

Обнаружение мошеннических операций на рынке ценных бумаг с использованием методов машинного обучения

Н.А. Стройкова, А.Э. Амшоков, К.С. Зайцев

Аннотация. Финансовые махинации с каждым годом представляют все большую угрозу для экономики и качества жизни. Они имеют серьезные последствия для стабильности и устойчивости экономики. В связи с этим финансовые учреждения вынуждены постоянно совершенствовать свои системы обнаружения мошенничества и подозрительных операций. Представленная статья посвящена применению методов машинного обучения для выявления подозрительных операций с точки зрения отмыывания доходов, полученных преступным путем, и финансированию терроризма на рынке ценных бумаг. Целью статьи является исследование эффективности различных моделей машинного обучения при анализе транзакций на фондовом рынке для выявления различных типов отмыывания доходов. В статье последовательно анализируются такие методы машинного обучения, как метод ближайших соседей, деревья решений, случайный лес, метод опорных векторов, логистическая регрессия, наивный байесовский классификатор, дискриминантный линейный анализ. Значительное внимание уделено процессу подготовки данных для проведения исследования, состоящему из нескольких этапов: отбор операций, кодирование параметров для загрузки в модели, нормализация параметров и их объединение в определенном формате. В качестве исходных данных были использованы сделки, совершенные на бирже в режиме анонимных торгов и неторговые операции (зачисления/списания денежных средств и ценных бумаг на брокерский и депозитарный счет). Лучшие результаты были получены при применении комбинированной модели из нескольких методов.

Ключевые слова – подозрительные операции, рынок ценных бумаг, отмыывание денег, машинное обучение, обнаружение сомнительных операций

I. ВВЕДЕНИЕ

Статья получена 20 июня 2022.

Стройкова Наталья Андреевна, Национальный Исследовательский Ядерный Университет МИФИ, аспирант, stroykova_nat@mail.ru

Амшоков Астемир Эдуардович, Банк ВТБ (ПАО), старший разработчик, AEAmshevik@vtb.ru

Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор, KSZajtsev@mephi.ru

Финансовое мошенничество оказывает огромное влияние как на финансовую индустрию, так и на повседневную жизнь. Факты мошенничества могут снизить в целом доверие к отрасли, дестабилизировать сбережения граждан и повлиять на стоимость жизни.

Для контроля подобных систем необходимо своевременно анализировать текущее состояние и превентивно оценивать динамику этого состояния в ближайшем будущем и на основе полученной картины данных принимать решения по управлению системой. Таким образом, ощущается всё большая необходимость в средствах анализа данных в потоке в режиме реального времени. Акцент этой работы смещен в сторону решения задачи предсказания значений временных рядов.

Финансовые учреждения используют различные модели предотвращения мошенничества для решения этой проблемы. Тем не менее, мошенники быстро адаптируются, и со временем используют новые способы проникновения в защищенные модели. Несмотря на все усилия финансовых организаций, надзорных и контролирующих органов, правительства, число случаев финансового мошенничества продолжает расти.

Актуальность этой проблемы подтверждается большим количеством исследований по указанному направлению. Так как количество участников Рынка ценных бумаг (РЦБ), а, соответственно, количество и объем проводимых ими сделок заметно выросли в последние несколько лет. Например, за последние два года объем проводимых сделок увеличился в 2.5 раза (см. рис.1), поэтому сегодня все чаще эту проблему пытаются решить, используя методы машинного обучения.



Рис.1. Статистика объемов торгов на Московской Бирже.

Отталкиваясь от существующих проблем в этой сфере, была сформулирована цель исследования. Это - поиск методов выявления незаконных операций с ценными бумагами, которые позволят сократить объем затрачиваемых ресурсов, на основе использования алгоритмов машинного обучения. Эти модели могут определить закономерности по совокупности проводимых операций на РЦБ и выявить клиентов, причастных к совершению подозрительных операций с позиции отмывания доходов и финансирования терроризма (ОД/ФТ). Достичь этой цели можно с помощью разработки обучаемой программной системы для автоматизированного построения такого типа прогнозов. В настоящем исследовании рассмотрена проблема выявления основных типологий отмывания доходов, полученных преступным путем и, финансирования терроризма на РЦБ, а также предложенное базовое решение задачи выявления подозрительных операций с использованием алгоритмов машинного обучения.

