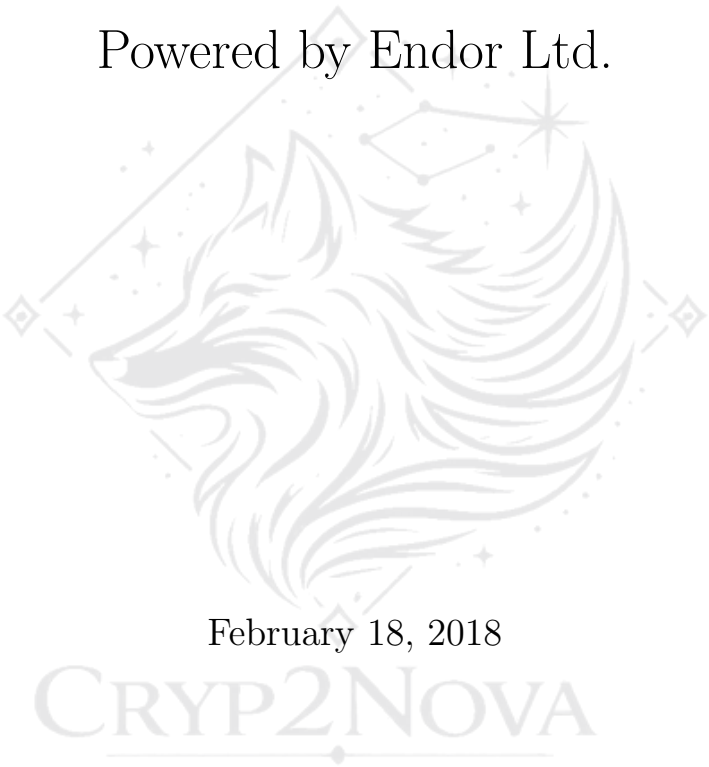
ENDOR.COIN PROTOCOL

Make Artificial Intelligence Predictions Accessible for All

Powered by Endor Ltd.



February 18, 2018

CRYP2NOVA

Archived copy · cryp2nova.com

Make Artificial Intelligence Predictions Accessible for All

Abstract

Endor.coin is reinventing predictive analytics by democratizing access to Artificial Intelligence data analysis, making it accessible, trustless, censorship-resistance and useful for all.

This is achieved through the *Endor.coin* Protocol, offering the world's first automated, self-served, predictive platform that allows business users and unprofessional crypto-token holders alike to ask complex predictive questions and obtain high-quality results in minutes. This aims to democratize the field of Data Science, that today is reserved mostly for Fortune-500 companies. *Endor.coin* is based on the novel science of *Social Physics* developed at *MIT* by the project's team members Prof. Alex Pentland and Dr. Yaniv Altshuler.

- **Constantly-expanding catalogue of predictions:** *Endor.coin* will be launched with a variety of pre-defined tokens-related predictions (e.g. tokens predicted to increase volume, decrease volatility, ...). These predictions would be accessible for purchase using the platform's dedicated *EDR* token. *Endor.coin* would then expand the selection of predictions it caters by allowing users to send *Requests For Predictions* (RFP) – suggesting new types of predictions, that would be implemented and become available for purchase.

- **Do-it-yourself API for advanced users:** tech-savvy users and professionals would be provided a self-serve interface, allowing them to easily provide a definition of any desired behavioral pattern, having the *Endor.coin* platform automatically generating a "look-a-likes" prediction of such pattern in return.

- **Automatic fusion of private and public data**: *Endor.coin* commercial customers (such as banks, retailers and insurers) will be able to easily integrate their proprietary data streams with the platform – producing high-quality predictive insights harvested from the fusion of private and public data. Thanks to the use of *Social Physics* data integration is automatic and friendly – requiring no cleaning or data-preparation.

- **Data Privacy**: Guaranteed, due to the unique ability of *Social Physics* to use data that is fully encrypted at the customer's side.

- **Predictions by the People, for the People**: using *Social Physics* data is processed once, and users pay only for the personalization component they require, offering the "99%" predictive capability reserved today for tech-giants, for 1% of its cost.

Endor.coin is rooted in the technological platform of Endor. A Gartner Cool Vendor, recently recognized by the World Economic Forum as a Technology Pioneer, Endor is an MIT spin-off financially backed by leading investors such as *Innovation Endeavors*, working with Fortune-500 companies such as Coca Cola, Walmart and MasterCard.

Archived copy · cryp2nova.com

Contents

1 Value Proposition **4**

2 Introduction **6**
2.1 An Inefficient and Troubled Market 6
2.2 Democratization Requires Decentralization and Data Separation 6
2.3 Technological Gap 7
2.4 *Endor.coin* Protocol — Project Overview 8

3 Democratizing the Prediction Science **10**
3.1 Social Physics – a New Science from *MIT* 10
3.2 1^{st} Phase : **Endor.com** – Automatic Prediction Engine for Enterprises . . . 13
3.3 2^{nd} Phase : ***Endor.coin* Protocol** – Data Science for the Masses 17
3.3.1 Definitions 18
3.3.2 Data Providers 18
3.3.3 Prediction Engines 19
3.3.4 Pre-Defined Predictions and Request for Predictions (RFPs) 19
3.3.5 Private-Data Analysis and Self-Serve API 19
3.4 Roadmap 19

4 Trustless, Censorship-Resistant and Accountable **22**
4.1 Accountability and Authenticity for Artificial Intelligence 22
4.2 Censorship-Resistance through a Decentralized Protocol 23
4.3 Privacy-Preserving Data Analytics: Using Encrypted Data 24
4.4 Network Effects 25

5 Enabling an Ecosystem **28**
5.1 Public Data Providers 28
5.2 Academic Research Groups 29
5.3 Catalysts – Application Developers 29
5.4 Data Sovereignty 30

6 Token Implementation **31**
6.1 Blockchain Structure 31

Archived copy · cryp2nova.com

6.2 Smart Contracts 32
6.3 Endor's Roles at Launch 34
6.4 Token Privileges and Economy 35
6.5 Use of Proceeds 36

7 Technological Advantages and Differentiation 38
7.1 A Scientific Revolution from the MIT Furnaces 38
7.2 Real Product. Proven Technology 38
7.3 Usability, Value to Users and Value for Token Holders 39

8 Team 40
8.1 Key Team Members 40
8.2 Advisors 43

Appendix A : *Social Physics* Explained 46

Appendix B : Endor's Common Use-Cases for Enterprises 82

Appendix C : Endor.coin Examples of Pre-Defined Predictions 94

Appendix D : Knowledge Sphere Class API 95

Bibliography 102

CRYP2NOVA

Archived copy · cryp2nova.com

Chapter 1

Value Proposition

The ability to understand, predict and influence consumer behavior quickly could give any business an unfair advantage over its competition. Smart business leaders have many ideas for influencing customer behavior to improve business performance. To implement them, they need to answer questions such as:

- Who are our top customers and how do we acquire more of them?
- Who is likely to try this newly-launched product?
- How can we reduce our reliance on promotions?
- Where should we open our next store?
- Who will switch from product A to B next month?

To answer these questions, organizations turn to powerful tools like data science and predictive analytics. Unfortunately, the current process for implementing these tools is slow, painful, and expensive.

- Requires 'unicorns': well-trained, expensive, rare data scientists and PhDs
- Requires 4-6 iterations, each taking a few days / weeks
- Each new business question requires building a new model over a few weeks
- When products and behaviors change, the model breaks

Powered by MIT's novel *Social Physics* technology the *Endor.coin* Blockchain-based Protocol is the first decentralized, trustless, censorship resistance behavioral prediction platform that provides high-quality results for any predictive question in minutes. No coding, data cleaning, or a team of PhDs required.

Archived copy · cryp2nova.com

Following are the key aspects of the *Endor.coin* project:

- **Technology:** Powered by MIT *Social Physics* technology [1], providing up to X10 higher accuracy for trends prediction, as well as the ability to generate predictive insights from fully-encrypted data (see our academic work on trends prediction in financial markets [2–6], relevant patents [7, 8] and additional reviews [9–11]).

- **Innovation:** Focuses on the automatic modeling of short and medium-range behavioral patterns (days to weeks), detecting such signals before they become observable by any other available technology (see Dr. Altshuler's talk at *FirstMark*'s "Data-Driven New York" talk [12], or a large-scale real-time analysis of financial investments [13]).

- **Industry Validation:** *Endor.coin* is based on the technology developed by *Endor.com* – an MIT spin-off [14], financially backed by leading investors [15], working with Fortune 500 companies such as *Coca Cola* [16], *MasterCard* [17], *Walmart* and others. See Endor's product featured at *Finnovate 2017* [18].

- **Awards and Recognition:** *Endor* is a *Gartner Cool Vendor* [19], and was recognized by the *World Economic Forum* as a "Technological Pioneer" [20]. The research done by the project's team at MIT had led to winning several additional prizes such as the prestigious *DARPA Network Challenge* [21], and the *McKinsey Award* [22].

- **Team:** Spearheaded by a team of world experts in Blockchain, digital banking technologies and predictive analytics from MIT's Computer Science and Artificial Intelligence Lab, the MIT Media Lab and MIT Sloan School of Management. Prof. Pentland (co-founder), a member of the *U.S. National Academy of Engineering*, is one of the world's most-cited scientists [23], and was recently declared by Forbes as the "7 most powerful data scientists in the world" [24]. The team has collectively published hundreds of scientific papers, dozens of commercial patents, as well as 8 books dedicated to Blockchain, machine intelligence, and data privacy [25–32].

- **Reinventing Predictive Analytics:** After years where Artificial Intelligence and Machine Learning capabilities were reserved for deep-pocketed companies, *Endor.coin* offers individuals and small businesses access to superior tech, for a fraction of the cost. Predictive insights are based on the collective analysis of the contributed data, offered at a low cost, while allowing data owners to control the privacy of their data. The *Endor.coin* Protocol enables the integration of new data sources, as well as new prediction engines, creating a double network effect – the more players join, cost per prediction decreases, while increasing prediction accuracy.

- **Trustless, Fully Decentralized, Accountable and Censorship Resistance:** The *Endor.coin* Protocol is fully-decentralized, providing complete accountability for the prediction results. This prevents any manipulation or bias during of the predictions. In addition, the decentralized and open nature of the Protocol enables the support of any prediction, preventing censorship by any single point of authority.

Archived copy · cryp2nova.com

Chapter 2

Introduction

2.1 An Inefficient and Troubled Market

In our world there many are organizations that gather, possess and carefully maintain data. There are Data Scientists and Machine Learning experts, who can process data and build predictive models. There are also many people with a desire to predict the future (these range from high ranking executives, through to middle range marketing or product managers in large companies, and individuals who try to decide on the right timing to buy a plane ticket). Today, in order for the latter to be offered predictions for their questions, all three entities must co-exist in the same organization. This means that 99% of the predictions are being generated by, and for, stakeholders in large companies. Furthermore, this process today is highly expensive, as Data Scientists are rare and expensive, and the process of producing predictions often requires months of hard-work, dedicated to every single project. This sets a very high entry barrier (and price tag) for anyone who is interested in prediction.

2.2 Democratization Requires Decentralization and Data Separation

The democratization of Artificial Intelligence and Predictive Analytics, making it accessible for the 'everyman', requires the development of a new paradigm, that would meet the following requirements:

Separation: Apart from large corporations and research labs, there is almost no commercial, academic or non-for-profit organization that can sustain high-quality activities that combines *data curation*, *data science* and the generation of *semantic oriented questions*. Individuals, NGOs and small-to-medium businesses usually only focus on one of those dimensions – they either possess (or produce) carefully maintained data, employ strong (and expensive) data scientists, or are experts in 'asking the right questions'. Therefore, in order to open the bottleneck and make predictions accessible outside the Fortune-500 club, these three basic elements need to be inherently separated: A truly

Archived copy · cryp2nova.com

democratized prediction Protocol must allow data-providers to freely contribute data (either marked public or private), while allowing technology experts to contribute AI and Prediction engines (that would seamlessly be plugged in, and integrated with the Protocol), all of this – to allow the end-users to easily consume predictions that are based on these data sources, and prepared for them by these engines.

Accountability: in a data science department of a tech-giant accountability is not required as a key feature, as it is automatically provided by the fact that "everyone is working for the same boss" (e.g. the relevant C-level executive in charge of business intelligence, marketing prediction, and so on). In a democratized platform, where data, intelligence and computation is constantly being rented on an ad-hoc basis, accountability becomes essential, for guaranteeing 'fair play', and tuning the merit function of all stakeholders for the 'long terms' rather then encouraging short-term revenues.

Decentralization: There are two critical contributions to a decentralized prediction framework – one being engineering related, and the other revolves around censorship resistance and bias prevention. As demonstrated in many examples, a decentralized solution tend to be easier to scale as well as to extend. Adding more data sources, computational resources and different types of prediction engines – all greatly benefit from a decentralized solution. In addition, Decentralized architecture is the *only* one that guarantees that these predictions would not be subject to explicit censorship, or implicit one, that is executed through the generation of biased results, or arbitration of monetary resources. Furthermore, efficient decentralization is the key to emerging network effects, that push cost-per-user down, while constantly increasing prediction accuracy, with the increase of the numbers of participants (both data-providers, prediction engines, and prediction consumers).

2.3 Technological Gap

Unfortunately, although the benefits of Data Science democratization have been clear for quite some time, implementing such a framework in reality has been challenging, to say the least. The main reasons is technological – the prevailing science of today simply cannot support a "generic decentralized behavioral prediction" paradigm... The technologies that exist today, be it Neural Networks (or Deep Learning), Genetic Programming, Decision Forests, SVMs and so on – all require a massive amount of data sanitation, processing and understanding, before any 'real' machine learning work can commence. This is the source of the bottleneck that the industry faces today, the rise of skilled Data Scientists' salaries, and their scarcity. A detailed discussion on this topic appears in Section 3.1 and Appendix A.

Without a scientific breakthrough that can *automatically* digest *any data*, allowing non-professionals and professionals alike to ask *any predictive question* – the industry was detained to the existing paradigm, heavily bottlenecked by the pace a company hires new data scientists, and the 6-figures salaries it is willing to pay them.

Archived copy · cryp2nova.com

2.4 *Endor.coin* Protocol — Project Overview

In order to transcend these limitations, a new Science had to be developed. *Social Physics*, developed by *Endor.coin* founders Dr. Yaniv Altshuler and Prof. Alex "Sandy" Pentland, is a mathematical theory that efficiently models the way human crowds behave. Through a set of mathematical equations that are shown to emerge in behavioral data sources, the *Social Physics* theory enables the automatic transformation of any behavioral data source to a set of behavioral clusters. This requires no cleaning, pre-processing, or understanding of the semantics of the data (or the questions to be asked). This collection of behavioral clusters is known as the "*Knowledge Sphere*".

The *Endor.coin* Protocol is based on the fact that when behavioral data is being handled in a canonic representation of behavioral clusters, the traditional process of Data Science can (finally) be broken down into it basic components, allocating each of them, in a decentralized way, to different executors. Following is the basic outline and main components of the *Endor.coin* Protocol:

Canonic Data Representation: Every data that is contributed to the *Endor.coin* network gets transformed to the "*Knowledge Sphere*" canonical representation. This can be done by the various Predictions Engines (see below), and the execution cost is paid by the engines. Once data undergoes this transformation the various behavioral clusters extracted from it can then be bundled together with clusters from other types of data, resulting in an efficient and automatic prediction process (see more details in Section 3.1 and Appendix A).

Separation of Data Providers: As data gets transformed into the canonic "*Knowledge Sphere*" representation, data providers are no longer required to actively take part in the later phases of the analysis. This enables data-owners to integrate their data (in full or in part) with the *Endor.coin* network, acting as an autonomous stakeholder in the ecosystem – focusing on maintaining the quality of their data, controlling who will have access to which part of it, and benefiting financially from future value it provides.

Separation of Prediction Engines: A known secret among data scientists is that around 90% of the time spent in a data science project is spent on data sanitation and pre-processing. Due to the revolutionary aspect *Social Physics* that for the first time automates these steps – various prediction engines can finally be seamlessly connected to data sources of various types. The only thing required for a provider of a prediction engine in order to connect to the *Endor.coin* network is to support the *Endor.coin* Protocol – defined as the ability to digest datasets (selectively, as defined by the engine), and provide output in the form of a "*Knowledge Sphere*" (see complete specification and API code in Appendix D).

Decentralized Execution: Using Blockchain. data-providers can donate data (stored on AWS) and accessible using the *Endor.coin* Protocol. Extraction of behavioral clusters is done in a decentralized way by the various Prediction Engines. Queries are being

Archived copy · cryp2nova.com

triggered by end-users by requesting the *Endor.coin* smart contract to issue a certain prediction (for *EDR* fee). For each prediction the best behavioral cluster is chosen by the *Endor.coin* prediction code (can be freely accessible on the project's GIT account [33]). Funds arbitration is being taken care of by the smart contract among data contributors and prediction engines, according to the clusters chosen for that prediction. This arbitration optimizes the quality of prediction, and unbiased results.

Data Sovereignty: Every data element that is contributed to the *Endor.coin* network can be flagged as either "public" or "private" (the same data source may contain certain columns marked public and others kept private). Public data sources are accessible to every prediction engine, being a source of behavioral clusters for any future predictions. In return, their providers are being compensated with *EDR* tokens when they are selected by the Protocol. Private data elements are still accessible by the various prediction engines, albeit in an encrypted way. The clusters extracted from such sources however can be selected as a source for predictions only when the user who requested these predictions provides the key for the data. See more details in Section 5.4.

Accountability and Censorship Resistance: Using Blockchain, predictions are stored indefinitely, and can be accessible by anyone who is interested in deducing the reputation of the platform, the data that was used for it, or the prediction engine that analyzed it. In addition, as the *Endor.coin* Protocol contains an open-source prediction code [33] that is in charge of selecting the behavioral clusters that are used for the generation of each prediction (and the arbitration of the funds paid for it), it is guaranteed to be free of bias, optimizing accuracy only. See more details in Section 4.2 and 4.1.

Prediction Efficiency for All: Ultimately, the *Endor.coin* Protocol enables end-users to obtain superior predictions at a low cost. This is based on the automatization of the process (saving the need to employ expensive full time data scientists), and the ability of *Social Physics* to seamlessly 'fuse' together various types of behavioral data sources. This means that even large commercial customers, such as Coca Cola, would be able to immediately benefit from migrating to the *Endor.coin* network – by flagging their data "fully private" they could be offered predictive insights that are based on the fusion of the proprietary data, fused together with public data that was contributed to the system, and pay a significantly lower fee that the alternative cost of obtaining this data on their own, and analyzing it. This also offers a positive network effect – lowering the cost as more users and data providers join the *Endor.coin* network (see more details in Section 4.4).

Archived copy · cryp2nova.com

Chapter 3

Democratizing the Prediction Science

3.1 Social Physics – a New Science from *MIT*

Social Physics is a revolutionary new science which uses big data analysis and the mathematical laws of biology to understand the behavior of human crowds, enabling Endor to overcome traditional Machine Learning limitations. This new science originated at MIT through research by Prof. Alex "Sandy" Pentland and Dr. Yaniv Altshuler. It was further developed by Endor using proprietary technology, resulting in a powerful engine that is able to explain and predict any sort of human behavior, even when the behavior is rapidly changing and evolving.

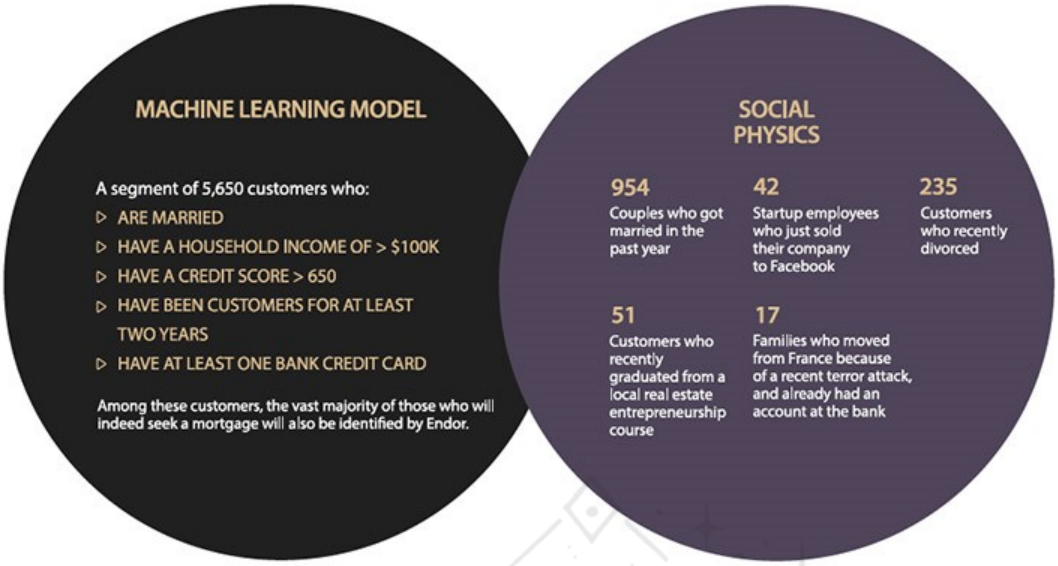
Simply put, Social Physics is based on the premise that all event-data representing human activity (e.g. phone call records, credit card purchases, taxi rides, web activity) is guaranteed to contain a special set of human activity patterns that are embedded within that data. These mathematical invariances, which are common to all human data-types, across all demographics, can then serve as a filter for detecting emerging behavioral patterns before they can be observed by any other technique.

Illustrating the Power of Social Physics: Imagine that the marketing department of a large bank constantly calls customers who are potentially in need of a loan in the near future. The revenues of the department are directly derived from the portion of the customers actually responding positively to the offer. As the direct-marketing costs involved in this ongoing campaign are significant, it is crucial to contact the right customers, at the right time: too late, and they might have already taken a loan from another source. Too soon – and the need has not materialized yet.

For this, the bank may consider two tools to predict who these customers are: A Machine Learning model developed in-house by the bank's data science team; and Endor's engine. Here is a simplified representation of what each tool recommended:

The group of customers that were detected by the Machine Learning model comprised of customers who will respond positively to a marketing offer by the bank (e.g. True Positives), as well as of customers who will not (e.g. False Positives). For example, let's assume that

Archived copy · cryp2nova.com



the True Positives are 10% of the model's results. Extensive experiments show that we can expect the vast majority of those 10% to also be detected by the Endor Social Physics engine, with two main differences: (a) many of the False Positives of the Machine Learning model will not be reported by the Endor engine; (b) Endor's results will contain many additional True Positives, not detected by the traditional model. The result was a significant improvement in the sales efforts, thanks to Endor's better precision / recall trade-off.

How? Detecting Temporal Patterns: Human reality is composed of many small temporary events and changes. Social Physics incorporates the underlying dynamics of human behavior and is therefore better equipped to uncover small groups in the population who are likely to behave in a certain way due to recent changes in their social environments. Social Physics is therefore uniquely capable of identifying dynamic signals in human behavior data: This is because without the aid of Social Physics such signals lack any sort of statistical significance, rendering them indistinguishable from noise for traditional Machine Learning and Deep Learning methods.

Machine learning and Deep Learning vs. Social Physics – Which one is better for which purpose? When solving a business query using data science and big data analytics tools, both Machine Learning and Social Physics are viable options. The table below can help identify the appropriate tool, based on its attributes.

