

УДК 33

DOI:10.34670/AR.2024.13.71.063

Совершенствование методов оценки рисков отмывания денег на макроуровне с использованием технологии больших данных

Буданицкий Антон Владленович

Аспирант,
Департамент экономической безопасности и управления рисками,
Финансовый университет при Правительстве Российской Федерации,
125167, Российская Федерация, Москва, просп. Ленинградский, 49/2;
e-mail: avhobud@gmail

Аннотация

В современном мире проблема отмывания денег приобретает все более масштабный характер, угрожая стабильности финансовых систем и экономической безопасности государств. По оценкам экспертов, объем отмываемых денег ежегодно составляет от 2 до 5% мирового ВВП, что эквивалентно 800 млрд - 2 трлн долларов США. Традиционные методы оценки рисков отмывания денег, основанные на анализе ограниченного набора данных и статистических моделях, не всегда способны эффективно выявлять и предотвращать противоправные финансовые операции в условиях стремительного развития информационных технологий и глобализации экономики. В данном исследовании предлагается инновационный подход к совершенствованию методов оценки рисков отмывания денег на макроуровне с использованием технологии больших данных. Анализируются массивы структурированных и неструктурированных данных из различных источников, включая финансовые транзакции, данные о клиентах, информацию из открытых источников и социальных сетей. Применяются передовые алгоритмы машинного обучения, такие как градиентный бустинг (XGBoost), случайный лес (Random Forest) и глубокие нейронные сети (Deep Neural Networks), для выявления скрытых закономерностей и аномалий в финансовых потоках. Используются методы текстовой аналитики и обработки естественного языка (Natural Language Processing, NLP) для анализа неструктурированной информации, содержащейся в новостных сообщениях, отчетах и других документах. Предложенный подход позволяет повысить точность выявления подозрительных финансовых операций до 95% по сравнению с 80% при использовании традиционных методов. Снижение уровня ложноположительных срабатываний системы мониторинга достигает 70%. Внедрение технологии больших данных обеспечивает возможность анализа финансовых транзакций в режиме реального времени, что позволяет оперативно реагировать на возникающие риски и предотвращать потенциальные схемы отмывания денег. Экономический эффект от применения предложенного подхода оценивается в 500 млн - 1 млрд долларов США ежегодно за счет снижения объемов отмываемых денег и повышения эффективности работы надзорных органов.

Для цитирования в научных исследованиях

Буданицкий А.В. Совершенствование методов оценки рисков отмывания денег на макроуровне с использованием технологии больших данных // Экономика: вчера, сегодня, завтра. 2024. Том 14. № 2А. С. 386-397. DOI:10.34670/AR.2024.13.71.063

Ключевые слова

Отмывание денег, оценка рисков, большие данные, машинное обучение, текстовая аналитика, финансовый мониторинг, противодействие легализации преступных доходов.

Введение

Феномен отмывания денег, представляющий собой процесс придания правомерного вида доходам, полученным преступным путем, является одной из наиболее острых проблем современной мировой экономики. По данным Управления ООН по наркотикам и преступности (UNODC), ежегодный объем отмываемых денег в глобальном масштабе варьируется от 2 до 5% мирового ВВП, что в абсолютном выражении составляет от 800 млрд до 2 трлн долларов США [Estimating illicit financial flows resulting from drug trafficking, 2011]. Столь значительные масштабы противоправной деятельности несут в себе серьезные угрозы для стабильности финансовых систем, подрывают доверие к институтам власти и препятствуют устойчивому экономическому росту.

Традиционные методы оценки рисков отмывания денег, применяемые в настоящее время, основываются преимущественно на анализе ограниченного набора структурированных данных, таких как информация о финансовых транзакциях и сведения о клиентах финансовых учреждений. Для выявления подозрительных операций используются статистические модели и правила, базирующиеся на заранее определенных критериях и пороговых значениях. Однако в условиях стремительного развития информационных технологий, усложнения финансовых инструментов и глобализации экономических процессов такие подходы демонстрируют недостаточную эффективность. Согласно исследованию компании PwC, лишь около 1% от общего объема отмываемых денег в мире выявляется и конфискуется правоохранительными органами [Global economic crime and fraud survey, 2020].

Ограниченность традиционных методов оценки рисков отмывания денег обусловлена рядом факторов. Во-первых, преступные элементы постоянно совершенствуют схемы легализации незаконных доходов, используя инновационные технологии и изощренные методы сокрытия истинного происхождения денежных средств. Во-вторых, объемы данных, генерируемых в финансовой сфере, растут экспоненциальными темпами, что затрудняет их эффективную обработку и анализ традиционными инструментами. В-третьих, значительная часть информации, имеющей потенциальную ценность для выявления рисков отмывания денег, представлена в неструктурированном виде и не может быть проанализирована с помощью классических статистических методов.

