



WHITEPAPER

10.30.2017

TABLE OF CONTENT

SUMMARY	3
INTRODUCTION	4
1.1. Problem	4
1.2. Solution	5
2. REVIEW OF USED TECHNOLOGIES	6
2.1. Technical Indicators	6
2.2. Machine learning	6
2.3. Artificial neural networks	8
2.4. Data Mining and Deep Learning	9
2.5. Ensemble of Neural Networks	9
2.6. Fundamental analysis	10
2.7. News analysis	11
2.8. Order book	11
2.9. Self-learning algorithms	12
3. OVERVIEW OF DATATRADING SYSTEM	13
3.1. Screener of stocks / crypto assets	13
3.2. Trading Advisor	14
3.3. Scoring IPO/ICO	14
3.4. Open constructor of machine learning models	15
3.5. Quality control of machine learning	16
3.6. Marketplace of machine learning models	17
3.7. External Modules	18
3.8. Blockchain infrastructure	18
4. ECONOMIC MODEL OF THE SYSTEM	20
5. ROADMAP	22
6. TOKEN SALE	23
7. TEAM	25
8. LEGAL DETAILS	27
9. RISKS	28
References	29

SUMMARY

DataTrading is an innovative project in the world of trading and consulting, which provides a set of analytical and forecasting tools for trading in stock and crypto exchange markets and is completely based on artificial intelligence.

Big Data Trading Limited was registered as a legal entity on January 23, 2017 with a representative office in Hong Kong. With the support of a professional team and investors, DataTrading already has a developed MVP that shows results, which exceed many trading strategies in terms of profitability.

There are many analytical and trading strategy services in the modern world. The fundamental difference between DataTrading and all other companies is that we use machine learning and neural networks to solve the tasks. These revolutionary instruments will also be available to traders and the community so that they can develop their own models for forecasting markets.

Trained models will be able to make a profit for each client of Data Trading: they can be used for trading as well as for selling to other market participants. The machine learning constructor will be easy to develop, so that every client and even those without specialist education can use it.

DataTrading service develops its own constructor of trading strategies and will also implement a full analytical tool for stock and cryptocurrency markets on neural networks, namely:

- Screener of shares / crypto assets;
- Trade advisor;
- Scoring of ICOs / IPOs;
- Constructor of trading strategies with the ability to connect and train neural networks available to the community; implementation of self-learning neural networks.

1. INTRODUCTION

The possibility to make money by forecasting the price movement of financial instruments has always attracted a large number of participants to the securities market. Many people earned good money by trading on stock exchanges, but many of them also went bankrupt. For centuries, mankind has been developing a mathematical model for forecasting these markets and with varying degrees of success different mathematical tools are used to make investment decisions nowadays.

In the middle of the twentieth century new technologies for the analysis and processing of information began to be actively developed, which were called an artificial intelligence (AI). Nowadays the potential of artificial intelligence seems to be comparable with the capabilities of the human brain, and in many cases exceeds it [1] [2] [3]. It is possible to provide automatic control of transport, to recognize visual and sound images, to identify individuals, to play intellectual games, to model engineering products, to create works of art, etc. with the help of AI. In addition, artificial intelligence copes well with the task of finding the implicit relationships between a huge number of factors and their influence on the object of study. For example, AI can diagnose patients on the basis of medical card data and predict the health of patients in the future [4] [5] [6].

1.1. Problem

The fact that there is a wide range of fields where artificial intelligence is applied raises the issue of the possibility of using this technology for the analysis of exchange markets and the formation of trade strategies on its basis. The success of using AI for trading on stock exchanges is confirmed by various researches [7]. A number of large hedge funds are actively using different instruments of artificial intelligence to make investment decisions. Profitability of investments of these funds in most cases exceeds the profitability of those investments that were made with using traditional analytical tools and technical indicators [8] [9] [10] [11].

Thus, it can be argued that artificial intelligence is very effective in forecasting exchange markets and can bring good profit. Nevertheless, most traders today do not have the opportunity to use this technology for their own trading. The problem is that in order to use AI effectively there is a need to study a large volume of mathematical apparatus and "spend a lot of time" in mastering the methodology of developing of AI. It is also necessary to find or purchase, select and correctly process a large amount of primary, inordinately diverse data from different sources, so that the results of analysis and trading strategies of AI are as accurate as possible. All these factors significantly complicate the access of ordinary traders for the use of artificial intelligence when trading on exchanges.

1.2. Solution

Our team has been working on DataTrading project since 2015. DataTrading aims to make the use of artificial intelligence affordable and convenient for traders so they can trade on the stock exchanges without the need to study of the mathematical foundations of this technology. We want to offer traders a ready-made toolkit that will help them trade on different stock exchanges and to receive the income, which is higher than the market level. We expect that even a novice trader will be able to get a good profit and increase his professionalism with DataTrading trading advisor. In addition, anyone who is interested in these technologies will be able to develop their own model of artificial intelligence on the DataTrading platform even without special education and use it for their own trade or for sale to other users of the system.

2. REVIEW OF USED TECHNOLOGIES

2.1. Technical Indicators

Technical indicators are traditional mathematical tools for assessing and forecasting trends in the behavior of the price of financial instruments, based on the values of statistical indicators of trading (price, time of transactions, trading volume, etc.). Nowadays there are hundreds of technical indicators (apart from the variations of the most famous ones).

Almost every trader is familiar with technical indicators, technical analysis and algorithmic trading are based on them. There has been a long debate on how effective technical indicators are and whether they can be used for decision-making. Usually experienced traders rarely make decisions based on one indicator only. In the majority of cases, each trader chooses several indicators and makes decisions to expand or reduce the position after carefully analyzing them and taking into account his own experience, knowledge of the market and intuition.

Technical indicators serve as the basis for most automated trading strategies in trading systems; trading signals on the opening or closing of trading positions are generated based on their combination. DataTrading system uses some technical indicators in its algorithms in order to aggregate incoming data and conduct primary analysis and selection, but it is not making trading decisions based only on technical indicators.