Why Social Physics? Rooted back in the 70s the various mathematical and statistical Machine Learning techniques were historically developed for 'static problems', such as image processing and text recognition. Such problems are dominated by a relatively small number of relatively stable 'signals'. A trained text recognition model would achieve similar perfor-

Archived copy · cryp2nova.com

	MACHINE LEARNING IS BETTER FOR...	SOCIAL PHYSICS IS BETTER FOR...	WHY?
Type of data	Mechanical / physical - driven data: **Examples:** • Monitoring an oil drill pump's control data to predict malfunction • Face recognition 	Human behavior data: **Example:** Analyzing financial transactions to predict who will purchase a premium service 	Human behavior is erratic, unpredictable, noisy, complex, and dynamic. Mathematically speaking, human behavior is dominated by a large number of "temporal" signals, each affecting a small group of individuals. Hence, it is very hard to "learn" human data producing consistent, stable models representing it. Endor uses Social Physics to detect such temporal signals, and therefore is specifically tailored to human-based data.

mance when processing the handwritten text of a 2016 MIT student and when analyzing Albert Einstein's personal letters. Similarly, neither *Siri* nor Google's speech recognition engine would find it difficult to transcribed a high-quality recording of J.F.K's famous "ich bin ein Berliner" speech.

Human behavior, however, is a different story. Governed by a multitude of 'dynamic signals' it is highly dynamic and highly 'fractured'. A traditional Machine Learning model trained to detect Millennials from credit-card purchases rapidly deteriorates in accuracy over time, requiring constant maintenance by a skilled expert continuously incorporating new semantics knowledge into it. As Millennials' behavior is subject to frequent (and constantly changing) trends, locating this in the data dictates not only a constant re-training of the model, but also the frequent development of new features intended to detect these trends (i.e. complex aggregative behavioral properties that are not part of the raw data). This can only be done through the combined work of a semantics domain expert working side by side with a data expert.



Following are the primary advantages of using Social Physics as a tool for behavioral

Archived copy · cryp2nova.com

prediction in human data:

	TRADITIONAL MACHINE LEARNING	DEEP LEARNING (W.O. SOCIAL PHYSICS)	SOCIAL PHYSICS	WHY?
Small data Sets	Able to analyze small data sets, but requires expert data scientists and is a time consuming process	Requires large amounts of data for every question	Requires very little data to answer any question related to human behavior. The results are generated automatically (no need for data scientists to be involved)	Endor does not require "big data" to generate results, due to the fact that Social Physics already incorporates the underlying dynamics of human behavior data. Hence, even with very small data sets it can immediately produce accurate predictions and actionable signals.

3.2 1^{st} Phase : Endor.com – Automatic Prediction Engine for Enterprises

As discussed in the previous sections, the state of Data Science today, as well as the available Machine Learning technologies, dictate that the use of such capabilities was preserved for deep pocketed tech giants. Companies who can afford a six-digit annual salary for these growing-in-rareness experts. However, even for those companies Data Science and Predictive Analytics were far from being a commodity. On the contrary – a project aimed at producing reliable predictions would typically block a team of 2-3 experts for 4-6 iterations of a few weeks each, and after the first model is produced – would usually require a constant maintenance. Industry standards therefore estimate the overall cost of an average prediction project at approximately $1.5 Million. This means that:

- Most companies cannot afford to fund more than a handful of projects.
- In order to have a positive ROI projects today must show huge margins, and focus only on the most central business aspects.

Archived copy · cryp2nova.com

	TRADITIONAL MACHINE LEARNING	DEEP LEARNING (W.O. SOCIAL PHYSICS)	SOCIAL PHYSICS	WHY?
Features vs. Raw data	Requires a skilled data scientist and/or a domain expert in order to define and select the right features representation of the raw data	Does not need features and can process raw data, but is limited to an extremely narrow type of problems (*)	Does not need features and can process raw data, for any type of predictive problem involving human behavior.	Machine Learning requires a long, often manual, process of transforming raw data into meaningful features. This is typically done from scratch for every problem, and for any new type of data. Although Deep Learning deals with feature crafting automatically, it still requires large amounts of data, and data requirements further increase with the complexity of the problem. Therefore, it is limited to "simple behaviors". In addition, Deep Learning is usually also confined to "static problems," as Deep Learning dynamics require a vast amount of data that is usually unavailable at the typical company. Social Physics automatically transforms any raw human behavior data into a canonical form of human behavioral clusters. Using this canonical representation, Endor can contend with all data types and all questions, regardless of data size, and to generate a unified human-behavior data set which then uses the power of Deep Learning to answer any predictive question.

	TRADITIONAL MACHINE LEARNING	DEEP LEARNING (W.O. SOCIAL PHYSICS)	SOCIAL PHYSICS	WHY?
Users and expertise needed	Machine Learning experts, usually with the assistance of domain experts, who help craft semantic features	Deep Learning experts	Business users. All you need is to provide an example of "people you want to find more of".	Machine Learning requires "learning" the underlying normal behavior of a large data set, or leveraging prior domain expertise. Endor already incorporates the underlying dynamic of human behavior data.
Pace of data change	Limited to slow changing data. Changes in the data further requires continuous intervention of domain experts, in order to tweak the features.	While it can deal with dynamics, it is limited to slow-changing data (a harsh limitation when it comes to human behavior data!)	Can easily analyze fast-changing data sources, and does so automatically (no need for domain experts).	Endor's engine is specifically tailored to human behavior data, and hence inherently works on data of a dynamic nature. As Social Physics is a set of mathematical invariances that are embedded in any human dataset, it can even detect signals representing extremely short time segments. In other words, it is able to identify emerging changes before they are observable by other techniques.
Scope of Analysis	Specific / Limited		Broad / Any question about human behavior	For Machine Learning, the learning process must be repeated for each dataset and question, since the automatically-selected model features need to be re-learned. Social Physics is based on underlying human behavior principles which are not question-specific.

Archived copy · cryp2nova.com

	TRADITIONAL MACHINE LEARNING	DEEP LEARNING (W.O. SOCIAL PHYSICS)	SOCIAL PHYSICS	WHY?
Data cleaning	Machine Learning is highly susceptible to noises and gaps in the data. Requires a long and expensive data-cleaning process Deep Learning often requires a careful process of transforming the data to a format acceptable by the Deep Learning tool		No data-cleaning required	Both Machine Learning and Deep Learning use data-driven mathematical patterns in order to deduce rules, extract signals, and produce predictions. This requires a delicate process of data cleaning. Social Physics, on the other hand, uses external patterns – mathematical invariances that are known to be embedded in every human behavioral dataset. This significantly reduces the effect of data noises. In addition, Social Physics transforms raw data into behavioral clusters, which further reduces the effect of data gaps and noise (mostly filtered out automatically).
Iterations and Tweaking	Each tweak in the data or the definition of the problem requires the joint work of a both domain expert (business user or analyst) and a Machine Learning \ Deep Learning expert. Each iteration can take weeks, and a typical project consists of at least 4-6 iterations.		Interface is designed for use by a business user or an analyst that is able to revise and modify the prediction query. The results are then be adapted automatically to the new definition.	Endor's engine is specifically tailored to human behavior data, and hence inherently works on data of a dynamic nature. As Social Physics is a set of mathematical invariances that are embedded in any human dataset, it can even detect signals representing extremely short time segments. In other words, it is able to identify emerging changes before they are observable by other techniques.

- Companies constantly have to prioritize, aiming the super-expensive weapon of Data Science only on the projects they think would result in (a) technological success + (b) high business returns.

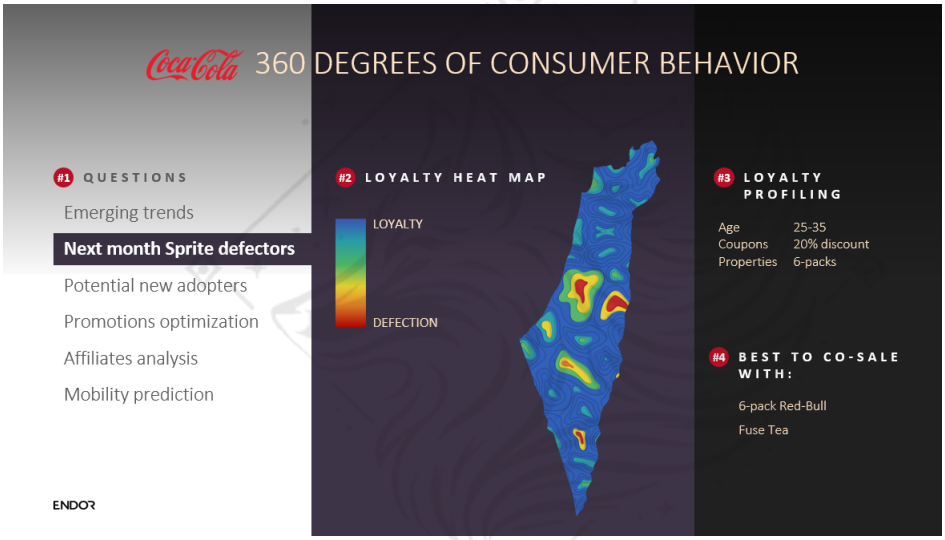
Endor, an MIT spin-off, was established around the novel science of Social Physics, in a bid to solve this problem, by disrupting the way Data Science is being perceived today. Based on 3 years of research at MIT, and an additional 3 years of development by an elite team of researchers and engineers, Endor has developed the world's first ever fully-automatic "Data Science as a Service" engine, that allows companies to onboard any behavioral data they possess, and after a quick integration (that typically requires a mere few hours) start asking an *unlimited* amount of prediction questions regarding the future behavior of any 'object' (users, products, coupons, locations, etc.) contained in the data.

For an annual cost of less than $1 Million, companies were given the opportunity to ask dozens of predictive questions, getting quick results within minutes. This is not just a linear improvement, but rather a paradigm change, as executing a 'prediction project' became as easy as 'Googling', virtually rendering the prioritization of potential projects unnecessary, as any decision could have now become prediction-oriented.

Example I – Coca Cola Joint Study: In a recent collaboration between Endor and Coca-Cola [16] the ability of Social Physics to provide accurate predictions regarding a wide variety of consumer behaviors was demonstrated. These included brand defection and loy-

Archived copy · cryp2nova.com

alty, adoption of new products, response to marketing campaigns and others. Using millions of point-of-sale transactions as its raw input (representing a 3-months period), Endor's Social Physics engine detected nearly 20 million 'correlated anomalies', each representing a single real-world behavioral group. Whereas the exact meaning of each group is unknown, the groups are then used for 'behavioral extrapolation': Given a sample-set (e.g. early-adopters of a new product), the system uses the collection of behavioral groups to find lookalikes – users who are *behaviorally similar* to the members of the sample-set (e.g. users likely to experience the new product soon). Using this methodology 15 different predictive questions were asked, each yielding a prediction report on demand. Reports high accuracy was demonstrated using out-of-sample data kept for validation.



"Social Physics is about behavioral analysis in big data, but it takes it to a completely new level. We were very fortunate to find Endor and work with it."

Dr. Alan Boehme, CTO, the Coca Cola Company

Example II – Automatic Analysis of Tweets: In a recent test 15 million Tweets' meta-data were provided to the Endor engine as raw-data for analysis. In addition, the customer revealed the identity of 50 Twitter accounts known to be *ISIS* activists that were contained in the input data, and tested Endor's ability to detect an additional 74 accounts that were hidden within the data. Endor's engine completed the task on a single laptop in only 24 minutes (measured from the time the raw data was introduced into the system until the final results were available), identifying 80 Twitter accounts as 'look-a-likes' to the provided example, 45 of which (56%) turned out to be part of the list of the 74 hidden accounts. Importantly, this provided an extremely low false alarm rate (35 False Positive

Archived copy · cryp2nova.com

results), so that the customer could easily afford to have human experts investigate the identified targets.

> "A revolutionary concept and a truly technological breakthrough. The results they presented are unmatched by any competing tool."
>
> *CIO of Israeli Intelligence Corps*

The Need for an Ecosystem-Enabling Decentralized Protocol: Using Social Physics and a strong team of professionals, Endor's engine has demonstrated how a large Fortune-500 company can pay significantly less, and get significantly more. However, this product was made available mainly for large banks and retailers, and it still comes at a cost ranging from $250K to $1.2M for an annual license. This is of course not a deal that is applicable to most 'long tail' businesses, not to mention individuals.

Therefore, the need for another solution became apparent. One that would allow *anyone* to benefit from the new technology of Social Physics, and to do so at a reasonable price. Such a solution must therefore:

- Be self-reliant, comprised of the resources made available by its ad-hoc participants.
- Generate a strong network effect, incentivizing participants to continuously join it, becoming better and cheaper as it grows.
- Preserve fairness, and be inherently unbiased and trustless. This requirement is unlike other services, as prediction is expected to have what can be called 'an external effect', caused by the action of the person or organization relying on it.

With this concept in mind, the Endor team is proud to present the next step in its revolution, the *Endor.coin* Protocol!

3.3 2^{nd} Phase : *Endor.coin* Protocol – Data Science for the Masses

Following the successful industry implementation of Social Physics inspired behavioral analysis by Endor, the *Endor.coin* Protocol was created with the aim of bringing this ability to the long-tail business, as well as professional individuals. Reinventing predictive analytics, the *Endor.coin* Protocol democratizes Artificial Intelligence for behavioral prediction – enabling the creation of an ecosystem that makes it accessible to all. Furthermore, the Protocol's fully-decentralized nature guarantees trustless, censorship-resistance and accountability. For the first time, behavioral predictions become available for all, for an affordable fee, in a secured framework, free from potential manipulation of tech-giants who control data and technology today.

Archived copy · cryp2nova.com

3.3.1 Definitions

Raw Data: the *Endor.coin* Protocol supports any time related, or transactional data source (e.g. Call Data Records, ERC20 Blockchain, in-app purchases, and so on). Data providers must indicate the column that is the basis for predictions. For example, taking the ERC20 it is possible to predict and discover addresses of interesting future behaviors, or tokens themselves. Tweets data can be used to find interesting Twitter IDs, or alternatively – interesting locations or hashtags.

Processed Data: Denotes data that has undergone extraction of behavioral clusters and was transformed to the Social Physics canonical representation as a 'Knowledge sphere'.

Knowledge Sphere: Denotes the collection of behavioral clusters extracted from one or more raw data sources. The *Endor.coin* Protocol enables the split or union of any number os Knowledge Sphere. As different behavioral clusters are of different relevant for different types of predictions, for a given prediction the *Endor.coin* Protocol selects the most relevant clusters at any given time, and generates the 'Knowledge Sphere' that is then used in order to generate the actual prediction.

Predictions: The *Endor.coin* Protocol support any question that can be phrased in the form of "Rank group X by its likelihood to be behaviorally similar to group Y". For example, if group X contains all the ERC20 tokens and group Y contains tokens that have recently increased volume dramatically, then the prediction result would contain the list of ERC20 tokens, tanked such that the top contains tokens most statistically likely to increase volume in the near future, and the bottom containing tokens least likely to display this behavior.

3.3.2 Data Providers

The *Endor.coin* Protocol supports the integration of any behavioral, time-associated structured data. Data onboarding is done using a simple API call, allowing the data owner to contribute data while controlling which columns remain private and which become accessible to the public analysis. Data defined by *Endor.coin* users to remain private is still automatically integrated with public data-streams, producing high-quality predictive insights harvested from the fusion of privacy and public data. Using the notion of Social Physics' *Knowledge Sphere* (see Appendix D) data integration is automatic and friendly – requiring no cleaning or data-preparation.

Data providers are required to pay *EDR* tokens for the analysis of their data, to the providers of prediction engines. They are, in turn, being rewarded *EDR* tokens when insights derived from their data are being used for predictions. This incentivizes the contribution and maintenance of high-quality data streams.

Archived copy · cryp2nova.com

3.3.3 Prediction Engines

The *Endor.coin* Protocol defines a 'prediction language' that is based on the projection of the data to a feature space of partially overlapping behavioral clusters. The current Endor engine would become the first prediction engine to be plugged in to that network, in order to make it usable immediately after launch. However, *Endor.coin* would facilitate, assist and fund the development of new prediction engines, aiming for the creation of an ecosystem that is comprised of multiple types of engines, providing complementary capabilities, boosting performance accuracy and increasing reliability. The growing number of prediction engines would also provide a guarantee for the unbias nature of the results, as predictions would be created based on the most relevant clusters, automatically chosen by the *Endor.coin* open source Protocol from the collection of clusters, extracted by the various prediction engines.

3.3.4 Pre-Defined Predictions and Request for Predictions (RFPs)

The *Endor.coin* Protocol will be launched with a large catalogue containing a variety of pre-defined predictions. These predictions would be accessible for purchase using the platform's *EDR* token. *Endor.coin* would then expand the selection of predictions it caters by allowing users to send *Requests For Predictions* (RFP) – suggesting new types of predictions, that would be implemented and become available for purchase.

This method of gradually expanding the predictions supported by the platform would utilize the wisdom of the crowd (as manifested by the requests of the Protocol users) to optimize the selection of newly supported predictions.

3.3.5 Private-Data Analysis and Self-Serve API

The later releases of the *Endor.coin* Protocol would include support in a complete 'do-it-yourself' API for advanced users: tech-savvy users and professional Data Scientists would be able to use a self-serve interface to easily onboard proprietary data and create on their own new types of predictions. These predictions can be defined as private, or – be shared with the public (rewarding the prediction develop with *EDR* tokens, if they become widely used.

3.4 Roadmap

In today's general software market, our approach can be compared to offerings such as platform-as-a-service or more recently Blockchain-as-a-service. The *Endor.coin* tokens (or *EDR*) will be used to power transactions on the platform. *EDR* serve as a key or software license, and more tokens can be used over time to increase performance and scale via a community of developers that will be enticed to expand the areas of applications. In addition, a dashboard will be created for our administrators to distribute tokens, monitor usage and purchase more tokens as necessary.

Archived copy · cryp2nova.com

The next step in the *Endor.coin* evolution will be to expand our ecosystem organically by an ever growing community of Catalysts. Everyone can provide RFPs ("request for prediction"), paid for in *EDR* , and the challenge of addressing the respective RFP can be undertaken by any Catalyst who will be rewarded from the respective payments through a smart contract for contributing to expand our range of predictions which we will embed as newly supported queries by *Endor.coin* (further accessible to everyone, using *EDR*).

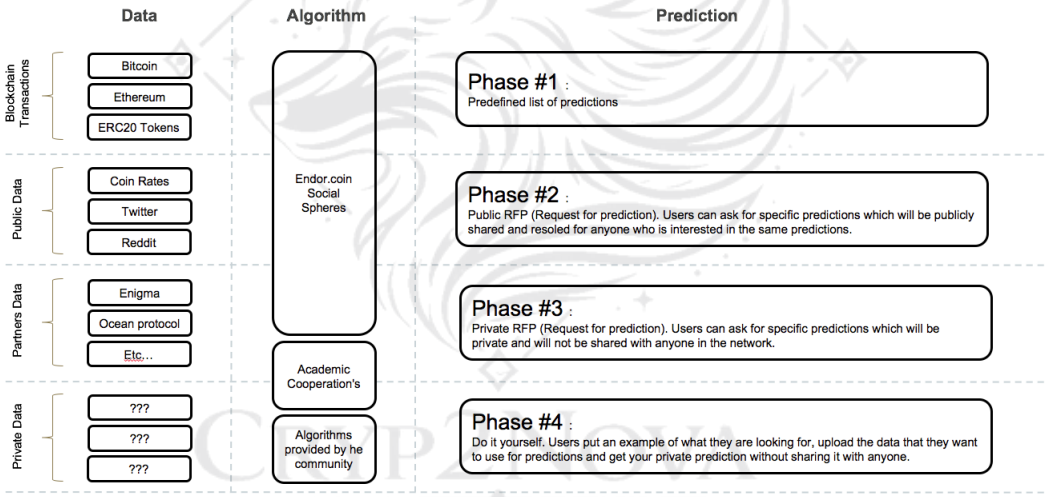
In this second phase *Endor.coin* will become a modular App platform on which Catalysts can expand the prediction domain through the power of Blockchain. To enable this *Endor.coin* is endowed by an open ended stream of micropayments to authors of reusable software components that can be perpetually combined and recombined to create an ever expanding library of useful, highly customizable query Apps. Catalysts are entitled to a Micro Use License for each component they add to *Endor.coin* platform. End users install the Apps of their choice. The license to be paid is the sum of all the Micro Use Licenses of the components used by that App. Through smart contracts *Endor.coin* is responsible for charging end users and distributing the payments to the respective Catalysts involved.

With time *Endor.coin* will become the premier development platform for entrepreneurial coders and enterprises looking to build data-rich web and mobile products on Blockchain thus democratizing Machine Learning which is so far accessible only to those highly skilled. Furthermore, existing Blockchains that today do not support the volume of transactions needed to efficiently execute data-analytics algorithms will be provided these capabilities, through the help of an ever growing community of Catalysts who will build applications using the engine continuously expanding the domain of queries.

Our vision is to create a fully automatic, trust-less decentralized predictions infrastructure, that will be fully transparent to end users. Incentives are paid in *EDR* tokens and applications are essentially a set of plugins. A positive feedback loop is foreseen: the more applications built, the more plugins are added to the system and more unique components are ready to be reused. This contributes to mutually self-reinforcing network effects across the *Endor.coin* ecosystem.

Ultimately we will deliver a Platinum platform for high end customers, where professional experts and businesses can submit more complex fine grained RFPs. To deliver on such complex queries we will expand our data access by rewarding providers with *EDR* as follows: when a Catalyst requests data from an entity (timed limited) the respective entity proposes a price (timed limited). If the Catalyst agrees to the payment the entity sends a link to the data encrypted with the Catalyst public key (at which point the money is being disbursed via a smart contract). Reviews can be published by both Catalysts and data providers (for future reputation). *EDR* will be generated from the arbitrage between the data providers payment and the funding received from the platinum application developer.

Archived copy · cryp2nova.com



Archived copy · cryp2nova.com

Chapter 4

Trustless, Censorship-Resistant and Accountable

4.1 Accountability and Authenticity for Artificial Intelligence

In his article "*Why you want Blockchain-based AI, even if you don't know it yet*" [34], the author Jeremy Epstein shares with his readers a conversation he and his young daughter had with Amazon's AI gadget Alexa. This conversation revolved around the topic of 'Network Neutrality', during which Alexa shared with its listeners an abundance of relevant information. Alas, given Epstein's familiarity with the topic, he had surprisingly discovered the information to be, albeit accurate, potentially one-sided. The experience highlights some of the risks of the AI-powered future into which we are hurtling at warp speed. It is also a reminder that big companies, such as Amazon, have traditionally had big advantages when it comes to big data and AI.

Quoting Epstein:

> If the race is about gathering, storing, and analyzing as much data as possible, then who is in the pole position to win? That's right, the FANGs in the U.S. (Facebook, Apple, Netflix, Google), the BATs in China (Baidu, Alibaba, Tencent), and the wealthy Fortune 1000 or so multinational corporations.
>
> They are the only ones with the reach and capital to get more data, store it, analyze it, and build AI models on top of it. What's more, they are the only ones who can offer starting salaries in the $300,000 to $500,000 range and top-tier salaries that extend into to seven and eight digits. Your son or daughter may not make it to the NBA or NFL, but become a top AI scientist and you're doing great.
>
> The net effect of all of this is that the rich become even richer and more powerful and the barriers to innovation become even higher.

Archived copy · cryp2nova.com

It is not only innovation that suffers, however. The closed nature of big-company AI means society must put its trust in "black boxes."

Providing "*prediction authenticity*" therefore relies on the availability of an infrastructure that would provide all of the following:

- **Accountability:** providers of predictions can prove, in retrospect, that their predictions were correct. Consumers of predictions can reliably deduce in retrospect the efficiency (or lack of) of any prediction provider to produce accurate predictions of any type. In other words, reputation should be impossible to manipulate.
- **Authenticity:** providers of predictions are who they say they are. Impersonation for the cause of misleading prediction consumers should be impossible.
- **No Bias:** there should be fair competition among prediction providers, that would utilize market forces to incentivize accuracy rather than implicit gain achieved through biased predictions.
- **Accuracy:** predictions should be accurate enough, generally speaking, to provide monetary inflow to the network, that would close the positive feedback loop with respect to the previous three requirements.