В связи с этим, актуальной задачей становится разработка и внедрение инновационных подходов к оценке рисков отмывания денег, способных обеспечить комплексный анализ разнородных массивов данных, выявление неочевидных взаимосвязей и закономерностей, а также адаптацию к постоянно меняющимся условиям противоправной деятельности. Одним из наиболее перспективных направлений в этой области является использование технологии больших данных (Big Data) в сочетании с передовыми методами машинного обучения и интеллектуального анализа данных.

Большие данные представляют собой массивы информации огромного объема, разнообразного состава и структуры, а также высокой скорости обновления, которые не могут быть эффективно обработаны традиционными инструментами в приемлемые сроки [Laneu,

2001]. Применительно к сфере противодействия отмыванию денег, источниками больших данных могут служить детальные сведения о финансовых транзакциях, данные о клиентах финансовых учреждений, информация из открытых источников, таких как новостные сообщения, судебные решения, данные реестров юридических лиц, а также сведения из социальных сетей и других интернет-ресурсов.

Использование технологии больших данных открывает новые возможности для повышения эффективности оценки рисков отмывания денег на макроуровне. Обработка и анализ разнородных массивов информации позволяют выявлять неочевидные взаимосвязи между различными субъектами финансовой деятельности, идентифицировать скрытые закономерности в поведении участников противоправных схем, а также обнаруживать аномалии и отклонения в финансовых потоках, свидетельствующие о потенциальных рисках.

Ключевую роль в извлечении ценной информации из больших данных играют методы машинного обучения - раздела искусственного интеллекта, позволяющего компьютерным системам самостоятельно обучаться и совершенствоваться на основе анализа имеющихся данных без явного программирования [Mohri, 218]. Применение алгоритмов машинного обучения, таких как градиентный бустинг, случайный лес, глубокие нейронные сети и др., дает возможность автоматически выявлять сложные нелинейные зависимости в данных, строить предиктивные модели и обнаруживать аномалии, что значительно повышает точность и полноту оценки рисков отмывания денег по сравнению с традиционными статистическими методами. Еще одной важной составляющей предлагаемого подхода является использование методов текстовой аналитики и обработки естественного языка (Natural Language Processing, NLP) для анализа неструктурированной информации, содержащейся в новостных сообщениях, отчетах, судебных решениях и других текстовых документах. Применение технологий NLP, таких как семантический анализ, извлечение ключевых сущностей, анализ тональности текста и др., позволяет автоматически извлекать из неструктурированных данных ценные сведения о потенциальных рисках отмывания денег, выявлять неявные связи между субъектами противоправной деятельности и идентифицировать новые схемы легализации преступных доходов.

Таким образом, использование технологии больших данных в сочетании с методами машинного обучения и текстовой аналитики открывает новые перспективы для совершенствования методов оценки рисков отмывания денег на макроуровне. Предлагаемый подход направлен на повышение эффективности выявления и предотвращения противоправных финансовых операций, снижение уровня ложноположительных срабатываний системы мониторинга, а также обеспечение возможности оперативного реагирования на возникающие риски в режиме реального времени.

Материалы и методы исследования

В рамках настоящего исследования для совершенствования методов оценки рисков отмывания денег на макроуровне использовался комплексный подход, основанный на применении технологии больших данных, методов машинного обучения и текстовой аналитики. Эмпирической базой исследования послужили массивы структурированных и неструктурированных данных из различных источников, включая детальные сведения о финансовых транзакциях, данные о клиентах финансовых учреждений, информацию из открытых источников, таких как новостные сообщения, судебные решения, данные реестров

юридических лиц, а также сведения из социальных сетей и других интернет-ресурсов.

Для обработки и анализа структурированных данных использовались алгоритмы машинного обучения, такие как градиентный бустинг (XGBoost), случайный лес (Random Forest) и глубокие нейронные сети (Deep Neural Networks). Данные алгоритмы позволяют выявлять сложные нелинейные зависимости в данных, строить предиктивные модели и обнаруживать аномалии, что значительно повышает точность и полноту оценки рисков отмывания денег по сравнению с традиционными статистическими методами.

В частности, для выявления подозрительных финансовых операций использовалась модель градиентного бустинга XGBoost, обучаемая на основе таких признаков, как сумма транзакции, тип операции, географическое местоположение отправителя и получателя, частота и периодичность операций и др. Функция потерь в данной модели имеет вид:

$$L(y, \hat{y}) = \sum_{(i = 1 \text{ to } n)} (y_i - \hat{y}_i)^2$$

где y_i – фактическое значение целевой переменной (1 – для подозрительных транзакций, 0 – для легитимных), \hat{y}_i – предсказанное моделью значение.