Thus, the use of technical indicators for the analysis of investment tools in DataTrading system will be a part of the first stage of data processing. Further, the results of application of technical indicators will be used in the machine learning module, where along with other feature detection they will serve as input layer for the process of Artificial intelligence learning.

We want to emphasize that while using data from technical indicators as one of many input layers for machine learning, the system can give much more accurate and relevant forecasts than standard trading strategies that would use these same indicators as the sole basis for trading signals.

2.2. Machine learning

Machine learning is a big subsection of the science of artificial intelligence, which involves the use of various data analysis algorithms, during which the system learns and independently finds interrelations between input parameters, and can make conclusions, decisions or predictions in the context of the tasks. Unlike the traditional approach in programming, in which the task is solved by creating certain set of rules and commands, the machines are trained on a large number of input data and this gives them the opportunity to learn how to perform the task.

Nowadays machine learning has many approaches and directions. Usually three sections of machine learning are distinguished: supervised learning, unsupervised learning and reinforcement learning.

DataTrading system will use all types of learning for different tasks, in this section we will stop on the description of the supervised learning, as the most obvious and understandable.

The simplified general task for supervised machine learning is as follows. There are many situations (experiments, observations) and the values of certain features that somehow influence the results of the experiment. The task is to identify the relationship between the set of signs and the results of observations (experiments)¹. The process of identifying and establishing this dependence is called the learning process. The data used for training, the values of attributes and the results of observations for which are known, called training samples. If during the learning on the training sample an explicit relationship between the signs and the results of the observations was determined, it is considered that the aim of the training was achieved and the developed model is used to work with data where the results of the experiments are unknown.

Here is a schematic example. Let us suppose, the task is to predict the price of a new car, depending on its parameters. To solve this problem, a training sample using the methods of machine learning is used, which consists of a number of observations (the more observations, the greater the accuracy of training). Each observation is the same and consists of a number of parameters (signs): the car's brand, the type of body, the type of engine, the capacity, the type of gearbox, the amount of horsepower, fuel consumption and so on. In the training sample, for each set of characteristics, the price of the car is known, for example:

- observation 1: Ford, sedan, gasoline engine, engine capacity 1.5 liters, manual transmission, 105 hp, fuel consumption 8 liters per 100 kilometers, price — 15,000 \$;
- observation 2: Ford, hatchback, gasoline engine, engine capacity 1.5 liters, manual transmission, 115 hp, fuel consumption 9.5 liters per 100 kilometers, price — 21,000\$;
- observation 3: Toyota, sedan, diesel engine, 1.8 liter engine, automatic transmission, 120 hp, fuel consumption 8.9 liters per 100 km, price — 19,000\$;
- ...

Then the method (or algorithm) of machine learning is chosen, data processing and model configuration are carried out, and the learning begins. If during the testing of the learning outcomes it was revealed that the relationship between the factors was not found, or it was very weak, then a new stage of training is conducted, for which a larger sample, another set of data (features) or different model settings is used. This process continues until the system finds the parameters and configuration of the model that reveal the relationship between the characteristics (characteristics of the car) and the results of observations (the price of the car). After successful training and testing the model is used for forecasting, i.e., predicting the cost of a new car depending on its characteristics. In our example, it looks like this: a set of features (for example, Mazda, crossover, gasoline engine, 3 liters, manual transmission, 205 hp, 11 l / 100 km) is used in the model and the model makes a prediction about the price. It is important to understand that a well-trained model will make an accurate forecast about the price of such a car even if such configuration was not in the training sample. This is because the model does not fit the results to the learning data (finds the closest configuration), but determines the relationship between the factors and how much each factor (body type, car brand, liter, etc.) affects the desired parameter (car price).

Machine learning is the core of DataTrading system. Many scientific works [7] and our developments show that the methods of machine learning can be successfully used to predict the dynamics of prices of

¹ In fact, this definition is not sufficiently generalized and is given in the text in this form in order to simplify the understanding of the essence of the process

financial instruments on stock, commodity and crypto exchanges, as well as to build trading strategies. In our system, we will use different algorithms and models of machine learning, as well as their combinations to achieve the most accurate and reliable forecasts. In addition, the constructor of learning algorithms will be available in the system, where each user of the platform will be able to choose a model for himself, train it and use either for his own trading, or selling to other users of the system (for more details see Section 3 “Overview of DataTrading System” and Section 4 “Economic model of the system”).

2.3. Artificial neural networks

Artificial neural networks are one of the methods of machine learning and serve to solve many tasks, such as image recognition problems, discriminant analysis, approximation, clustering methods, decision making, forecasting, etc. Artificial neural networks are built on the principle of the organization and functioning of biological neural networks (networks of nerve cells of a living organism). Neural networks can find and identify relationships between input parameters (even if these relationships are not known in advance) and make very accurate forecasts based on the found patterns.

The mathematical model for artificial neural networks was proposed in the 50-60s of the twentieth century, but for a long time it did not find its practical application due to the fact that even the most basic neural networks required very powerful computer calculations and were for a long time unfeasible or unreasonably expensive for application. In the second half of the first decade of the 21st century, rapid technological progress made parallel computing of neural networks on graphic cards possible and effective and a new era of practical application and development of machine learning began.

Due to its ability to identify non-linear mathematical patterns of time series and quickly adapt to changes in market trends, neural networks are one of the most effective and accurate tools for predicting the behavior of markets in general and their specific components in particular. Traditional technical indicators usually take into account only historical data on the volume and price level of orders of one investment instrument in their forecasts, while a neural network can take into account the movement of prices throughout the market as a whole, by industry and by specific companies in particular. In addition, the neural network can take into account the financial and operational performance of companies to build the forecast, as well as information from news channels, which is almost impossible to implement in the technical analysis. On the basis of the revealed interrelations, the trained model can make extremely accurate forecasts for the price of the company’s stock, goods or crypto currency (depending on the market in question) and generate profitable trading strategies.