II. РОЛЬ РЫНКА ЦЕННЫХ БУМАГ В СОВЕРШЕНИИ ПОДОЗРИТЕЛЬНЫХ ОПЕРАЦИЙ

РЦБ занимает лидирующую позицию на мировом финансовом рынке. Популярность отрасли рынка ценных бумаг стремительно возрастает на фоне достаточно невысоких процентов по депозитам, предлагаемых вкладчикам кредитными организациями, а также простоты и скорости проведения сделок на биржах. Для РЦБ характерна быстрая смена старых продуктов и услуг новыми, чтобы удовлетворить потребности всех инвесторов – и любителей и квалифицированных инвесторов, которые приходят на РЦБ для целей спекуляции,

хеджирования рисков, а также инвестирования с различным горизонтом.

Наиболее привлекательными для лиц, преследующих незаконные цели, являются такие характеристики рынка ценных бумаг, как высокая скорость заключения сделок, ликвидность финансовых инструментов, возможность конвертировать денежные средства в другой вид активов, международная природа рынков и т.д. Все это не только является преимуществом, но и дает возможность сокрытия и изменения идентификационных параметров первоначального актива, который был получен преступным путем. В связи с чем количество преступных схем, связанных с отмыванием денежных средств, постоянно растет.

Таким образом, РЦБ становится все более популярным и доступным инструментом для обычного обывателя в нашей стране, и количество преступных схем, связанных с отмыванием денежных средств, растет ежедневно. Преступники ухищряются разрабатывать все более запутанные и сложные схемы для их выявления организациями, подконтрольными Росфинмониторингу и Центральному Банку Российской Федерации.

Сложности выявления операций и схем, направленных на отмывание денег через этот сектор заключаются в недостатке информации.

Профессиональные участники фондового рынка в соответствии с Законом №115-ФЗ обязаны сообщать о подозрительных транзакциях, однако учитывая количество проводимых операций и сложность схем для отмывания денежных средств, а также отсутствие возможности обмениваться информацией с другими участниками (биржа, банки, финансовые посредники, иные профессиональные участники и проч.) в виду наличия банковской тайны, это еще больше затрудняет процесс выявления подозрительных операций и их участников.

К популярным схемам незаконного поведения лиц на РЦБ относятся:

1) *Транзит денежных средств (ДС) или ценных бумаг (ЦБ)* - целью данной схемы является быстрый оборот ДС или ЦБ через финансовые организации. Основная суть заключается именно в коротком сроке оборачиваемости актива. Например, лицо зачисляет ДС на брокерский счет, совершает несколько сделок по приобретению ЦБ на бирже (покупает ЦБ в «стакане» по рыночным ценам), заключает обратные сделки на продажу ЦБ, т.е. в конечном итоге у лица на руках снова ДС. Учитывая короткий срок между заключением

прямых и обратных сделок, рыночные цены на один и тот же вид актива могут либо не измениться, либо измениться незначительно. После проведения нехитрой цепочки операций клиент выводит ДС на счет в другой кредитной организации, не той, из которой изначально были зачислены ДС. Транзакции можно повторять неограниченное число раз. Аналогичные действия можно осуществить и с ЦБ. В результате цепочки операций, кредитная организация, в которую выводятся ДС, видит ДС на счете клиента, источником которых является продажа ЦБ через Брокера на бирже, по рыночной цене (в «стакане»). В случае возникновения вопросов об источниках происхождения ДС лица информация из отчета Брокера подтвердит указанную информацию. Аналогично для депозитария.