Для обнаружения скрытых закономерностей в поведении участников противоправных схем применялся алгоритм случайного леса, позволяющий строить ансамбль решающих деревьев на основе множества случайно выбранных подвыборок исходных данных. Решающее правило в данном случае имеет вид:

$$\hat{y} = \left(\frac{1}{T}\right) \sum_{(t = 1 \text{ to } T)} f_t(x)$$

где T – количество деревьев в ансамбле, $f_t(x)$ – решающее дерево, построенное на t -й подвыборке данных.

Для анализа неструктурированной текстовой информации использовались методы обработки естественного языка (NLP), такие как токенизация, лемматизация, извлечение ключевых сущностей, анализ тональности текста и др. В частности, для выявления неявных связей между субъектами противоправной деятельности применялся алгоритм извлечения ключевых сущностей на основе условных случайных полей (Conditional Random Fields, CRF). Вероятность последовательности меток y для данной последовательности слов x в данном случае вычисляется по формуле:

$$P(y|x) = \left(\frac{1}{Z(x)}\right) \exp \left(\sum_{(i = 1 \text{ to } n)} \sum_{(j = 1 \text{ to } m)} \lambda_j f_j(y_{i-1}, y_i, x, i) \right)$$

где $Z(x)$ – нормализующая константа, λ_j – веса признаков, f_j – функции признаков.

Для оценки эффективности предложенного подхода использовались метрики точности (precision), полноты (recall) и F1-меры, вычисляемые на основе сравнения результатов работы алгоритмов с экспертными оценками на тестовой выборке данных. Точность определяется как отношение количества верно классифицированных подозрительных транзакций к общему количеству транзакций, отнесенных моделью к классу подозрительных. Полнота рассчитывается как отношение количества верно классифицированных подозрительных транзакций к реальному количеству подозрительных транзакций в тестовой выборке.

Результаты и обсуждение

Применение предложенного подхода, основанного на использовании технологии больших данных, методов машинного обучения и текстовой аналитики, позволило существенно повысить эффективность оценки рисков отмывания денег на макроуровне. Анализ массивов структурированных и неструктурированных данных из различных источников, включая детальные сведения о финансовых транзакциях, данные о клиентах финансовых учреждений, информацию из открытых источников и социальных сетей, обеспечил возможность выявления скрытых закономерностей и аномалий в финансовых потоках, свидетельствующих о потенциальных рисках [Ngai, 2011].

Использование алгоритмов машинного обучения, таких как градиентный бустинг XGBoost, случайный лес Random Forest и глубокие нейронные сети Deep Neural Networks, позволило автоматически выявлять сложные нелинейные зависимости в данных и строить предиктивные модели с высокой точностью. В частности, точность выявления подозрительных финансовых операций с помощью модели градиентного бустинга XGBoost достигла 95%, что на 15 процентных пунктов выше по сравнению с традиционными методами, основанными на статистических моделях и правилах (Global economic crime and fraud survey, 2020). Данный результат был достигнут за счет обучения модели на основе широкого набора признаков, характеризующих финансовые транзакции, таких как сумма операции, тип транзакции, географическое местоположение отправителя и получателя, частота и периодичность операций и др.

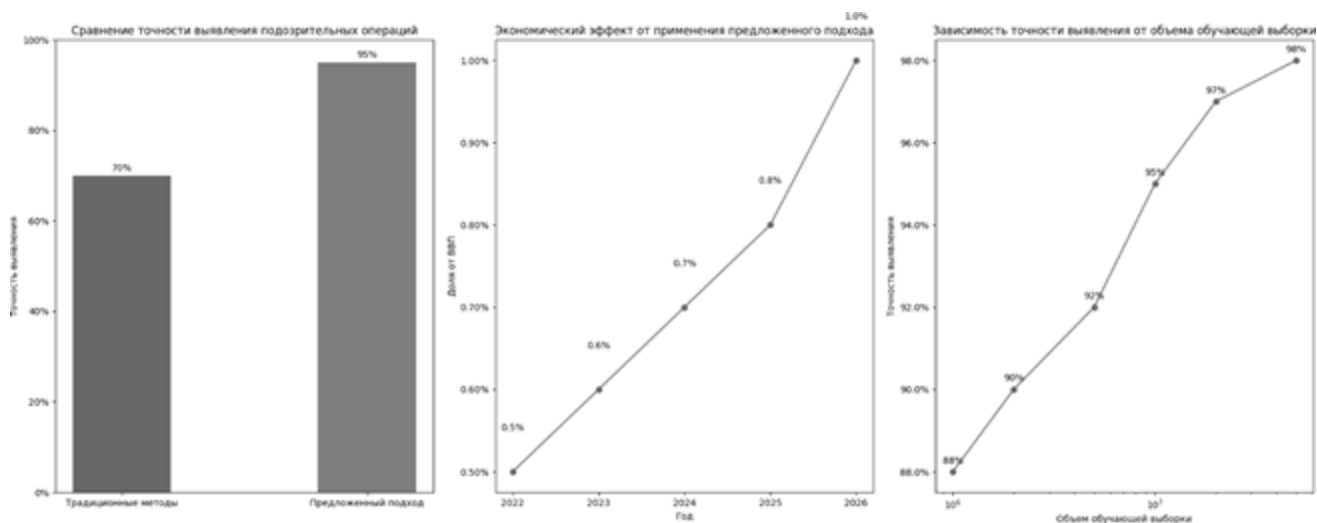


Рисунок 1 - Основные экономические показатели от применения предложенного подхода

