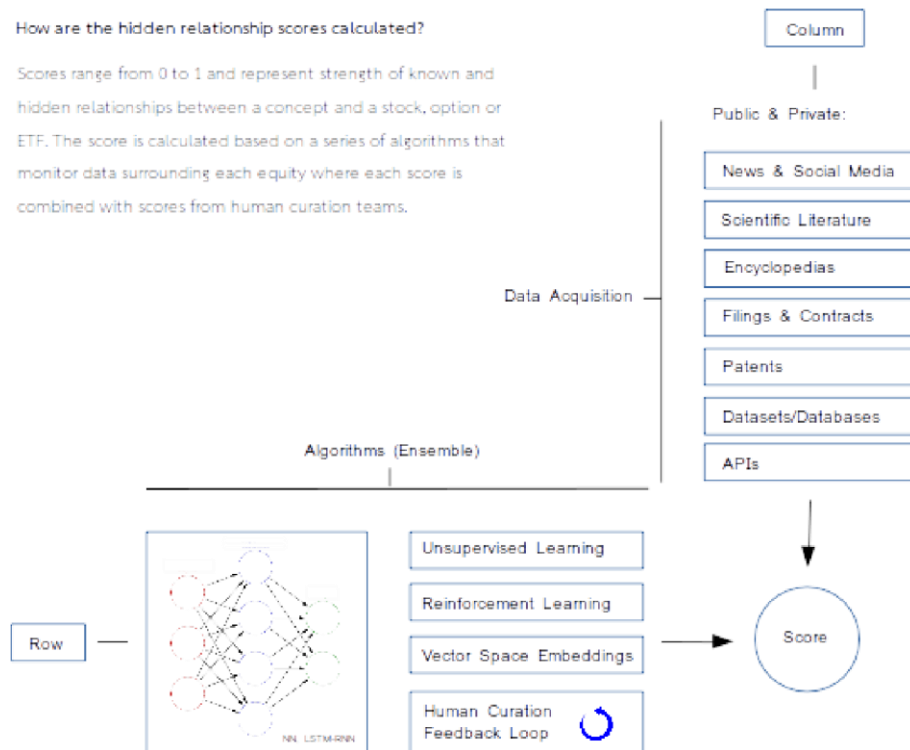
The system also provides a recommendation engine to guide ML and AI systems in the automated process of selecting features or labels.

Operational supercolumn dataset pipeline in use today for the financial and cryptocurrency markets:



#### How are the hidden relationship scores calculated?

Scores range from 0 to 1 and represent strength of known and hidden relationships between a concept and a stock, option or ETF. The score is calculated based on a series of algorithms that monitor data surrounding each equity where each score is combined with scores from human curation teams.



Starmine is enabled with an interface that allows human curators or experts to automatically engage in feature engineering in terms of generating supercolumns for unique customized anonymous datasets. These supercolumns, once created by an individual or a machine, are then ready to be placed in the blockchain with an associated smart contract.

Once this process is complete, supercolumns can be traded and exchanged anonymously for augmentation to any other datasets as automatically requested by any other ML or AI operation.



## Revenue Model:

Our revenue model includes three tiers of pay-as-you-go and on-demand pricing models that can fit the budget of an individual or institution:

### On-Demand Price Tiers

Free	Data Vendor	Institutional
<b>TIER 1: FREE LIMITED</b>	<b>\$0.99 PER ON-DEMAND UPDATE</b>	<b>\$1,950.00/MO + \$0.99 PER ON-DEMAND UPDATE</b>
1 Free on-demand update	100 Free on-demand update	10,000 Free on-demand updates
Equity Types: NYSE stocks	Equity Types: Bitcoin & CryptoCurrencies, NYSE, Nasdaq, OTC	Equity Types: Any
Data Streams: Featured only	Data Streams: Any data streams	Data Streams: Any
	Custom Features	Custom Features
	Support	Support

### Starmine 3 – Year revenue projections:

	Year 1		Year 2		Year 3	
	Market: Nasdaq/NYSE	Cryptocurrency	Market: Nasdaq/NYSE	Cryptocurrency	Market: Nasdaq/NYSE	Cryptocurrency
<b>Customers:</b>	100	75	220	260	380	500
<b>Trading hours per day:</b>	6	24	6	24	6	24
<b>Trading days per year (avg.):</b>	252	365	252	365	252	365
<b>On-Demand Updates per minute:</b>	9,072,000	39,420,000	19,958,400	136,656,000	34,473,600	262,800,000
<b>Number of Custom Data Sources:</b>	5	5	5	5	5	5
<b>Cost per On-Demand Update:</b>	\$0.99	\$0.99	\$0.99	\$0.99	\$0.99	\$0.99
<b>Revenue:</b>	\$44,906,400.00	\$195,129,000.00	\$98,794,080.00	\$676,447,200.00	\$170,644,320.00	\$1,300,860,000.00
<b>Total Revenue:</b>	\$240,035,400.00		\$775,241,280.00		\$1,471,504,320.00	

Starmine currently uses Stripe payment processing for transactions, however, Starmine Tokens will be used at a discount to any fiat currency. Revenue projections as Starmine tokens are used to transact dataset assets combined with standard fiat currency:



## **Customer Acquisition, Sales & Marketing Strategy:**

Our use of funds will include building out a sales and marketing team based on current product-to-market fit. Our partnership strategy, which includes using our currently released products and technology to open doors for us, is closely related to our sales and marketing strategy. Our team has had direct hands-on experience in this area for many years. We think long term and consider every customer as a potential partner of ours.

Quickly expanding into other vertical markets including Life Sciences/Pharmaceutical, Food, Energy, Travel, Automotive, Entertainment, Advertising etc. will be the primary focus of our customer acquisition, sales and marketing strategy.

## **ICO Details & Participation:**

**Token name:** SME (Starmine Tokens)

**Token value:** 1 ETH = 2000 SME (Starmine Tokens)

**Tokens created:** 2 Billion coins pre-mined

**Tokens held by the management team:** 310 Million

**Tokens held by pre-ICO participants:** 1.090 Billion

**Tokens available during the ICO:** 600 million coins or 30%

Starmine plans to launch the ICO on the 30<sup>th</sup> of September, 2017. It will end on the 16<sup>th</sup> of October, 2017 or until the maximum cap is reached.

**Accepted Currencies:** ETH



## Transaction of Currencies:

ETH can be sent to the list of addresses, which will be published on the Starmine website before the Starmine ICO.

## Allocation & Use of Funds:

Funds from the Starmine ICO will be used as growth capital for the purpose of scaling Starmine operations globally. This includes expanding dataset and automated supercolumn feature engineering construction and transactions across all industries. In addition, capital will be used to ramp up a sales and marketing organization along with data engineering operations including security, customer support, maintenance, joint development agreements, legal expenses and PR.

We currently have team of engineers in Silicon Valley and abroad which we intend to expand significantly including acquiring additional offices.

## Team



### **Kasian Franks** ([HN](#) | [LinkedIn](#) | [Medium](#))

As a 25-year Silicon Valley veteran and pioneer in digital content streaming 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 Berkeley Lab and the 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 advising hedge funds on Machine Learning, Natural Language Processing (NLP), Artificial Intelligence and data science strategies.



### **Mike Muldoon** ([LinkedIn](#))

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 pushing 1.6Gb/s of traffic.



### **Caleb Pate** ([LinkedIn](#))

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.