2) *Конвертация одного вида активов в другой* - Целью данной схемы, что очевидно из ее названия, является конвертация одного вида актива в другой. Например, конвертация ДС в ЦБ или наоборот. ЦБ, обращающиеся на бирже, в большинстве являются высоколиквидными, и в последующем не возникает проблем с их конвертацией обратно в ДС. Обратная ситуация. В депозитарий лицо зачисляет ЦБ, затем через Брокера продает их на РЦБ и получает на руки ДС. Учитывая скорость проведения сделок, а также возможность дистанционно проводить каждую из вышеописанных операций риск вовлечения профессиональных участников в схемы отмывания ДС крайне высок.

3) *Вывод активов за рубеж* - для целей вывода активов за рубеж используются как ДС, так и ЦБ. Но на практике, процедура вывода ЦБ несколько проще, поскольку в РФ контроль за банками гораздо сильнее, чем за иными участниками финансовой системы. В случае вывода с брокерского счета ДС в иностранный банк, такой платеж должен согласовать «валютный контроль» банка, в котором у Брокера открыт специальный брокерский счет. Могут задать дополнительные вопросы, а также отклонить перевод. В случае вывода ЦБ в иностранный депозитарий, перевод осуществляется со счета депозитария в Национальном расчетном депозитарии (далее – НРД) в Euroclear и Clearstream — это системообразующие европейские клиринговые центры, которые предоставляют депозитарные услуги. И затем уже переводятся на счет депозитария, где у клиента открыт счет для перевода ЦБ. Из представленного процесса очевидно, что, как только ЦБ уходят из НРД, контроль за этим активом невозможен. Клиент

может любым удобным способом распорядиться этим активом, например, продать ЦБ, заключив внебиржевую сделку и обналить ДС.

III. ПОСТАНОВКА ЗАДАЧИ

Перед тем, как приступить к выбору алгоритмов, подходящих для достижения поставленной цели, сформулируем задачи, решаемые в рамках проводимого исследования:

1) Разработка методики получения обучающей выборки из операций отмывания преступных доходов для использования методов машинного обучения.

2) Формирование выборки операций с признаками подозрительных операций. Признаки определяются на основе изучения свойств этих операций.

3) Построение формальных моделей типологий операций отмывания преступных доходов на рынке ценных бумаг.

4) Проведение экспериментальной проверки разработанной системы, выбор наиболее эффективной технологии машинного обучения и сравнение результатов обнаружения с имеющимися.

IV. ВЫБОР ИНСТРУМЕНТОВ ДЛЯ РЕШЕНИЯ ЗАДАЧИ

Конечной целью исследования является выявление подозрительных операций, т.е., решение задачи бинарной классификации. Поэтому для деления операций на «подозрительные» и «неподозрительные», нам потребуются следующие алгоритмы машинного обучения.

1. Метод ближайших соседей (K-nearest neighbor) – как простейший метрический кластеризатор, основанный на оценивании сходства объектов. Классифицируемый объект относится к тому кластеру, к которому принадлежат ближайшие к нему объекты обучающей выборки.

2. Деревья принятия решений – так как при правильном обучении может решить поставленную задачу классификации.

3. Случайный лес (Random Forest) – так как они являются хорошим семейством базовых классификаторов; они сложнее простого дерева, но значительно точнее его.

4. Метод опорных векторов (Support vector machines) – включают комбинацию алгоритмов управляемого машинного обучения, применяемого для решения задач классификации. Отличительным свойством данного метода

считается возможность сокращения эмпирической ошибки классификации.

5. Логистическая регрессия (Logistic Regression) – быстро обучается и хорошо подходит для задач бинарной классификации. Статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой..

6. Наивный байесовский классификатор (Naive Bayes) – основным его преимуществом считается малое количество входных данных, необходимых модели для обучения, определение значимых параметров, их оценки и проведения классификации.

7. Дискриминантный линейный анализ (Linear discriminant analysis) – представляющий собой расширение линейного дискриминанта Фишера. Линейный дискриминант Фишера – метод, нацеленный на задачи распознавания образов, обучения компьютерных систем для определения линейных комбинаций признаков и параметров, подробно описывающих или разделяющих данные на N классов или типов событий.

V. ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных принимаются операции клиентов на брокерских и депозитарных счетах, включая их торговую деятельность на бирже. Также результаты исследования должны быть применимы к любым сферам деятельности, в частности к банковской деятельности, в которых данные во временных интервалах обладают характеристиками схожими с финансовыми операциями. Финансовые операции клиента внутри определенного интервала времени мы рассматриваем как временной ряд.

В качестве исходных данных были использованы торговые и неторговые операции. Под торговыми операциями понимаем сделки, совершенные лицом на бирже в режиме анонимных торгов, под неторговыми операциями – зачисления/списания денежных средств и ценных бумаг на брокерский и депозитарный счет соответственно, открытый у профессионального участника рынка ценных бумаг.

Исходные данные для исследования были сформированы на основании реальных торговых сделок клиентов, которые получены с сайта московской биржи и самостоятельно сгенерированных неторговых операций клиентов, на основании изученных схем проведения

подозрительных с позиции отмыывания денег операций на РЦБ (см. табл. 1 и 2).

Таблица 1. Структура данных торговых операций.

Наименование	Описание
Сделка No	Номер сделки
Клиент No	Номер клиента в системе учета профессионального участника рынка ценных бумаг
Дата	Дата проведения сделки
Время	Время проведения сделки
Инструмент	Наименование ценной бумаги
Цена (1 акция)	Цена за 1 акцию
Лотность	Количество ценных бумаг в одном лоте
Объем, лот.	Количество лотов, участвующих в сделке
Объем сделки	Количество ценных бумаг, участвующих в сделке
К/П	Направление сделки (покупка или продажа)

Таблица 2. Структура данных неторговых операций.

Наименование	Описание
Клиент No	Номер клиента в системе учета профессионального участника рынка ценных бумаг
Дата	Дата проведения сделки
Время	Время проведения сделки
Направление операции	Зачисление или списание
Кредитная организация / Депозитарий (реестродержатель)	Кредитная организация или депозитарий зачисления/списания актива
Объем ДС	Объем денежных средств, участвующих в операции
Инструмент (ЦБ)	Наименование ценной бумаги
Объем ЦБ	Количество ценных бумаг, участвующих в сделке
Оценка ЦБ	Оценка ценных бумаг в денежном эквиваленте на дату совершения операции (оценивается по средней цене торгов на бирже в этот день)

Исходные данные были сформированы таким образом, чтобы в них помимо стандартных транзакций, присутствовали транзакции, попадающих под критерии сомнительных операций с позиции отмыывания денег. К основным критериям которых относятся:

1) Короткий срок размещения актива на брокерском/депозитарном счете, т.е. между зачислением и списанием актива проходит до 5-7 рабочих дней.

2) Быстрая оборачиваемость активов – т.е. клиент в течение короткого срока (3 -) изменяет форму актива.

3) Использование нескольких кредитных организаций/депозитариев с целью снижения риска выявления противозаконных действий.

4) Многократность и однотипность совершаемых операций.

Исходные данные содержат информацию об операциях 103 клиентов, 8 из которых однозначно содержат признаки совершения подозрительных с операций. Половина из «плохих» клиентов попадает в обучающую выборку, половина - в целевую. Все операции «плохих» клиентов признаются плохими.

В исходных данных операции клиентов попадают под следующие схемы отмывания денег (см. табл. 3):

Таблица 3 - Клиенты с подозрительными операциями.

Схема ОД/ФТ	Операции клиентов
Транзит ДС (зачисление ДС и списание ДС)	Клиент №5 Клиент №101
Транзит ДС-ЦБ (зачисление ДС и списание ЦБ)	Клиент №31 Клиент №102
Транзит ЦБ-ДС (зачисление ЦБ и списание ДС)	Клиент №63 Клиент №80
Транзит ЦБ (зачисление ЦБ и списание ЦБ)	Клиент № 65 Клиент №103

Клиенты №5, №31, №65, №80 включены в обучающую выборку. Клиентов № 63, №101, №102, №103 считаем целевыми.