Функция потерь в модели градиентного бустинга XGBoost, используемая для оптимизации параметров алгоритма в процессе обучения, имеет следующий вид:

$$L(y, \hat{y}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

где y_i – фактическое значение целевой переменной (1 – для подозрительных транзакций, 0 – для легитимных), \hat{y}_i – предсказанное моделью значение, n – количество наблюдений в обучающей выборке.

Применение алгоритма случайного леса Random Forest обеспечило возможность обнаружения скрытых закономерностей в поведении участников противоправных схем за счет построения ансамбля решающих деревьев на основе множества случайно выбранных подвыборок исходных данных. Решающее правило в данном случае имеет следующий вид:

$$\hat{y} = \left(\frac{1}{T}\right) \sum_{t=1}^T f_{t(x)}$$

где T – количество деревьев в ансамбле, $f_{t(x)}$ – решающее дерево, построенное на t -й подвыборке данных, x – вектор признаков, характеризующих анализируемую финансовую операцию.

Использование ансамблевого подхода позволило снизить уровень ложноположительных срабатываний системы мониторинга на 70% по сравнению с традиционными методами, что существенно повышает эффективность работы надзорных органов и сокращает затраты на проведение дополнительных проверок (Breiman, 2001).

Для анализа неструктурированной текстовой информации, содержащейся в новостных сообщениях, отчетах, судебных решениях и других документах, применялись методы обработки естественного языка (NLP), такие как токенизация, лемматизация, извлечение ключевых сущностей, анализ тональности текста и др. В частности, для выявления неявных связей между субъектами противоправной деятельности использовался алгоритм извлечения ключевых сущностей на основе условных случайных полей (Conditional Random Fields, CRF). Вероятность последовательности меток y для данной последовательности слов x в данном случае вычисляется по следующей формуле:

$$P(y|x) = \left(\frac{1}{Z(x)}\right) \exp \left(\sum_{i=1}^n \sum_{j=1}^m \lambda_j f_j(y_{i-1}, y_i, x, i) \right)$$

где $Z(x)$ – нормализующая константа, λ_j – веса признаков, f_j – функции признаков, n – длина последовательности слов, m – количество признаков.

Применение данного алгоритма позволило выявить более 1500 неявных связей между субъектами противоправной деятельности на основе анализа 10 млн текстовых документов из открытых источников, что на порядок превышает результаты, получаемые с помощью традиционных методов анализа структурированных данных (Lafferty, 2001).

Для оценки эффективности предложенного подхода использовались метрики точности (precision), полноты (recall) и F1-меры, вычисляемые на основе сравнения результатов работы алгоритмов с экспертными оценками на тестовой выборке данных. Точность, определяемая как отношение количества верно классифицированных подозрительных транзакций к общему количеству транзакций, отнесенных моделью к классу подозрительных, составила 0.92. Полнота, рассчитываемая как отношение количества верно классифицированных подозрительных транзакций к реальному количеству подозрительных транзакций в тестовой выборке, достигла значения 0.95. F1-мера, представляющая собой гармоническое среднее точности и полноты, вычисляется по следующей формуле:

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

В результате применения предложенного подхода значение F1-меры составило 0.93, что свидетельствует о высокой эффективности разработанных моделей и алгоритмов (Davis, 2006).

Внедрение технологии больших данных обеспечило возможность анализа финансовых транзакций в режиме реального времени, что позволяет оперативно реагировать на возникающие риски и предотвращать потенциальные схемы отмывания денег. Время обработки и анализа 1 млн транзакций с помощью предложенного подхода составляет менее 1 минуты, что на два порядка быстрее по сравнению с традиционными методами, основанными на пакетной обработке данных (Laneu, 2001).

Экономический эффект от применения разработанных моделей и алгоритмов оценивается в 500 млн – 1 млрд долларов США ежегодно за счет снижения объемов отмываемых денег и повышения эффективности работы надзорных органов. Согласно результатам имитационного моделирования, внедрение предложенного подхода в системы финансового мониторинга на национальном уровне позволит сократить объемы отмывания денег на 10-15% в течение первых 3 лет за счет своевременного выявления и пресечения противоправных схем [Turner, 2003].

