

## Криптоэкономика: пилотный анализ инвестиций в ICO-стартапы с использованием нейронных сетей

**Андрей Александрович Козлов** (e-mail: study@demontroual.com), магистр НИУ ВШЭ (г. Москва)

**Андрей Васильевич Власов** (e-mail: study@demontroual.com), преподаватель НИУ ВШЭ (г. Москва); консультант DeMontroyal (г. Пенза)

### Аннотация

В сфере криптоэкономики проект Ethereum (Ethereum Foundation) предоставил всем желающим возможности для выпуска «собственной» криптовалюты, точнее сказать, нового токена на основе одноименной блокчейн-платформы смарт-контрактов без навыков низкоуровневого программирования. Стало очевидно, что данные токены можно использовать для краудфандинга, как это сделал тот же Ethereum в 2014 г. Нерегулируемый и довольно простой по своей реализации способ привлечения альтернативных инвестиций стал популярным среди технологических стартапов, которые связали свою деятельность с блокчейн-технологией. Данный способ был назван первоначальным предложением монет (Initial Coin Offering, ICO). Несмотря на свое название, схожее с IPO, этот инструмент обычно используется в венчурном финансировании (ранней стадии привлечения инвестиций) новых проектов, стартапов, а не для расширения устоявшегося «рабочего» бизнеса. В исследовании авторы использовали машинное обучение для классификации ICO-проектов и их оценки с точки зрения финансового коэффициента ROI (прибыльность инвестиций) на основе обработки и анализа цифровых данных, которые публичны и доступны в интернет-пространстве. Целью исследования является разработка устойчивой и эффективной модели, которая может прогнозировать доходность (или тренд доходности) ROI отдельного ICO-стартапа. В результате работы была построена прогностическая модель, разработан сервис (приложения) по анализу данных для отбора ICO-стартапов.

### Ключевые слова:

криптовалюта, токены, инвестиции, стартап, машинное обучение, нейронные сети, ICO, ITO, ROI, риск, криптоэкономика

**JEL:** C45, C53, C60, C80, D81, M13, O31, P49

### ВВЕДЕНИЕ

В настоящее время технологии распределенного реестра (англ. сокр. DLT) широко используются в различных криптопроектах<sup>1</sup>. Капитализация самой популярной криптовалюты биткойн (Bitcoin) резко выросла примерно с \$15 млрд в январе 2017 г. до \$174 млрд в январе 2018 г.

<sup>1</sup> Distributed Ledger Technology: beyond block chain. A report by the UK Government Chief Scientific Adviser / Government Office for Science. 2016. URL: <https://www.gov.uk/government/publications/distributed-ledger-technology-blackett-review>.

Блокчейн-проект Ethereum предоставил простейшую возможность создания (выпуска) нового токена на базе платформы смарт-контрактов для всех желающих без навыков низкоуровневого программирования. Подобная организация работы с выпуском токенов дала импульс для стремительного роста рынка новых криптовалют<sup>2</sup>. Учитывая очень высокую волатильность топовых криптовалют (в целом — всего рынка), альткоины также имели высокую волатильность. Так, в 2017 г. крипторынок стал востребован для проведения экстремально спекулятивных операций с криптоактивами. Рискованный, манипулируемый держателями больших портфелей различных валют рынок, иногда совершенно непредсказуемый, начал привлекать внимание огромного количества трейдеров-любителей, институциональных инвесторов, фондов; это был рынок<sup>3</sup> с огромными возможностями «выигрыша» и низким уровнем входа на него.

Тогда стало очевидно, что токены могут быть использованы для краудфандинга, как это сделал проект Ethereum в 2014 г. Здесь краудфандинг понимается нами как «альтернативный способ привлечения новых средств (фондов) в технологический стартап» [Власов А. В., 2018; Власов А. В., Абреков М. М., 2018а]. Он был назван ICO, что означает «первоначальное предложение монет» (используется также другое его название — ИТО, где слово «монета» заменено на «токен»); далее будет использоваться термин ICO. Нерегулируемая, простая и доступная процедура ICO стала популярной среди блокчейн-стартапов. Несмотря на свое название, схожее с IPO, это инструмент обычно использовался для финансирования нового проекта, а не для расширения уже устоявшегося, рабочего бизнеса. После нескольких больших историй успеха ICO-проектов и их инвесторов, прибыль которых составила более 200 %<sup>4</sup>, эта процедура стала рассматриваться как альтернативная возможность привлечения ранних (посевных) инвестиций в технологические стартапы, а также расширения капитала компаний из консервативных отраслей (например, производственных). Между тем многие люди, у которых не было опыта и знаний в инвестициях и торговле на фондовом рынке, начали переводить свои сбережения из классического портфеля инвестиций с низким доходом (например, состоящего из банковских вкладов, облигаций, владения недвижимостью) в рискованный, но потенциально высокодоходный рынок альткоинов, в т. ч. покупая токены во время ICO с надеждой заработать суммы, в несколько раз превышающие размер своих вложений.