The *Endor.coin* Protocol provides an infrastructure that would satisfy all of the above requirements:

- ***Endor.coin* Protocol Accountability and Authenticity:** using Blockchain, predictions are stored indefinitely, and are accessible to everyone. When time-sensitive predictions are only sent to consumers, they are also stored in an encrypted-yet-public version, with a key that is released after a certain period of time.
- ***Endor.coin* Protocol Zero Bias:** the Protocol selects for each prediction the most relevant behavioral clusters (regardless of the analytics engine, or data source). Therefore, as the Protocol is a separated entity, uncontrolled by data providers, as well as analytics engines providers, bias is inherently prevented.
- **Accuracy:** *Endor.coin* is based on MIT's *Social Physics* technology, extensively proven by the industry to consistently provide accurate predictions for a large variety of use-cases.

4.2 Censorship-Resistance through a Decentralized Protocol

Censorship is a tricky business. When exists in its explicit manifestation it is easy to detect, and therefore – to bypass, through the use of "low level" technological solutions such as IP-proxies, and so on. Implicit censorship on the other hand is a different issue altogether. It is

Archived copy · cryp2nova.com

known that companies such as Google and Facebook block certain types of search queries. Many times this is done for the right moral or legal reasons, but – how can we be assured that this is always the case? Can we "force" Google to provide a relevant webpage, if it has decided that we "should not have an easy access to it"? Can we "force" Alexa to comply with some of our requests it would deem inappropriate? The answer is simple – no.

Prediction in this context is very similar to Search, in the sense that it is subject to censorship by the operating entity, or the regulatory agencies they are subject to. If we look at how AI and predictive analytics work, they have three essential layers with respect to potential censorship:

- **Data Repository:** guaranteeing the integrity, completeness and security of the data (are the inputs accurate and reliable, and can they be manipulated or stolen?)
- **The Algorithm/Machine Learning Engine:** making sure predictions are not inspected by a centralized authority, and that all prediction requests are being executed fairly, with quality-of-service considerations unrelated to the topic of prediction.
- **Queries Interface:** reliably representing the output of the prediction query, effectively capturing new data, and having any limitations on supported predictions being unrelated to topic of prediction.

If one is going to trust one's decision-making to a centralized prediction source, one implicitly assumes with absolute confidence that all the above requirements are met. In a centralized, closed model prediction scheme, when one is asked to trust in each layer without knowing what is going on behind the scenes, such confidence is difficult to justify (if not plain gullibility).

The *Endor.coin* Protocol inherently provides these assertions by allowing any prediction to be executed, in a fully decentralized, trustless way. Once data is contributed and onboarded to the network, any relevant prediction can be executed, with the *Endor.coin* Protocol optimizing it automatically.

4.3 Privacy-Preserving Data Analytics: Using Encrypted Data

The use of the novel new science of Social Physics offers a key advantage when it comes to the formation of a democratized ecosystem for data analysis and prediction: it allows computing across data silos without compromising data privacy or integrity, because the Social Physics computation can be done on encrypted data. Specifically, the unique way in which *Endor.coin* analyzes data enables data contributors to monetize the analytics of their data in encrypted form without exposing the data itself. Instead, the system enables the extraction of 'implicit behavioral clusters' (mapped onto a hashed or encrypted data space) that will subsequently be used as the basis for 'look-a-like prediction' – resulting in

Archived copy · cryp2nova.com

an accurate behavioral prediction that is oblivious to the nature of the source data, and the semantics of the prediction question.

Similar technology is already used by *Endor* today, working with large banks and financial customers, allowing the latter to onboard their data, in a fully-encrypted version, rendering the compromise of the data privacy and integrity impossible.

4.4 Network Effects

The full-decentralization of the *Endor.coin* Protocol results among others in several network effects, that increase its merits as its usage widens. Following are the main expected thrusts that will be triggered by an increased adoption of the Protocol:

Adding users decreases the cost for each of them: Unlike large commercial customers, individuals and small businesses require predictions that are based mainly on public datasets. The *Endor.coin* Protocol enables such datasets to be analyzed *once* – generating a collection of behavioral clusters known as a "Knowledge Sphere". This data structure can then support *all* of the prediction use-cases that are based on this data source, requiring end-users to pay a small fee that encapsulates their "personalization factor" – the delta between the extraction of the behavioral clusters, and the specific use-case they are interested in.

In other words, for a given dataset (i.e. the Bitcoin Blockchain), and for a specific use-case, the cost is calculated as follows:

- **Knowledge Sphere Calculation:** Consists of roughly 99% of the overall cost, and is calculated once. Resources required for this task are amortized among all users of this data-steam.
- **Prediction Personalization Factor:** Consists of roughly 1% of the overall cost, and is calculated (and paid for) for each user. In addition, for similar predictions made by different users, the *Endor.coin* Protocol automatically reuses insights derived from the prediction that is calculated first, during the calculation of the later ones, further improving prediction accuracy for the same cost.

This means that the governing cost equation of a network of N end-users would be:

$$\frac{0.99 \times C}{N} + 0.01 \times C$$

for C denoting the cost of the same prediction by a single commercial player. This ultimately means that as the number of users N increases, the accuracy of the predictions increases, while its cost aspires to approximately 1% of the cost of the same system, used by a large commercial customer.

Ultimately, the more people asking questions – the better the results they receive, and for a lower cost, as the vast majority of the cost is divided across all active users.

Archived copy · cryp2nova.com

Adding data-providers increases accuracy: Per the definitions of the *Endor.coin* Protocol, data providers are required to fund the execution of their data using *EDR* tokens, for creating the "Knowledge Sphere" that is based on the behavioral clusters that are extracted from the data. This initial execution cost is then repaid to the data-providers in part, equally, or with significant interest, by the end-users – all according to the quality of the data, and its contribution to the various prediction queries. This funds arbitration is being taken care of by the *Endor.coin* Protocol, whch is able to access the entire collection of behavioral clusters detected by the various analytic engine, on the various datasets, and select the top clusters of highest-relevance. Execution tokens are then delivered to the providers of these datasets, pro-rata, with respect to their contribution to the final prediction.

The result of this mechanism is that owners of high-quality data are incentivized to continue supporting their data sources (and even further increase their quality and availability), while providers of data of poor quality are being rooted out, simply because their costs are not being repaid. This economy utilizes the market forces to automatically guarantee that the available data-sources to the prediction Protocol remain of the highest possible fit. New data-providers can therefore only increase overall prediction accuracy, without increasing their overall costs.

Adding prediction-engines increases prediction efficiency: The first phase of the *Endor.coin* project relies on the Endor prediction engine to act as the first provider of behavioral clusters extraction. However, in time additional prediction engines are expected to support the *Endor.coin* Protocol. The introduction of such new prediction engines is expected to have a significant positive effect on both data-providers as well as end-users: engines that utilize different technologies are expected to produce different types of clusters from the same data source. This means that as the variety of technologies used for predictions engines that support the *Endor.coin* Protocol increases, so will the variety of clusters which will become available for the Protocol to select from, when new predictions are being requested.

This is expected to have three main effects:

- **Improved Accuracy, and Increased Support for New Predictions:** As new types of clusters become available, the *Endor.coin* Protocol will be able to better select clusters to be used for the generation of each requested prediction. For existing predictions this would result in an increased accuracy (due to the availability of "orthogonal insights", derived from the use of a clustering technology). However, this would also imply that new predictions, that until that point were not supported by the system due to suboptimal accuracy, would now become cost-effective to execute, increasing the variety of predictions supported by the system, and subsequently further boosting the overall accuracy, due to the sharing of insights among predictions executed at adjacent times.

- **Reduced Cost per Prediction:** The increase in the amount of predictions supported by the system would also result in an increase in the number of end-users who pay for

Archived copy · cryp2nova.com

predictions – immediately increasing the overall pool of available *EDR* tokens paid for the initial data analysis, and subsequently increasing the amount of funds received by the data-providers, while decreasing the cost-per-prediction for the end-users.

- **Economic Sustainability of New Data Sources:** Finally, the availability of new clustering technologies may also render the contribution of certain types of data sources economically feasible, in cases where these new technologies are better compatible for the analysis of such data sources then the available ones. In such cases, such data sources would suddenly become a viable source of information for the *Endor.coin* Protocol, repaying the cost of their initial integration to the platform, and subsequently also increasing the amount of available data sources, with the benefits that are derived from it.

CRYP2NOVA

Archived copy · cryp2nova.com

Chapter 5

Enabling an Ecosystem

The *Endor.coin* project aims towards the establishment of a multi-faceted ecosystem, creating a synergic collaboration between data owners, developers, data science professionals, small businesses, and individual users. By enabling each of these players to contribute their assets (be it data, funds, ideas regarding new predictions that should be added, and so on) – the *Endor.coin* community will and enable and facilitate a self-perpetuating positive-feedback business loop. Following is a short description regarding the main aspects of this ecosystem.

5.1 Public Data Providers

The *Endor.coin* Protocol would incentivize from day one the contribution of public data streams, provided at first phase by 'registered data providers' (undergoing a brief compliance and quality assurance process, making sure the ecosystem gets bootstrapped with the highest quality data), and later on – by any person or company that wishes to enrich the data available for analysis, and being rewarded by *EDR* for doing so.

Examples for potential data contributors are :

- **Data Partners:** projects such as *Enigma*, *Twine*, *Thasos*, *Ocean Protocol* and others, connecting their data stream (either partially or in full) to the *Endor.coin* platform for analysis.
- **Social Channels:** such as *Twitter*, *Reddit* and others.
- **Data Scrapers:** harvesting, cleaning, structuring and semantically enriching data from a variety of publicly available sources.
- **Blockchain Protocols:** such as *ORBS* or similar, by establishing a node that would download and parse the data, and make it accessible for analysis.

Endor.coin is already negotiating partnerships with several such potential contributors.

Archived copy · cryp2nova.com

5.2 Academic Research Groups

As the *Endor.coin* Protocol supports the integration of multiple analytics engines for clustering of raw data, the company intends to utilize some of the proceeds in order to encourage the collaboration with leading academic research groups in the field. Technologies that would be developed during this activity could be easily integrated to the growing *Endor.coin* ecosystem, further rewarding their developers with *EDR* tokens whenever they are used to enrich the accuracy of predictions.

This effort would be spearheaded under the guidance of *Endor.coin* Scientific Advisory Board (see a detailed list in Section 8.2) that is comprised of world leaders in data science, machine learning and blockchain technologies from the industry and academia.

As an *MIT* spin-off that relies on the scientific revolution of *Social Physics* that was developed at *MIT* by the project's founders, the company intends to strive for the creation of a strong and sustainable alliance, led by *Endor.coin* , encompassing leading research groups at the academia, Blockchain projects in the fields of data analysis and data marketplaces, and infrastructures for the proliferation and usage of data insights and predictions.

5.3 Catalysts – Application Developers

To increase the range of questions that *Endor.coin* can address, the domain is infinite and limited only by those who want to build applications using the Social Physics engine, the community of "Catalysts". Fueled by them, *Endor.coin* will become a modular App platform on which Catalysts can expand the prediction domain through the power of Blockchain. To increase the range of questions that *Endor.coin* can address, the domain is infinite and limited only by those who want to build applications using the engine, the community of "catalysts" *Endor.coin* will become a modular App platform on which Catalysts can expand the prediction domain through the power of Blockchain. *Endor.coin* enables an open ended stream of micropayments to authors of reusable software components that can be perpetually combined and recombined to create an ever expanding library of useful, highly customizable Apps. Catalysts are entitled to a Micro Use License for each component they add to the Framework. End users install the Apps of their choice. The license to be paid is the sum of all the Micro Use Licenses of the components used by that App. Through smart contracts *Endor.coin* is responsible for charging end users and distributing the payments to the respective Catalysts involved. Everything is automated and transparent to end users. Incentives are paid in tokens. Applications are essentially a set of plugins, a positive feedback loop is foreseen: the more applications built, the more plugins are added to the system and more unique components are ready to be reused. This contributes to mutually self-reinforcing network effects across the *Endor.coin* ecosystem.

Archived copy · cryp2nova.com

5.4 Data Sovereignty

In a world which consumes our privacy for the benefit of greedy corporations, *Endor.coin* is committed to changing the game and to return the wealth back to those who provide access to their data. A simple sign-up on *Endor.coin* starts the process which can unfold according to various scenarios enabled by the *EDR* tokens. Members can receive individual insights based on their own data or team up as a group to aggregate a larger pool of data working together with a small business to get more personalized services, for example zooming in on the right product, at the right time, for the best price.

Based on the fundamental belief that people have the right to own, control and benefit from the data they generate *Endor.coin* opens the power of insights currently available only to those with deep pockets, to the so far disadvantaged long tail of small businesses which can now aggregate *Endor.coin* Members' digital data for their own use, with the data owners' permission, to provide high value data analysis that brings back to its members both highly efficient services resulted from that analysis, as well as monetary gain – when the data is utilized in other ways. This Personal Data Independence based approach, championed by *Endor.coin* unlocks tremendous value for both companies and people, without loss of value or trust for either. *Endor.coin* supports businesses by co-opting membership and brand participation at a negative acquisition cost as companies and other organizations sign up to eliminate data liability, access superior data about their customers, and develop a customer relationship.

As membership will continue to increase, *Endor.coin* will also provide a suite of tools and services that will help members gain insights and understanding regarding many aspects of their lives from their own data. With the mission "Data of the People, by the People for the People" *Endor.coin* gives back to the individuals their inherent right to capitalize on their lives which have been so far violated by the likes of Facebook, Google, Uber and many other data-based conglomerates, thriving on the "platform economy". Through a multi-faceted approach that integrates proven enterprise expertise, forward legal thinking, technical know-how and value creation for consumers, *Endor.coin* offers a unique business model which resolves the "data piracy" problem by creating a consensual data relationship which enables companies to generate good will with their customers by rewarding members directly for access to their data insights.

To ensure the integrity of the process we are creating *Endor.coin* Trust, an independently managed entity that works in a synergistic relationship with *Endor.coin* to maximize the benefits, value creation and data security for members. As people sign up to *Endor.coin* their data is overseen by the Trust and they automatically become Trust members. The *Endor.coin* Trust represents its members in many ways to ensure and maximize the safety, security, privacy and value of their data. The Trust protects members' data by setting and enforcing standards of personal data control, value realization and privacy.

Archived copy · cryp2nova.com

Chapter 6

Token Implementation

6.1 Blockchain Structure

Data Storage: Blockchains are not general-purpose databases. *Endor.coin* has a decentralized off-chain distributed hash-table storage that is accessible through the Blockchain, which stores references to the data but not the data itself. Private data is encrypted (using either AES-256 – Amazon S3 server-side encryption, or AWS IAM mechanism) prior to transfer and storage, having the access-control Protocols programmed into the Blockchain. *Endor.coin* is designed to connect to an existing Blockchain as well as to private and public datasets in an off-chain network (stored in any form of database, central storage, etc., that can be exported into a structured format that is eventually uploaded to *Endor.coin* AWS infrastructure). In future releases the Data Layer will be opened to external players which will be able to sell data for *EDR* tokens. The data will be certified and uploaded to the *Endor.coin* off-chain infrastructure, where it will be available to the owners of the data, and if marked 'public' – for any customers for which the prediction algorithm deem it as relevant. The pricing will be determined by the data provider, and adjusted automatically by the *Endor.coin* Protocol dynamically, according to the demand for this prediction, incentivizing users to share the cost for the same data source.

Consuming Predictions: The Blockchain cannot handle heavy complex transactions. The same off-chain computational network is used to run heavy computations (required by the various predictive engines). Once results are available, they are broadcast throughout the public Blockchain for end users use (authenticated using the key of the user who requested the prediction). In parallel, the same results are encrypted using a different,. temporary key, and are broadcasted publicly. This key is released following a pre-defined time, allowing results to be authenticated at a later stage, even by users who did not require them originally. This mechanism can be adjusted in a way that creates such multiple temporary keys, with varying cost that depends on the 'freshness' of the prediction.

Archived copy · cryp2nova.com

Processing Data: Code execution is divided for execution on the Blockchain (public predictions, public RFP orders, private RFP orders) and on *Endor.coin* infrastructure based on using proprietary hyper-elastic computational layer running on AWS, or similar environments (for example, the *GOLEM* project). As processing data and extracting behavioral clusters requires a complex and expensive execution environment, later releases of the *Endor.coin* Protocol would open the execution layer of the Protocol as well. This would enable a new type of stakeholder to join the *Endor.coin* ecosystem – one that specializes in intensive execution. The larger the data to be analyzed and the heavier (computation-wise) the clusters-extraction algorithm – the higher the price in *EDR* tokens needed to run it will be.

Payment: a user can pay for every prediction request an amount of *EDR* depending on the prediction complexity, dynamically defined by the *Endor.coin* Protocol, per the available resources and the demand at the time of request, and the number of users asking similar questions. Pre-defined queries are expected to remain at a relatively steady price, whereas RFPs (Request for Prediction) would start with a relatively high cost (as by definition they start with a small number of users), gradually decreasing as a community of users for which it is relevant, is formed. The larger this community the less everyone eventually pays, incentivizing the creators of new prediction queries to 'spread the word' among people in their network.

6.2 Smart Contracts

The *Endor.coin* Protocol provides several basic primitives for its end users:

- GetPrediction(prediction_def) – to be implemented at Phase 1
- PutPredictionReq(prediction_def) – to be implemented at Phase 2
- PutData(data_def) – to be implemented at Phase 2
- RunCustomPrediction(data_alg_prediction_def,price) – to be implemented at Phase 3

These primitives allow customers to retrieve predictions available at the *Endor.coin* platform for a dynamic cost and later on to upload data and sell it to be used for prediction purposes. While the primitives cover the default use cases for the *Endor.coin* Protocol, future releases would enable far more complex operations to be designed on top of Get and Put by supporting a deployment of smart contracts. As the Protocol progresses towards private data and custom private predictions, it will enable additional layers of security and complexity on top of the basic smart contract that will be introduced as the first public predictions version. Smart contracts would enable *Endor.coin* users to write stateful programs that can spend tokens, request predictions and be used for retrieval of data in the markets, as well as validate data quality proofs. Users will be able interact with the smart contracts by sending transactions to the ledger that would trigger function calls in the contract. The Smart Contract system will be extended to support *Endor.coin* specific operations (proof

Archived copy · cryp2nova.com

verification), supporting contracts specific to data upload (which will be used for public or private use at later on stages), as well as more generic smart contracts.

getPrediction(prediction_def): Allowing users to retrieve predictions stored on the *Endor.coin* network by paying *EDR* tokens. Clients initiate the Get Protocol by submitting a bid order to the Retrieval Predictions Market order book (by propagating their order to the network). When a matching ask order from prediction provider, the client receives private temporal link to the prediction. When received, both parties sign a deal order and submit it to the Blockchain to confirm that the exchange succeeded. The prediction consumer will pay the equivalent price which was attached to the prediction that he asked to get. The predictions will be accessible for download from the *Endor.coin* webportal.

putPredictionReq(prediction_def): Allowing users to suggest additional prediction types by submitting it to the *Endor.coin* platform. The interface will be accessible in the *Endor.coin* webportal. The suggested predictions will be visible to all *Endor.coin* users and they will be able to rank them. The most highly ranked predictions will be added to the continuously growing predictions catalog. Prediction requests will be identified with a public address of the prediction issuer, once the prediction is added to the catalog, the issuer of the prediction will be compensated each time the prediction is consumed by other users.

putData(data_def): Used for data providers and partners. Data providers are rewarded by *EDR* tokens whenever their data is being used for predictions, having the pricing automatically calculated by the *Endor.coin* Protocol prediction algorithm. This variant of the function will be accessible in the second phase of the *Endor.coin* project. During the third release of the Protocol an API for a 'do-it-yourself' mode, allowing customers to upload their proprietary data streams, marking them as 'private', and having them fused together with the public data streams that are accessible on the platform. Client initiates the Put Protocol by submitting a bid order to the Storage Market order book (by submitting their order to the Blockchain), waiting for a matching Ask order to be placed from data validator (e.g. the prediction engines providers). Clients are required to fund the extraction of behavioral clusters by the prediction engines, but are then able to decide on the price they want to get for each usage by the end users.

runCustomPrediction(data_alg_prediction_def, price): This method would be implemented at the third release of the Protocol, supporting the creation of predictions that are not included in the *Endor.coin* catalogue, on a 'do-it-yourself' mode. Input contains the description of the desired behavior, by the way of example, or description of a logic that refers to publicly or proprietary data. Maximal price is defined, as the request can be handled by any available engine. The fees will be arbitrated between the different players that contributed to the prediction (e.g. the prediction engine and data providers).

Archived copy · cryp2nova.com

6.3 Endor's Roles at Launch

As described in details in the previous sections, the *Endor.coin* Protocol enables the inter-communication among multi-faceted players (e.g. data providers, analytics engines developers, and so on). In this sense, the Protocol is not a source of predictions per-se, but rather – but rather an *enabler* and a *language*, that allows the creation of value at the 'edges'. Similar to the TCP/IP protocol that provides an easy communication between various value-offerers, the *Endor.coin* Protocol would enable data-providers to offer their goods, prediction-engines to offer their analytics services, and catalysts to offer ways that these would be able to used for the creation of a synergic value offering.

For this, *Endor.coin* would operate under a separate legal entity, focused on the implementation of the Protocol, and securing the optimal composition of the its surrounding ecosystem, and maximizing its growth acceleration.

However, in order to guarantee a swift establishment of a production-grade workflow of the Protocol, and allow customers to consume a variety of accurate predictions as early as possible (and even – on day one), the *Endor.coin* project would optionally feature *Endor* as an actor playing several key roles. Later on, as new players adopt the *Endor.coin* Protocol, Endor's roles are expected to decrease in dominance, making way to innovation and value that would be introduced by new participants in the *Endor.coin* ecosystem.

Following are the main roles to be provided by *Endor* upon *Endor.coin* launch. Each of these activities would require dedicated funding, that would be allocated from the proceeds received during the ICO in exchange for *EDR* tokens. The companies are now negotiating the exact details of this collaboration.

Public Data Provider: *Endor* would use its Dev team to implement production-grade infrastructures that would harvest and clean a variety of data streams, uploading them to the *Endor.coin* infrastructure as data ready for analysis. These will include the Bitcoin Blockchain, the ERC20 Blockchain, and several other proprietary Blockchains. In addition, this will later support social network feeds such as *Twitter*, *Reddit*, etc. *Endor* would provide this service on a 'cost plus' basis, covering its HR expenses.

Processed Data Provider: *Endor* would act as the first analytics engine that supports the *Endor.coin* Protocol, contributing its capability for extracting behavioral clusters from transactional data streams. *Endor* would provide this service on a 'cost plus' basis, covering its *AWS* expenses.

Demand Provider: *Endor* would act as a demand provider, channeling hunger-for-predictions from its existing (and new) enterprise customers. *Endor*'s customer-success crews will use USD paid by enterprise customers and purchase *EDR* tokens, generating the desired predictions.

Archived copy · cryp2nova.com

Applications and Partnerships Development : In order to further expand the *Endor.coin* ecosystem and proliferate the penetration of *Endor.coin* based predictions, *Endor* would actively strive for the creation of new prediction-based businesses, either through partnerships, or by forming new types of predictions and releasing them to the public – acting as a catalyst. This would be utilizing the business-development team of *Endor*, as well as its large and high quality advisory board.

6.4 Token Privileges and Economy

As discussed in previous chapters, the *EDR* tokens issued by *Endor.coin* would be utilized as the payment mechanism for a variety of services offered by the *Endor.coin* ecosystem. Noting that some of these service would be rendered free-of-charge at the initial phase of the project in order to facilitate its growth, following are the main usages of the *EDR* tokens :

Purchasing pre-defined or DIY predictions: The main usage of the *EDR* tokens is expected to be the payment mechanism for the consumption of predictions. These would include pre-defined predictions available at the ever-growing catalogue, as well as (at a later stage) "Do-it-Yourself" predictions, to be ordered by the users using a dedicated self-serve API. The payment would mainly cover the cloud-computation resources (e.g. *Amazon AWS*, *Golem*, etc.), with 10% to 25% to be allocated as payment to the providers of the data used for the creation of the predictions, on a pro-rate basis. Cost of predictions would be pegged to the cost of cloud resources required to generate it.