VI. ПРИМЕНЕНИЕ МОДЕЛЕЙ И ОЦЕНКА ИХ ЭФФЕКТИВНОСТИ

Применение различных алгоритмов машинного обучения показало следующие результаты:

SVM - при выборе вероятности обнаружения от 0.85 (85%) - записи, которые обнаружил этот метод относятся к 20-ти клиентам, среди которых попали все 4 «плохих» клиента: [102, 101, 103, 81, 4, 62, 68, 69, 52, 58, 63, 1, 97, 29, 57, 43, 99, 35, 59, 42].

K-nearest - при выборе вероятности обнаружения от 0.80 (80%) записи, который нашёл этот метод относятся к 11-ти клиентам, среди которых все 4 «плохих» клиента: [101, 103, 102, 63, 34, 97, 6, 67, 26, 13, 81].

RF - при выборе вероятности обнаружения записи от 0.50 (50%), который нашёл этот метод относятся к 62-ти клиентам, среди которых все 4 «плохих» клиента: [102, 101, 103, 81, 41, 12, 23, 24, 72, 35, 8, 4, 11, 17, 13, 96, 69, 47, 9, 71, 58, 63, 40,

43, 45, 1, 88, 84, 54, 68, 30, 33, 42, 83, 57, 2, 16, 29, 3, 67, 38, 82, 19, 77, 97, 18, 99, 20, 48, 34, 14, 25, 94, 79, 87, 89, 95, 75, 60, 46, 28, 93].

NB – при достаточно высокой точности данный метод упускает из вида один из искомым объектов. Модель находит только 3 записи с искомыми параметрами, это объекты [102, 63, 101]. То есть клиента №103 мы упускаем из вида, поэтому данную модель в сравнении использовать не будем.

LR - модель смогла выявить всех целевых клиентов, но при этом также в результирующую выборку попали еще 81 клиент. С такой точностью данная модель не представляет интереса, так как она скорее указывает небольшую выборку, не попадающую в искомый класс.

LDA - общая точность метода линейного дискриминантного анализа также не высокая, он выделяет большой класс записей клиентов, которых относит к искомому классу (74). Аналогично модели логистической регрессии не представляет для нас интереса, так как не сужает круг искомым объектов.

Результаты моделей нужно оценить. При конвертации вещественного ответа алгоритма (как правило, вероятности принадлежности к классу) в бинарную метку, мы должны выбрать какой-либо порог, при котором 0 становится 1. Естественным и близким кажется порог, равный 0.5, но он не всегда оказывается удачным.

Одним из способов оценить модель в целом, не привязываясь к конкретному порогу, является AUC-ROC (или ROC AUC) — площадь (Area Under Curve) под кривой ошибок (Receiver Operating Characteristic curve). Эта кривая представляет из себя линию от (0,0) до (1,1) в координатах True Positive Rate (TPR) и False Positive Rate (FPR). Для ROC-кривых каждая точка на графике соответствует выбору некоторого порога, и чем больше площадь кривой, тем лучше работает алгоритм.

Для сравнения всех проверенных алгоритмов бинарной классификации рассмотрим их ROC-кривые в совокупности (рис. 2). По обобщенному графику видно, что наибольшую площадь имеют кривые для моделей SVM, LR и LDA.