Учитывая специфику рынка, описанного выше, стало очевидно, что он стал привлекать множество различных мошеннических проектов, особенно с помощью инициирования проведения ICO. При этом почти все добросовестные и хорошо известные проекты старались избегать законодательного регулирования со стороны правительств (в т. ч. используя существующие пробелы в праве той или иной страны) и трактовки (квалификации) токенов, выпускаемых стартапом для получения инвестиций, включая в свои документы оговорку (*disclaimer*) об отказе от ответственности, сообщая участникам ICO о том, что токены — это высокорискованные инвестиции, что получение прибыли от них ничем не гарантируется. Следует также отметить ряд мошеннических проектов, обычно называемых «скамом» [Власов А. В., Абреков М. М., 2018а; Власов А. В., Абреков М. М., 2018b], которые принимали «инвестиционные деньги» и закрывались без каких-либо фактических причин и возврата этих инвестиций. Проведение некоторых ICO сопровождалось значительным ростом случаев фишинга (*phishing*), реализации Понци-схем (*Ponzi scheme*) [Bartoletti M. et al., 2017] и других видов финансового мошенничества,

---

<sup>2</sup> Как правило, их называют альткоины — альтернативные монеты.

<sup>3</sup> На 31 января 2019 г. капитализация биткойнов составила около \$60,7 млрд при общей капитализации рынка криптовалют и токенов около \$113,9 млрд.

<sup>4</sup> «Прибыль» в данном случае — это разница в стоимости токенов в момент их покупки во время ICO и последующей продажи.

на которые на тот момент приходилось около 10 %<sup>5</sup> от общего объема собранных средств ICO.

В данном исследовании авторы применили алгоритмы машинного обучения для классификации ICO-проектов, оценки показателя ROI с использованием сбора, обработки и анализа общедоступных данных веб-страниц (в интернете). Тексты технического описания проекта (из *white paper*, *yellow paper* и др.), информация о членах команды и участниках (учредителях), консультантах и эдвайзерах (*advisors*), детали эмиссии токенов, наличие pre-ICO, длительность процесса ICO, возможные скидки и бонусы (*bounty*, прочая информация о лояльности) и др. были применены в качестве параметров (набора) для построения модели. Помимо прогнозируемых (будущих) значений стоимости токенов созданная модель данных определила значимость параметров, которые могут быть использованы экспертами в выборочном аналитическом мониторинге (например, динамическом просмотре агрегированных данных) и индивидуальной экспертной оценке ICO-проектов [Власов А. В., 2018; Власов А. В., Абреков М. М., 2018b].

### **КРАТКИЙ ОБЗОР ПО ТЕМАТИКЕ ИССЛЕДОВАНИЯ**

Научных работ, непосредственно напрямую связанных с тематикой данного исследования, обнаружено не было. Однако есть отдельные научные статьи, которые рассматривают сферу инвестиций в токены ICO-проектов.

В работе «Первичный выпуск монет (ICO) и построение платформы» [Li J., Mann W., 2018] авторами была разработана модель предоставления рекомендаций для регуляторов. В ней исследователи анализируют обстоятельства, которые указывают, следует ли запретить или разрешить ICO. Они рационализируют использование ICO для запуска одноранговых (p2p) ИТ-платформ.

Кристиан Каталини и Джошуа С. Ганс описывают преимущества ICO перед традиционным акционерным финансированием в статье «Первичный выпуск монет (ICO) и стоимость крипто токенов» [Catalini C., Gans J. S., 2018]. Данная статья — это обзор ключевых аспектов процесса ICO и их преимуществ для использования технологическими предпринимателями. Исследователи показывают, что механизм ICO позволяет генерировать конкуренцию покупателей за токен, которая раскрывает его потребительскую ценность без того, чтобы говорить о готовности конечных потребителей продукта платить за него.

В статье [Bian S. et al., 2018] авторы предлагают собственный подход к идентификации мошенничества при организации ICO. Группа исследователей разработала рейтинговую систему на основе моделей глубокого обучения. Была построена модель обработки естественного языка для доступной текстовой информации из таких источников, как «белая книга» (*white paper*), веб-сайты проектов, информация о команде и основателях и др.

Следующая работа [Fenu G. et al., 2019] наиболее близка к проводимому исследованию. В ней ученые изучают различные данные, собранные с ICO-трекеров (информационных веб-ресурсов с различными данными об ICO-проектах), и данные котировок токенов с веб-ресурса CoinMarketCap (<https://coinmarketcap.com>). Они делают попытку поиска факторов, которые, скорее всего, влияют на успешность процедуры ICO; применяют многофакторный анализ к таким факторам, как категории, рейтинги от ICObench (<https://icobench.com>), размер команды, блокчейн-платформа, страна происхождения. Некоторые инсайты, такие как «успешные страны» (например, США и Словения), были выделены авторами, но не рационализированы и не формализованы<sup>6</sup>.

<sup>5</sup> Прим.: экспертная оценка.

<sup>6</sup> Ранее один из соавторов данной статьи также участвовал в аналогичном исследовании [Jin S., Ali R., Vlasov A. V., 2017] факторов успеха проведения ICO.