As the main component of the prediction generation process is shared among different predictions, this would create a cost function that asymptotically decreases with the increase in the number of users, with temporal spikes associated with sudden increase in demand:

Payment for data providers Between 10% to 25% of the prediction cost.

Cloud resources for data analysis (*AWS* or similar) Approximately 80% to 90% of Cloud cost, divided by the number of active users (namely, between 55% and 80% from the overall cost of prediction, divided by n, the number of active users).

Cloud resources for personalized prediction (*AWS* or similar) Approximately 10% to 20% of Cloud cost (namely, between 10% and 15% from the overall cost of prediction).

A qualitative payment model for prediction cost would therefore be :

$$0.7 \cdot O\left(\frac{C_{PREDICTION}}{n}\right) + 0.3 \cdot C_{PREDICTION}$$

such that $C_{PREDICTION}$ denotes the cost of a single prediction at the launch of the *Endor.coin* platform and n denote the number of active platform users.

Archived copy · cryp2nova.com

Submitting RFPs (Request for Predictions): Users interested in specific predictions that are not yet supported by the catalogue, could apply for the creation of a new predictions, using the *RFP* mechanism. In such a case, users will pay *EDR* tokens covering the cost of time required for the creation and optimization of the prediction by a skilled *Endor.coin* team-member, as well as the cloud resources it requires. In this sense, *RFPs* would be pegged to the combination of cloud-cost and the average market salary for skilled data engineers.

Requesting QOS preference for predictions: When purchasing a prediction, or requesting a new prediction through the *RFP* mechanism, a user may select to purchase a premium access to the prediction, either receiving it several hours before it is being released to the non-premium purchasers, or in the case of the *RFP* mechanism – retaining exclusive access to it for a pre-defined period of time. Such *Quality-of-Service* elements would be possible to purchase by multiplying the cost of the prediction by a pre-defined factor. In this sense, this premium-service would also be pegged to the cost of cloud resources and market salaries, as it is a multiplication of the basis cost of prediction.

Submitting new data streams for the platform: A data provider who wishes to contribute a new data stream for the *Endor.coin* ecosystem will be required to pay the cost of the cloud resources required for the initial analysis of the data, and its adaptation to the *Endor.coin* Protocol. This is required in order to provide positive incentives for the integration of high-quality data sources, in order to enhance the overall quality of predictions. This initial process fee would grant the data contributor the right to have their data supported by the *Endor.coin* Protocol as a source of behavioral clusters for predictions, and subsequently, also the right to be compensated in *EDR* tokens when it is being used to create predictions.

6.5 Use of Proceeds

Proceeds received from the purchase of *EDR* tokens will be used to implement the *Endor.coin* Protocol software infrastructure, support its initial growth and adoption through marketing and strategic collaborations, fund its required cloud resources, and cover required legal and administrative expenses.

The main part of the proceeds will be allocated for *Endor.coin* R&D team that comprised of cutting edge Cloud and Blockchain engineers, as well as world experts in Data Science, Machine Learning and Social Physics.

An important role of *Endor.coin* project and team would be to guarantee the quick adoption of the *Endor.coin* Protocol by as many leading players as possible. One of the ways this will be achieved is by forming a joint research activity with the academia, providing leading research access to the *Endor.coin* Protocol and infrastructure at day one. Up to 10% of the proceeds will be allocating to support such activity, with a world leading research institute.

Archived copy · cryp2nova.com

Up to 30% of the proceeds will be allocated for the possibility of purchasing relevant proprietary technologies such as prediction engines or ETL connectors, as these are required for facilitating and accelerating the proper bootstrapping of the *Endor.coin* Protocol and its surrounding ecosystem.

Archived copy · cryp2nova.com

Chapter 7

Technological Advantages and Differentiation

7.1 A Scientific Revolution from the MIT Furnaces

Powered by MIT's novel *Social Physics* technology [1], *Endor.coin* utilizes the world's most advanced behavioral prediction technology. This scientific breakthrough that started at the MIT Media Lab at the beginning of this decade revolutionized the field of behavioral data analysis, and was at the forefront of technological accomplishments which included winning the prestigious *DARPA Network Challenge* [21], boosting the returns of a community of retail investors [9, 13] and successfully forecasting the existence of efficient unknown cyber-attacks [35]. This technology was developed by a team of academic and industry experts, that have collectively published hundreds of scientific papers, dozens of patents, and several books dedicated to these subjects.

And now, this revolution reaches the public, utilizing Blockchain technology in order to provide professionals and laymen alike access to capabilities that hitherto were the sole privilege of large retailers, banks and tech giants. A detailed discussion about Social Physics appears in Section 3.1 and Appendix A.

7.2 Real Product. Proven Technology

As previously outlined in Section 3.2, Endor is an MIT spinoff [14] that took upon itself to carry out the implementation of Social Physics as a product, designed to offer large banks and retailers a SaaS solution that would open their predictions bottleneck. The company is backed by leading investors [15], and works with Fortune-500 companies such as *Coca Cola* [16], *MasterCard* [17], *Walmart* and others, successfully demonstrating its ability to automatically generate accurate predictions for a variety of use-cases.

The value of the technology for Blockchain and cryptocurrency analysis was demonstrated in [27] and others. Endor's product has been featured at leading venues, such as *Money-2020* and*Finnovate 2017* [18]. *Endor* is a *Gartner Cool Vendor* [19], and was recognized by the

Archived copy · cryp2nova.com

World Economic Forum as a "Technological Pioneer" [20].

7.3 Usability, Value to Users and Value for Token Holders

A key uniqueness of *Endor.coin* is that the *EDR* Token will be usable on day 1 of the token launch – offering token holders complete access to the pre-defined predictions. In addition, a group of beta users will be selected shortly after token launch, and will be given the opportunity to request predictions in addition to the pre-defined ones.

Furthermore, token owners who believe they might require affordable access to predictions in the future would be incentivized to buy and hold *EDR* tokens, as expected increased usage will result in increased cost of predictions, driving token value up, rewarding prediction users, who purchased *EDR* tokens early on.

CRYP2NOVA

Archived copy · cryp2nova.com

Chapter 8

Team

8.1 Key Team Members

Dr. Yaniv Altshuler

Dr. Altshuler is the CEO of *Endor*, and a research affiliate at the MIT Media Lab, were together with Prof. Pentland he developed "Social Physics", a new science that models crowds behavior. At MIT, the technology was used to win the prestigious *DARPA Challenge* [21], help the Government of Singapore improve its ability to predict traffic jams, and assist a community of thousands of financial investors to improve their financial returns [13]. At *Endor*, the technology was used to accurately predict the behavior of crowds in a large variety of use-cases, efficiently catering large banks and retailers [14].

Prior to his position at MIT and the incorporation of *Endor*, Altshuler was a researcher at *IBM*, developing a novel optimization technique used to boost the performance of supercomputers. Active in Blockchain research since 2011, Dr. Altshuler has authored over 60 scientific papers and filed 15 patents. Altshuler's works have been featured in popular venues such as the *Financial Times* [10], *Harvard Business Review* [11] and others. His recent published books are '*Security and Privacy in Social Networks*' [26] and '*Swarms and Network Intelligence in Search*' [25].

Prof. Alex "Sandy" Pentland (Chairman of the Scientific Advisory Board)

Director of the MIT Media Lab Entrepreneurship Program, as well as the MIT Connection Science and Human Dynamics labs. Prof. Pentland is one of the world's most-cited scientists [23], and was recently declared by Forbes as the "7 most powerful data scientists in the world" along with Google founders and the Chief Technical Officer of the United States [24]. He has received numerous awards and prizes such as the *McKinsey Award* from *Harvard Business Review* [22], the *40th Anniversary of the Internet* from *DARPA* [21], and the *Brandeis Award* for his work in privacy.

He is a founding member of advisory boards for *Google*, *AT&T*, *Nissan*, and the *UN Secretary General*, a serial entrepreneur who has co-founded more than a dozen com-

Archived copy · cryp2nova.com

panies, as well as social enterprises such as the Data Transparency Lab, the Harvard-ODI-MIT *DataPop Alliance* and the Institute for Data Driven Design. He is a member of the *U.S. National Academy of Engineering* and a leader within the *World Economic Forum*. His most recent books are '*Social Physics*', and '*Honest Signals*'.

Stav Grinshpon

Mr. Grinshpon is a veteran tech-industry expert, with 18 years experience of product and management in companies such as *SAP* and *AT&T*. Grinshpon is a world expert in cyber defense, serving 8 years as a technological leader at the Israeli 8200 technological unit. Grinshpon is the author of 3 patents focusing on data analytics, and headed the R&D activities at *Endor*.

David Shrier

David Shrier is a globally recognized authority on financial innovation, and leads the *University of Oxford*'s online programmes *Oxford Fintech* and *Oxford Blockchain Strategy*, as well as *MIT's Future Commerce*. He has published multiple books on fintech, Blockchain and cybersecurity, including *New Solutions for Cybersecurity*, *Frontiers of Financial Technology*, and *Trust::Data* [27, 29, 32].

Shrier is the CEO of *Distilled Analytics*, a Machine Learning company derived from MIT research that is transforming the financial services industry with behavioral analytics, and Chairman of *Riff Learning*, an AI-driven collaboration technology platform provider. David is an Associate Fellow with the *Said Business School*, *University of Oxford*, and Lecturer at the *MIT Media Lab*. He also counsels the Government of Dubai on Blockchain and digital identity; *Millennium Advisors*, a middle market credit liquidity provider, on technology trends; and *Cleer.digital*, a Blockchain-based digital commodities exchange, on strategy.

David is on the advisory board of *WorldQuant University*, a program offering a totally-free, accredited, online Masters degree in financial engineering. He previously advised the European Commission on commercializing innovation with a focus on digital technology. He is presently a member of the *FinTech Industry Committee* for *FINRA*, the U.S. securities industry's self-regulatory body, counseling on new developments impacting financial services.

Prof. Mihaela Ulieru

Expert in Computational Intelligence, is a *Blockchain Champion* at the *World Eco-*

Archived copy · cryp2nova.com

nomic Forum [36]. Her research in distributed intelligent systems created a strong foundation for governance on Blockchain as an institutional technology after it revolutionized manufacturing, logistics and homeland security. Prof. Ulieru, a California Berkeley Alumna, is a member of the World Economic Forum's Global Agenda Council, the Science and Engineering Research Council of Singapore and the Canadian Science Technology and Innovation Council.

Dr. Goren Gordon

Dr. Gordon is the head of the Curiosity Lab at the Tel Aviv University, where he develops state-of-the-art models of computational curiosity. Gordon, a leading expert in Deep Learning and Neural Networks Optimization, holds a PhD in Quantum Physics, and another PhD in Neurobiology. Dr. Gordon had utilized this experience, together with his additional degree in Medical Sciences, during his work at MIT Media Lab with Prof. Cynthia Breazeal, where he studied how curious robots interact with curious children. Gordon is also interested in scientific education, a cause for which he developed "Quantum Computer Games" that teach Quantum Physics to children via play. Gordon is also a popular lecturer on the topics of quantum physics, the brain and inter-disciplinary thinking.

Dr. Arie Matsliah

Dr. Matsliah is a world expert in the theory of graphs analysis, and spent 16 years working for industry giants such as *Google*, *IBM*, *Intel*and *Lyft* as well as being the Chief Architect at *TripActions*, a successful Menlo-Park based Start-Up. Dr. Matsliah has also published over 40 scientific papers, focused on fundamental research in algorithms, complexity, and quantum computing.

Shahar Somin-Gavrielov

Ms. Somin-Gavrielov is an expert in Statistical Learning Theory, holding a Masters Degree (with honors) from the Hebrew University. Somin-Gavrielov is a seasoned researcher, combining deep theoretic realization of data science with hands-on industry experience. In her past, she was a decorated analyst at the Israeli 8200 intelligence unit.

Edo Eisenberg

Financial Risk Management professional, previously managed retail credit portfolios at *Morgan Stanley* and *Barclays*. Additional 9 years experience at *NICE*, designing and implementing fraud prevention solutions deployed in 7 of top 10 U.S. banks. Alumnus of the Duke MBA Program.

Lior Regev

Mr. Regev is a seasoned software engineer with a great passion for technology. Regev is an ex-intelligence tech-leader, with vast experience in distributed systems, cloud architectures and SaaS products.

Archived copy · cryp2nova.com

Liat Yitshaki

Ms. Yitshaki is an expert in public law litigation. She holds an MA in Law and Ethics and an LLB with honours from the University of London. In the past Yitshaki worked for *McKinsey & Co.*, as well as a senior legal adviser to the British Government.

8.2 Advisors

Prof. Alexander Lipton

Professor Lipton served as the Managing Director of the Global Quantitative Solutions for *Bank of America*, as well as the Managing Director of Global Quantitative and Credit Analytics groups at *Merrill Lynch*. Prior to this, Prof. Lipton was the head of Quantitative Research at *Citadel*, and Head of Equity Derivatives at *Credit Suisse*. Lipton is currently a professor of Financial Engineering at *EPFL* as well as a Fellow at *MIT* Connection Science Center, and recent author of the Scientific American article titled "*How technology could help fix our broken financial system*" [37].

Ron Gross

Ron Gross is the Founder and a Board Member of the Israeli Bitcoin Association. He has been active in the Bitcoin and Blockchain worlds since 2011, and was the Executive Director of *Mastercoin* (now *Omni*) – the world's first ICO. Gross in an ex-Googler, and has also served as the Chief Architect at *Commerce Science*.

Dr. Nuria Oliver

Dr. Oliver, an MIT Alumna, is the Director of Research in Data Science at *Vodafone* and the Chief Data Scientist at the *Data-Pop Alliance*. Formerly, Dr. Oliver was the Scientific Director of *Telefonica R&D* and a researcher at *Microsoft*. Oliver is a world expert on Data Philanthropy – the use of big data analytics as a form of collaboration in which private sector companies share data for public benefit. Oliver has authors over 90 scientific papers and book chapters and has filed over 40 patent applications.

Dr. Daniel Tunkelang

A World expert in Data Science, Tunkelang, an ex-Googler as well as ex-IBM-er, is an advisor to tech giants such as *Apple*, *Salesforce Etsy*, *Yelp* and *Pinterest*. Previously, Tunkelang had served as the Director of Engineering in Search at *LinkedIn* as well as the latter's *Data Scientist in Residence*, as well as being the Chief Scientist at *Endeca* (acquired by Oracle). Tunkelang holds an Masters Degree from MIT, and a PhD from Carnegie Mellon University.

Guy Zyskind

Founder and CEO of *Enigma*, the company behind the *Enigma Protocol* for creating Secret Smart Contracts, and *Catalyst* - a platform that allows anyone to start a crypto hedge fund using sophisticated tools and data. Prior to Enigma, Guy was a graduate student at *MIT*, researching and teaching Blockchain technology. Guy authored several

Archived copy · cryp2nova.com

academic papers, most recently on privacy and the Blockchain, including the *Enigma Whitepaper* and "Decentralizing Privacy: Using Blockchain to Protect Personal Data". Guy holds an M.S. from *MIT* and a B.S. in Electrical Engineering and Computer Science from Tel-Aviv University.

Prof. Michael Bronstein

Serial entrepreneur and a leading researcher, Prof. Bronstein is a Research Fellow at Harvard University, and a Professor of Informatics at the University of Lugano. Bronstein has authored over 100 publications in leading journals and conferences, over 20 patents, the research monograph "*Numerical Geometry of Non-Rigid Shapes*", and edited 4 books. Bronstein is one of a few researchers who received 3 *European Research Council* (ERC) grants, and was also awarded the *Google Faculty Research* award, the *Radcliffe fellowship* from *Harvard University*, and the *Rudolf Diesel industrial fellowship*. He was selected by the *World Economic Forum* as one of the world's 40 leading researchers under the age of Forty.

Prof. Bronstein is actively involved in industrial applications. He co-founded and served as Vice President of technology in *Novafora* (licensed to Turner Broadcast) and was a co-founder and one of the main developers of the 3D sensing technology company *Invision* (acquired by *Intel*, where Bronstein currently serves as a Research Scientist for Perceptual Computing). Bronstein is a co-founder and technical advisor of *Videocites*.

Dr. Wei Pan

Dr. Pan is an MIT Alumnus, ex-Googler, and a world expert on big data analytics, machine learning and complex systems. He is the inventor of the analytic approach known as "Reality Hedging", that tries to understand financial markets dynamics and macro economics through the understanding and models of social systems, and big data measurements of people and crowds. Currently, Dr. Pan is a Co-founder and Chief Scientist at *Thasos Group*, a New York based start-up. Pan previously worked at *Fidelity Investments* under the Chief Economist where he focused on systemic research and the Flash Crash.

Igor Gonta

An MIT Alumni in Engineering and Computer Science, Mr. Gonta is the Commodity and Risk Management at *BlackRock*. In his past Gonta served as the CEO of *Market Prophit* – a real-time stocks sentiment generation engine, that is based on social-media conversation analysis. Gonta was the Founder of the company, which he then sold to a large hedge-fund. In his early career Gonta served as the Managing Director of Commodities Sales at *Barclays*, as well as Vice President of Commodities Sales at *Goldman Sachs*.

Thomas Hardjono

Hardjono is a security expert, specializing in decentralized identity, Blockchains and smart contracts. Hardjono was the Executive Director of *VeriSign* and the MIT Ker-

Archived copy · cryp2nova.com

beros Consortium, and has published 5 books dedicated to computer security and cryptography [38–42].

Archived copy · cryp2nova.com

Appendix A – *Social Physics* Explained

Archived copy · cryp2nova.com

1. HOW SOCIAL PHYSICS WORKS?

Clarification: This document intends to provide a comprehensive view of Social Physics, the high level principles behind it, as well as its technical implementation by Endor. However, certain technical details regarding the specific mathematical formulation of the Social Physics Laws that are used by Endor, as well as the specific implementation of the mechanisms used in order to detect violations of these Laws, were intentionally omitted, due to IP considerations.

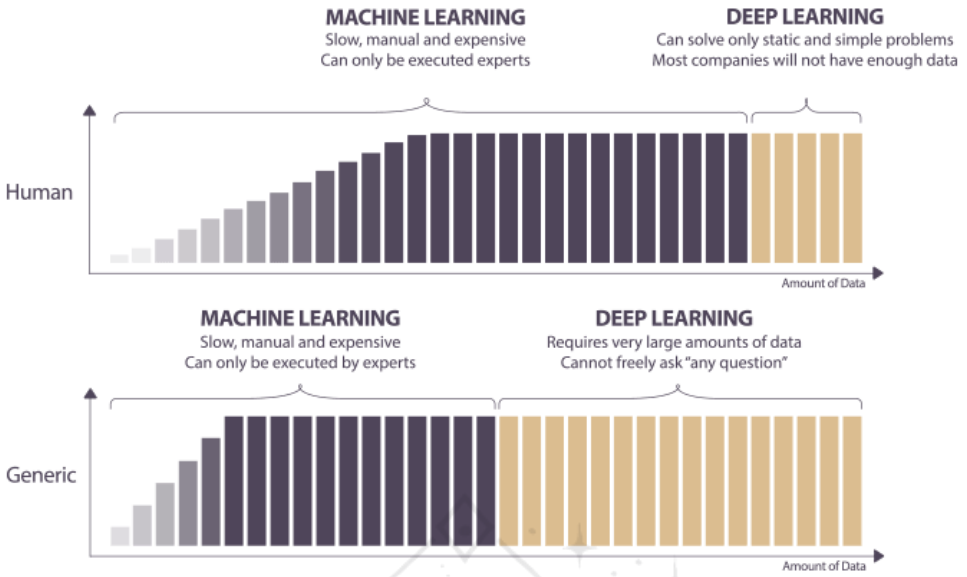
In the Information Age, companies gather data of all types and from numerous sources about their businesses operations. Data encompasses images and videos, text and tweets, transactions and usage logs. However, the majority of data originates from a single underlying source: People.

Thus, for example, tweets and blog posts are written by humans for humans; purchase transactions and phone call information convey human desires for things and other people; usage and app logs report on how people interact with computers and mobile devices.

Data derived from human behavior is "messy": it is dynamic, complex and extremely versatile. Humans' behavior, as recorded in such digital data channels, changes drastically over time, is influenced by underlying complex social networks, and is conveyed in highly multimodal data streams. These characteristics pose significant challenges for companies that wish to analyze, understand, and predict their customer behavior in order to improve their business operations.

In recent years, data scientists have started to employ "heavy-weight" statistical methods and Machine Learning algorithms to try and cope with this complexity. These powerful tools, including the new "Deep Learning" techniques, collect data and analyze its attributes in order to be able to classify behavioral patterns, detect anomalies, and predict future trends. However, such tools – historically developed for "static problems" such as image processing and text recognition – cannot easily cope with human behavior data: learning dynamic, complex, and versatile data streams is extremely hard and sometimes nearly impossible.

Archived copy · cryp2nova.com



Endor's Social Physics engine works in a completely different way. Instead of deriving patterns from input data itself, it is based on the discovery that all human behavioral data is guaranteed to contain within it a set of common "social behavioral laws" - mathematical relationships that emerge whenever a large enough number of people operate in the same space. These laws govern the way various statistical properties of crowd behavior evolve over time, regardless of the type of data, the demographics of the users who created it, or the data size. Endor has integrated these laws into its data analytics engine, which efficiently extracts the underlying social attributes of all people contained in the raw data being provided as input (e.g. phone calls, taxi rides, financial investments, etc.).

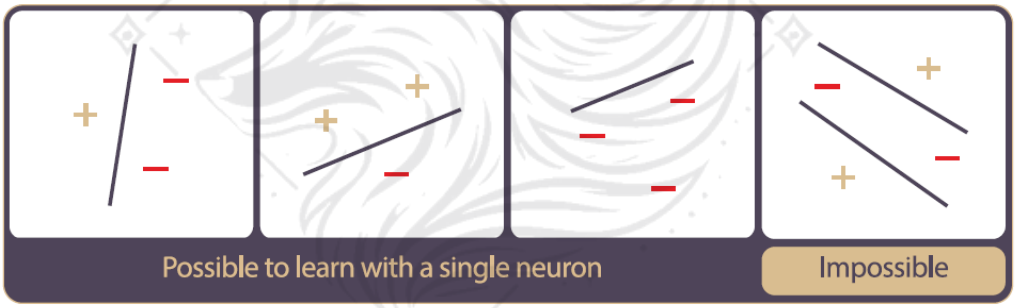
CRYP2NOVA

Archived copy · cryp2nova.com

2.1. WHY SOCIAL PHYSICS IS NEEDED?

Abstractly, Learning Problems, or the ability to classify objects or to produce predictions for future events, requires data to be analyzed, and an algorithm that analyzes it. The quantity of the required data depends on several factors:

1. **Internal Complexity of the Problem** – problems come in different shapes and sizes, and some are undoubtedly harder than others. The "hardness" or "complexity" of a problem refers to the minimal "strength" a learning algorithm must possess in order to successfully learn the problem: if a learning algorithm is not strong enough, it will simply not be able to learn an instance of the problem correctly. For example, it can easily be shown that a "Perceptron" (i.e. the simplest Neural Network, comprised of a single neuron) can never learn the "XOR" function (namely, the Boolean "Exclusive-Or" function). The reason is embedded in the way a Perceptron works (which can be imagined as linearly dividing the input space) whereas the XOR function is simply "too complicated" to be represented this way:



An example of the XOR function appears in the right chart (Source: Wikipedia)

CRYP2NOVA

The learning-complexity of the underlying model of a problem is often referred to as the Vapnik–Chervonenkis Dimension (or VC Dimension) of the problem. The higher the learning-complexity of a problem is, the stronger a prediction algorithm needs to be in order to be able to learn it successfully, the more data such algorithm requires in order to properly model it. For example, a function y = f(x) that has a linear complexity (meaning, it can be well-modeled using a polynomial of rank 1) requires, by definition, less sampling points than a function that requires a polynomial of a higher degree in order to be accurately modeled.