Thus, it is possible to build trading strategies based on well-trained neural networks that, in terms of their profitability, will far exceed traditional algorithmic strategies, based on the use of technical indicators (see Section 3.2 “Trading Advisor”). Also, neural networks can be successfully used to select investment tools from many others available on the exchange, which can bring maximum profit in the short or long term (depending on the chosen strategy, see Section 3.1 “Screener of stocks/crypto assets”).

2.4. Data Mining and Deep Learning

Data mining is a set of methods designed to search hidden and nontrivial knowledge in a large amount of data that was previously unknown and which can be used in subsequent analysis or decision making. The purpose of data mining is the extraction of information from a set of data and their transformation into understandable structures for further use (through various interpretations, visualizations, etc.).

Deep training is a set of machine learning methods for solving complex problems of modeling of high-level abstractions with a large amount of input data. An example of such problems can be recognition of images, “understanding” of computer algorithms for texts, finding relationships and regularities in a vast amount of disparate information, etc. Among other things, deep learning methods are also used to solve the data mining tasks.

Data Mining and deep learning methods will be an essential part of the DataTrading system. The main algorithms that will be used include convolutional neural networks, recurrent neural networks, networks with long short-term memory (LSTM networks). It is also planned to visualize the found dependencies of the results of the mining date (including the results of the fundamental analysis).

2.5. Ensemble of Neural Networks

Ensemble of Neural Networks is a set of neural network models that collectively decide on the formulated problem.

A simplified model of this architecture looks as follows. There is a certain number of neural network models in the system that are differently trained (possibly on different incoming data) and give different forecasts for the same parameter (for example, the company’s stock price). The final decision is made by a separate neural network that takes into account the accuracy of prediction of a model in the past and corrects its influence on the forecasted parameter as a whole, thus combining the forecasts into one and making it more accurate.

An example. Let us suppose that there are three neural networks that predict the price of the Ethereum, with such differences²

- neural network 1: the input receives data on the history of the change in the price of the etherium as well as data on the movement of prices of 10 other cryptocurrencies
- neural network 2: as an input receives data on the price of the Ethereum and the volumes of transactions, general indexes of the crypto-currency market, data on the volumes of orders placed for each price cluster
- neural network 3: uses the same data as the neural network 2, but has different settings (another number of neurons, the number of hidden layers, the learning rate, etc.)

² It is also possible to use one neural network that will receive the same input data that the first and second neural network of the example

Obviously, all three neural networks will give a different forecast for the price of Ethereum. When using ensemble of neural networks in DataTrading system, the final decision will be made by a separate neural network that takes into account the accuracy of the forecasts of each network in the past and corrects the overall forecast of all networks.

The work on the development of the application of Ensemble of Neural Networks in DataTrading system is planned immediately after the release of the first version of DataTrading 1.0. We expect that in 6 months after the release of the first version of the system the ensemble of neural networks will be available for the users of the platform (see Section 5 “Road map”).

2.6. Fundamental analysis

Fundamental analysis is the estimation of the company’s internal value, stock, currency, derivative or product based on an analysis of the main influencing external and internal factors.

Different methods are used to estimate intrinsic value of various types of financial instruments. For example, the main indicators of financial and production activity of a company and indices of its business activity can be analyzed to find the value of a company and its shares. The main macroeconomic factors such as nominal and real interest rate, economic growth rates, GDP, trade balance, inflation, etc. are evaluated for the analysis of exchange rates. Adoption and real use of the technology by business and the ordinary people, legal regulation at national levels, the emergence and development of competing projects play an important role for the evaluation of crypto-currencies. To assess the value of the product on commodity exchanges, the main factors affecting the value of the commodity are estimated, such as the volumes of production (for raw materials markets), the weather (for agricultural goods), the dynamics of the cost of competing and competing goods, the change in the cost of the resources necessary for the extraction or production of this commodity goods, the state of technological progress in the industry, etc.

Nowadays there is no univocal methodology how to conduct a fundamental analysis — each analyst based on his experience takes into account certain factors, conducting a fundamental analysis of an financial instrument. Although a certain mathematical model can be used in the process of fundamental analysis, the analyst’s subjective influence on the results of the analysis is very high: he chooses the factors, determines the influence of each indicator on the final results, outlines the formulas and coefficients used. And although in some cases it is possible to partially algorithmize certain evaluation processes and aggregate them, it can be argued that the fundamental analysis was not amenable to automation.

Nevertheless, the use of various methods of machine learning can partially or completely replace the role of the analyst in the fundamental analysis. In addition, it is likely that the fundamental analysis carried out by artificial intelligence can produce more accurate results and forecasts than the traditional one, since machine algorithms can better locate and determine hidden regularities between factors.

Almost all forecasts of DataTrading system will be built taking into account the fundamental analysis carried out by the methods of machine learning. Thus, the artificial intelligence of the DataTrading system will combine the results of technical and fundamental analysis for investment tools and, based on these

findings, various trading strategies will be formed. According to our estimates, the use of such strategies for trading on exchanges should bring profit exceeding the market average.

2.7. News analysis

Any trader knows that the behavior of the price of financial instruments is influenced, among other things, by the news flow, directly or indirectly related to this instrument. Positive news about the company's activities (for example, the introduction of new technologies or the acquisition of competitors, or promising trends in the industry) leads to an increase in the share price of this company, while negative news reduces the cost of shares.

With the development of machine learning technologies and the development of methods of deep learning (using semantic analysis, convolutional neural networks, recurrent neural networks, networks with long short-term memory, etc.), it became possible to analyze arbitrary texts by computer algorithms and transfer the obtained analysis results to forecasting modules as input layers. In the DataTrading platform, specially trained neural networks will be used to continuously monitor the entire news flow and to identify information signals that can affect the price of stocks, crypto-currencies and other financial instruments and, based on these signals, the strategies of the trade advisors will be immediately adjusted.