## **МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ**

Прежде всего для решения задач исследования необходимо было определить, какие данные необходимо собрать и могут быть собраны, исходя из их доступности [Власов А. В., Абреков М. М., 2018b]. Ниже приводятся различные виды данных, которые авторы смогли формализовать на основе полуавтоматизированной работы (*data mining*) с различными веб-ресурсами:

- общая информация (название проекта, страна, токен, члены команды, документы, категория проекта, и т. д.);
- информация о процедуре ICO (дата открытия и закрытия продаж, объем выпуска (эмиссии) токенов, целевые показатели продажи токенов, в т. ч. пресеил (или pre-ICO), цена токенов и т. д.);
- информация из социальных сетей (статистика из Twitter, обратные гиперссылки (*backlinks*) на веб-сайт и оценки (*score*), упоминания в социальных сетях и т. д.) [Jin S., Ali R., Vlasov A. V., 2017];
- информация об ИТ-разработке («дорожная карта» проекта, активность разработки на *github.com*, выбор блокчейн-платформы, детали смарт-контракта и т. д.);
- результаты процедуры ICO (объем привлеченных средств при продаже, цена на токены на криптобирже, ROI, объем торгов и т. д.).

В настоящее время все необходимые данные для исследования не могут быть найдены на одном ресурсе (веб-странице). Нет единого реестра, который бы содержал все ICO-проекты, в т. ч. для традиционного рынка нет такого веб-сайта (например, для IPO [Власов А. В., Абреков М. М., 2018a]). Это отчасти следствие децентрализованного характера природы ICO, в том числе из-за этого порой невозможно определить место происхождения проекта (место его регистрации). Поэтому авторы для выполнения задач исследования решили, что лучше найти как можно больше различных данных, чтобы наиболее полно увидеть всю картину в целом по рынку и правильно подойти к оценке инвестиций и рисков<sup>7</sup>.

В то же время в интернет-сети существовало множество так называемых ICO-трекеров. Обычно ICO-трекер имеет список некоторого количества ICO-проектов, выбранных по нераскрываемому провайдером данного сервиса публично принципу; данный список проектов может быть сформирован из токенов с «платным размещением» (листинг для целей рекламы проекта) или являться фильтром проектов с некоторым количеством параметров. На данных трекерах была представлена информация только об ICO-проектах; каждый трекер имел свой функциональный набор сервисов. Не было найдено ни одного трекера, который удовлетворял бы все потребности по количеству размещенных проектов, по своим функциональным возможностям для решения задач данного исследования.

Из-за этого было принято решение использовать различные источники и объединить все данные для получения удовлетворительного количества требований (параметров) в собственный, единый набор данных.

## **ПОДГОТОВКА ДАННЫХ**

Очистка и предварительная обработка данных — наиболее важный шаг для этого проекта и, как правило, для машинного обучения в целом. Ошибки, допущенные на этом этапе, влияют на все дальнейшие расчеты и могут не только снизить эффективность модели, но и сделать ее несостоятельной. Также следует учитывать как техническую, так и «бизнес-логику» решения. Первый момент касается статистической информации о таких характеристиках, как распределение, выбросы, нормализация и отсутствие данных.

---

<sup>7</sup> ISO 31000:2018. *Risk management — Guidelines*. URL: <https://www.iso.org/iso-31000-risk-management.html>.

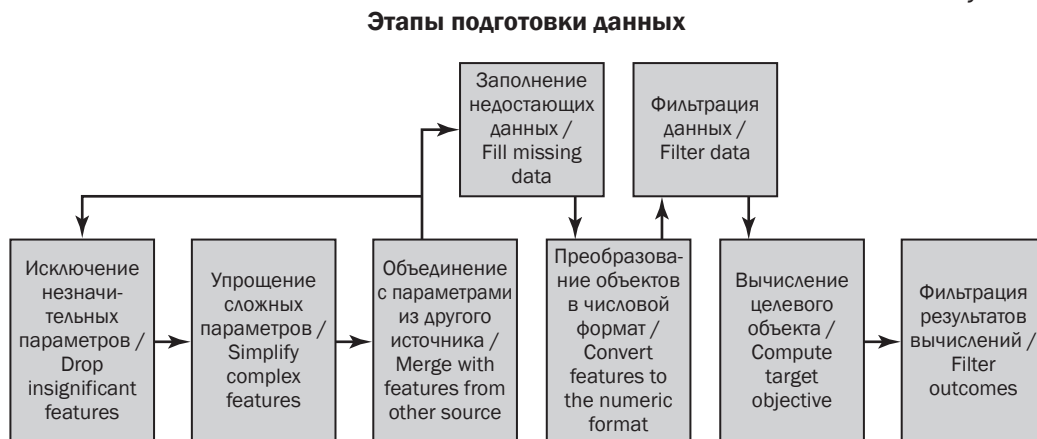
Стоит отметить, что использование в расчетах цен на токены ICO и торговых цен (на биржах) в одной валюте, долларах США, нельзя считать лучшим вариантом, поскольку биткойн имеет высокую волатильность и может сильно повлиять на показатель ROI при его расчете в долларах США, даже если сам токен не был стабильным относительно биткойна. Из-за этой причины необходимо было вычислить дополнительные значения цены биткойна (BTC) и цены эфира (*ether* — ETH) во время начала ICO и первого торгового дня на криптобиржах.

Чтобы подготовить набор данных, потребовалось исключить наблюдения (фрагменты набора), которые не включали схожий по составу и структуре объем информации (для объединения в общий набор данных), как и большинство других, и эти недостающие данные (наблюдения) не могли быть вычислены. Поэтому к набору данных были применены следующие фильтры:

- процедура ICO должна быть завершена;
- есть достоверная информация о цене токена во время проведения ICO в валютах USD и BTC;
- есть достоверная информация о цене токена в первый торговый день (на бирже) в валютах USD и BTC.