Archived copy · cryp2nova.com

Learning Efficiency of the Algorithm

There are many learning algorithms, each requiring different quantities of training data and expert domain knowledge to properly ascertain the model parameters. For example, simple regression requires large amounts of data and many problem-specific features to work well, whereas deep learning, while requiring vast amounts of data, can automatically learn the domain features. For our discussion we can therefore quantify the overall efficiency of the learning process, that is – how much data would be required in order to learn the underlying model, and how much does the algorithm learn by itself the proper representation:

$$\eta_A = \frac{R_A(M)}{D_A(M)}$$

M = bits of model, is the number of bits required to formulate or model the data, e.g. number of parameters a perfect description of the model would require. In this sense, **M** is akin to the Kolmogorov Complexity of the problem – a known theoretical measure of the data complexity, referring to the shortest Turing Machine that can generate a given data. The model is determined by the problem and cannot be changed, i.e. if it's a simple problem then the model has a few bits.

$R_A(M)$ = bits of model-specific representation, is the number of bits learned by the algorithm **A**, to represent the underlying model **M**. This number represents the automatic feature detection of the algorithm and is inversely proportional to the number of domain-specific knowledge experts must manually program into the algorithm.

$D_A(M)$ = bits of data, is the number of bits required by the algorithm **A**, to learn model **M**.

η_A = learning efficiency of the algorithm **A**, that is, the ratio between the automatic representation learning of the algorithm and the amount of data required to preserve its learn quality. A high efficiency algorithm can learn the proper representation with small amounts of data, whereas a low efficiency algorithm requires manually crafting features and vast amounts of data to tune them.

For example, given a problem class **M**, some algorithms would require more data than others, to preserve the learning quality:

- **Logistic Regression** algorithms usually require manual coding of features by experts and large amounts of data to fine-tune them to the specific problem at hand:

Archived copy · cryp2nova.com

$$R_M(A) << 1 \quad D_M(A) >> 1 \quad \eta_A << 1$$

- **One-Shot learning** algorithms also require expert-domain features, but can use only a few examples to fine-tune the underlying model and have predictions:

$$R_M(A) << 1 \quad D_M(A) << 1 \quad \eta_A \sim 1$$

- **Deep-Learning** algorithms can automatically learn the most informative features, but require vast amounts of data to do so:

$$R_M(A) >> 1 \quad D_M(A) >> 1 \quad \eta_A \sim 1$$

- **Endor's algorithm uses Social Physics** to automatically extract the relevant behavioral features from only a small sample of the data:

$$R_M(A) >> 1 \quad D_M(A) << 1 \quad \eta_A >> 1$$

2. **Rate of Change of the Problem** – another factor that influences the amount of data required to produce accurate prediction is the rate of change of the problem's underlying model. Some problems are static, wherein their underlying parameters do not change, or change rarely. For example, faces in images do not change over months; faces are faces. On the other hand, underlying behavior patterns that lead to the churning out of a paid service may change over time, either gradually via social changes over months, or rapidly in a matter of days, as a response to a successful marketing campaign by a competitor. We quantify this dynamics as follows:

$$\tau = \frac{\partial T}{\partial M}$$

CRYP2NOVA

τ is the problem's persistence or tenacity, denoting the rate at which the underlying model changes, where **∂T** represents the duration over which the model changes by **∂M** bits.

For example, **τ = 1 day / 10 bits** means that the model changes drastically over a period of one day (i.e. fast rate of change) compared to **τ = 1 month / 10 bits** which refers to a much slower change of the model. Effectively, a dynamic model presents a different model every period, requiring re-training or re-learning the model.

Archived copy · cryp2nova.com

Operational Implementation of Predictive Analytics: Chasing the Changing Model

The main challenge in a feasible implementation of predictive analytics for a given problem is therefore obtaining enough information to cope with the behavioral change of the modeled objects. The amount of information bits that one can accumulate per one time unit is denoted as $\mathbf{I_t}$.

Operators of extremely-large social networks or search engines (e.g. Google or Facebook) can often accumulate vast amounts of information in relatively short periods of times. This is however impossible for the majority of the companies interested in predicting their customers' behavior.

In addition, even large players that acquire vast amounts of information every day would find it challenging to accurately models problems that are either (a) too complex, or (b) change too quickly, or of course (c) a combination of the former two.

We can therefore point to a simple equation that determines the ability of a company to implement an operational predictive analytics solution. Companies that are able to satisfy this principle using the learning algorithms they employ and new data they continuously acquire, for the problems they are interested in predicting, will be able to successfully construct an operational process that achieves this goal, whereas companies who fail to do so, will not be equally successful.

The Fundamental Operation Learning Principle:

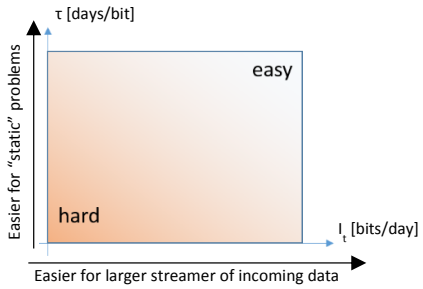
$$I_t \times \tau \times \eta_A > 1$$

The practical implications of this principle dictate that companies who fail in their attempts to construct an operationally functioning prediction systems should either:

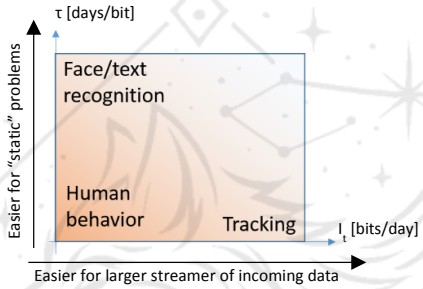
1. Improve their data collection bandwidth, acquiring larger amounts of relevant data per day; or
2. Focus on more static problems; or
3. Resort to more efficient learning algorithms.

Archived copy · cryp2nova.com

This simple relation means that one needs more information per day as the problem's complexity increases and its persistence decreases as illustrated in the following chart:



Here is how a few common predictive problems look when analyzed with this



- **Face recognition:** This problem is characterized by a very low change rate, namely **τ >>1**, as generally speaking – faces do not change. Therefore, even for inefficient learning algorithm for which η_A is relatively small, I_t can still be very small, as one needs a lot of information to initially learn the problem, but not a lot of information to re-train it, since it does not change much.

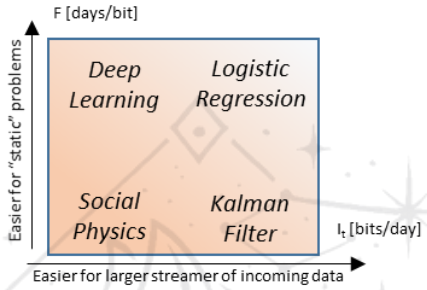
- **Tracking a maneuvering mobile target:** Tracking a moving target is characterized by **τ << 1** since the fast movement of the target (or better yet – the fast changes in its trajectory or dynamic local changes that are introduced by its navigation algorithm to avoid detection) renders any non-immediate past data useless. However, such problems are also usually characterized by a large stream of incoming data, providing I_t **~ 1/τ >> 1**, which means that the dynamic nature of the target's location is fully conveyed to the learning algorithm by the input data stream. This enables relatively simpler algorithms with η_A **~ 1** to solve the problem efficiently.

- **Human behavior:** Human behavior is extremely dynamic, having **τ << 1**, meaning it contains elements that change relatively fast. Moreover, although companies who wish to predict human behavior can (and do) obtain additional information

Archived copy · cryp2nova.com

about their customers, it is always a partial derivation of the actual behavioral change. Metaphorically, this is similar to Plato's Cave allegory, where the information perceived by our sensors passes through a very crude lens that captures only several aspects of it. In our formalism this translates to I_t **<< 1/τ** , which subsequently means I_t **≥ 1**. This means that in order to efficiently predict human behavior, one must employ extremely efficient algorithms of η_A **>>1**.

Similarly, we can see how different solution techniques can be best used for each problem, modeled using this relation:



The chart illustrates the previously mentioned Fundamental Operational Learning Principle $I_t \times \tau \times \eta_A > 1$. The more static the problem is, and the more data about it we have – the closer we are to the upper-right corner of the chart (and the more accurate our predictions will likely be). And the closest we are to the upper-right corner – the less-efficient algorithms we need to employ to produce accurate predictions.

Techniques that require vast amounts of information to train do not work well with dynamics problems, hence Deep Learning works best for problems whose underlying structure does not change fast, such as image processing and gesture recognition. In contrast, simple algorithms that can process information quickly, e.g. Kalman Filter, can deal with dynamics problems but require a high throughput of information to successfully predict. Traditional Machine Learning approaches such as Logistic Regression would be efficient in scenarios where we have relatively high volumes on incoming training data provided that the problems are fairly static as well.

Archived copy · cryp2nova.com

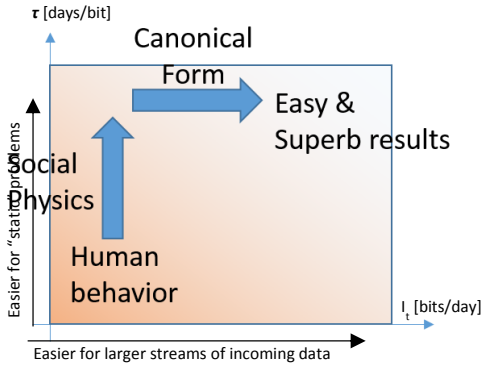
Social Physics as the Solution to the Never-ending Chase

Endor tackles the need for constantly obtaining large quantities of data for changing models from two orthogonal points of view:

- **Transforming dynamic problems to a static model:** The laws of Social Physics are immutable and agnostic to data-type and origin. Therefore, using them to project raw data into a Social Physics canonic representation space transforms the original problem into an instance of a new class of problem, whose underlying model is static (thanks to the inherent static nature of the Social Physics laws). This transforms the actual $\tau << 1$ of the original problem to be extremely large $\tau_{SP} >> 1$.

- **Creating "Big Data" from Small Data:** as mentioned above, when faced with the challenge of predicting their customers' behavior, most companies face a major hurdle – the amount of relevant data they possess is often insufficient. This is specifically true for dynamic problems (frequent in marketing use-cases) or for problems that involve the introduction of a new element (such as predicting the response to a new product, or utilizing a new kind of input data). Endor transforms all the data streams it receives from its customers into a Social Physics canonical form, regardless of type, size and source. By using this canonical form, Endor is able to unify and consolidate all the data from all clients and all their queries into a single extremely large data base, growing at a consistently high pace (namely, with $I_t >> 1$). This is then used to train a single, immutable deep-learning network that is trained on analyzing instances of Social Physics Canonical Forms (and not a specific query or customer). Hence, even if each customer provides very limited data, Endor is able to accumulate a vast amount of (canonically formed) data.

  Endor's engine transforms the most difficult problems of human behavior predictions into slowly changing ones (via Social Physics), and to Big-Data (via the transformation to a canonical form), resulting in an "easy and efficient" problem, then solved using Deep Learning tools. This is illustrated in the following chart:



Archived copy · cryp2nova.com

2.2. SOCIAL PHYSICS: OVERVIEW

In the previous section we have briefly described how Social Physics can be used to overcome the inherent challenge of predicting human behavior, using semantics agnostic static mathematical invariances. In order to better understand how this is feasible, consider physical laws – for example Newton's second Law or the Law of Momentum Conservation. Any object maintains its initial path unless an external force acts upon it. Note that in order to deduce the existence of such hidden force there is no need to "learn the data", understand its statistical attributes, or test many systems that behave in a similar way. Since the physical law is a given, any violation of it is abnormal and can immediately be detected and interpreted as the outcome of some "invisible hand." If one detects objects that suddenly change their direction, it is possible to immediately deduce that a force has acted upon them. If their change is similar, then it is most probable that the same force was exerted on all of them. This simple realization is only possible due to the understanding of the physical law.

While Social Physics is far less absolute and rigorous than physical laws, the concept is similar. If something violates Social Physics laws, it can immediately be qualified as "interesting," being the data manifestation of some valuable property or attribution in the real world. This does not require learning, benchmarking, baselines or any other data science or machine learning tools. A violation of the social physics laws can be detected extremely fast and in a very robust way, regardless of the data type that generated it.

2.3. ENDOR: POWERED BY SOCIAL PHYSICS

2.3.1. Data Transformation into a Canonical Representation

By extracting the previously discussed "behavioral clusters" (namely, the detected violations of Social Physics invariances) and aggregating them into a "Knowledge Sphere," the raw input data (of whatever shape or form, as long as it originates from human actors) is being transformed into a canonical form. This form represents clustering of people who violated a Social Physics Law "together", in other words – people that display a "too high synchronous change" in their behavior, vis-à-vis a certain invariance. This is akin to physical objects that would change direction at a specific time in a similar manner – while the force that caused these changes is not visible, we can deduce with high likelihood that the objects were all affected by a single source. Similarly, Endor's canonical representation of data in the form of behavioral clusters contains groups of people who most probably were influenced by the same hidden "social forces" and thus share common "real world traits". Whenever new raw data is available, it is sent to Endor by the customer (usually on a daily or weekly basis), allowing for additional behavioral clusters to be automatically extracted.

Archived copy · cryp2nova.com

The benefit of this representation is threefold:

- **Automatic:** using the mathematical invariances of Social Physics, as any data that originates from human behavior can automatically be transformed into a collection of behavioral clusters, regardless of the type of input data, that does not need to be declared or analyzed (e.g. phone call records, credit card purchases, taxi rides, or any other type of proprietary data the customer may possess). This, combined with an extremely high resilience to noises and gaps in the input data, means that the process of transforming dirty raw data of an unknown type into uniform behavioral clusters becomes, for the first time, fully automatic (see "Robustness to Noise" section for additional information).

- **Uniform:** by stripping the data of any domain, demographics or semantics aspects, the remaining information containing the behavioral clusters is ideally shaped for the following "querying" phase. In fact, using this uniform representation, Endor can "create big data" when there is none, allowing the querying phase of the process to utilize deep-learning techniques that were impossible for the original owner of the data! This is enabled as the Endor deep-learning engine has access to behavioral clusters originating from many types of data, and from many customers, all transformed into a single-form.

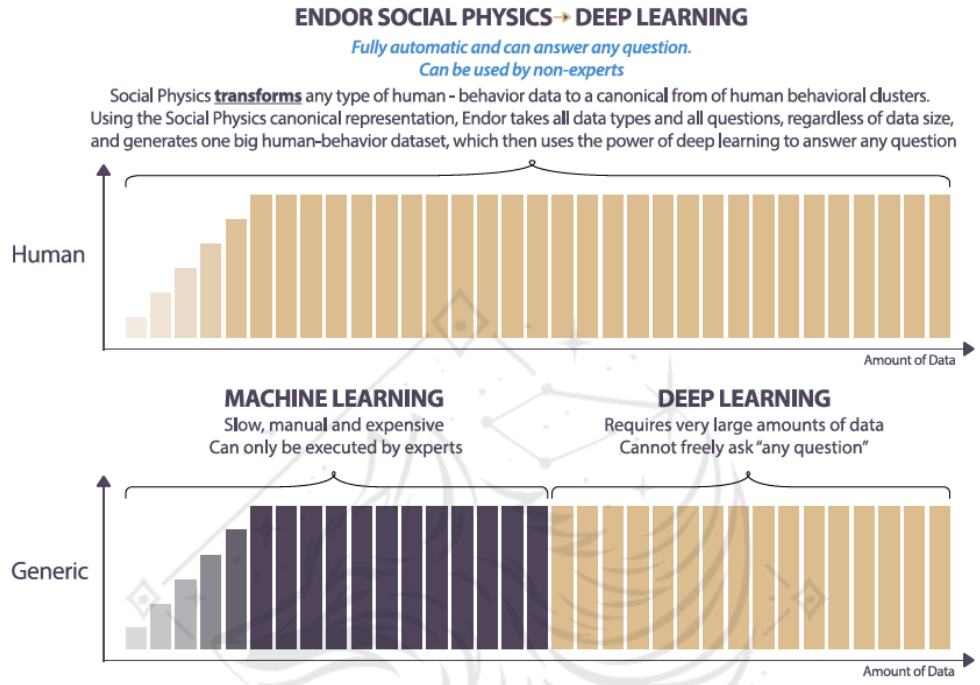
- **Emerging trends:** simply put, the Social Physics invariances describe the way certain statistical properties of human crowds evolve over time. This time-oriented aspect enables Endor to easily detect emerging behavioral changes – dynamics that only recently occur, and that in most cases did not have enough time to generate enough observable data that would enable them to be detected with a high enough statistical significance using traditional methods. In addition to the fact that through the use of Social Physics these additional signals can be detected, these are usually the very signals that are of high importance for a variety of business questions – as they contain information about recent trends.

"Old-school" Machine Learning worked with pre-defined features and could extract relevant information from relatively small amounts of generic data. However, much of the results depended on the selected features. Deep Learning identifies the most relevant features by itself, but requires huge amounts of data. Each data type and question asked require finding the relevant features again, thus requiring more data.

Social Physics transforms any type of human behavior data to a canonical form of human clusters based on their behavior. This works with both small and large amounts of data. In addition, thanks to the Social Physics' canonical form, Endor can ingest all data types and all

Archived copy · cryp2nova.com

questions, regardless of data size, and generate one huge human behavior data set that uses the power of Deep Learning to answer any question.



Current predictive analytics paradigm (top charts) vs. new paradigm enabled through *Social Physics (bottom charts)*.

CRYP2NOVA

Archived copy · cryp2nova.com

2.3.2. Querying the Canonical Representation ("Knowledge Sphere")

As the Knowledge Sphere contains the overall information detected about all users, it can provide predictions for any question with the ease and speed of a simple data search: The initial creation of the Knowledge Sphere usually takes 1-4 hours for a typical data consisting of a billion records. After this process is complete, the same Knowledge Sphere can be used to answer dozens of questions, in minutes. There is no need for prior domain knowledge or extraction of relevant features.

A query is submitted by providing an "example" (positively labeled IDs), of any size. Endor's engine then uses the Knowledge Sphere to generate an answer: a ranked list of users, most probable to be "behaviorally similar" to the relevant query. There is no "training" / "learning" phase; there is no need to "interpret the result." The equivalent to this automatic process for each specific question would have required anything from weeks to months of effort for data scientists using conventional methods.

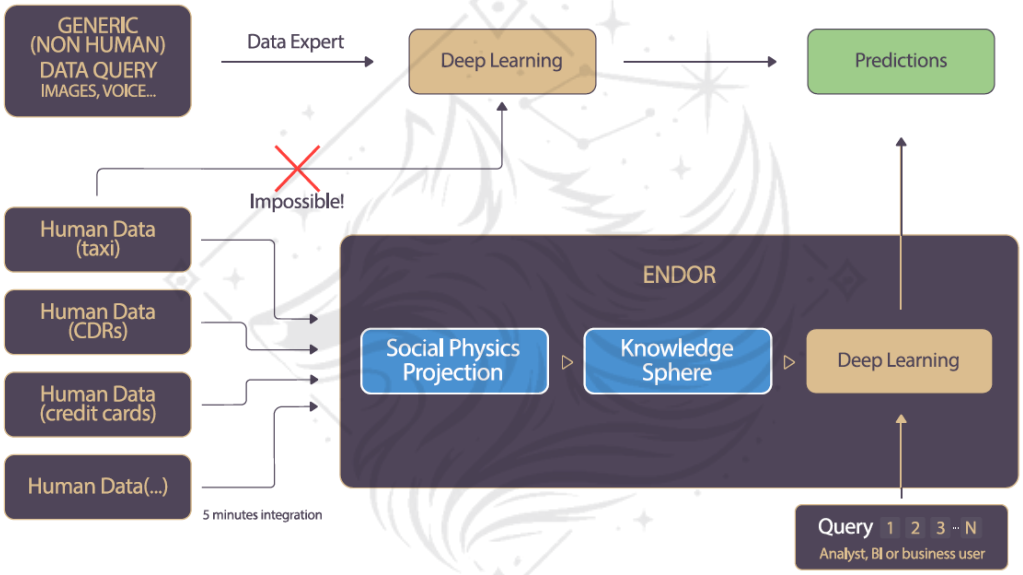
For example, a query regarding users who are mostly likely to churn in the near future is described by a list that contains the identity of previous churners. Alternatively, a query that tried to identify which new customers are likely convert to a premium account would be described by a list containing customers who recently converted to this premium service. Both queries, however, would use the same Knowledge Sphere, requiring no re-training of any sort.

Notice that whereas most of such behavioral predictive questions would be extremely hard to resolve using Deep Learning (due to lack of sufficient data, and the frequent change in the underlying model), Endor circumvents this issue by using the collection of all Knowledge Spheres and queries. Thus, Endor "creates" the big data that is required for its internal Deep Learning component, which is in charge of providing the actual predictions. These predictions are based on the particular Knowledge Sphere, even if it is based on an extremely small amount of data.

The following chart illustrates the flow of traditional Deep Learning (upper flow) in comparison to Endor's one (lower flow). For non-human behavioral data (e.g. millions of images) Deep Learning can likely produce high quality predictions, given a proper training done by a data expert – as current Deep Learning tools are designed to be used by Engineers, who are in addition, experienced with the use of such tools. Human data on the other (e.g. taxi data) has an underlying model that changes so frequently that it cannot be easily resolved by Deep Learning modeling (see Section 2.1 for an in-depth discussion on this issue).

Archived copy · cryp2nova.com

However, when transformed into the form of a set of behavioral clusters, taxi data is stripped of its semantics and becomes virtually identical in representation to behavioral clusters obtained from phone call records, credit card purchases, or any other type of human data. In addition, as there are many different customers who contribute data of each of those types (namely, many e-commerce platforms, each uploading their own purchase and web-activity data), training a Deep Learning model now becomes a possibility. This is enabled as the model would not be trained on the raw data or problems, but on the multitude of behavioral clusters – a large collection of data in a uniform representation, that is also characterized by a static underlying model (i.e. the Social Physics Laws, that are mathematical invariants and are therefore static, compared to the behavioral dynamics of the raw data, that is highly dynamic).



CRYP2NOVA

Archived copy · cryp2nova.com

2.4. DATA SECURITY AND ANONYMITY

2.4.1. Data Stored by Endor

In addition to its ability to automatically resolve an endless variety of behavioral prediction questions, Endor's solution provides a high level of data security and the ability to anonymize any required data field. As detailed in the previous sections, the prediction is done using the Knowledge Sphere – a collection of semantics-free behavioral clusters, containing simply a large number of groups-of-users (each guaranteed to have a common social or behavioral trait). It is easy to see that any sensitive or personal information, if such information existed in the original raw data, is no longer part of any data used by the system at this stage. The only information that is available to the system is IDs of users, as shown in the following example:

$$Behavioral_cluster_{1} = (ID_{1}, ID_{17}, ID_{23}, \ldots)$$

$$Behavioral_cluster_{2} = (ID_{142}, ID_{4287}, ID_{9711}, \ldots)$$

$$\ldots$$

$$Behavioral_cluster_{748,329} = (ID_{5}, ID_{37}, ID_{218}, \ldots)$$

It should be noted that even this information can easily be hidden by hashing the IDs at the raw-data level by the customer, upon data onboarding (see below).

2.4.2. Data Onboarding to Endor

As presented in previous sections the methods used for extracting the behavioral clusters from raw input data rely on the detection of groups of users who display data-patterns that violate a Social Physics invariance. This is done by tracking the dynamics of certain statistical properties that portrays the synchronous nature of the activity of the users. For tis implementation, the aforementioned does not require the actual values contained in the data, but alternatively can be done by a fully hashed replacement. This enables the customer to provide Endor with a fully hashed dataset, while still benefitting from its superior predictive capability. In addition, as Endor is 100% semantics agnostic, the names of the data-fields can be hashed as well.