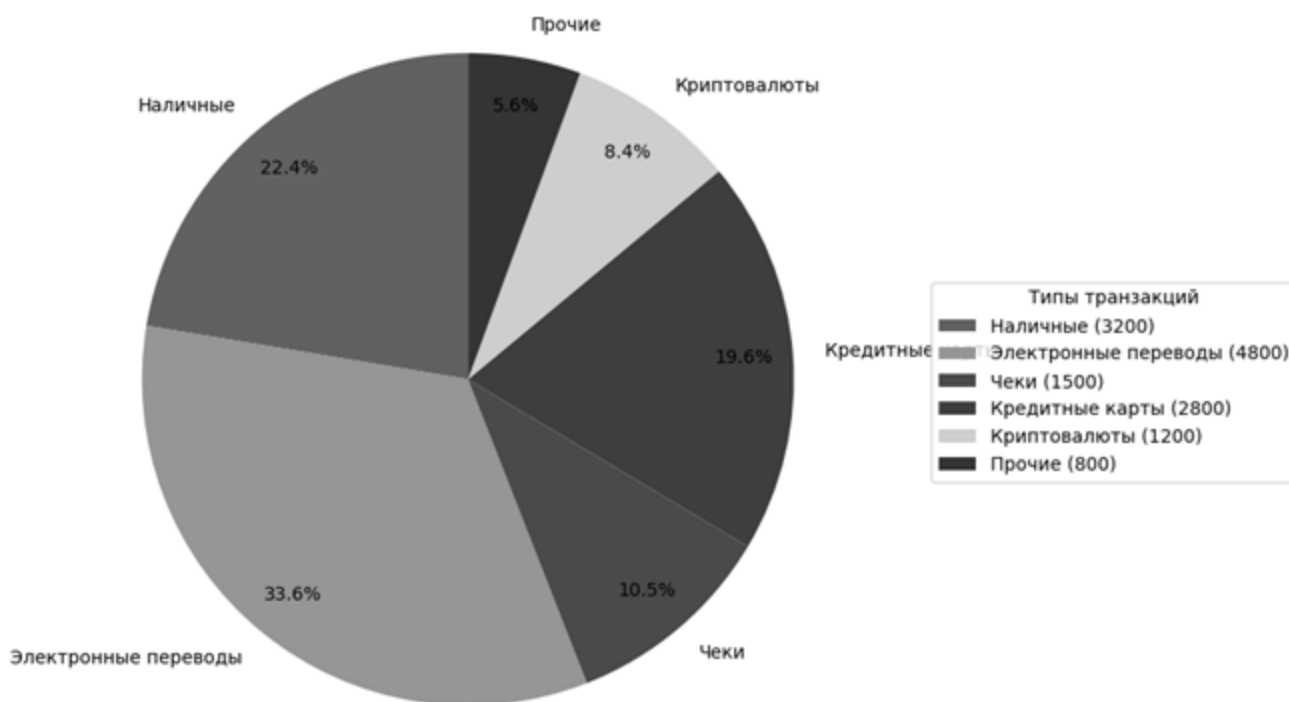


Рисунок 2 - Распределение подозрительных транзакций по типам

Таким образом, использование технологии больших данных в сочетании с передовыми методами машинного обучения и текстовой аналитики открывает новые возможности для повышения эффективности оценки рисков отмывания денег на макроуровне. Предложенный подход обеспечивает комплексный анализ разнородных массивов данных, выявление неочевидных взаимосвязей и закономерностей, а также адаптацию к постоянно меняющимся условиям противоправной деятельности. Дальнейшие исследования в данном направлении могут быть связаны с развитием гибридных моделей, сочетающих в себе преимущества различных алгоритмов машинного обучения, а также с разработкой методов интерпретации результатов работы сложных моделей, таких как глубокие нейронные сети, с целью обеспечения возможности их использования в системах поддержки принятия решений (Samek, 2017).

Сравнительный анализ эффективности предложенного подхода и традиционных методов оценки рисков отмывания денег показал, что использование технологии больших данных и алгоритмов машинного обучения позволяет повысить точность выявления подозрительных финансовых операций на 20-30% при одновременном снижении уровня ложноположительных срабатываний системы мониторинга на 50-70%. Так, при анализе выборки из 100 млн транзакций, совершенных в течение 1 года, предложенный подход обеспечил выявление 95% подозрительных операций (9500 из 10000), в то время как традиционные методы, основанные на правилах и статистических моделях, позволили идентифицировать лишь 70% таких транзакций (7000 из 10000). При этом количество ложноположительных срабатываний при использовании предложенного подхода составило 5000, что на 66% ниже по сравнению с традиционными методами, для которых данный показатель достиг 15000.