Этапы подготовки данных схематично показаны на рис. 1.

Рисунок 1



Источник: составлено авторами.

Информация о ROI не была представлена в наборе первоначальных (сырых) данных. Данный показатель рассчитывался на основе данных из других полей набора, таких как: цена токена на дату начала процедуры ICO и цена токена в первый торговый день при размещении на криптобирже (т. е. при первоначальном появлении токена на бирже) по формуле (1).

$$ROI = \frac{Total\ Revenue - Total\ Cost}{Total\ Cost} \tag{1}$$

Целью данного исследования является прогноз получения прибыли или риска потери инвестиций, сделанных на этапе ICO. Предполагается, что инвестор (хедж-фонд) может купить токены проекта только по официальной цене во время ICO, это было одним из ограничений исследования, и продать их только на криптовалютной бирже, получая при этом прибыль или убыток. В данном случае общая выручка (потенциальный доход) — это цена токена на одной из криптобирж в день их первого листинга (средняя цена может быть использована в случае, если размещение токена произошло на нескольких биржах

одновременно, в один день). Общая сумма расходов (первоначальных инвестиций) — это цена токена во время проведения ICO.

Следует учитывать еще одно ограничение: показатель ROI рассчитывался по цене в день листинга на «первой» из криптобирж. Есть несколько причин для данного выбора:

- самый короткий период времени проведения, как правило, уменьшает влияние внешних факторов, таких как цена биткойна, рыночный цикл, новое государственное регулирование;

- период времени должен быть фиксированным для всех ICO-проектов (наблюдений) для того, чтобы произвести поиск паттернов;

- цена в первый день листинга должна быть ближе к цене токена во время ICO и базироваться в основном на особенностях процесса ICO. С другой стороны, анализ цены после более длительного периода времени торговли на бирже должен быть основан на техническом анализе; не стоит забывать о рыночных манипуляциях, например схемах «пампа» (*pump*) и «дампа» (*dump*). Анализ на основе паттернов биржевой торговли и данных временных рядов хорошо известен, но выходит за рамки данного исследования;

- это предоставило возможность использовать большее количество наблюдений, чем в случае анализа набора с данными более длительных периодов, потому что часть наблюдений пришлось бы отфильтровать, т. е. убрать;

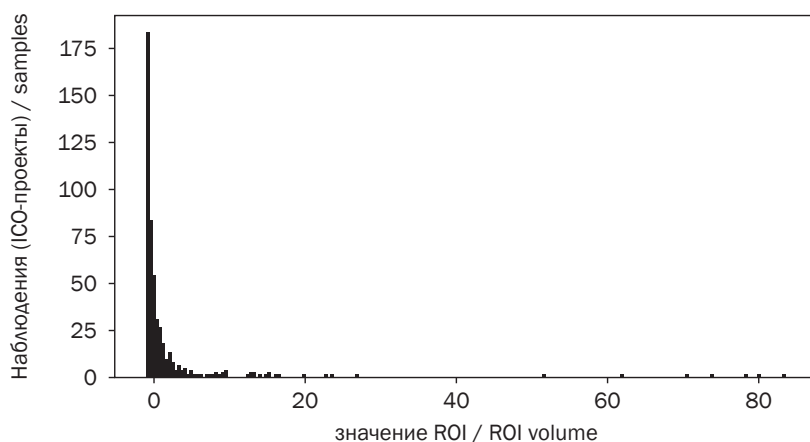
- ICO-проекты имеют различные «дорожные карты» и скорость прогресса их развития. Поэтому сложно определить оптимальный период времени, по истечении которого стоимость токена должна (может) вырасти.

Поскольку данные были получены из различных источников, преобразованных из нестрогих форматов, в наборе данных для анализа были выбросы, которые могли быть ошибками преобразования, ошибками исходных данных или нетипичными (необычными) проектами. Даже тюнинг модели для устойчивости к выбросам мог негативно отразиться на качестве модели, использованных в ней весов.

В расчетах исследования показатель ROI может принимать отрицательные значения, где  $-1$  (полная потеря). Большинство значений этого набора данных лежали между значениями  $-1$  и  $1$ . Согласно гистограмме (рис. 2) распределение положительно искажено. В связи с тем что огромные положительные значения данного показателя могли бы слишком сильно повлиять на веса нейронной сети, наблюдения со значением показателя ROI больше  $10$  ( $1000\%$ ) были удалены из набора данных для дальнейшего анализа.

Рисунок 2

### Распределение ROI



Источник: составлено авторами.