Example. Suppose that a trader uses the DataTrading system to monitor the commodity market of wheat. Most likely, the model used for the forecast at the training stage will reveal the relationship between the price of wheat and the price of fuel materials. If during the analysis of the news flow the system finds news that will lead to an increase in the price of fuel (for example, the decision of the OPEC countries to reduce oil production), it will connect this input signal with an increase in the price of wheat in the near future and advise the client: to buy wheat at the actual price (because of the probable increase in the price and the opportunity to play on the growth of the market), or to hold the position before the price increases to a certain level. Of course, this is a greatly simplified and idealized example. In fact, the factors affecting the price of goods, shares or crypto currency are much larger, in addition, the relationship between the two factors may not always be permanent, so all forecasts are made with an indication of the probability of implementing predicted events.

2.8. Order book

Order book — all orders for the purchase and sale of an investment instrument or commodity at a certain point in time and their dynamic change on a particular exchange. Information includes the price and volume of orders. Depending on the exchange, orders with the same price level can be combined into one order (without the possibility of knowing the number of participants behind this application).

You can evaluate the supply and demand for an financial instrument on the market at a given moment in time after analyzing the information from the order book. There is a number of algorithms and indicators that use an Order book to develop a trading strategy. The disadvantage of such algorithms is that they take into account only the current status of the Order book or its change in a short period of time, they do not take into account the medium and long term dynamics to elaborate strategies.

DataTrading system will comprehensively use information from the Order book during the process of machine learning: neural networks and algorithms will find the relationship between the state of the Order book and the dynamics of price changes over the entire period of quotations and form a trading strategy on the basis of the identified relationship and the current state of the bids. It should be noted that trading strategies will be based not only on the analysis of the Order book, the results of training will also be influenced by many other factors, such as market analysis in general, price movements of related investment tools, news analysis, etc.

2.9. Self-learning algorithms

Machine learning methods such as “supervised learning” are usually used to solve the problems of detecting trends and dependencies. In such methods, features are indicated, the system’s response to these features are known for each observation and the system should establish the relationship between the features and the results of observations. The disadvantage of this approach is the complexity of the initial configuration of the system: it is required to go through a lot of parameters and conduct a large number of learning experiments in order to choose the optimal configuration of the model.

Self-learning algorithms solve the above-mentioned problem: such algorithms can independently sort out the settings of their system and the types of data on which training is conducted, in order to identify the optimal parameters and fix them. If in case of ordinary systems it requires constant observation and participation of the experimenter during training, the role of the human being in such systems is minimized, the system is very autonomous.

Self-learning algorithms will be available in DataTrading system and can be used to build screens or trade advisers. Thus, a person can upgrade the system to their needs regardless of the level of knowledge of technology.

3. OVERVIEW OF DATATRADING SYSTEM

DataTrading is a cloud with a set of open and customizable analytical tools for trading, provided on a subscription or purchase basis, consisting of the following modules:

- screener of financial instruments;
- trading advisor;
- scoring of ICO/IPO;
- open constructor of machine learning models;
- quality control of machine learning;
- a marketplace of trained machine learning models for use in market screeners, trading advisers, scoring, forecasting, etc.;
- external modules (integration with broker platforms);
- blockchain infrastructure for transparency.

3.1. Screener of stocks / crypto assets

More than 45,000 of companies' shares are being traded on the stock exchanges of the world; the market of derivatives is even greater [12]. The number of crypto instruments is more than 1 thousand [13]. How to choose those financial instruments (stocks, crypto-currencies, derivatives, etc.), which at the moment give the best result, from such a variety of options? Screeners will solve the abovementioned task.

The screener is an tool for selecting financial instrument (usually, shares of companies) with the help of the specified filters. Indicators of financial performance, any indexes or technical indicators can serve as filters in the modern screeners.

The main task of DataTrading screener is to find and show financial instrument that will bring maximum profitability in the short, medium or long term. The screener will recommend a betting strategy (play on raising or lowering), the expected profitability in the chosen time interval, the riskiness of investments and the likelihood of implementing the proposed strategy. Thus, choosing an financial instrument will be extremely simple; this screener will be successfully used by both experienced traders and beginners.

DataTrading screeners are based on machine learning and artificial intelligence. When forecasting and selecting financial instrument, not only the dynamics of the price of an individual instrument is taken into account, but also the movement of the whole market, the industry, the fundamental analysis, the order book and its change, news analysis, etc. Technical indicators will also be used during the training, but only for the primary aggregation of information and will be one of many parameters of the input data layer.

3.2. Trading Advisor

The trading advisor is one of the key services of DataTrading system. The task of the service is to help traders efficiently trade on exchanges with any financial instruments. The trading adviser monitors the status of the selected instruments in real time and gives trading signals for buying or selling.

Just like in case of the stock screener, the trade advisor works with the results of machine learning algorithms. When developing trading signals, not only the trading data for the selected stock, commodity or crypto currency is taken into account, but the state and dynamics of the entire market. When setting up or using trading advisors, you do not need to know or assign links between different market indicators or to know patterns of market movement — the system will find and identify them. The trading signals of the adviser will be based on the identified trends.

Each trading signal (for purchase or sale) will be accompanied by a probabilistic evaluation of the success of trading action and its profitability. In addition, individual parameters (warning thresholds) can be set for each instrument, such as acceptable profitability, riskiness, and so on. After the integration of DataTrading system into the trading platforms, the trading advisor can be used for automated trading (for those instruments that will be available in these sites).

3.3. Scoring of IPO/ICO

Scoring is the classification of the researched series of objects into different groups according to implicit factors. For example, scoring is widely used in the banking sector to identify creditworthy clients, as well as investment-attractive projects or companies. Multiple studies [14] [15] [16] [17] [18], as well as our own experience gained in the DataScoring project, shows that the use of neural networks for scoring in comparison with linear algorithms gives a significant increase in accuracy.

At the moment, the cryptocurrency world is oversaturated with ICO-projects, many of which are scams. We are confident that it is possible to use machine learning for scoring ICO projects in order to identify potentially successful or failed projects even before the sale of tokens. At the moment, a sufficient amount of data on projects with different histories has already been accumulated. It is possible to train neural networks and make forecasts for new projects on their basis. Also, machine learning can be used not only to classify projects, but also to forecast the behavior of the price of the token after entering the exchange.