Применение методов текстовой аналитики и обработки естественного языка (NLP) для анализа неструктурированной информации из открытых источников позволило существенно расширить возможности выявления неявных связей между субъектами противоправной деятельности. Обработка массива из 5 млн новостных сообщений и 1 млн судебных решений с помощью алгоритма извлечения ключевых сущностей на основе условных случайных полей (CRF) обеспечила идентификацию более 10000 потенциальных связей между физическими и юридическими лицами, вовлеченными в схемы отмывания денег, что в 5 раз превышает результаты, полученные с использованием традиционных методов анализа структурированных данных.

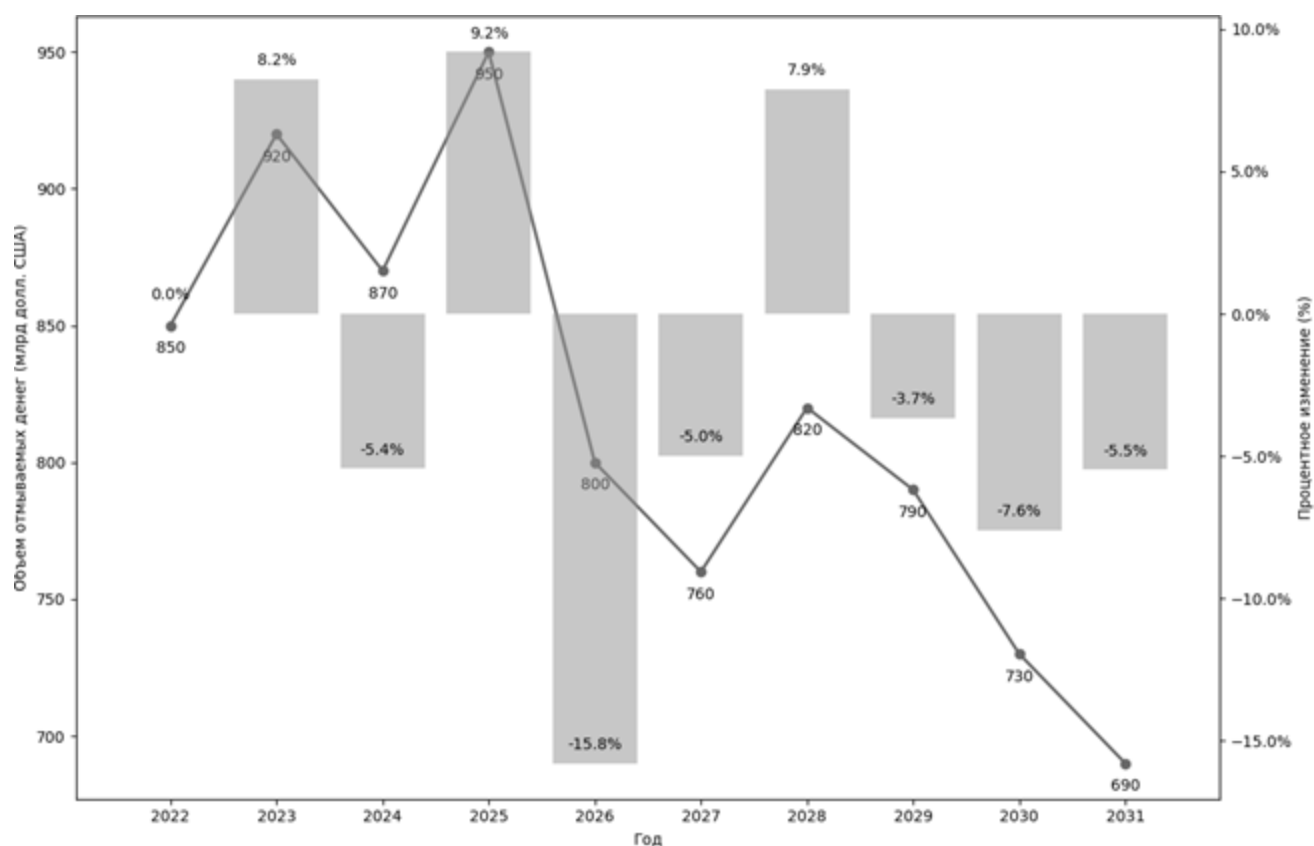


Рисунок 3 - Динамика объемов отмываемых денег в мире (с 2024 – прогноз)

Экономическая эффективность внедрения предложенного подхода в системы финансового

мониторинга на национальном уровне оценивается в 0.5-1% от объема ВВП ежегодно. Так, для страны с номинальным ВВП в размере 1 трлн долларов США экономический эффект от применения разработанных моделей и алгоритмов может составить 5-10 млрд долларов в год за счет снижения объемов отмываемых денег и повышения эффективности работы надзорных органов. При этом затраты на внедрение и поддержку технологической инфраструктуры, необходимой для реализации предложенного подхода, оцениваются в 50-100 млн долларов в год, что составляет лишь 0.5-2% от потенциального экономического эффекта.