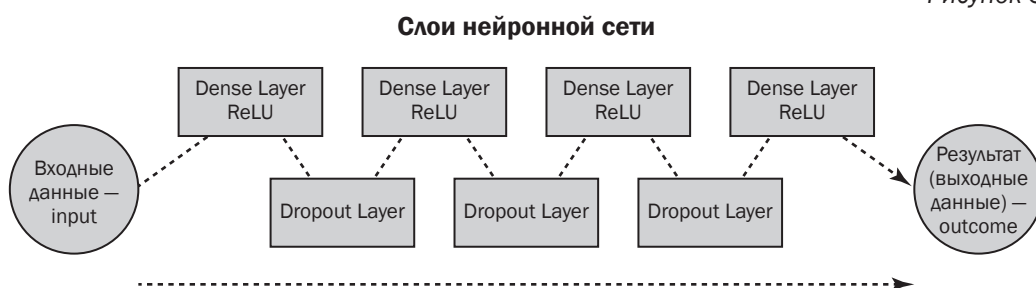
Как правило, при решении задач с использованием машинного обучения набор данных разделяется на обучающие и тестовые наборы. Для тестового набора был выбран 20-процентный размер. Набор данных для проверки (*validation set*) не использовался, т. к. полученный набор данных был относительно невелик. В наборе данные были распределены неравномерно, существовала вероятность того, что случайный сплитинг (*splitting*) приведет к некоторому дисбалансу между тестовым и обучающим наборами. Перекрестная проверка (*cross validation*) с KFold использовалась для окончательной настройки и оценки модели.

Вместо прогнозирования конкретных (числовых) значений показателя ROI было выбрано построение прогноза прибыльности инвестиций как таковых, т. е. были сделаны прогнозы типа «да или нет». Если ROI имеет положительное значение, то инвестиции являются прибыльными, в противном случае инвестор теряет свои деньги. В связи с этим рассматривался диапазон прогнозируемых значений в интервале от 0 до 1. В расчетах за ориентир (целевое пороговое значение) было взято значение ROI больше 0,5<sup>8</sup>, при котором субъективно считается, что модель прогнозирует «достаточный уровень» прибыльности вложений инвестора в ICO-проект.

### РЕАЛИЗАЦИЯ МОДЕЛИ

Для решения задачи была разработана искусственная нейронная сеть с использованием библиотеки *keras*. Слои данной нейронной сети<sup>9</sup> показаны на рис. 3.

Рисунок 3



Источник: составлено авторами.

Создаваемая модель будет предсказывать бинарную классификацию, поэтому для использования была выбрана бинарная функция потерь кроссэнтропии. Для решения этой проблемы хорошо работает оптимизатор (оптимизационный алгоритм) Адама (*Adam optimizer*, или *adaptive moment estimation*). Функция активации PReLU (*Parametric Rectified Linear Unit activation function*) показала лучшие результаты по сравнению с другими модификациями на ReLU. В данной модели двоичная точность является основной метрикой. Поскольку нужно было найти, чтобы значение выхода (*output*) попадало

<sup>8</sup> Размерность 0,5 выбрана как целевой ориентир по доходности потенциального инвестора в криптоактивы (токены).

<sup>9</sup> Слой *Dense — ReLU*. На данном слое запускается функция активации под названием «выпрямитель» (*rectifier*, по аналогии с однополупериодным выпрямителем в электротехнике). Нейроны с данной функцией активации называются ReLU (*rectified linear unit*). ReLU имеет следующую формулу  $f(x) = \max(0, x)$  и реализует простой пороговый переход в нуле. Слой *Dropout* (дропаут). На данном слое контролируется емкость нейронной сети, что позволяет предотвратить переобучение. Используется метод регуляризации [Srivastava N. et al., 2014]. В процессе обучения из общей сети случайным образом многократно выделяется подсеть, обновление весов выполняется только в рамках этой подсети. Нейроны попадают в подсеть с вероятностью  $p$ , которая называется коэффициентом дропаута (0,5).

в диапазон  $[0, 1]$ , для выходного слоя использовалась сигмоидная функция активации<sup>10</sup> (*sigmoid activation function*).

Перед началом работы с моделью был просмотрен текущий набор данных. Первому классу ( $ROI < 0$ ) присвоено 296 наблюдений, второму ( $ROI \geq 0$ ) — 189 образцов.

Классы были не сбалансированы: в текущем положении речь шла о двух неудачных ICO-инвестициях против одной прибыльной. Если искусственная нейронная сеть должна была обучаться на таком наборе данных, то ее точность не смогла бы отразить реальную производительность модели. Модель могла точно предсказать, что каждое тестовое наблюдение (проект) относится к первому классу, и могла выдавать точность намного больше, чем 50 %, но только из-за такого дисбаланса.

Прежде всего модель не должна оцениваться с использованием только одной метрики точности. Точность — это простая и понятная метрика, однако она не делает различий между классами и проблемами, присущими самому набору данных. Также существуют более сложные метрики, такие как:

- площадь под ROC-кривой AUC (площадь под кривой) является агрегированной характеристикой качества классификации, которая не зависит от соотношения цены ошибки. Большее значение AUC-ROC<sup>11</sup> означает лучшую модель классификации. Этот показатель часто используется для сравнительного анализа нескольких моделей классификации;

- метрика конверсии (*precision*) показывает, как много наблюдений соответствовали предсказаниям (т. е. построенной модели), были отнесены к определенному классу, фактически принадлежат этому классу. Результат находится в диапазоне  $[0, 1]$ ;

- метрика полноты (*recall*) показывает, сколько образцов, которые принадлежат к определенному классу, были предсказаны, чтобы принадлежать к этому классу по модели. Результат находится в диапазоне  $[0, 1]$ ;

- F1-мера (*f1-score*) совмещает средние гармонические (*harmonic mean*) метрики конверсии и полноты<sup>12</sup>.