What is more, artificial intelligence can be successfully used for scoring IPO projects: an even larger amount of historical data will provide more accurate scoring results and a forecast of the dynamics of stock prices on stock exchanges.

It is planned that all data on IPO and ICO projects will be constantly kept in the system and updated. If the user is subscribed to this service, the scoring data of all the future IPOs and ICO projects will be available to him. If the project is not in the system (for example, the project is only in the process of preparing an IPO/ICO), the user can enter all necessary data about the project and get the results of the analysis. In addition, in order to prevent the adjustment of the project parameters for obtaining good scoring results, different defense mechanisms will be provided (for example, deliberately slowing down the receipt of the analysis results, for the impossibility of enumeration of the parameter).

3.4. Open constructor of machine learning models

About the constructor

Machine learning is a great technology; its application for the exchange analysis shows the results far exceeding traditional tools. Unfortunately, at the moment, access to these technologies is limited to a certain number of scientists, data-scientists and developers. The reason is in the lack of good tools for developing and teaching prediction models, rather than in the complexity of the mathematical apparatus that underlies the machine learning.

One of the tasks of DataTrading system is to overcome the existing barrier and to make it possible to use the available mathematical model without its in-depth study. To implement this goal, we are developing a machine learning constructor.

The Constructor of DataTrading Machine Learning is one of the components of the system, which is an interface that allows any member of the system to design a machine learning model, select and process the necessary data, train the model and perform the test of the results. The system will be designed in such a way that no special knowledge will be required to complete all these operations, only the understanding of the general principles of machine learning.

This Constructor will be effectively used by both professionals in artificial intelligence, as well as ordinary traders or anyone who is interested in the topic of machine learning. Thus, we want to create a community of professionals and enthusiasts around DataTrading platform who will share their experience as well as develop and promote the ideas of blockchain and machine learning.

In most cases you do not need to input raw data into the system for the implementation of training, all market data for machine learning is constantly available and will be updated in real time (the data preparation will be handled by system administrators) in DataTrading platform. The user only has to choose those values that are necessary for training a particular algorithm. However, if the user wants to use some specific or unique data, he will have the opportunity to upload and connect them to the model.

The resulting models can be used both for the developer's own trading or for its sale to other platform participants. The system will implement automatic quality control, the user will be confident in the properties, accuracy and capabilities of the model, which he will purchase from a third-party developer.

Cost of development

The machine learning process consumes a lot of computing power, the complexity of the calculations is directly proportional to the volume of incoming data and the configuration of the model (for example, for neural networks this is the number of neurons, the number of hidden layers, the activation function, etc.). In order to make the computing capacities of the system fair and economically justified, we will use the DataTrading Token (DTT) token as a calculated inner system unit.

Here is an example of one of the scenarios for using the system. In order to start developing models, the user needs to purchase a subscription to gain access to the Constructor and should have a non-zero DTT balance. The system will set the price of computing (training) per hour (in DTT). After setting all the necessary parameters and starting calculations after each hour of training, the system will take a certain amount of DTT from the user's balance.

Suppose that the cost of a monthly subscription to the machine learning Constructor costs 100 DTT, and the cost of training per hour is 10 DTT. The user has 500 DTT and he is ready to spend all of them on developing his own model of forecasting a certain market. After using 100 DTT for subscription to access the constructor, the user has 400 DTT left. Suppose that the user spent 3 experiments for 12 hours of training each, using 360 DTT ($3 \times 12 \times 10$). Having on the balance of 40 DTT, the user starts another experiment (with an estimated calculation time of about 10-12 hours). After 4 hours of calculation, the user's balance is reset to zero ($40 - (4 \times 10)$) and the calculations for the fourth experiment terminate prematurely. Let us suppose that despite the lack of results for the fourth experiment, the user will find the results for any of the three experiments acceptable. If the model obtained as a result of the experiment can pass the tests of internal quality control, then the user has the opportunity to expose this model for sale, setting any price. If in our case, the user sets a price of 100 DTT for buying a model he has trained, then 6 customers are enough to make the user's expenses pay off and bring profit to him and to users of his model.

3.5. Quality control of machine learning

It is expected that DataTrading system will be interesting not only for traders who want to get a reliable forecast of the dynamics of markets and investment tools, but also for developers in the field of machine learning who will train algorithms and thereby earn money by offering trained models to other participants of the system.

In order to maintain the high level of quality of all forecasting tools, the module of quality control of machine learning will be included in the DataTrading system. The marketplace will be able to get only those models that give an acceptable level of errors and which show a high accuracy of forecasts on historical data. All data of testing and quality control will be entered in the blockchain and will be available for monitoring and auditing. In addition, the models will be checked for similarity to other existing developments, to ensure the preservation of intellectual property developers.

Based on the results of automatic testing multifactorial rating will be formed for each model. The rating will include an assessment of accuracy, profitability, riskiness, and other parameters. Thus, the rating will give a possibility to evaluate adequacy of the model and make a decision about the expediency of buying or renting. In addition, after the publication of the model in the marketplace, regardless of its use, its forecasts and strategies will be constantly monitored and checked with real data, thus the rating of the model will be constantly updated depending on its efficiency. There are scenarios which claim that some models will be removed from the marketplace after a certain time (for example, if new trends appeared on the markets and they were not present during the training of an old model).

It is not true that models with a lower rating will not bring their developers any income. It must be remembered that the developer sets the cost and terms of use of each model independently. It is possible

that models with high ratings will be too expensive for some traders and they will be ready to use cheaper models with a lower rating (for example, working in non-priority markets).

3.6. Marketplace of machine learning models

Each user of DataTrading system has the opportunity to develop his/ her model of machine learning. This model will be further used in screeners, trade advisors or for scoring ICO / IPO. If desired, these models can be published in the marketplace of machine learning models and sold to other participants of the system.

The developer sets the conditions for using each model, for example:

- cost of use;
- time of use at the specified price (one month, year, all the time);
- the maximum number of traders who can purchase this model (from one (exclusive use) to an unlimited number).