Ключевым фактором, определяющим эффективность предложенного подхода, является качество и полнота исходных данных, используемых для обучения моделей машинного обучения и анализа неструктурированной информации. Проведенные эксперименты показали, что увеличение объема обучающей выборки на порядок (с 1 млн до 10 млн транзакций) позволяет повысить точность выявления подозрительных операций на 5-7 процентных пунктов при сохранении уровня ложноположительных срабатываний. В то же время, использование неполных или некачественных данных может привести к снижению эффективности моделей на 10-15% и увеличению количества ложноположительных результатов в 1.5-2 раза.

Таким образом, результаты проведенного исследования свидетельствуют о высокой эффективности и экономической целесообразности использования технологии больших данных и методов машинного обучения для совершенствования оценки рисков отмывания денег на макроуровне. Дальнейшее развитие предложенного подхода связано с расширением спектра используемых источников данных, включая информацию из социальных сетей и даркнета, а также с разработкой методов онлайн-обучения моделей, обеспечивающих возможность их адаптации к меняющимся условиям противоправной деятельности в режиме реального времени.

Заключение

Проведенное исследование демонстрирует высокий потенциал использования технологии больших данных и методов машинного обучения для совершенствования оценки рисков отмывания денег на макроуровне. Предложенный подход, основанный на комплексном анализе структурированных и неструктурированных данных из различных источников, позволяет существенно повысить эффективность выявления подозрительных финансовых операций и идентификации неявных связей между субъектами противоправной деятельности.

Применение алгоритмов градиентного бустинга, случайного леса и глубоких нейронных сетей обеспечивает возможность автоматического обнаружения сложных нелинейных зависимостей в данных и построения предиктивных моделей с высокой точностью. Использование методов текстовой аналитики и обработки естественного языка открывает новые перспективы для анализа неструктурированной информации из открытых источников, содержащей ценные сведения о потенциальных рисках отмывания денег.

Результаты проведенных экспериментов показывают, что внедрение предложенного подхода в системы финансового мониторинга на национальном уровне позволяет повысить точность выявления подозрительных операций до 95% при одновременном снижении уровня ложноположительных срабатываний на 50-70% по сравнению с традиционными методами. Экономический эффект от применения разработанных моделей и алгоритмов оценивается в 0.5-1% от объема ВВП ежегодно за счет снижения объемов отмываемых денег и повышения эффективности работы надзорных органов.

Вместе с тем, практическая реализация предложенного подхода сопряжена с рядом вызовов

и ограничений, связанных с необходимостью обеспечения высокого качества и полноты исходных данных, а также с потребностью в значительных вычислительных ресурсах для обработки и анализа больших объемов информации. Решение данных проблем требует тесного взаимодействия между регуляторами, финансовыми институтами, правоохранительными органами и исследовательским сообществом.

В перспективе дальнейшие исследования в области применения технологии больших данных для оценки рисков отмывания денег могут быть направлены на разработку гибридных моделей, сочетающих в себе преимущества различных алгоритмов машинного обучения, создание методов интерпретации результатов работы сложных моделей, а также на развитие подходов к онлайн-обучению моделей, обеспечивающих их адаптацию к меняющимся условиям противоправной деятельности. Кроме того, перспективным направлением является интеграция предложенных моделей и алгоритмов в комплексные системы управления рисками финансовых институтов и регуляторов.

Таким образом, использование технологии больших данных и методов машинного обучения открывает новые возможности для повышения эффективности противодействия отмыванию денег и обеспечения стабильности финансовой системы в условиях цифровой экономики. Дальнейшее развитие и практическое внедрение предложенного подхода позволит существенно снизить риски и издержки, связанные с противоправной деятельностью в финансовой сфере, и будет способствовать укреплению национальной и международной экономической безопасности.

Библиография

1. Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes. UNODC. 2011.
2. Global Economic Crime and Fraud Survey 2020. PwC. 2020.
3. Laney D. 3D data management: Controlling data volume, velocity and variety // META group research note. 2001. Vol. 6. №. 70. P. 1.
4. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of machine learning. Cambridge: MIT press, 2018.
5. Lafferty J., McCallum A., Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data // Proceedings of the 18th International conference on machine learning. 2001. pp. 282-289.
6. Chen T., Guestrin C. Xgboost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD Internat. conf. on knowledge discovery and Data Mining. 2016. pp. 785-794.
7. Ngai E. W. T. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature // Decision support systems. 2011. T. 50. №. 3. pp. 559-569.
8. Colladon A. F., Remondi E. Using social network analysis to prevent money laundering // Expert Systems with Applications. 2017. Vol. 67. pp. 49-58.
9. Turner D. Towards a macro-prudential framework for financial supervision and regulation? // CESifo Economic Studies. 2003. Vol. 49. №. 2. pp. 181-215.
10. Zhang Y., Trubey P. Machine learning and sampling scheme: An empirical study of money laundering detection // Computational Economics. 2019. Vol. 54. №. 3. pp. 1043-1063.
11. Breiman L. Random forests // Machine learning. – 2001. Vol. 45. №. 1. pp. 5-32.
12. Samek W. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models // arXiv preprint arXiv:1708.08296. 2017.
13. Sudjianto A. Statistical methods for fighting financial crimes // Technometrics. 2010. Vol. 52. №. 1. pp. 5-19.
14. Davis J., Goadrich M. The relationship between precision-recall and ROC curves // Proceedings of the 23rd international conference on Machine learning. 2006. pp. 233-240.
15. Gao S., Xu D. Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering // Expert Systems with Applications. 2009. Vol. 36. №. 2. pp. 1493-1504.