Затем данный несбалансированный набор данных должен быть сбалансирован. Есть разные способы сделать это. Используем технику искусственной модификации датасета для выравнивания соотношения классов. Их можно разделить на чрезмерную выборку по классам меньшинства, недостаточную выборку по классам большинства и сочетание этих подходов.

В нашем случае комбинация SMOTE и Tomek использовалась для балансировки набора данных. Была взята библиотека *imblearn* (*imbalanced learning* — библиотека для несбалансированных классификаций, несбалансированного обучения), которая имеет SMOTETomek ресемплинга (*resampling*). Перед пересчетом набор данных состоял из 485 наблюдений (ICO-проектов), а после ресемплинга был сокращен до 460 наблюдений. Но это все еще намного меньше, чем удаление половины наблюдений класса большинства.

Модель I была натренирована с такими же параметрами, как предыдущая: 10 наблюдений в 1 поднабор, 100 эпох, 10 KFold сплитов.

<sup>10</sup> Гладкая монотонная возрастающая нелинейная функция.

<sup>11</sup> Прим.: естественным кажется порог, равный 0,5, но он не всегда оказывается оптимальным (это субъективный уровень порога).

<sup>12</sup> *Precision* можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющихся положительными, а *recall* показывает, какую долю объектов положительного класса нашел алгоритм. *Precision* и *recall* не зависят от соотношения классов и потому применимы в условиях несбалансированных выборок. Существует несколько различных способов объединить *precision* и *recall* в агрегированный критерий качества. F-мера — это среднее гармоническое *precision* и *recall*.



## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ОБСУЖДЕНИЕ

Метрика перекрестной проверки подтвердила точность модели. Рассчитанная модель показала 79 % точности (это среднее значение по всем результатам расчетов), которое дала наилучшая модель).

Чтобы доказать результаты данного расчета, был выполнен отдельный прогноз и рассчитаны следующие метрики:

- AUC составил 0,78 для тестового набора;
- Precision\*: 0,82 — для убыточных наблюдений (отрицательный ROI); 0,76 — для прибыльных (положительный ROI);
- Recall\*: 0,64 — для убыточных наблюдений, 0,9 — для прибыльных;
- F1-score: 0,72 — для убыточных наблюдений, 0,82 — для прибыльных.

Учитывая полученные результаты и сбалансированный набор данных, можно сделать вывод, что модель предсказывает целевые значения гораздо лучше, чем модель случайного выбора (*random guess*). На основе модели с точностью 79 % стало возможно определение вектора инвестиции в ICO-стартапы, их прибыльность или убыточность.

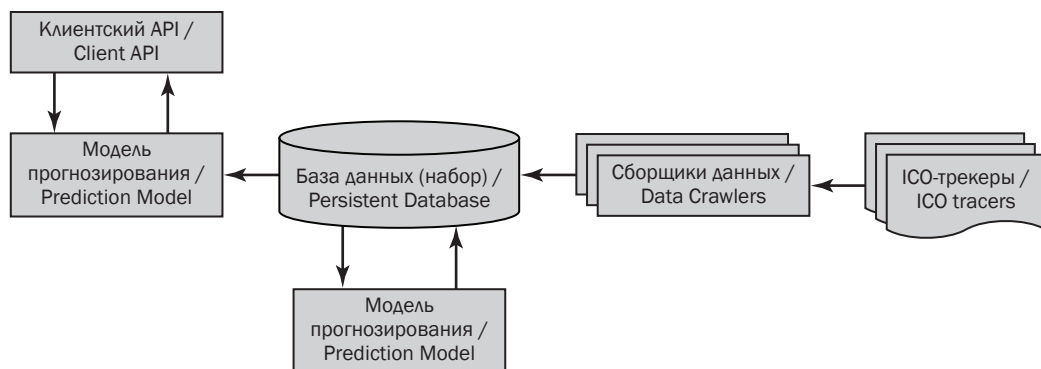
Заметим, что исходный набор данных содержал более 3 тыс. наблюдений (ICO-проектов)<sup>13</sup>. После очистки и удаления выбросов осталось только 518. Из них 234 проекта имеют положительный ROI, что означает, что эти ICO были прибыльными (в рассматриваемом «инвестиционном» периоде).

Результаты исследования могут быть использованы в качестве основы для прогнозирования. Были реализованы приложения, сборщики данных, скрипт предварительной обработки данных, модель прогнозирования. В реализованном сервисе использовалась база данных MongoDB в качестве постоянного хранилища, но в случае повышения требований к обработке данных она может быть заменена.

Функциональность сервиса может быть расширена путем добавления клиентского API для прогнозирования нового потока данных от пользователей. Внутренние приложения сервиса должны быть взаимосвязаны по протоколам http или gpc. Реализация таких процессов, как сбор данных и обучение (переобучение) модели, может быть доработана и автоматизирована. Пример описанной архитектуры сервиса (приложений) приведен на рис. 4.

Рисунок 4

### Архитектура сервиса (приложений)



Источник: составлено авторами.

<sup>13</sup> Рассчитывались как доли. Это максимально возможное количество проектов, по которым удалось собрать и проанализировать данные в период проведения исследования (февраль–май 2018 г.).