In addition, it is planned that the system will recommend the optimal competitive pricing strategy and terms of use for the DataTrading Marketplace based on the results of back-testing, which the developer can accept or reject.

Traders are sometimes interested in using unique models that are not available to other participants of the system, so reducing the number of customers who can use this strategy simultaneously increases the competitiveness of this model in the marketplace. Also, the developer can create exclusive right to use the model (only one trader can use this model). Since the system will keep a constant record of quality control and settings of all models, developers will not have the opportunity to cheat and sell the same model or its easy modification.

What is more, traders will be able to prepare collective or individual applications for the development of machine learning models for exclusive use, and the system will guarantee the interests of all parties (a user who make a purchase as a result will receive a model of appropriate quality, and the developer will receive the appropriate payment). In addition to development orders, any other applications may also be submitted to the marketplace, for example, open challenges with or without a prize fund, applications for consulting services and so on.

3.7. External Modules

Our team will begin integrating the platform into the most popular brokerage platforms immediately after the release of the first version of DataTrading. If integrated successfully, the trading advisors of DataTrading will be used for the placement of orders in these systems based on the received trading signals.

Thus, our team will do everything to make DataTrading system as convenient as possible and beneficial for its users.

3.8. Blockchain infrastructure

Blockchain is a technology of distributed ledger, usually used for the decentralization of information and management systems. The technology of blockchain in its various implementations underlies all cryptocurrencies.

DataTrading system will use the blockchain to provide:

- transparency of agreements between all users of the platform;
- quality control of artificial intelligence;
- control of intellectual property (without disclosing technological features of the implementation).

Having considered all the advantages and disadvantages of implementing the blockchain systems, we came to the conclusion that the most relevant decision would be to create a private blockchain of the Ethereum network and make its public monitoring service (such as etherscan.io) available. The hash of all the last “unsaved” blocks from private blockchain will be recorded in the public blockchain Ethereum in order to ensure reliability and avoid the situation of “double waste”. In this way, we can ensure the unchangeability of the entire history of the DataTrading blockchain, while making all transaction costs inside the blockchain zero for the system participants.

Ensuring transparency of mutual settlements between all users of the platform

As in traditional cryptocurrencies, all mutual settlements between users of the system will be recorded in the blockchain and anyone will be able to conduct their audit. Also, the blockchain will provide pseudonymity of the participants. It means that the whole history of operations of any address will be available in the registry of the blockchain due to openness and transparency, but the users will not know which participant is hidden behind which address.

Providing quality control of artificial intelligence and preservation of intellectual property

Each trained model of artificial intelligence will undergo automatic quality control. All key information on this model will be recorded in the blockchain without a possibility to be changed or tampered. In the blockchain will be recorded:

- results of back-testing and quality control;
- the author of the model;
- version of the model;
- date of creation;
- description;
- hash parameters and settings of machine learning algorithms;
- other data.

The recording and availability of this information in the blockchain not only guarantee quality control, but also preserve the intellectual property of the author.

4. ECONOMIC MODEL OF THE SYSTEM

DataTrading token (DTT) will be used to ensure the efficient operation of the economic model of the DataTrading system. The initial distribution of DTT will start during the token sale (for more details see Section 6 “Token sale”).

DTT will serve as an inner system currency and will be used to pay for services of the system, or to reward developers. At the moment, there are several scenarios of using DTT:

- subscription payment for using the services of the system;
- subscription payment for access to the machine learning constructor, payment of the use of computing power to create and train own models;
- purchasing models and strategies developed by other participants of the system in the marketplace;
- payment of personal or collective orders.

Subscription to use the services of the system

Each user should pay for subscription in order to gain access to the system’s services, screeners of financial instruments, trading strategies, ICO / IPO scoring modules, performing operations in the marketplace, etc. There will be different options for subscription, so a trader can choose the package of services that he needs.

Payment for the use of machine learning constructor

If a developer or trader wants to use the DataTrading system to develop his own trading advisor, screener, scoring or forecasting model, he can purchase a subscription to access the machine learning constructor. In addition to the subscription, the user will be asked to pay for the use of computing power of the system (calculated as the number of hours spent on the calculation multiplied by the cost of one hour of calculation). The resulting training models can be used either for the user’s own trading or for sale in the marketplace. For more information on using the machine learning constructor, see Section 3.4. “Open Constructor of Machine Learning Models”.

Purchasing of models and strategies developed by other system participants in the marketplace

Each user has the opportunity to purchase alternative models developed by other system participants in addition to using standard models of forecasts, scoring, and scanning available in various versions of the DataTrading subscription. Depending on the terms of use, created by the developer, the models may have:

- limitation on the maximum number of users (in order to ensure the uniqueness of the result of the forecast and, as a consequence, to limit the circle of traders who can benefit from the use of this model, thus increasing the interest of traders in this model);
- limited or unlimited time of use (the trader can either purchase the eternal right to use a model or for a certain period of time, for example, for a month with the possibility to extend the right to use in the next period)

- the developer's ability to receive a part of the profit obtained as a result of using his model (for example, the developer can set a low price for using the model, but indicate a certain percentage of the profit that his model brings to a trader).

Payment for personal or collective orders

It is planned to implement an order placement unit on the basis of the DataTrading marketplace, which can be used in different ways. For example, a trader or a group of traders can place an application for the development of an exclusive forecasting model or screener for a particular market. Its infrastructure will ensure the preservation of the rights of both customers and performers: if the trained model meets the conditions of the order, the developer will have a guarantee of receiving a reward. In addition, this module can be used to organize challenges, competitions, etc. In such a scenario, any competition is an order and the developer whose model will show the best performance on the quality control module will be considered the winner of the competition and will receive a reward. For more details on using the machine learning marketplace, see section 3.6.

5. ROADMAP

2012: getting Started with Artificial Intelligence and Neural Networks.

2014: the beginning of work with the technology of blocking and cryptocurrencies.

Q1 2015: development of Neural Networks for the analysis of the stock market and the market of CO2 quotas.

Q1 2016: implementation of the first version of the DataTrading system in the form of a plug-in for Bloomberg Terminal.