Archived copy · cryp2nova.com

An example of such hashing for financial records appears below:

Header before hashing:				
ACCOUNT NUMBER	**BRANCH**	**GENDER**	**TYPE OF TRANSACTION**	**DESTINTATION ACCOUNT**
Data records before hashing:				
183972	*291/30*	*Male*	*Transfer*	*382732*
183972	*291/30*	*Male*	*Balance Inquiry*	*N/A*
382732	*291/30*	*Female*	*Transfer*	*439001*
…				
Header after hashing:				
Field1	**Field2**	**Field3**	**Field4**	**Field5**
Data records after hashing:				
AjF32sdx	*Q2KPbv3A*	*Wsqp289X*	*q8Vb3MAs*	*Je2qx92n*
AjF32sdx	*Q2KPbv3A*	*Wsqp289X*	*q8Vb3MAs*	*x3PNm78A*
Je2qx92n	*Q2KPbv3A*	*m28SbA12*	*q8Vb3MAs*	*yL19B4GQ*
…				

Archived copy · cryp2nova.com

2.5. SOCIAL PHYSICS: MATHEMATICAL EXPLANANATION

2.5.1. Framework

We first introduce the basic principles of Endor's engine with their generic mathematical formalism. This is followed by two examples of possible implementations: (1) computer-vision-oriented and (2) social-graph-based. The first example demonstrates Endor's drawback when used with sensor-based data (that is human-unrelated), whereas the latter illustrates the concepts of Social Physics, and its benefits in predicting human behavior.

Note: throughout this discussion we provide numerous mathematical illustrations for the principles of Social Physics and the way it is used by the Endor engine. However, certain mathematical details regarding the Social Physics Laws were omitted from this discussion due to IP considerations.

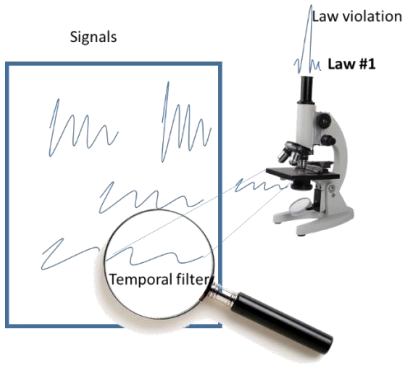
Let $d(x,t)$ be a temporal data stream, where x represents a single data point.

Let $L(\cdot)$ be a Law Operator which transforms the raw data $d(x,t)$ into a Law Representation:

$$L_{X,T}(d) = \frac{1}{|X||T|}\int_X \int_T L\left(d(x,t)\right)dxdt$$

The Law itself is formulated as an equation that equates the Law Operator to an a-priori constant **C** (which can be a number, a distribution class, such as a Power Law, etc.). This **C** represents the invariant represented by the Law.

For the purpose of illustration, we can imagine a hypothetical Law that represents the understanding that the change of the output of a white noise signal over time is limited to a very small threshold. A violation of this Law would take the form, for example, of a signal that displays a sudden strong output-spike. A mechanism that filters signals and can detect such anomalous outputs can then be built, in order to find violations of this Law. Naturally – using such a mechanism would only make sense in the case of signals known to be governed by the Law in question (i.e. origins of white noise), as for other signals such spikes cannot be classified as "violations".



Archived copy · cryp2nova.com

Another example can be an X-ray machine, built in order to detect anomalous "chunks" of radiation-absorbing materials. In this example the X-ray device (and the technician operating it and deciphering the resulting images) serve as a "violation detector" for analyzing 2-dimensional information streams and locating violations of an invariance that asserts that a coherent X-ray beam that hit a film should create an image of equal absorption (more or less, depending on the quality of the film, and the coherence of the beam).

This invariance, or Law, is a known physical fact, and it is used for medical applications by artificially inducing coherent X-ray beams onto high-quality films, while passing through a third substance – with the aspiration of finding "violations" that would indicate the existence of an X-ray absorbing element in this substance. Bones for example are such a material, representing in this case a "real world phenomenon" that is detected in the "data" (that is the exposed film), manifested by its unequal absorption of radiation, that in itself is a violation of the aforementioned Law.



Concrete formal mathematical examples for the implementation of the above principles appear in Sections 2.5.2 and 2.5.3.

In the case of Social Physics such violations would comprise a group of people, having generated a certain dynamic in the input data, analytically known to be statistically highly improbable, under the assumption that the input data was originated from human activities, and therefore adheres to the Social Physics Laws.

2.5.2. Validating that a Data-Subset is a Violation of Laws

Endor's engine implements Law Operators and Constants of several Social Physics Laws (the specific details of which are not described due to IP considerations). In this section we describe the validation of a Law Violation. It is important to note that as in many other computation problems, the validation that a specific signal consists of a Law Violation is fundamentally different to the detection of such interferences. In this section only the former is discussed, namely – given the validation whether a "potential interference" is indeed a Law Violation or

Archived copy · cryp2nova.com

not. The details of Endor's search algorithm that can efficiently scan a large scale high-dimensional data source and perform such on-the-fly validation will not be discussed here.

The mathematical formulation of the Laws we validate is always given in the following form:

$$L_{X,T}\left(d\right)=C$$

Given the explicit formulation of a Law, local deviations from it can be validated by measuring their deviation from it, denoted by ξ, as follows:

$$\xi\left(\Delta x,\Delta t\right)=L_{\Delta x,\Delta t}\left(d\right)-C$$

Here Δx represents a subspace of X, whereas Δt represents a temporal window. This deviation can thus be calculated for every subspace of X and any period of time, and generates a measure of how much that subspace violates the relevant Law, during the given time period.

By comparing this measure to a pre-defined threshold $\xi_{threshold}$ the subspaces that violate the Law can be detected:

$$r\left(\Delta t\right)=\left\{\Delta x:\left|\xi\left(\Delta x,\Delta t\right)\right|>\xi_{threshold}\right\}$$

The violation threshold $\xi_{threshold}$ is selected such that the spontaneous emergence of a signal that would defer from the Law further than the threshold is highly improbable. This enables automatic verification that a certain data subset is a violation of a Law, with a high-enough statistical significance, without any prior knowledge of the semantics of the data itself.

Notice that as the signal changes both in time and space, different temporal windows can create different subspaces that are detected as Law Violations. Endor uses a pre-defined fixed set of temporal windows that (derived from the Laws and not from the data) : Δt = 1-day, 7-days, 30-days, 90-days.

If the data is highly dynamic, the longer temporal windows are unlikely to generate any deviation groups; if the data is static, the shorter temporal windows are unlikely to generate any deviation groups. Regardless, none of the windows generate "junk-groups", because by definition – noise cannot generate a consistent Law Violation (or in more formal terms, the probability that noise will generate a large enough violation of the Law, is close to zero, due to the fact that this is the way the threshold $\xi_{threshold}$ is selected).

Archived copy · cryp2nova.com

The Knowledge Sphere is an aggregation of all group deviations from all Laws, for all relevant temporal windows:

$$K_{sphere} = \left\{ r\left(\Delta t\right) : \forall \Delta t, L \right\}$$

This Knowledge Sphere is calculated once per data-set, as this process is unaffected by the queries being asked, but rather the internal behavioral structure originating from the raw data. From an abstract point of view, Endor's Social Physics engine "compresses" the anonymized raw data into behaviorally relevant canonical representation.

2.5.3. Examples

Following are two examples that demonstrate the use of Laws for detection violating-patterns in data. By way of illustration we use known mathematical phenomena for demonstrating this mechanism.

Example 1: Vision

Let d(x,t) denote the color of a specific pixel x at a specific frame t. We may now define the Same Color Law, dictating that every color of any sub-region must be some pre-defined color C:

$$L_{X,T}\left(d\left(x,t\right)\right) = \frac{1}{|X||T|}\int_X \int_T d\left(x,t\right) dxdt$$

This Operator takes region x and time window t and calculates the average color for this input data.

This example illustrates the concept of Laws by negation: there is no inherent a-priori known law about pixels in a video (certainly not one that assumes every sub-region is of the same average color…). Considering this Low Operator and a constant C = Light Blue, we can now define Δx as a square window of size N×N pixels, and use a single frame Δt = 1.

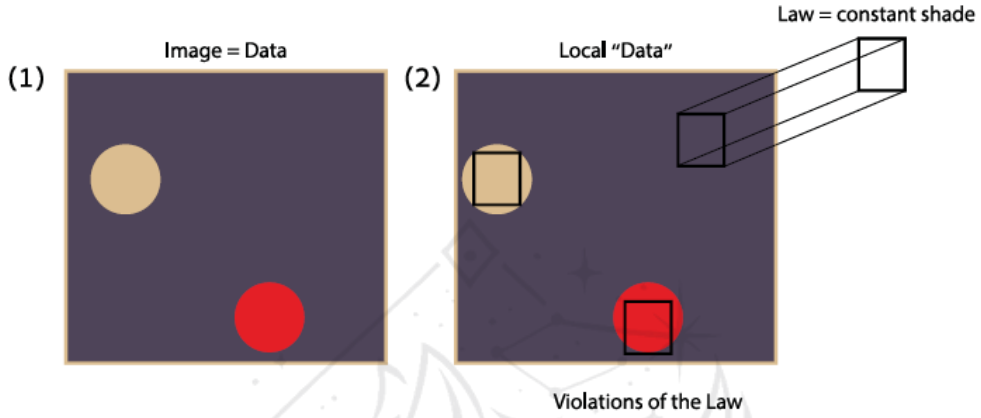
This yields the following as the local deviation of N×N windows from Light Blue:

$$\xi\left(\Delta x, \Delta t\right) = L_{\Delta x, \Delta t}\left(d\right) - C$$

In 2-dimenional images this can easily be applied to measure the deviation of patches from Light Blue, comparing this deviation to a pre-defined threshold. Patches whose deviation surpass this threshold will be classified as "clusters of pixels with a similar property", representing all the pixels in the square that violated the "Light Blue law".

Archived copy · cryp2nova.com

We can illustrate this using a slightly more generic hypothetical Law: let's assume that images are always "smooth" (namely, they are monotonous in smaller scales, lacking "peaks", or local maxima\minima). Under this assumption, we can observe the following "data" – an image that contains a "lawful" background of smooth color, with two patches (one Blue, and the other Red). A sampling of a large number of small square-shaped random regions would easily locate these two "violations":



Then, a "query" can be asked, in the form of Red pixel. Such a pixel would be identified as being contained in the Red violation, returning the Red patch as the result.

Note again that the intention of these examples is to illustrate the mathematics and mechanics of Social Physics rather to suggest that it is advantageous for a simple vision-based application, as such applications can be well addressed with traditional computational vision or deep learning methods.

Example 2: Scale-Free Networks

In this example $\boldsymbol{d}(\boldsymbol{x},\boldsymbol{t})$ abstractly represents a graph with x being the graph's nodes. The Law Operator is the degree-distribution operator, formulated as:

$$\bar{L}(x) = l_n(x) = \begin{cases} 1 & x \text{ has degree } n \\ 0 & \text{otherwise} \end{cases}$$

This vector operator generates 1 for the degree of each node. The summation of the result of this operator over all the graph's nodes yields a cardinality vector for the graph's degrees (equivalent to the degrees distribution, when dividing by the number of nodes).

Archived copy · cryp2nova.com

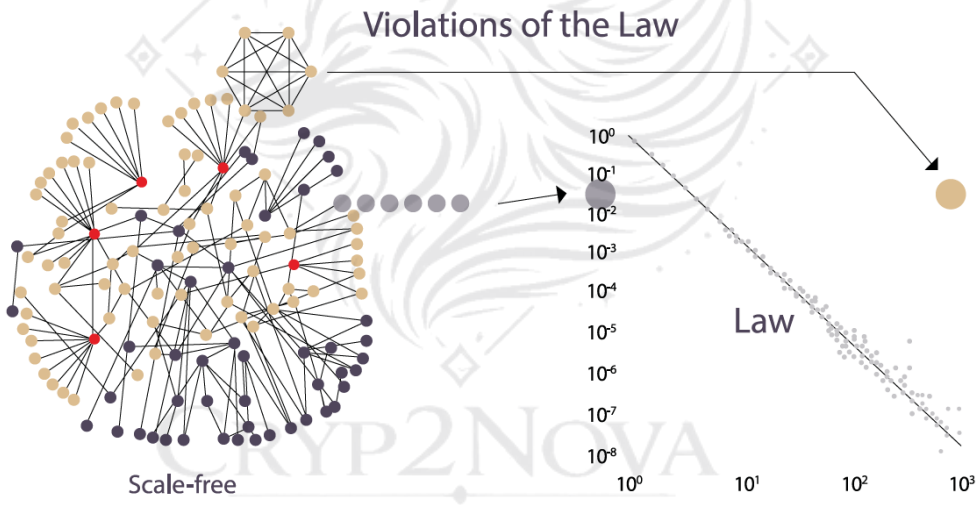
In this example we shall assume that the graph is a Scale-Free network. Therefore a Law Constant that assumes the power-law degree distribution can be applied (for some normalization constant α):

$$\bar{C} = c_n = \alpha \cdot n^{-\gamma}$$

This Law Constant can then be formulated as:

$$L_{X,T;n}(d) = \frac{1}{|X||T|}\int_X \int_T L_n\left(d(x,t)\right)dxdt = C_n$$

This Law implies that the overall graph should obey a power law distribution of the degrees of all its node. However, in many large real-world scale-free graphs there could be significant local deviations from such distribution. This may occur for example around cliques (i.e. fully-connected sub-graphs) or chains (i.e. sub-sets of the nodes that form a connected tree with no node having more than 2 neighbors). Such deviation, or violations, of the Law are illustrated in the following chart, containing their manifestation in both a structural representation (left) and as an adjacency matrix (right):



Note that given the Law, such violations can easily be validated by a variety of measures, such as:

$$\xi(\Delta x, \Delta t) = \sum_n \left| L_{\Delta x, \Delta t;n}(d) - C_n \right|^2$$

Archived copy · cryp2nova.com

This deviation measures the cumulative square of the differences (whereas another example for such measure can be the *KL-divergence* of both probability distributions). Here, Δx represents all possible subgraphs of the graph. Obviously, scanning all possible subgraphs in an input graph is not feasible, as it is a member of a class of "difficult problems" known as "*Non-Polynomial Hard problems*". In this sense, it is important to distinguish between validating a Law Violation (that requires knowing the details of the Law) and detecting a Law Violation (that requires a set of proprietary techniques that are specifically developed for each Law). Such detection techniques are part of Endor's proprietary technology, tailored-made for the Social Physics mathematical laws.

Returning to our scale-free example, assuming we possessed an efficient technique for finding such local interferences in graphs, they would have resulted in a collection of sub-graphs that can be formulated as follows:

$$r\left(\Delta t\right)=\left\{\Delta x:\left|\xi\left(\Delta x,\Delta t\right)\right|>\xi_{threshold}\right\}$$

And the Knowledge Sphere implied by this Law is defined as:

$$K_{sphere}=\left\{r\left(\Delta t\right):\forall\Delta t,L\right\}$$

Once a Knowledge Sphere containing the collection of violations of Social Physics Laws is available, it can be used to detect "lookalikes" for any given labeled exemplar, defined as a list of objects from the same domain. In our example, given a list of graph nodes all of the other graph nodes can be scored according to how many clusters they share with the labeled exemplar. Alternatively, different scoring metrics can be used, as long as they solely rely on the detected clusters and the labeled list as input, in order to produce an output in the form of a population scoring. We refer to such metrics, or scoring mechanisms as "Scorers", and will elaborate on these in the next section.

Note that different temporal windows can generate different Knowledge Spheres, representing very different associations among the graph nodes. In addition, these clusters represent behavioral connections that are not generated from external data sources such as social media or social networks, but rather from the customer's own internal transactional data source(s). This enables Endor to detect implicit behavioral clusters that are not explicitly manifested in any available data.

Archived copy · cryp2nova.com

2.5.4. Answering Questions: Simple "Scorer"

Once the Knowledge Sphere is available the user can start asking business-relevant questions by providing Queries – lists (possibly very small) of labeled data with semantic meaning:

$$y \in X.$$

In this section we give two examples of "Scorers" – functions that use the behavioral clusters + query to generate a ranked population list as output.

The first scorer we discuss is a simple co-clusters aggregation scorer that for each candidate $\tilde{y} \in X$ calculates the following score:

$$score_{\tilde{y}} = \sum_{\Delta x \in K_{sphere}} \begin{cases} 1 & \tilde{y} \in \Delta x, y \in \Delta x \\ 0 & \text{otherwise} \end{cases}$$

Each possible "candidate" is scored according to the number of clusters in the Knowledge Sphere that they co-inhabit with members of the labeled data. This simple scorer aims for counting the number of behavioral similarities of the members of X with the input labeled list. The list of objects that share clusters with objects in the labeled input list is represented as:

$$\tilde{y} = \left\{ x \in \Delta x : y \in \Delta x, \Delta x \in K_{sphere} \right\}$$

For example, given a colored point and the "*Same Color Law*" from Example 1, the colored area around a point that is given as input. In the "*Scale Free Graphs Law*" given a node the output would be the nodes in all subgraphs containing that node that locally violate the degree distribution.

CRYP2NOVA

Archived copy · cryp2nova.com

2.5.5. Answering Questions: Deep Learning Scorer

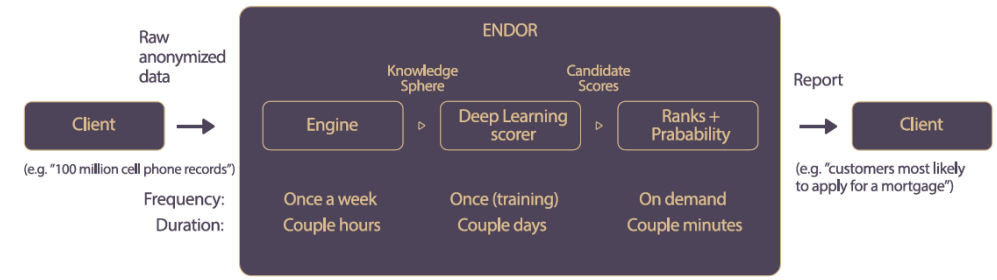
The scorer discussed in the previous section is an intuitive example of the way the clusters comprising the Social Physics Knowledge Sphere can be used in order to produce high quality "lookalikes prediction" for any "examples input" (i.e, a Query), without knowing in advance the nature of this query. This simple example illustrates the advantages of the social physics canonical data representation. However, in order to produce an accurate prediction Endor developed a more robust Scorer based on state-of-the-art deep learning algorithms.

As noted above, where human behavior is concerned, an efficient use of deep learning requires vast amounts of data for its training phase. This is hard to achieve using the raw data, as each labeled query usually contains a small number of labels, and there are typically only a handful of instances per query, and also a relatively small number of different queries per client. This type of complex, dynamic and small-sized label data is not well suited to deep learning. Endor's Social Physics engine overcomes this limitation by transforming all datasets from all clients, and all instances of all the queries into a large collection of a single canonical form: Knowledge Sphere that is composed of clusters of people, and examples that refer to these clusters. Combining all data from all these sources enabled Endor to create a large enough labeled training-set to train a deep-learning network that scores each person, given the labels and the Knowledge Sphere. This process is done once, resulting in a high-quality trained deep-learning model, that can then be used to efficiently process any new clusters set that is produced from new data sources, and new queries.

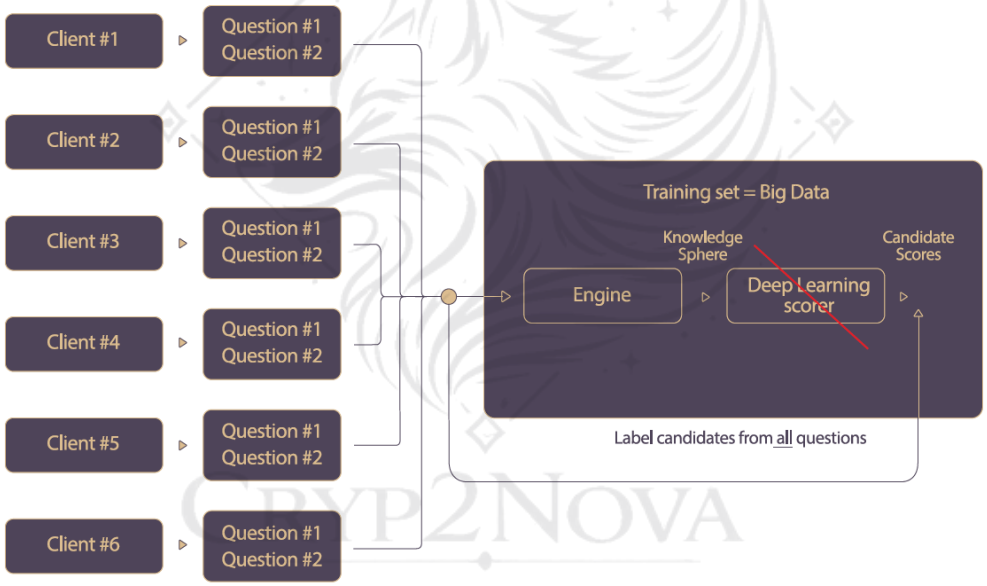
Note that this trained model does not need to be periodically retrained – not for new queries, nor for new types of datasets, as it was trained on (many) instances of data represented as a collection of clusters (i.e. the Social Physics canonical representation form). Using this method, Endor combines the advantages of both Social Physics and Deep Learning: Social Physics transforms anonymized raw data into canonical form based on violations of Social Physics Laws, whereas Deep Learning algorithms then score candidates based on the created Knowledge Sphere and the (possibly few) labeled data.

Archived copy · cryp2nova.com

This flow is illustrated in the following charts:







Archived copy · cryp2nova.com

2.6. ROBUSTNESS TO NOISE

One of the main strengths of the Endor predictive platform is its high resilience to data-gaps and noisy data. Traditionally, every data-analytics project begins with a "data sanitation" phase, consisting of the detection and desired mitigation of undesired data segments such as:

- Data gaps (i.e. periods of missing data, either full or partial)
- Gibberish insertions in the raw data
- Semantic ambiguities, such as a category name that may appear in the data in more than one form
- Normalization issues of numeric values
- Required binning of numeric values
- …

Machine Learning algorithms are usually highly susceptible to noise and given that data is normally noisy, the phase of mitigating those data problems is typically costly and protracted. The reason for this is that whereas in the real world data is generated in its raw form (financial transactions, phone calls, etc.) in order for it to be analyze-able by traditional Machine Learning techniques it must be converted to an aggregative form, referred to as "features", or "properties". This aggregation can in turn be significantly affected by a small number of wrong data values, or different quantities of values for different users.

Endor overcomes this requirement by analyzing the raw data itself (as described in detail in the previous sections). In addition, Endor's engine does not perform statistical analysis of the data in the search of patterns that can be used for prediction, but rather – uses the Laws of Social Physics – mathematical invariances, external to the data, and unaffected by it. This approach has several significant advantages, stemming from the following basic notion:

Noise cannot create a data-pattern that "cannot arbitrarily emerge"

(where the latter is defined as data-patterns that can analytically be shown to exist in negligible probability)

In order to understand how this observation allows for the automatic extraction of data insights from any human-data without any prior cleaning, let us recall that the extraction process searches for groups of objects in the data that violates one of the Social Physics Laws. Namely, groups that display data patterns that we can prove cannot naturally emerge in the data (using the mathematical analysis enabled by Social Physics).

This means that whereas noise can indeed hide insights from our engine, it can, by definition, (almost) never create a data pattern that would be detected as a Law violation. Noisy data cannot violate a Social Physics Law, only human-driven signal data can.

Archived copy · cryp2nova.com

3. RESULTS

This section presents a variety of use-cases that illustrate how the Endor prediction system can be utilized. A detailed description of the overall prediction process, using data which comprises 7 days' worth of activity of a large financial investment platform to accurately answer 4 different predictive questions (including a comparative analysis with Google's Tensor Flow deep learning platform);

1. An example of using Endor in a fully automatic way in order to crack a Kaggle Challenge.
2. A Coca-Cola case-study in which Endor provides accurate predictions for a multitude of business questions using point-of-sale data, in less than 24 hours.