Это является логическим продолжением полученных результатов исследования с учетом их практического характера. Помимо разработки сервиса (приложений) сама модель прогнозирования может быть улучшена как минимум путем добавления новых данных (параметров) из дополнительных источников информации. Автоматизация процессов анализа рисков также позволит непрерывно (в режиме реального времени) осуществлять мониторинг, изучать и производить настройку модели более точно.

## **ЗАКЛЮЧЕНИЕ**

Для реализации задач исследования были разработаны различные ИТ-компоненты архитектуры сервиса (приложений) по мониторингу, анализу и прогнозированию рисков ICO-стартапов.

Первый компонент — это приложение по сбору данных. Сбор достаточных и структурированных данных сам по себе является сложной задачей. Она была решена путем просмотра различных источников данных, выбора наиболее полезных и выгрузки-загрузки этих данных различными способами, что было реализовано в виде приложения, написанного на Golang и Python.

Следующий компонент — это очистка и предварительная обработка данных, где Python-скрипт загружает данные из постоянного хранилища (базы данных) MongoDB в библиотеку Pandas DataFrame. Затем применяется запуск различных фильтров для отброса (удаления) наблюдений, которые имеют недостаточно полную информацию или не удовлетворяют другим минимальным требованиям для дальнейшего анализа. Наборы данных из разных источников объединяются в более общие наборы с помощью применения общих функций, таких как токен или имя. Наконец, функции преобразуются в формат, который может быть применен для работы с нейронной сетью.

Для решения задачи классификации и прогнозирования прибыльности инвестиций (положительный ROI) была разработана искусственная нейронная сеть. Средний ROI среди прибыльных ICO (по выборке исследования) составил 47 %, что является относительно высоким показателем для данного типа инвестиций.

Учитывая то, что крипторынок высоковолатильный, и вероятность того, что подобные инвестиции не принесут никакой прибыли, намного выше, чем вероятность инвестиционного успеха, такая модель прогнозирования может оказаться очень ценной для целей критического отбора (исключения) ряда ICO-проектов от потенциального инвестирования в них. Разработанные компоненты могут быть использованы в качестве основы для создания сервиса по мониторингу ICO-проектов, архитектура которого описана в предыдущем разделе (рис. 4). Сама модель прогнозирования рисков может быть улучшена в первую очередь использованием наиболее полного (более широкого) набора данных. При этом решение отдельных задач сбора и обработки данных может проводиться вручную, что потребует дополнительных ресурсов.

Необходимо заметить, что могут быть разработаны другие типы нейронных сетей как для анализа текста, так и для анализа торговых данных временных рядов [Alessandretti L. et al., 2018]. Это, возможно, приведет к логике использования ансамблей моделей, использование которых потенциально поможет обеспечить наиболее точные прогнозы.

## **Благодарности**

*Основная часть исследования была подготовлена в результате работы над магистерской диссертацией (весной 2018 г.) в Национальном исследовательском университете «Высшая школа экономики» (НИУ ВШЭ) А. А. Козловым под руководством А. В. Власова. Авторам хотелось бы выразить благодарность анонимному эксперту, который принял участие в консультациях и рецензировании статьи.*

**Оговорка (Disclaimer)**

Текст данной статьи носит исключительно справочный характер. Любое решение купить токены или ценные бумаги в ходе предполагаемого размещения следует принимать исключительно на основе информации, представленной в официальных документах (*white paper*, проспекте эмиссии ценных бумаг и пр.), опубликованных в целях указанного размещения. Текст данной статьи не является рекламой криптовалют, токенов и ценных бумаг. Точка зрения авторов может отличаться от точек зрения аффилированных с ними организаций или противоречить им. Инвестиции в неликвидные криптовалюты, токены и ценные бумаги связаны с высокой долей риска и подходят исключительно для профессиональных инвесторов, имеющих повышенную терпимость (толерантность) к риску и не ожидающих немедленного и быстрого вывода инвестиций в денежные средства. Указанная в тексте информация не может быть воспроизведена, повторно распространена или опубликована полностью или частично для любой цели без письменного разрешения авторов. Она не может быть использована с целью создания финансовых инструментов, продуктов или любых индексов.

**Список источников / References**

Власов А. В. Криптоэкономика: идентификация рисков стартапов в процессах управления инновациями // Управление финансовым риском. 2018. № 4 (56). С. 308–322 [Vlasov A.V. (2018). Cryptoeconomics: risk identification of startups in the processes of innovation management. *Upravlenie finansovym riskom — Financial Risk Management*, 2018, no. 4 (56), pp. 308–322. (In Russ.)].

Власов А. В., Абреков М. М. IPO российских компаний: классический и альтернативные способы привлечения инвестиций // Модели, системы, сети в экономике, технике, природе и обществе. 2018. № 1 (25). С. 18–31 [Vlasov A.V., Abrekov M.M. (2018a). Russian IPO: Classical and Alternative Ways of Raised Funds. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve — Models, Systems, Networks in Economics, Technology, Nature and Society*, no. 1 (25), pp. 18–31. Available at: <https://mss.pnzgu.ru/mss218> (In Russ.)].