При комбинировании результирующей выборки всех представленных алгоритмов классификации (без байесовской модели, так как она не обнаруживает один из обязательных искомым объектов) мы получили следующий набор из 6 клиентов, отнесённых к искомому классу: [63, 81, 97, 101, 102, 103]

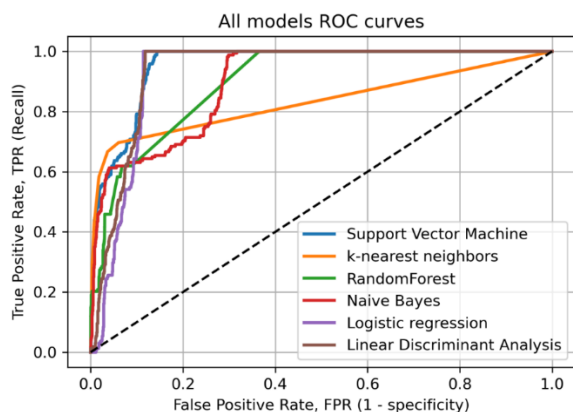


Рис.2. Сравнение ROC-кривых примененных моделей.

Лишними в этой модели выглядят клиенты №81 и №9, поскольку их деятельность не была проанализирована изначально. Но в результате углубленного изучения операций клиентов 81 и 97 стало очевидно, что деятельность данных клиентов обладает признаками типологий отмывания денег, сформулированных ранее, а именно операции клиента 81 попадают под схему *транзит ЦБ-ДС*, а операции клиент 97 – под схему *транзит ДС*.

VII. ЗАКЛЮЧЕНИЕ

В данной работе исследовались методы, решения задачи автоматического выявления подозрительных операций с точки зрения отмывания денег, полученных преступным путем, проводимых на рынке ценных бумаг.

В работе проведен сравнительный анализ эффективности применения алгоритмов машинного обучения, таких, как метод ближайших соседей, деревья решений и случайный лес, метод опорных векторов, логистическая регрессия, наивный байесовский классификатор, дискриминантный линейный анализ.

Для решения поставленной задачи выявления подозрительных операций на рынке ценных бумаг автором предложено применять комбинированный алгоритм как наиболее эффективной.

Для реализации предложенного решения в среде Python Jupyter Notebook был разработан пилотный программный комплекс с модулем оценки качества результатов используемых моделей. Разработанная программа в настоящее

время проходит опытную эксплуатацию на стендах одного из банков.

Исходя из результатов сравнительного анализа применения указанных моделей помимо целевых клиентов, чьи операции были проанализированы изначально, были выявлены и операции иных клиентов, в результате углубленного изучения операций которых удалось подтвердить, что деятельность данных клиентов обладает признаками типологий отмывания денег.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Высшей инженеринговой школе НИЯУ МИФИ за помощь в возможности опубликовать результаты выполненной работы.

БИБЛИОГРАФИЯ

- [1] Макаева, К. И. Рынок ценных бумаг как составная часть финансового рынка / К. И. Макаева, З. Г. Манджиева // Современная наука: проблемы и перспективы: сборник материалов VI Международной научно-практической конференции, Ставрополь, 2017. – Ставрополь: Центр научного знания "Логос", 2017. – С. 40-45.
- [2] Колесова, И. В. Искусственный интеллект и противодействие легализации преступных доходов с использованием криптовалют / И. В. Колесова, Т. А. Стась // Вызовы и возможности финансового обеспечения стабильного экономического роста (ФИНАНСЫ-2019) : Материалы Всероссийской научно-практической конференции: Сборник научных трудов, Севастополь, 2019. – Севастополь: Федеральное государственное автономное образовательное учреждение высшего образования "Севастопольский государственный университет", 2019. – С. 174-177.
- [3] Национальная оценка рисков легализации (отмывания) преступных доходов // Публичный отчет. – 2018. [Электронный ресурс]. — Режим доступа — URL: <https://www.fedsfm.ru/content/files/documents/2018/%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0%20%D1%80%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%20%D0%BE%D0%B4.pdf> (дата обращения 01.04.2022).
- [4] Бекетнова, Ю. М. Сравнительный анализ методов машинного обучения при идентификации признаков вовлеченности кредитных организаций и их клиентов в сомнительные операции / Ю. М. Бекетнова // Финансы: теория и практика. – 2021. – Т. 25. – № 5. – С. 186-199. – DOI 10.26794/2587-5671-2020-25-5-186-199.
- [5] Калагина, Д. С. Применение новых технологий в сфере под/ФТ / Д. С. Калагина, В. С. Логинская // Вектор экономики. – 2019. – № 11(41). – С. 61.
- [6] Алифанова, Е. Н. Выявление зон уязвимостей к риску отмывания денег на различных секторах финансового рынка / Е. Н. Алифанова, Ю. С. Евлахова // Финансовые исследования. – 2015. – № 3(48). – С. 42-52.
- [7] Воронцов, В. К. Прогнозирование временных рядов / Курс «Машинное обучение» [Электронный ресурс] / Режим доступа — URL: <http://www.machinelearning.ru/wiki/images/archive/c/cb/2016041217491Voron-ML-forecasting-slides.pdf> (дата обращения: 24.04.2022)
- [8] Waleed Hilal, S. Andrew Gadsden, John Yawney. Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances – 2022. [Электронный ресурс] / Режим доступа —