Q2 2016: the beginning of funds management based on the forecasts of the DataTrading system in the New York Stock Exchange.

Q3 2016: development of the concept of a trade advisor, a generator of trading strategies, stocks and cryptocurrencies screener on neural networks with self-training.

Q1 2017: registered legal entity Big Data Trading Limited, successful use of the DataTrading system to manage funds on the New York Stock Exchange, start of managing funds in the cryptocurrency markets.

Q2 2017: expansion of the team and preparations for the token sale.

November 2017: token pre-sale.

December 2017: token sale.

Q1 2018: expansion of the team of developers, acceleration of the development process. Development of a Stock screener and crypto assets on the basis of neural networks.

Q2 2018: creation of a trading advisor, forecasting of stock quotations/ crypto assets, development of analytical tools for fundamental and news analysis of the stock and crypto market on the module of artificial intelligence.

Q3 2018: development of an open constructor of trading strategies with the ability to connect and train neural networks available to users of the system / community (with the possibility of further sale of trained models to other users for tokens).

Q4 2018: implementation of self-learning neural networks (in the future will be available to the community for making their own forecasts). Scoring ICO / IPOs by artificial intelligence. Start of active marketing promotion of the system.

Q1 2019: DataTrading 1.0 release: developed analytics for stock and cryptocurrency markets based on machine learning and artificial intelligence. Integration of DataTrading 1.0 with the key brokerage platforms. Beginning of implementation of DataTrading 1.0 in the Neural networks ensemble system, preparation of infrastructure for working with Level2 data (work with data in the context of orders).

Q2 2019: active involvement of experts and professional market participants, marketing, further promotion of the system.

Q3 2019: DataTrading 2.0 release: analytics based on a neural network ensemble. Getting started with Level2 data.

Q1 2020: DataTrading 3.0 release: analytics on neural networks based on Level2 data.

6. TOKEN SALE

DataTrading Token (DTT) is a token issued on the Ethereum platform and conforming to the ERC20 standard [19].

Crowdsale of DataTrading Tokens will be conducted in three stages:

- pre-sale 11.20.2017–11.30.2017
- token sale: round A, for the eastern region 12.18.2017–01.08.2018
- token sale: round B, for the western region 02.01.2018–02.22.2018

Terms of the pre-sale:

- pre-sale opens 11.20.2017 9:00 UTC, closes 11.30.2017 22:00 UTC;
- the cost of one token is 0.1 USD;
- 30% bonus for all pre-sale participants;
- available ways to purchase DTT — BTC, ETH;
- the token's rate in cryptocurrencies will be dynamically adjusted relative to the rates of these cryptocurrencies on the exchange;
- the maximum amount of tokens to purchase is \$30,000 at the current exchange rate;
- hard cap (the maximum amount the pre-sale will receive) — 1 million USD;
- if the maximum target amount is received before the official end of the token pre-sale, the pre-sale will be terminated ahead of schedule;
- tokens will be mint within a few days after the end of the pre-sale.

Terms of the token sale:

- will be held two rounds of the token sale: round A, for the eastern region 12.18.2017–01.08.2018, round B, for the western region 02.01.2018–02.22.2018;
- the conditions for round A and round B are identical;
- 360,000,000 DTTs will be issued during the pre-sale and two rounds of a token sale, which will compose 100% of all system tokens;
- the cost of one token is 0.1 USD;
- available ways to purchase DTT — BTC, ETH, LTC, DASH;
- the token's rate in cryptocurrencies will be dynamically adjusted relative to the rates of these cryptocurrencies on the exchange;
- hard cap (the maximum amount the crowdsale will receive) — 12.5 million USD for each round;
- if the maximum target amount is received before the official end of any round of the token sale, this round will be terminated ahead of schedule;
- the minimum target amount (soft cap) for two rounds — 5 million USD;
- if the minimum amount is not received at the end of the second round of the token sale, all funds will be returned to the participants of the token sale (minus the transaction commissions received by the miners);
- if more funds than the minimum target amount are received, but the maximum target amount is not reached, excess tokens will be burned;

- tokens will be released within a few days after the end of the token sale.

Distribution of DTT tokens:

- crowdsale — 80% (288 000 000 DTT);
- bounty program — 1% (3 600 000 DTT);
- fund for the promotion of the platform after the release (free trial subscriptions and services) — 10% (36 000 000 DTT);
- team and advisors — 9% (32 400 000 DTT).

Use of received funds:

- platform development — 30%
- purchase of equipment for building its own cloud computing infrastructure (servers and their maintenance) — 35%
- marketing and promotion — 20%
- legal costs — 6%
- administrative costs — 3%
- development of the community, development of training materials available to customers, organization of conferences — 5%

7. TEAM

Anton Vokrug

CEO

Visionary person, business analyst, serial entrepreneur.

Co-founder of DataHolding (DataProm, DataScoring, DataTrading), mentor on BigData and Blockchain technologies.

Alexander Gandzha

CTO

Data scientist, Cybernetic scientist.

Co-founder of DataHolding (DataProm, DataScoring, DataTrading), mentor on BigData and Blockchain technologies.

Li Yundzhin

Lead Partner

Entrepreneur, Candidate of Technical Sciences, Associate Professor.

Expert in the field of investment, trade, business management.

Li Yungan

Lead Partner

Entrepreneur, financier, founder, expert in the preparation of companies with Chinese capital to enter the IPO of America and Hong Kong.

Dmitry Tomchuk

Managing partner

Founder of the investment fund Fison, a serial entrepreneur, co-founder of DataHolding (DataProm, DataScoring, DataTrading), mentor, financier, founder of financial theory 40/60.

Lena Khlevnaya

Business development

Co-founder of DataHolding (DataProm, DataScoring, DataTrading), mentor on strategic planning, PR and marketing

Julia Chaschina

Development of international relations partner

Expert in international business, entrepreneur, translator, economist.

Awarded the Order of Friendship of the Weifang City Administration for success in work, the Chilu Friendship Order of the Shandong Provincial Government for a special contribution.