Власов А. В., Абреков М. М. Криптоэкономика: методика экспертной оценки ICO стартапов в процессах управления финансовыми инновациями // Актуальные вопросы экономики и права. 2018. Т. 12. № 4. С. 745–761 [Vlasov A.V., Abrekov M.M. (2018b). Cryptoeconomics: expert evaluation methodology of ICO start-ups in the processes of financial innovation management. *Aktual'nye voprosy ekonomiki i prava — Actual Problems of Economics and Law*, vol. 12, no. 4, pp. 745–761. Available at: <http://dx.doi.org/10.21202/1993-047X.12.2018.4.745-761> (In Russ.)].

Alessandretti L. et al. (2018). Anticipating cryptocurrency prices using machine learning. Cornell University Library — arXiv.org. Available at: <https://arxiv.org/abs/1805.08550>.

Bartoletti M., Carta S. et al. (2017). Dissecting Ponzi schemes on Ethereum: identification, analysis, and impact. Available at: <https://arxiv.org/abs/1703.03779>.

Bian S., Deng Z., Li F., Monroe W. et al. (2018). IcoRating: A Deep-Learning System for Scam ICO Identification. Cornell University Library — arXiv.org. Available at: <https://arxiv.org/abs/1803.03670>.

Catalini C., Gans J.S. (2018). Initial Coin Offerings and the Value of Crypto Tokens. MIT Sloan Research Paper No. 5347-18; Rotman School of Management Working Paper No. 3137213. Available at: <http://dx.doi.org/10.2139/ssrn.3137213>.

Fenu G. et al. (2018). The ICO Phenomenon and Its Relationships with Ethereum Smart Contract Environment. Cornell University Library — arXiv.org. Available at: <https://arxiv.org/abs/1803.01394>.

Jin S., Ali R., Vlasov A.V. (2017). Cryptoeconomics: Data Application for Token Sales Analysis. International Conference Information Systems 2017. Special Interest Group on Big Data Proceedings. Available at: <http://aisel.aisnet.org/icis2017b/1>.

Li J., Mann W. (2018). Initial Coin Offering and Platform Building. WFA. Available at: <http://dx.doi.org/10.2139/ssrn.3088726>.

Srivastava N., Hinton G., Krizhevsky A. et al. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958.

Дата поступления статьи в редакцию: 31 мая 2018 г.

A. A. Kozlov, A. V. Vlasov

## Cryptoeconomics: Pilot Study on Investments in ICO Startups Using Neural Networks

### Authors' affiliation:

**Andrey A. Kozlov** (e-mail: study@demonroyal.com), ORCID 0000-0003-3811-9560, National Research University Higher School of Economics (HSE), Moscow 101000, Russian Federation

**Andrey V. Vlasov** (e-mail: study@demonroyal.com), ORCID 0000-0001-9227-1892, National Research University Higher School of Economics (HSE), Moscow 101000, Russian Federation; DeMontroyal, Penza 440015, Russian Federation

### Abstract

In the field of cryptoeconomics the Ethereum (Ethereum Foundation) project gave opportunity to create "own" cryptocurrency – new token based on its smart-contract platform to everyone without low-level programming skills. Then it became obvious that tokens could be used for crowdfunding as the Ethereum did in 2014. Unregulated and easy to access such scheme became popular among related to the blockchain tech startups. It was named Initial Coin Offering (ICO/or ITO). Despite its name, which is similar to IPO, this scheme is usually used for venture funding of a new project instead of expanding already well-established working business.

The authors use machine-learning algorithms to classify ICOs and estimate ROI based on public digital data and web-sources. The goal of the research is to develop sustainable and efficient model, which will predict target profit ROI (profit trends) of ICO startup. Data collection and analysis period: Feb-Mar 2018. The prediction model and the application (service) of ICO startups' selection are developed as the result of the study.

Results. There were over 3000 samples of ICO-startups in the research dataset. After cleaning and elimination of outliers, it contained only 518. The number of samples with positive ROI (which means that these ICOs were profitable) was 234. Cross validation metric was confirmed to be accurate. The model achieved 79 % accuracy (average value). To prove this score separated prediction was executed the metrics: for test dataset AUC is 0.78; for profitable samples Precision: 0.76; Recall: 0.9 for profitable; F1-score: 0.82.

Discussions. In order to achieve the objectives of this study, various IT components of the service architecture (applications) were developed to monitor, analyze and predict the risks of ICO startups. An artificial neural network was developed to solve the problem of ROI classification and prediction. The average ROI among profitable ICOS was 47 %.

Taking into consideration that the crypto market is highly volatile and that there is a possibility that such investments will not bring any profit, this model of monitoring, analysis and prediction can be very valuable for the purposes of critical selection (exclusion) of a number of ICO projects from potential investment.

Conclusion. The developed components can be used as a basis of monitoring service of ICO startups. The risk-forecasting model can be improved, foremost, by using the most complete (and wider) set of data. In this case, individual data collection and processing tasks can be performed manually, which will require additional resources. It should be noted that other types of neural networks can be developed for both text analysis and trading data analysis. This may lead to the logic of using a combination of models, which will potentially help to provide the most accurate predictions.

### Keywords:

cryptocurrency, tokens, investment, machine learning, neural networks, ICO, ITO, ROI, risk, crypto-economics

**JEL:** C45, C53, C60, C80, D81, M13, O31, P49

<https://doi.org/10.31107/2075-1990-2019-1-76-87>