3.1. USING FINANCIAL ACTIVITY OVER 7 DAYS TO ACCURATELY AND AUTOMATICALLY ANSWER FOUR PREDICTIVE QUESTIONS

In this section we demonstrate the overall prediction process using the Endor system:

- Description of the raw data used
- The transformation of the data to the Social Physics representation (namely, the knowledge sphere that comprised a set of behavioral clusters)
- The definition of four predictive questions
- The manifestation of the queries in the Social Physics form
- The overall predictive accuracy
- A comparison to Google's Tensor-Flow deep learning platform.

3.1.1. Data:

The data that was used in this example originated from a retail financial investment platform and contained the entire investment transactions of members of an investment community. The data was anonymized and made public for research purposes at MIT (the data can be shared upon request).

Following is an overall summary of the dataset:

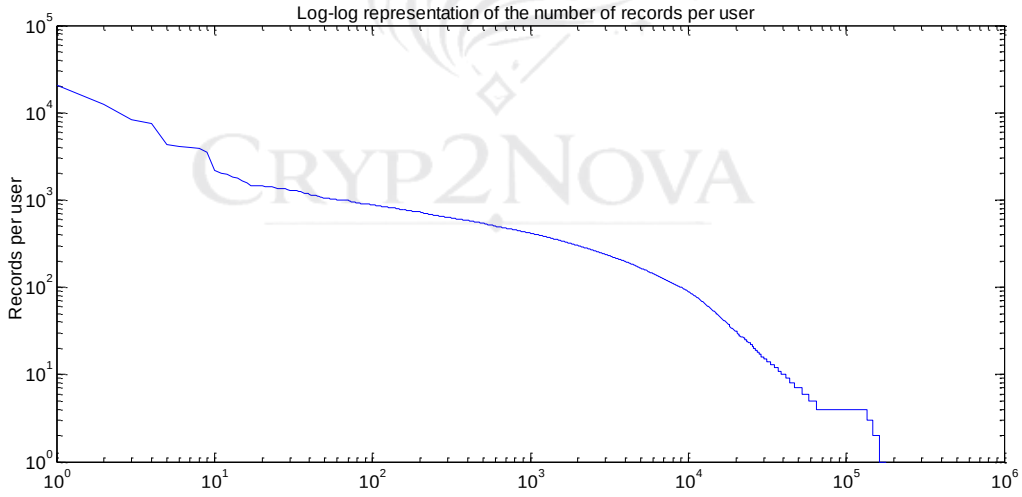
- 7 days of data
- 3,719,023 rows
- 178,266 unique users
- No contextual or semantic interpretation of the data was given
- 12 data fields:

Archived copy · cryp2nova.com

FIELD NAME	TYPE		# UNIQUE VALUES
Time1	Time		501573 unique
Time2	Time		4 unique
User ID	INT32	Categorical	178266 unique
Record ID	INT32	Categorical	1053574 unique
Property1	INT32	Categorical	24 unique
Property2	INT32	Categorical	7 unique
Property3	INT32	Categorical	134 unique
Property4	INT32	Categorical	77527 unique
Property5	INT32	Categorical	10 unique
Property6	INT32	Categorical	27 unique
Property7	INT16	Categorical	9 unique
Property8	Double	Numeric	3772 unique

The following important aspects of the dataset should be noted

- The data was given in its raw non-aggregated form, and contained events on a user-level.
- No contextual or semantic interpretation of the data was provided.
- No data-sanitation was done. The data contained noises, gaps, duplicate records, and so on.
- The data contained an extremely uneven distribution of records-per-user (78% of the users has 10 or less transactions, but there are 4,000 with >200 records. Median is 4 records per user):



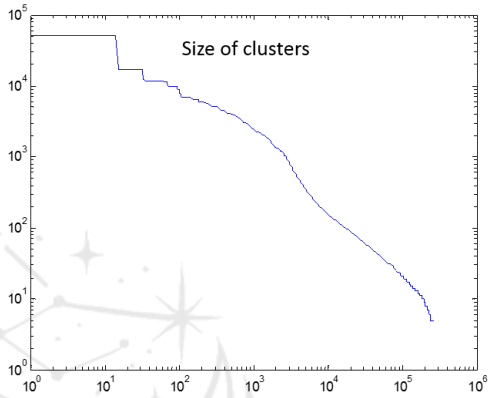
Archived copy · cryp2nova.com

3.1.2. Automatic Clusters Extraction

Upon first analysis of the data the Endor system detects and extracts "behavioral clusters" – groups of users whose data dynamics violates the mathematical invariances of the Social Physics. These clusters are based on all the columns of the data, but is limited only to the last 7 days – as this is the data that was provided to the system as input.

Following is a summary of the behavioral clusters from the dataset that were detected by the system:

Number of clusters:	268,218
Clusters sizes:	62 (Mean), 15 (Median), 52508 (Max), 5 (Min)
Clusters per user:	164 (Mean), 118 (Median), 703 (Max), 2 (Min)
Users in clusters:	102,770 out of the 178,266 users
Records per user:	6 (Median), 33 (Mean): applies only to users in clusters (Compared to 4 and 21 in general population)



3.1.3. Prediction Queries

The following prediction queries were defined:

- **New users to become "whales":** users who joined in the last 2 weeks that will generate at least $500 in commission in the next 90 days
- **Reducing activity :** users who were active in the last week that will reduce activity by 50% in the next 30 days (but will not churn, and will still continue trading)
- **Churn in "whales":** currently active "whales" (as defined by their activity during the last 90 days), who were active in the past week, to become inactive for the next 30 days
- **Will trade in Apple share for the first time:** users who had never invested in Apple share, and would buy it for the first time in the coming 30 days

As can be seen, all of the above questions refer to data extending beyond the 7 days dataset – both past data (used to generate the "search population", or "examples"), and future data (used for validation of the predictions). In order to avoid providing the system with any information external to

Archived copy · cryp2nova.com

the 7-day period, the queries were formulated in the form of lists containing users_IDs values. For example:

- **Query name:** "New users to become "whales"
- **Search population:** a list of user_IDs containing users who joined in the past 2 weeks (prior to *End_of_data*).
- **"Examples":** a list of user_IDs containing users known to be "whales" (i.e. users who generated more than $500 in commission in the 90 days prior to *End_of_data*).
- **List of targets that need to be found:** a list of user_IDs, that is a subset of "Search population", containing users who generated more than $500 in commission in the 90 days after *End_of_data*. This list was only used for validation purposes, and was not provided to the system.

3.1.4. Knowledge Sphere Manifestation of Queries

It is again important to note that the definition of the search queries is completely orthogonal to the extraction of behavioral clusters and the generation of the Knowledge Sphere, which was done independently of the queries definition. Therefore, it is interesting to analyze the manifestation of the queries in the clusters detected by the system: do the clusters contain information that is relevant to the definition of the queries, despite the fact that:

- The clusters were extracted in a fully automatic way, using no semantic information about the data, and
- The queries were defined after the clusters were extracted, and did not affect this process.

This analysis is done by measuring the number of clusters that contain a very high concentration of "samples"; In other words, by looking for clusters that contain "many more examples than statistically expected". A high number of such clusters (provided that it is significantly higher than the amount received when randomly sampling the same population) proves the ability of this process to extract valuable relevant semantic insights in a fully automatic way.

The following table illustrates this observation, comparing the number of behavioral clusters that contain a certain amount of "Samples", compared to the number of "Random Clusters" that contain the same amount of samples. Random clusters refers to a set of N groups of users randomly sampled from the customers population, such that N equals the number of behavioral clusters detected by the system, and that the sizes of these randomly sampled groups equals the sizes of the clusters detected by the system. This is used to demonstrate the information encapsulated by the behavioral clusters, as they contain significantly more clusters of high consistency of "target users" compared to the random samples (and recalling they were detected by the system prior to the definitions of the "questions").

The number of "samples" requested is given in units of "Baseline". Namely – "X5 of baseline" for the "Reducing Activity" query (for which the baseline is approximately 11%) means clusters that have 55% overlap with the "samples" (namely, that 55% of their members also appear in the list of "Previous users who reduced activity").

Archived copy · cryp2nova.com

Clusters containing <u>many</u> target customers		*#Random Clusters*	*#Behavioral Clusters*
Reducing activity	X2 from baseline	0	98
	X5 from baseline	0	11
Churn in "Whales"	X2 from baseline	212	1678
	X5 from baseline	67	525
	X10 from baseline	21	114
Never Bought	X2 from baseline	3962	60044
	X10 from baseline	1332	25542
	X20 from baseline	415	9090
New Whales	X2 from baseline	38	1898
	X5 from baseline	0	65

3.1.5. **Prediction Results**

The following table illustrates the accuracy of the predictions for the four queries

- Baseline: the average portion of requested target customers in a random sample of the population, representing the accuracy of a random guess.
- Candidates: the size of the search population. For example, in the "New users to become whales", the number of candidates refers to the number of new users.
- Top-100: the portion of requested targets customers in the top-100 members of the prediction report (similarly, for Top-250 and Top-500).

	Baseline	*Top 100*	*Top 250*	*Top 500*
New users to become "Whales" Users who joined in the last 2 weeks that will generate at least $500 in commission in the next 90 days	**7.5%** (170 out of 2270 candidates)	**37%**	**28.8%**	**21%**
Reducing activity Who among the current active users will reduce activity by 50% in the next 30 days (but will not churn)	**11.4%** (255 out of 2233 candidates)	**21%**	**23%**	**20.2%**
Churn in "Whales" Currently active "whales", who were active in the past week, to become inactive for the next 30 days?	**1.66%** (69 out of 4141 candidates)	**10%**	**10.8%**	**9.2%**
Will trade in Apple for the first time Users who had never bought Apple share, and will buy it for the first time in the coming 30 days.	**0.5%** (839 out of 161382 candidates)	**14%**	**12%**	**10%**

As can be seen, and as expected, accuracy decreases as we reach deeper to the predictions list.

Archived copy · cryp2nova.com

3.1.6. Comparison to Tensor-Flow

In this section a comparison between prediction results obtained by the Endor system and Google's Tensor Flow is presented. It is important to note that Tensor Flow, like any other Deep Learning library, faces some difficulties when dealing with data similar to the one under discussion:

- An extremely uneven distribution of the number of records per user requires some canonization of the data, which in turn requires:
    - Some manual work, done by an individual who has at least some understanding of data science.
    - Some understanding of the semantics of the data, that requires an investment of time, as well as access to the owner or provider of the data
- A single-class classification, using an extremely uneven distribution of positive vs. negative samples, tends to lead to the overfitting of the results and require some non-trivial maneuvering. This again necessitates the involvement of an expert in Deep Learning (unlike the Endor system which can be used by Business, Product or Marketing experts, with no perquisites in Machine Learning or Data Science).

We have asked an expert in Deep Learning to spend 2 weeks crafting a solution that would be based on Tensor Flow and has sufficient expertise to be able to handle the data. The solution that was created uses the following auxiliary techniques:

- Trimming the data sequence to 200 records per customer, and padding the streams for users who have less than 200 records with neutral records.
- Creating 200 training sets, each having 1,000 customers (50% known positive labels, 50% unknown) and then using these training sets to train the model.
- Using sequence classification (RNN with 128 LSTMs) with 2 output neurons (positive, negative), with the overall result being the difference between the scores of the two.

CRYP2NOVA

Archived copy · cryp2nova.com

The table below compares the results obtained using these techniques (red) to Endor's predictions (blue):

	Baseline	*Top 100*	*Top 250*	*Top 500*
New users to become "Whales" Users who joined in the last 2 weeks that will generate at least $500 in commission in the next 90 days	**7.5%** (170 out of 2270 candidates) (2135 Examples)	**37%** 21%	**28.8%** 27.2%	**21%** 19.6%
Reducing activity Who among the current active users will reduce activity by 50% in the next 30 days (but will not churn)	**11.4%** (255 out of 2233 candidates) (366 Examples)	**21%** 8%	**23%** 18.8%	**20.2%** 19.4%
Churn in "Whales" Currently active "whales", who were active in the past week, to become inactive for the next 30 days?	**1.66%** (69 out of 4141 candidates) (21156 Examples)	**10%** 11%	**10.8%** 12.4%	**9.2%** 8.4%
Will trade in Apple for the first time Users who had never bought Apple share, and will buy it for the first time in the coming 30 days.	**0.5%** (839 out of 161382 candidates)	**14%** 1%	**12%** 0.8%	**10%** 1%

Observations:

- Endor outperforms Tensor Flow in 3 out of 4 queries, and results in the same accuracy in the 4th.
- The superiority of Endor is increasingly evident as the task becomes "more difficult" – focusing on the top-100 rather than the top-500.
- There is a clear distinction between "less dynamic queries" (becoming a whale, churn, reduce activity" – for which static signals should likely be easier to detect) than the "Who will trade in Apple for the first time" query, which are (a) more dynamic, and (b) have a very low baseline, such that for the latter, Endor is X10 times more accurate!
- As previously mentioned – the Tensor Flow results illustrated here employ 2-weeks of manual improvements done by a Deep Learning expert, whereas the Endor results are 100% automatic.

Archived copy · cryp2nova.com

3.2. AUTOMATICALLY CRACKING A KAGGLE CHALLENGE IN 3 HOURS

In another example, we have tested Endor's system with the publicly available data from the Kaggle competition known as "Acquire Valued Shoppers Challenge". The data contained nearly 300 million point-of-sale records, referring to hundreds of thousands of customers, whereas the challenge entailed predicting which users who received a certain promotional coupon would become a recurring customer.

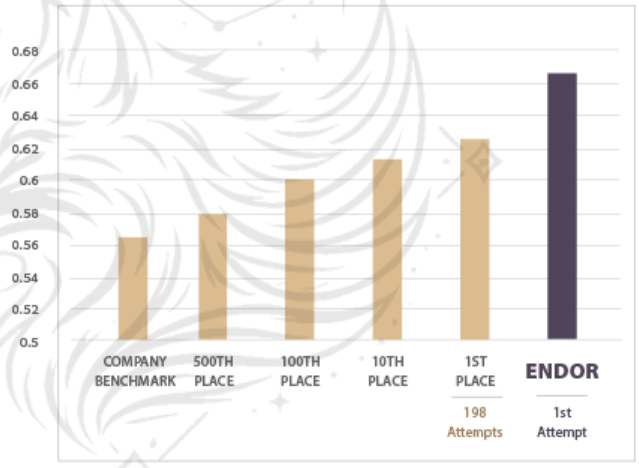
The original challenge encompasses 952 teams and lasted 3 months.

The Endor machine, running fully automatically on the raw data given to the challenge's participants, was able to produce predictions that outperformed the team who originally won first place.

kaggle

- 952 Competing data science teams
- 3 months challenge

ENDOR:
Few clicks, 1st Place



CRYP2NOVA

Archived copy · cryp2nova.com

Appendix B – Endor's Common Use-Cases for Enterprises

What would it be like to have your own personal Oracle? What would it be like to have access to the powerful data engines which only the likes of Google and Facebook exploit? What would it be like if you could use the most powerful such engine to date Endor to get the most reliable answer any available tech can give? With the vision to make Blockchain prediction accessible and useful for all by democratizing advanced machine learning, *Endor.coin* empowers you with that ability, which until now was reserved only to technological giants, equipped with internal teams of professional data-experts.

Following is a summary of the common prediction use-cases requested by Endor's enterprise customers. It categorizes the use-cases by the main segments Endor currently serves:

- Retail Banks
- Insurance Companies
- Retails and e-Commerce

However, thanks to the use of *Social Physics*, new use-cases can easily be supported – on average requiring a few hours of an Endor's sales engineer.

CRYP2NOVA

Archived copy · cryp2nova.com



Archived copy · cryp2nova.com

COMMON USE CASES



SELL MORE

- Propensity to buy
- Cross sell
- Up sell
- Big spenders
- New services - early adopters

QUESTION EXAMPLES

- Propensity to buy: out of all qualified active customers in the past 3 months, who will take a loan above $X given a call next week? who will take a loan without having a credit card? short term loan? Long term loan?
- Cross sell: out of customers who used product X in the past 6 months, who will start using product Z given promotion Y next month
- Big spenders: out of new customers joined in the past 30 days, who is likely to spend above $X per month, in the next 3 months?
- New service: out of all active customers in the past 3 months, who will register to a new service that we launched 2 weeks ago?

ENDOR

CRYP2NOVA

COMMON USE CASES

SELL MORE



ADDITIONAL EXAMPLES

- Out of all active credit card customers in the past 90 days, who will increase spend above $/%X if given a second card?
- Out of non-active credit card customers in the past 90 days , who will start using the card if we change from product A to B
- Out of all active customers in the last 6 months, who will carry a balance in the next 3 months ?
- Out of all active customers in the last 6 months, who will be willing to take Tasa 0 next month?
- Out of all active gold and platinum accounts, who will be willing to include a card on file charge next month?
- Out of interest generating customers in the last 6 months, who will pay off their balances in the next 3 months?
- Out of people purchasing plane tickets in our travel website, who will be interested in buying hotel nights as well?
- Which customers from Platinum and Gold portfolios will redeem points in high cost categories in the next 2 months?
- Out of platinum card high value customers, who will call to cancel their card in the next month?
- Out of active customers in the past 12 months, who is likely to spend above $Y, across all products, in the next 2 months
- Out of active customers of product Y in the past 6 months, who is likely to spend above $Z, on product Y, in the next 2 months
- Out of customers who didn't taken any credit in the last year, who will take a new credit next month if contacted next week

Archived copy · cryp2nova.com

COMMON USE CASES



MARKETING EFFICIENCY

- Promotions efficiency
- Stickiness across channels
- Customer loyalty

QUESTION EXAMPLES

- Promotions efficiency: out of active customers, who will register to premium service X, given a text message? call? promotion Y?
- Digital card: out of customers who bought at least Y times on the listed websites, who will buy using the payback service next month
- Out of customers who made at least one Duty Free purchase and didn't buy anything abroad in the past X months, who will buy abroad using the credit card in the coming 3 months?
- Out of all active customers, who will fly to NY or UK in the next 30 days
- Out of customers who joined in the last 90 days, who have not activated their card, who will activate if offered a $10 bonus? $20?
- Out of all active card holders in the last 6 months, who is likely to stop using our card next month?

ENDOR

CRYP2NOVA

COMMON USE CASES

MARKETING EFFICIENCY



ADDITIONAL EXAMPLES

- Out of customers who didn't open the email with the monthly promotion of Consumer Credit, who will take a new credit in the next 3 weeks if we (1) call them (2) send SMS next week
- Out of married customers who have 3.000 to 24.000 points, who will redeem points for fragrances in the next 4 days if we send them an email tomorrow?
- Out of customers who use the card in restaurants, and have a credit limit above $1 million, and live in Region X, who will increase spending on product X by over Y% next week, given an email this week
- Which customers will use our card to pay in gasoline stations in the next month if we send them a special discount for gasoline next week?
- Out of customers who renegotiated their outstanding debts in the last 6 months, who will become delinquent in the next month, given a SMS tomorrow reminding them to pay their account balance?
- Out of customers that didn't use their card for the last 2 months, who will start using it again if we send them a special discount for gasoline next week?

ENDOR

Archived copy · cryp2nova.com

COMMON USE CASES



DIGITAL CONVERSION

- Increase online activity
- Transition from branches to online
- Retargeting website visitors
- First time buyers, Repeating buyers, Retention
- Big spenders

QUESTION EXAMPLES

- Out of online customers who were inactive for 2 months, who will become active in the next 7 days given a call? Incentive X?
- Out of all customers who do traditional transactions (branch, call center), who will use digital given promotion X next week?
- Out of new online users in the last 7 days, who is likely to do in aggregate transactions above $X in the first 3 months?
- Out of all the customers who signed up for a new product through a digital channel in the past 3 months, who will sign up for a second product within the next month, given promotion X?
- Which website visitors in the last 3 months are likely to register to our online banking if we will retarget them?

ENDOR

CRYP2NOVA

COMMON USE CASES

DIGITA CONVERSION



ADDITIONAL EXAMPLES

- Out of new online users, who registered last month, who will continue using online in the next 3 months? who won't?
- Out of all our online banking customers, who will stop using online banking in the next month?
- Out of all new online users who registered last month, who will become a recurring online user given incentive X next week
- Out of all online customers, who use functionality X but do not use Y, who will use Y in the next month if contacted
- Out of all customers who signed up for a new digital product (X,Y,Z) in the past 3 months, who will sign up for a second product within the next month, given promotion X
- Out of all customers who acquired a product through a digital channel in the past 3 months, who will leave in 1 month
- Out all website visitors (cookie ID) in the last 3 months, who is likely to register to our online banking if retargeted
- Out of customers who have only one banking product, who is likely to purchase a second one in the next month?
- Out of customers who never logged into our website, who will do it on the next week if we send them an email tomorrow?

ENDOR

Archived copy · cryp2nova.com



Archived copy · cryp2nova.com

THE CHALLENGE

On a daily basis, how do you Predict? Decide? Boost performance?

You need to answer questions such as



SALES

- Who will add an ancillary coverage X
- Who is likely to upsell/cross sell if approached next week
- Among the ones who cancelled pre effective date, who will rebuy
- Who will buy a new product / service launched a week ago
- Who will apply for a new plan next week
- Out of new customers, who will become a big spender / premium



MARKETING

- Who will respond to promotion X
- Who will terminate a contract in first 90 days
- Who will use our mobile service (app)
- Who will travel to NY & Mexico in 2m'
- Who is likely to refer a friend?



DIGITAL

- Who will increase activity above $/%X
- Who will use a new digital service
- Which non-active will become active
- Who will convert from branch to online
- Who will respond to promotion X
- Which website users will convert if retargeted

ENDOR

CRYP2NOVA

COMMON USE CASES



SELL MORE

- Propensity to buy
- Cross sell
- Up sell
- Big spenders
- New services - early adopters

QUESTION EXAMPLES

- Propensity to buy: out of all potential customers, who is likely to add policy X? Out of all existing customers, who is likely to add ancillary coverage?
- Cross sell: out of customers of plan X, who is likely to purchase product/service Y, given promotion Z?
- Big spenders: out of new customers joined in the past 30 days, who is likely to spend above $X per month, in the next 3 months?
- New service: out of all active customers in the past 3 months, who will register to a new service that we launched 2 weeks ago?

ENDOR

Archived copy · cryp2nova.com

COMMON USE CASES



DIGITALIZATION

- Increase online activity
- Transition from branches to online
- Retargeting website visitors
- First time buyers, Repeating buyers, Retention
- Big spenders

QUESTION EXAMPLES

- Out of online customers who were inactive for 2 months, who will become active in the next 7 days given a call? Incentive X?
- Out of all customers who engage offline (branch, call center), who will use digital given promotion X next week?
- Out of new online users in the last 7 days, who is likely to spend above $X in the first 3 months?
- Out of all the customers who signed up for a new product through a digital channel in the past 3 months, who will sign up for a second product within the next month, given promotion X?
- Which website visitors in the last 3 months are likely to add coverage online if we retarget them?

ENDOR

CRYP2NOVA

COMMON USE CASES



MARKETING EFFICIENCY

- Promotion efficiency
- New products
- Customer loyalty and referrals
- Targeted promotions

QUESTION EXAMPLES

- Promotion efficiency: Out of active customers, who will register to premium service X, given a text message? call? promotion Y?
- Targeted promotions: Out of all active customers, who will travel to destination X in the next 30 days?
- Referrals: Out of all active customers, who is likely to refer a friend, given incentive X?
- New products: Out of all active customers, who is likely to use new digital service X, if approached next week?