- URL:
<https://www.sciencedirect.com/science/article/pii/S0957417421017164> (дата обращения: 14.06.2022)
- [9] Python-school [электронный ресурс] // TensorFlow vs PyTorch: что и когда выбирать для Machine Learning / Режим доступа — URL: <https://python-school.ru/tensorflow-vs-pytorch/> (дата обращения 20.04.2022)
- [10] Mark Eshwar Lokanan, Kush Sharma. Fraud prediction using machine learning: The case of investment advisors in Canada / Machine Learning with Applications – 2022. [Электронный ресурс]. — Режим доступа — URL: <https://www.sciencedirect.com/science/article/pii/S2666827022000111> (Дата обращения 01.06.2022).
- [11] Muhammad Atif Khan Achakzai, Peng Juan. Using machine learning Meta-Classifiers to detect financial frauds, Finance Research Letters Advances – 2022. [Электронный ресурс] / Режим доступа — URL: <https://www.sciencedirect.com/science/article/pii/S1544612322001866> (дата обращения: 14.06.2022)
- [12] Xuting Mao, Hao Sun, Xiaoqian Zhu, Jianping Li. Financial fraud detection using the related-party transaction knowledge graph. Procedia Computer Science – 2022. [Электронный ресурс] / Режим доступа — URL: <https://www.sciencedirect.com/science/article/pii/S1877050922000928> (дата обращения: 14.06.2022)
- [13] Международные стандарты по противодействию отмыванию денег, финансированию терроризма и финансированию распространения оружия массового уничтожения // Рекомендации ФАТФ. – 2012. [Электронный ресурс]. — Режим доступа — URL: <https://www.fatf-gafi.org/media/fatf/documents/recommendations/pdfs/FATF-40-Rec-2012-Russian.pdf> (дата обращения 16.06.2022).

Detection of fraudulent transactions in the securities market using machine learning methods

N. A. Stroykova, A.E. Amshokov, K.S. Zaytsev

Annotation — Financial fraud is a serious problem and an increasing threat to the economy. It has serious consequences for the stability and sustainability of the economy. In this case financial institutions are forced to constantly improve their systems for detecting fraud and suspicious transactions. This article is devoted to the use of machine learning methods to identify suspicious transactions in money laundering and financing of terrorism, provide in the area of security markets. The purpose of the article is to study the effectiveness of various machine learning models in the analysis of transactions in the stock market to identify various typologies of money laundering. The article consistently analyzes such machine learning methods as K-nearest neighbors (K-nearest), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes Classifier (NB), Discriminant Linear Analysis (LDA). Considerable attention is paid to the process of preparing data for research, which includes several stages such as selecting operations, encoding parameters for loading into the model, normalizing the parameters and combining them in a specific format. Transactions made on the stock exchange in the anonymous trading mode and non-trading operations (crediting/debiting funds and securities to the brokerage and depository accounts) were used as initial data. The best results were obtained using the SVM, LR and LDA models, but the combined model is the most effective.

Keywords — Suspicious Transactions, Securities Market, Money Laundering, Machine Learning, Security Fraud.