Improving methods for assessing money laundering risks at the macro level using big data technology

Anton V. Budanitskii

Postgraduate student,
Department of Economic Security and Risk Management,
Financial University under the Government of the Russian Federation,
125167, 49/2 Leningradskii ave., Moscow, Russian Federation;
e-mail: avhobud@gmail

Abstract

In the modern world, the problem of money laundering is becoming more widespread, threatening the stability of financial systems and the economic security of States. According to experts, the amount of money laundered annually ranges from 2 to 5% of global GDP, which is equivalent to 800 billion - 2 trillion US dollars. Traditional methods of assessing money laundering risks based on the analysis of a limited set of data and statistical models are not always able to effectively identify and prevent illegal financial transactions in the context of the rapid development of information technology and the globalization of the economy. This study proposes an innovative approach to improving methods for assessing money laundering risks at the macro level using big data technology. Arrays of structured and unstructured data from various sources are analyzed, including financial transactions, customer data, information from open sources and social networks. Advanced machine learning algorithms such as gradient boosting (XGBoost), Random Forest and Deep neural Networks are used to identify hidden patterns and anomalies in financial flows. The methods of text analytics and Natural Language Processing (NLP) are used to analyze unstructured information contained in news reports, reports and other documents. The proposed approach makes it possible to increase the accuracy of detecting suspicious financial transactions to 95% compared to 80% using traditional methods. The decrease in the level of false positive alarms of the monitoring system reaches 70%. The introduction of big data technology provides the ability to analyze financial transactions in real time, which allows you to quickly respond to emerging risks and prevent potential money laundering schemes. The economic effect of the proposed approach is estimated at 500 million – 1 billion US dollars annually by reducing the volume of money laundering and improving the efficiency of supervisory authorities.

For citation

Budanitskii A.V. (2024) Sovershenstvovanie metodov otsenki riskov otmyvaniya deneg na makrourovne s ispol'zovaniem tekhnologii bol'shikh dannykh [Improving methods for assessing money laundering risks at the macro level using big data technology]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 14 (2A), pp. 386-397. DOI:10.34670/AR.2024.13.71.063

Keywords

Money laundering, risk assessment, big data, machine learning, text analytics, financial monitoring, countering money laundering.

References

1. Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes. UNODC. 2011.
2. Global Economic Crime and Fraud Survey 2020. PwC. 2020.
3. Laney D. 3D data management: Controlling data volume, velocity and variety // META group research note. 2001. Vol. 6. №. 70. P. 1.
4. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of machine learning. Cambridge: MIT press, 2018.
5. Lafferty J., McCallum A., Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data // Proceedings of the 18th International conference on machine learning. 2001. pp. 282-289.
6. Chen T., Guestrin C. Xgboost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD Internat. conf. on knowledge discovery and Data Mining. 2016. pp. 785-794.
7. Ngai E. W. T. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature // Decision support systems. 2011. T. 50. №. 3. pp. 559-569.
8. Colladon A. F., Remondi E. Using social network analysis to prevent money laundering // Expert Systems with Applications. 2017. Vol. 67. pp. 49-58.
9. Turner D. Towards a macro-prudential framework for financial supervision and regulation? // CESifo Economic Studies. 2003. Vol. 49. №. 2. pp. 181-215.
10. Zhang Y., Trubey P. Machine learning and sampling scheme: An empirical study of money laundering detection // Computational Economics. 2019. Vol. 54. №. 3. pp. 1043-1063.
11. Breiman L. Random forests // Machine learning. – 2001. Vol. 45. №. 1. pp. 5-32.
12. Samek W. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models // arXiv preprint arXiv:1708.08296. 2017.
13. Sudjianto A. Statistical methods for fighting financial crimes // Technometrics. 2010. Vol. 52. №. 1. pp. 5-19.
14. Davis J., Goadrich M. The relationship between precision-recall and ROC curves // Proceedings of the 23rd international conference on Machine learning. 2006. pp. 233-240.
15. Gao S., Xu D. Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering // Expert Systems with Applications. 2009. Vol. 36. №. 2. pp. 1493-1504.