ENDOR

Archived copy · cryp2nova.com

COMMON USE CASES



RETENTION

- Policy renewal
- Termination
- Returning customers

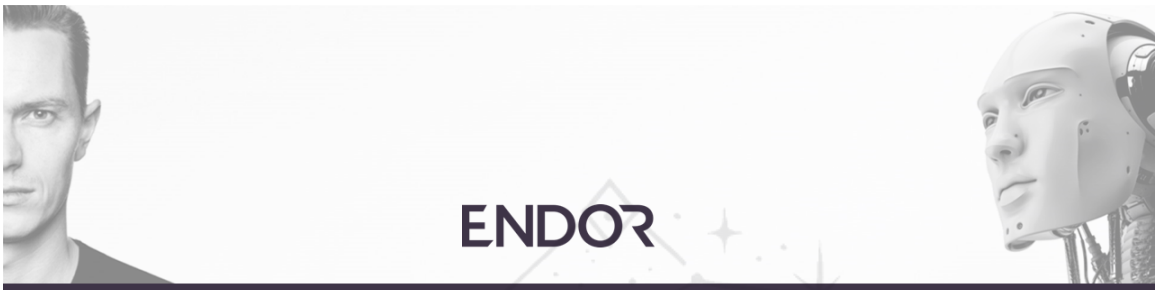
QUESTION EXAMPLES

- Termination: Out of all new customers of policy X, who is likely to terminate their contract in the first 90 days?
- Returning customers: Among all customers who cancelled before the effective date, who will re-sign?
- Renewal: Out of all customers with new policies in the last 60 days, who will cancel before the effective date?

CRYP2NOVA

ENDOR

Archived copy · cryp2nova.com



Use Cases : Retail and e-commerce

ENDOR

CRYP2NOVA

EXEMPLE USE CASES

On a daily basis, how do you Forecast? Decide? Boost performance?

You need to answer questions such as

SALES

- Who will buy X next month
- Who is likely to buy a product launched a week ago
- Who will convert to premium
- Who will become a big spender
- Which X customers will buy Y

MARKETING

- Who will respond to promotion X
- Which campaigns to maintain
- Which customers will buy online
- Who will use our mobile app
- Who will increase activity above $X

ECOMMERCE

- Which cookies IDs will convert
- Who will respond to promotion X
- Who will become first time buyer
- Who will increase activity
- Which inactive customers become active given promotion X

STORE/PRODUCT

- Which stores will reach $X sales of product Y
- Which stores will underperform
- Which products will experience >20% increase next month

STRATEGY

- Early Warning System
- Where we will get defection
- Where we will get tailwind for product X

ENDOR

Archived copy · cryp2nova.com

COMMON USE CASES



SELL MORE

- Propensity to buy
- Cross sell
- Up sell
- Future Big spenders
- New products - early adopters

QUESTIONS EXAMPLES

- Out of all active customers in the past 3 months, who is likely to buy 352553 in the coming 30 days?
- Out of all customers who bought product 219 in the past 30 days, who is likely to buy product 890 in the coming 90 days?
- Out of active customers of product 32236 in the past 3 months, who is likely to upgrade to premium in the next 7 days?
- Out of all active customers in the past 6 months, who is likely to buy a new product that we launched 3 weeks ago?
- Out of all new customers in the last 30 days, who will purchase in aggregate more than $10K in the first 3 months?

ENDOR

CRYP2NOVA

COMMON USE CASES



MARKETING EFFICIENCY

- Better segmentation
- Promotions efficiency
- Stickiness across channels
- Customer's loyalty

QUESTIONS EXAMPLES

- Out of customers spending in the top 10% for the last 6 months, who will reduce spend by at least 20% in 2 months?
- Out of customers spending in the bottom 20% for the last 6 months, who will increase activity by at least 30% in 3 months?
- Out of all customers, who will purchase product 6347, given a coupon of 20% discount next week?
- Out of all customers, who will purchase product 6347, if we will send them a text message next week?
- Out of all customers who bought through the website in the last 90 days, who is likely to use the mobile app in the next month?

ENDOR

Archived copy · cryp2nova.com

COMMON USE CASES



INCREASED ACTIVITY

- Efficient users acquisition (retargeting)
- First time buyer
- Repeating buyers
- Increased activity

QUESTIONS EXAMPLES

- Out of all website visitors (cookies), who is likely to become a customer, if retargeted via Google next week
- Out of all website visitors (cookies), who will purchase in aggregate of more than $10K in the first 3 months?
- Out of all customers which were not active in the past 6 months, who will become a FTB in the upcoming month?
- Out of all customers who placed their first grocery order in the last 7 days, who will place 2 new orders within 60 days?

ENDOR

CRYP2NOVA

MANY OTHER...



PRODUCTS / STORES

- Which services will experience >20% increase next month
- Which services are likely to experience >50% decrease in the next 3 months
- Which stores will reach >50% sales of category X in the next 3 months



OPERATIONS

- Who will contact the call center / technical support for X tomorrow
- Who will make transactions in branch X tomorrow



STRATEGY

- Where we will get defection from brand X next month (headwind)
- Where we will get tailwind for brand Y

THERE IS NO "PRE DEFINED MENU"
ANY BUSINESS USER ASKS WHATEVER NEEDED TO GROW THE BUSINESS

ENDOR

Archived copy · cryp2nova.com

Appendix C – Endor.coin Examples of Pre-Defined Predictions

While *Endor.coin* 's grand vision aims at any sector being transformed by Blockchain, such as Insurance, Banking, eCommerce or Health, to start with – we are offering an unprecedented prediction platform for cryptocurrency insights to support cryptoholders seeking trading leads. On Blockchain's trust proven decentralized infrastructure, allowing anyone to test any hypothesis, on any data source, without disclosing their actual trading strategies, *Endor.coin* empowers you to see the future before it becomes observable through the lenses of any other existing technologies.

Following is an example of some of the first pre-defined Blockchain predictions to be supported by the *Endor.coin* platform upon launch. New use-cases would gradually be added, as requested by the community, using the *RFP* (Request for Predictions) mechanism.

Cryptocurrency Addresses Predictions: These predictions receives a pre-defined list of addresses (i.e. "Bitcoin addresses that have had at least one outgoing transaction in the past month") and rank them according to their behavioral similarity to a pre-defined behavior in the recent past (i.e. "addresses that have received at least 0.1 Bitcoin in the past week"). The resulting list would contain at its top the addresses who most resemble the pre-defined behavior (and therefore are statistically more likely to display the same behavior in the future), whereas the bottom of the results list would contain addresses that least resemble the pre-defined behavior.

- **Active addresses:** From all addresses who were active at least once in the past month, which most resemble the addresses which significantly increased their number of transactions recently?
- **Heavy-trading addresses:** From all addresses who were active at least once in the past month, which most resemble the addresses whose overall transactions volume in the past month passed 10 BTC?
- **Becoming inactive addresses:** From all addresses who were active at least once in the past month, which most resemble the addresses which decreased their overall transactions volume by 50% recently?

Archived copy · cryp2nova.com

Token Predictions: These predictions receives a pre-defined list of token (i.e. "Tokens that had at least $10M trading volume during the last month") and rank them according to their behavioral similarity to a pre-defined behavior in the recent past (i.e. "tokens that increased their average monthly volume trading by 2"). The resulting list would have tokens that most resemble the pre-defined behavior (and therefore are statistically more likely to display the same behavior in the future) at the top, whereas tokens that least resemble the pre-defined behavior would be places at the bottom of the list.

- **Profitable tokens:** From all tokens with at least $1M USD trading volume in the last month, which ones most resemble tokens that have increased their price with respect to BTC by over 50% in the last month?
- **Non-Profitable Tokens:** From all tokens with at least $1M USD trading volume in the last month, which ones most resemble tokens that have decreased their price with respect to BTC by over 50% in the last month?
- **Volatile Tokens:** From all tokens with at least $5M USD trading volume in the last month, which ones most resemble the 10 most volatile tokens of last month?
- **Stable Tokens:** From all tokens with at least $5M USD trading volume in the last month, which ones most resemble the 10 least volatile tokens of last month?
- **Increasing Volume Tokens:** From all tokens with at least $20M USD trading volume in the last month, which ones most resemble tokens that have doubled their monthly trading volume in the last month?
- **Decreasing Volume Tokens:** From all tokens with at least $20M USD trading volume in the last month, which ones most resemble tokens that have decreased their monthly trading volume by 50% in the last month?

CRYP2NOVA

Archived copy · cryp2nova.com

Appendix D – Knowledge Sphere Class: API Access

This Appendix contains the specification of the "Knowledge Sphere" data structure – the basic building block of the *Endor.coin* protocol – as well as the complete description of the class that provides access to it, for implementation of future analytics engines, to be plugged to the *Endor.coin* network.

Usage Explanation: The clusters object consists of 3 entities:

- Sparse matrix M, of dimensions ($|\text{searchable Objects}| \times |\text{ Behavioral Clusters}|$), representing the connectivity between Searchable Objects and Behavioral Clusters, having $M_{i,j} = 1$ iff searchable object of index $i \in$ cluster j. Searchable objects refer to tokens, wallet addresses, locations, phone numbers – or any other type of objects that is included in the data, and serve as the basis of the prediction.
- Array A_M mapping each Searchable Object SO to an index in the sparse matrix M.
- DataFrame D_M containing miscellaneous clusters properties, as defined and calculated by each prediction engine. Such properties can be for example the size of the cluster, ratio of internal connectivity vs. external connectivity, the internal module used by the prediction engine to generate this cluster, and so on.

In order to extract and build the clusters object for usage, the following files are required:

- In order to build the sparse matrix M:

  ```
  newFormat_unified_blocks_connectivity_matIrFile.spmat
  newFormat_unified_blocks_connectivity_matJcFile.spmat
  newFormat_unified_blocks_connectivity_matMtFile.spmat
  ```

- In order to build the mapping array A_M:

  ```
  newFormat_TranslationTabletmp_numbers.spmathlp
  ```

- In order to build clusters properties DataFrame D_M:

  ```
  newFormat_unified_blocks_blk_data1.mat
  ```


Archived copy · cryp2nova.com

The above mentioned files are to be placed in a specific path, referred to by the clustersExtractor class depicted below, as:

```
<cluster_files_path>
```

Usage Example: After placing the relevant 5 clusters files in ¡cluster_files_path¿, clusters can be extracted using the following code:

```
>>> extractor = clustersExtractor(cluster_files_path)
>>> pop_to_clusters_map, clusters_props, translation_pop = extractor.retrieveClusters()
```

In this example "pop_to_clusters_map" points to the sparse matrix M. "clusters_props" points to the DataFrame D_M containing the clusters properties. "translation_pop" points to the array mapping each searchable object to its index in the sparse matrix, A_M.

The Clusters Extractor Class: Following is the complete description of the Knowledge Sphere API:

```
import os
from cStringIO import StringIO
import struct
import tempfile
import numpy as np
import re
import scipy.io as scio
import pandas as pd
from tempfile import NamedTemporaryFile
from scipy.sparse import csr_matrix

class clustersExtractor(object):
    def __init__(self, cluster_files_path):
        self.__path = cluster_files_path
        self.__mat_init_name = 'blk_data'
        self.__mat_fields = {
            'src': 2,
            'type': 1,
            'blk_type': 1,
            'field': 1,
            'fieldby': 1,
            'N': 1,
            'WN': 1,
            'deg': 1,
            'thrs': 2,
            'SUB_CLUSTERS_FILE': 1,
            'percInternal': 1
        }
    def get_path(self):
        return self.__path
    def retrieveClusters(self):
        pop_to_clusters_map = self.__build_pop_to_clusters_map()
        clusters_props = self.__build_mat_props_df()
```


Archived copy · cryp2nova.com

```
    translation_pop = self.__get_translation_pop()
    return pop_to_clusters_map, clusters_props, translation_pop

def __build_pop_to_clusters_map(self):
    dims_for_mat = self.__get_dimensions_for_matrix()
    indices = self.__get_ir_list()
    indptr = self.__get_jc_list()
    pop_cluster_map = self.__build_pop_clust_matrix(dims_for_mat,
                                                    indices, indptr)
    return pop_cluster_map


def __get_dimensions_for_matrix(self):
    dims_file_name = "MtFile.spmat"
    all_files = self.__get_files_in_path(dims_file_name)
    full_path = all_files[0]
    a = self.__open(full_path)
    f = StringIO(a.read())
    mat_sizes = struct.unpack('<2IQ', f.read(16))
    return {'n_rows': mat_sizes[0], 'n_cols': mat_sizes[1],
            'nnz': mat_sizes[2]}

def __get_ir_list(self):
    ir_file_name = "IrFile.spmat"
    all_files = self.__get_files_in_path(ir_file_name)

    full_path = all_files[0]
    remote_file = self.__open(full_path)
    data = remote_file.read(10 * 1024 * 1024)
    local_temp_path = os.path.join(tempfile.mkdtemp(),
                                   ir_file_name)
    with open(local_temp_path, 'w') as f:
        while data != '':
            f.write(data)
            data = remote_file.read(10 * 1024 * 1024)

    ir = np.fromfile(local_temp_path, dtype=np.int32)
    os.unlink(local_temp_path)
    return ir

def __get_jc_list(self):
    jc_file_name = "JcFile.spmat"

    all_files = self.__get_files_in_path(jc_file_name)

    full_path = all_files[0]
    data = self.__open(full_path)
    local_temp_path = os.path.join(tempfile.mkdtemp(), jc_file_name)
    with open(local_temp_path, 'w') as f:
        f.write(data.read())
```


Archived copy · cryp2nova.com

```
    jc = np.fromfile(local_temp_path, dtype=np.int64)
    os.unlink(local_temp_path)
    return jc

def __build_pop_clust_matrix(self, mat_dims, indices, indptr):

    nrows = mat_dims['n_rows']
    ncols = mat_dims['n_cols']
    nnz = mat_dims['nnz']

    data = np.ones(nnz)
    try:
        mat = csr_matrix((data, indices, indptr),
                         shape=(ncols, nrows))
    except Exception as e:
        msg = """Couldn't build population to
        cluster match due to: %s, aborting.""" % str(e)

        raise ValueError(msg)
    return mat

def __build_mat_props_df(self):
    mat_names_list = self.__get_mat_files_names()

    if len(mat_names_list) > 1:
        prop_names = 'temp_'
        mat_files = [scio.loadmat(StringIO(self.__open(mat_name).read()))
                     for mat_name in mat_names_list]
        props_df = self.__build_multiple_clusters_props(mat_files, prop_names)
    else:
        prop_names = 'blk_data'
        mat_file = scio.loadmat(StringIO(self.__open(mat_names_list[0]).read()))
        props_df = self.__build_single_clusters_props(mat_file, prop_names)
    return props_df

def __build_multiple_clusters_props(self, mat_files, prop_names):

    all_props_df = pd.DataFrame()
    for mat_file in mat_files:
        df = self.__build_single_clusters_props(mat_file, prop_names)
        all_props_df = all_props_df.append(df)
    all_props_df.index = range(len(all_props_df.index))
    return all_props_df

def __build_single_clusters_props(self, mat_file, prop_names):
    try:
        clusters_props = mat_file[prop_names]
    except Exception:
        msg = """Field %s doesn't exist in mat file,
        but expected. Cannot continue""" % prop_names
        raise ValueError(msg)
```


Archived copy · cryp2nova.com

```
        cluster_props = {}
        for prop, counts in self.__mat_fields.iteritems():
            try:
                mat_values = clusters_props[prop][0][0]
            except (KeyError, IndexError):
                ValueError("""Field %s doesn't exist in mat file,
                please remove it from config file.""" % prop)
                break
            if counts == 1:
                cluster_props[prop] = mat_values.flatten()
            else:
                for i in np.arange(counts):
                    cluster_props[prop + '_' + str(i)] = mat_values[:, i]
        cluster_prop_df = pd.DataFrame(cluster_props)
        # noinspection PyUnresolvedReferences
        cluster_prop_df.index.names = ['cluster']
        wn_sizes = cluster_prop_df['WN'].astype(float)
        n_sizes = cluster_prop_df['N'].astype(float)
        cluster_prop_df['W_prcntg'] = wn_sizes / n_sizes
        del mat_file
        return cluster_prop_df

    def __get_translation_pop(self):
        spmat_help_name = ".spmathlp"
        all_files = self.__get_files_in_path(spmat_help_name)

        full_path = all_files[0]

        f_spmathlp_data = self.__open(full_path).read()
        local_temp_path = os.path.join(tempfile.mkdtemp(), spmat_help_name)
        with open(local_temp_path, 'w') as f:
            f.write(f_spmathlp_data)
        with open(local_temp_path, 'rb') as f_spmathlp:
            num = struct.unpack('<Q', f_spmathlp.read(8))
            # noinspection PyTypeChecker
            ids = np.fromfile(f_spmathlp, dtype=np.double)

            if num[0] != len(ids):
                msg = """translating ids went wrong. Found %d ids,
                where expected %d ids, aborting""" % (len(ids), num[0])
                self.__logger.error(msg)
                raise ValueError(msg)

        os.unlink(local_temp_path)

        return ids

    def __get_files_in_path(self, name):
        all_files_in_dir = list(self.__list_dir())
```


Archived copy · cryp2nova.com

```
    relevant_files = [file_name for file_name in
                      all_files_in_dir if name in file_name]
    return relevant_files

def __open(self, path):
    real_path = os.path.expanduser(path)
    if not os.path.isfile(real_path):
        raise LookupError('"{}"  does not exist'.format(real_path))

    return open(real_path, 'rb')

def __get_mat_files_names(self):
    mat_names_list = self.__get_files_in_path(self.__mat_init_name)
    mat_names_list.sort(key=lambda x: int(re.search(r'\d+', x).group()))
    return mat_names_list

def __list_dir(self):
    return [os.path.join(self.__path, f) for f in
            os.listdir(os.path.expanduser(self.__path))]
```

CRYP2NOVA

Archived copy · cryp2nova.com

Bibliography

[1] Wikipedia – *Social Physics* (2017).
URL https://en.wikipedia.org/wiki/Social_physics

[2] W. Pan, Y. Altshuler, A. Pentland, Decoding social influence and the wisdom of the crowd in financial trading network, in: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), IEEE, 2012, pp. 203–209.

[3] Y.-Y. Liu, J. C. Nacher, T. Ochiai, M. Martino, Y. Altshuler, Prospect theory for online financial trading, PloS one 9 (10) (2014) e109458.

[4] Y. Altshuler, W. Pan, A. Pentland, Trends prediction using social diffusion models, arXiv.org, 2011.

[5] P. M. Krafft, J. Zheng, W. Pan, N. Della Penna, Y. Altshuler, E. Shmueli, J. B. Tenenbaum, A. Pentland, Human collective intelligence as distributed bayesian inference, arXiv preprint arXiv:1608.01987.

[6] Y. Altshuler, A. S. Pentland, G. Gordon, Social behavior bias and knowledge management optimization, in: Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, 2015, pp. 258–263.

[7] Y. Altshuler, A. Pentland, Methods and apparatus for tuning a network for optimal performance, uS Patent 8,914,505 (Dec. 16 2014).
URL https://www.google.com/patents/US8914505

[8] W. Pan, Y. Altshuler, A. Pentland, N. Aharony, Methods and apparatus for prediction and modification of behavior in networks, uS Patent 9,098,798 (Aug. 4 2015).
URL https://www.google.com/patents/US9098798

[9] Tuning social networks to gain the wisdom of the crowd (*MIT Media Lab Website* (2017).
URL https://www.media.mit.edu/research/highlights/tuning-social-networks-gain-wisdom-crowd

[10] Markets insight: Wake up to the twitter effect on markets (*Financial Times*) (2017).
URL http://web.media.mit.edu/~yanival/Markets_Insight.htm

[11] Beyond the echo chamber (*Harvard Business Review*) (2017).
URL https://hbr.org/2013/11/beyond-the-echo-chamber

[12] Rethinking predictive analytics (*FirstMark's Data Driven*) (2017).
URL http://firstmarkcap.com/insights/rethinking-predictive-analytics/

[13] *MIT*s $1 million test to see if social media can make investors money (2013).
URL https://tinyurl.com/MIT-1M-USD

[14] Endor – inventing the *"Google for Predictive Analytics"* (2017).
URL http://news.mit.edu/2017/endor-inventing-google-predictive-analytics-1220

[15] Endor leading investor – *Innovation Endeavors* (2014).
URL http://www.innovationendeavors.com

Archived copy · cryp2nova.com

[16] A. Boehme, Y. Altshuler, Using social physics to predict consumers behaviors, in: Network Science (NetSci), 2017.

[17] Mastercard brings 5 new startups into start path accelerator program (2016).
URL https://tinyurl.com/MasterCard-Endor

[18] Endor – *Finnovate Fall 2017* (2017).
URL https://www.youtube.com/watch?v=69rUQloq-qA

[19] Endor – a *Gartner Cool Vendor* (2017).
URL https://www.gartner.com/doc/3727117

[20] Endor – a technological pioneer acknowledgement of the *World Economic Forum* (2017).
URL http://widgets.weforum.org/techpioneers-2017/

[21] *DARPA Network Challange* (2011).
URL http://archive.darpa.mil/networkchallenge/

[22] The 2012 mckinsey award winners (2012).
URL https://hbr.org/2013/04/the-2012-mckinsey-award-winners

[23] *Google Scholars – Professor Alex "Sandy" Pentland* (2017).
URL https://scholar.google.com/citations?user=P4nfoKYAAAAJ&hl=en

[24] *Tim Oreilly: the World's 7 Most Powerful Data Scientists* (2017).
URL http://www.forbes.com/pictures/lmm45emkh/6-alex-sandy-pentland-professor-mit/

[25] Y. Altshuler, A. Pentland, A. M. Bruckstein, Swarms and Network Intelligence in Search, Springer, 2017.

[26] Y. Altshuler, Y. Elovici, A. B. Cremers, N. Aharony, A. Pentland, Security and privacy in social networks, Springer Science & Business Media, 2012.

[27] H. Shrobe, D. L. Shrier, A. Pentland, New Solutions for Cybersecurity, MIT Press, 2018.

[28] J. Clippinger, D. Bollier, From Bitcoin to Burning Man and Beyond: The Quest for Identity and Autonomy in a Digital Society, ID3 and Off The Common Books, 2014.

[29] T. Hardjono, D. Shrier, A. Pentland, TRUST:: DATA: A New Framework for Identity and Data Sharing, 2016.

[30] A. Pentland, T. Heibeck, Honest signals: how they shape our world, MIT press, 2010.

[31] A. Pentland, Social physics: How good ideas spread-the lessons from a new science, Penguin, 2014.

[32] D. Shrier, A. Pentland, Frontiers of Financial Technology: Expeditions in future commerce, from blockchain and digital banking to prediction markets and beyond, Publisher: Visionary Future, 2016.

[33] *Endor.coin protocol GIT* (2017).
URL https://github.com/orgs/EndorCoin

[34] Why you want blockchain-based ai, even if you dont know it yet (2017).
URL https://tinyurl.com/blockchain-based-AI

[35] Y. Altshuler, N. Aharony, A. Pentland, Y. Elovici, M. Cebrian, Stealing reality: When criminals become data scientists (or vice versa), Intelligent Systems, IEEE 26 (6) (2011) 22–30. doi:10.1109/MIS.2011.78.

[36] M. Ulieru, Blockchain: what it is, how it really can change the world, World Economic Forum.

[37] How technology could help fix our broken financial system (2017).
URL https://tinyurl.com/technology-fixing-our-financia

Archived copy · cryp2nova.com

[38] J. Pieprzyk, T. Hardjono, J. Seberry, Fundamentals of computer security, Springer Science & Business Media, 2013.

[39] T. Hardjono, L. R. Dondeti, Multicast and group security, Artech House, 2003.

[40] T. Hardjono, L. R. Dondeti, Security in Wireless LANS and MANS (Artech House Computer Security), Artech House, Inc., 2005.

[41] J. Seberry, T. Hardjono, Towards the Cryptanalysis of Bahasa Indonesia and Malaysia, 1989.

[42] S. G. Ong, J. Seberry, T. Hardjono, A. D. F. Academy., Towards the cryptanalysis of Mandarin (Pinyin), 1991.

DISCLAIMER: This White Paper is for discussion purpose only.
Endor.coin does not guarantee the accuracy of the conclusions reached in this white paper.
Copyright © 2018 Endor.coin.

CRYP2NOVA

Archived copy · cryp2nova.com