REFERENCES

- [1] Makaeva, K. I. The securities market as an integral part of the financial market / K. I. Makaeva, Z. G. Mandzhieva // Modern science: problems and prospects: collection of materials of the VI International Scientific and Practical Conference, Stavropol, 2017. - Stavropol: Center for Scientific Knowledge "Logos", 2017. - P. 40-45.
- [2] Kolesova, I. V. Artificial intelligence and countering the legalization of criminal proceeds using cryptocurrencies / I. V. Kolesova, T. A. Stas // Challenges and opportunities for financial support of stable economic growth (FINANCE-2019): Proceedings of the All-Russian Scientific and Practical Conference: Collection of scientific papers, Sevastopol, 2019. - Sevastopol: Federal State Autonomous Educational Institution of Higher Education "Sevastopol State University", 2019. - P. 174-177.
- [3] The National risk assessment of legalization (laundering) of criminal proceeds // Public report. – 2018. [Online resource]. - [website] - URL: <https://www.fedsfm.ru/content/files/documents/2018/%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0%20%D1%80%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%20%D0%BE%D0%B4.pdf> (Accessed 04/01/2022).
- [4] Beketnova, Yu. M. Comparative analysis of machine learning methods for identifying signs of involvement of credit institutions and their clients in questionable transactions / Yu. M. Beketnova // Finance: theory and practice. - 2021. - T. 25. - No. 5. - S. 186-199. – DOI 10.26794/2587-5671-2020-25-5-186-199.
- [5] Kalagina, D. S. The use of new technologies in the field of AML / CFT / D. S. Kalagina, V. S. Loginskaya // Vector of Economics. - 2019. - No. 11(41). – S. 61.
- [6] Alifanova, E. N. Identification of vulnerability zones to the risk of money laundering in various sectors of the financial market / E. N. Alifanova, Yu. S. Evlahova // Financial research. - 2015. - No. 3(48). - S. 42-52.
- [7] Vorontsov, V. K. Time series forecasting / Machine learning course [Online resource] / [website] — URL: http://www.machinelearning.ru/wiki/images/archive/c/cb/20160412121749!Voron-ML_forecasting-slides.pdf (Date of request 04/24/2022)
- [8] Waleed Hilal, S. Andrew Gadsden, John Yawney. Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances – 2022. [Online resource] / [website] — URL: <https://www.sciencedirect.com/science/article/pii/S0957417421017164> (Date of request: 06/14/2022)
- [9] Python-school [Online resource] // TensorFlow vs PyTorch: что и когда выбирать для Machine Learning / [website] — URL: <https://python-school.ru/tensorflow-vs-pytorch/> (Date of request 04/20/2022)
- [10] Mark Eshwar Lokanan, Kush Sharma. Fraud prediction using machine learning: The case of investment advisors in Canada / Machine Learning with Applications – 2022. [Online resource]. — [website] — URL: <https://www.sciencedirect.com/science/article/pii/S2666827022000111> (Date of request 06/14/2022).
- [11] Muhammad Atif Khan Achakzai, Peng Juan. Using machine learning Meta-Classifiers to detect financial frauds, Finance Research Letters Advances – 2022. [Online resource] / [website] — URL: <https://www.sciencedirect.com/science/article/pii/S1544612322001866> (Date of request: 06/14/2022)
- [12] Xuting Mao, Hao Sun, Xiaoqian Zhu, Jianping Li. Financial fraud detection using the related-party transaction knowledge graph. Procedia Computer Science – 2022. [Online resource] / [website] — URL: <https://www.sciencedirect.com/science/article/pii/S1877050922000928> (Date of request: 06/14/2022)
- [13] The International Standards for Combating Money Laundering, Financing of Terrorism and Financing Proliferation of Weapons of Mass Destruction // FATF

Recommendations. – 2012. [Online resource]. — [website]
— URL: <https://www.fatf-gafi.org/media/fatf/documents/recommendations/pdfs/FATF-40-Rec-2012-Russian.pdf> (Date of request 06/16/2022)