Eugene Potemskyi

Data scientist / developer, investment analyst

Expert in the field of investment and technical analysis, mentor on BigData and Blockchain technologies.

Roman Sobko

Head of marketing

Expert in digital marketing, worked with a large e-commerce marketplace.

Bitcoin enthusiast, Miner.

8. LEGAL DETAILS

We have prepared all the necessary standards for the safe sale of DataTrading tokens and understand all responsibility assigned to our company.

Given the uncertain status of cryptocurrencies and digital tokens at the global level, we sell tokens without guarantee of our company in this or that jurisdiction. Despite this, our team will respond to any regulatory request and will seek for full openness and transparency.

Tokens DTT are created specifically for functional use on the DataTrading platform. DTT Tokens are not securities. DTT tokens are sold as a digital asset.

Terms of possession of DTT tokens preclude the right of ownership or the right to the property of DataTrading. It also excludes any influence on decision-making in business development. However, personal opinion and feedback of users as members of the community will be taken into account when discussing development-related issues. DTT Tokens are used exclusively for purchasing services on the DataTrading platform.

We do not recommend buying DTT tokens for investment speculation.

In some jurisdictions, forecast products are treated as financial services regulations, in some cases they are not regulated at all. We will make decisions depending on the legal compliance. We will conduct a legal analysis of the current regulatory rules in this jurisdiction before choosing a specific one. Our company will receive all necessary licenses or permits for activities, or refuse to provide activities on the specific territory.

DTT Tokens have not been registered and will not be registered under the laws of the United States of America, and will not be offered for sale in the United States of America or in the interests of US citizens. That is, the Company assumes no responsibility if the DTT token buyer is a person residing in the USA or its territories or possessions or is a corporation, partnership or other legal entity established in accordance with the USA law, or an agency, branch or an office located in the United States of America relating to a corporation, partnership or other legal entity established under laws other than US law; or a trust, any manager of which is the person described above; or a legal entity whose shares are not sold on the stock exchange, with more than 45% of whose shares are owned or managed in the interests of the individual or legal entity described above; or a member of any type of US troops; for an agent or trustee acting on behalf of or in the interests of the individual or legal entity described above. The decision to buy DTT tokens was not based on information received in the US or from sources in the United States of America.

In case of misunderstanding of one of the language versions of the Whitepaper, the English version is considered as the standard.

Big Data Trading Limited is a legal entity of the company, registered on January 23, 2017 at Shop T18, 3 / F., Cathay Pacific 88 Malls, NO. 125, Wanchai Road, Wanchai Hong Kong.

9. RISKS

Funds received in the process of ICO are not insured. There is no private or public insurance representative, to which the user can apply in case of loss or loss of value.

It is possible that DataTrading platform and subsequent marketing activities associated with it will not achieve success in this pre-sale and token sale for various reasons, including the insolvency of business arrangements or marketing strategies.

Crypto-tokens, including DTT, are a new and completely untested technology. There are additional risks that can not be foreseen at present and which can arise in other forms of risk, as indicated above.

You agree that the use or inability to use DTT tokens is carried out solely at your own risk, thereby removing all responsibility from the DataTrading platform.

DTT Tokens will be sent to you without warranty, either expressed or implied. Agreeing with the above stated conditions, you waive any guarantees of commercial value for the implementation of a specific purpose and take the name without violating anyone's intellectual property rights. The abovementioned exceptions to implied warranties may not apply to you, because some jurisdictions do not allow the exclusion of implied warranties.

References

1. Multi-column Deep Neural Networks for Image Classification (Dan Ciresan, Ueli Meier, Jurgen Schmidhuber; 2012 IEEE Conference on Computer Vision and Pattern Recognition; June 2012)
2. ImageNet Classification with Deep Convolutional Neural Networks (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton; Neural Information Processing Systems Conference; 2012)
3. What I learned from competing against a ConvNet on ImageNet (Andrej Karpathy; Sept. 2014)
4. Artificial intelligence 'as good as cancer doctors' (James Gallagher; bbc.com; Jan 2017)
5. IBM Is Counting on Its Bet on Watson, and Paying Big Money for It (Steve Lohr; nytimes.com; Oct. 2016)
6. IBM's Watson gives proper diagnosis for Japanese leukemia patient after doctors were stumped for months (Alfred NG; nydailynews.com; Aug. 2016)
7. Stock Market Prediction with Artificial Neural Network and Machine Learning: A Review of Literature (Nishchal Sharma, Dr. Chaman S. Chauhan; International Journal of Innovative Research in Computer and Communication Engineering; June 2016)
8. The Massive Hedge Fund Betting on AI (Adam Satariano, Nishant Kumar; bloomberg.com; Sep. 2017)
9. David Siegel on Innovation, Machine Learning, and the Future of Work (twosigma.com; Sep. 2017)
10. World's largest hedge fund to replace managers with artificial intelligence (Olivia Solon; theguardian.com; Dec. 2016)
11. Inside a Moneymaking Machine Like No Other (Katherine Burton; bloomberg.com; Nov. 2016)
12. The World Federation of Exchanges, Market Highlights (H1 2017)
13. coinmarketcap.com
14. Neural network credit scoring models (David West; Computers & Operations Research; 2000)
15. Credit scoring and rejected instances reassigning through evolutionary computation techniques (Mu-Chen Chena, Shih-Hsien Huang; Expert Systems with Applications; 2003)
16. Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Network, and Decision Tree Models (Marijana Zekic-Susac, Natasa Sarlija, Mirta Bencic; 26th International Conference on Information Technology Interfaces; June 2004)
17. Two-stage genetic programming (2SGP) for the credit scoring model (Jih-Jeng Huang, Gwo-Hshiung Tzeng, Chorng-Shyong Ong; Applied Mathematics and Computation, 2006)
18. A Better Comparison Summary of Credit Scoring Classification (Sharjeel Imtiaz, Professor Allan J. Brimicombe; International Journal of Advanced Computer Science and Applications; 2017)
19. github.com/ethereum/EIPs/blob/master/EIPS/eip-20-token-standard.md