International Scientific and Practical Conference Samarkand Institute of Economics and Service
Vol. 1 Issue 1 | pp. 281-284 **| ISSN:** XXXX-XXXX
Available online @ https://globalconf.net/index.php/samisi

Article

Применение Методов Машинного Обучения Для Выявления Рисков Теневой Экономики (Экономические Вопросы Использования Цифровой Экономики, Инноваций И Технологий Искусственного Интеллекта)

Серик Акбота Мейрбеккызы*[1] Муханова Гаини Казибаевна[2]

1. Магистрант Академии государственного управления при Президенте Республики Казахстан, г. Астана, Казахстан.
2. Научный руководитель, кандидат экономических наук, исполняющий обязанности профессора Института цифровизации и инноваций Академии государственного управления при Президенте Республики Казахстан.

* Correspondence: a.serik@apa.kz

Citation: Мейрбеккызы С. А. Применение Методов Машинного Обучения Для Выявления Рисков Теневой Экономики (Экономические Вопросы Использования Цифровой Экономики, Инноваций И Технологий Искусственного Интеллекта). Institute of Economics and Service 2026, 1(1), 281-284.

Received: 12th Aug 2026
Revised: 26th SSep 2026
Accepted: 18th Oct 2026
Published: 22th Nov 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

Аннотация; В статье рассматривается развитие зелёного ландшафта на территории комплекса «Вечный город» в Самарканде в контексте устойчивого развития городской среды и сохранения культурного наследия. Особое внимание уделяется озеленению, созданию тематических садов на основе исторических традиций, внедрению ресурсосберегающих технологий и формированию экологически устойчивой инфраструктуры. Анализируется роль зелёных насаждений в улучшении качества воздуха, снижении городских температур, сохранении биоразнообразия и развитии туристического потенциала Самарканда. Обосновывается значимость проекта «Зелёный сад» для экологической устойчивости и повышения качества городской среды.

Ключевые слова: зелёный ландшафт, озеленение, Самарканд, зелёная инфраструктура, устойчивое развитие, экологическая устойчивость, ландшафтная архитектура, туризм.

1. Введение

Я рассматриваю машинное обучение как инструмент, полезность которого нужно проверять на конкретной задаче. В случае теневой экономики такая проверка начинается с выбора показателя. Региональная доля ненаблюдаемой экономики доступна в открытой статистике, однако её нельзя читать как долю нарушителей. В подходе Международного валютного фонда понятия неформальной и ненаблюдаемой экономики разграничиваются в зависимости от целей измерения [1].

В этой статье я проверяю, можно ли точнее прогнозировать региональную долю, если заменить прошлогоднее значение методом машинного обучения. Для сравнения выбрана регрессия пяти ближайших соседей. Исходная гипотеза состоит в том, что она даст меньшую ошибку на данных за 2025 год. Мне представляется, что без такого сравнения обсуждать пользу алгоритма для мониторинга рисков преждевременно.

3. Обзор литературы

Felix, Alexandre и Lima сравнили двенадцать моделей на данных 122 стран за 2004–2014 годы. На их выборке методы машинного обучения показали преимущества перед линейными моделями [2]. В моей задаче данных значительно меньше, а объясняющий признак всего один. Поэтому этот результат даёт основание проверить алгоритм, но не позволяет заранее ожидать того же выигрыша.

Battaglini и соавторы используют административные данные Италии для отбора налоговых проверок. Здесь возникает проблема избирательности меток, поскольку результаты известны только для уже проверенных налогоплательщиков [3]. Savić и соавторы ищут аномалии в налоговых данных Сербии, сочетая кластеризацию и обучение представлений [4]. Обе работы важны для темы рисков, но переход от региональной статистики к такому анализу потребует других данных.

4. Методология

Для расчётов я использовала динамический ряд Бюро национальной статистики Республики Казахстан «Доля ненаблюдаемой экономики в ВВП по регионам» [5]. В сохранённом XLS-файле от 9 ноября 2026 года выбран лист «доля ННЭ в ВРП» за 2013–2025 годы. Несмотря на название ряда, на этом листе доля рассчитана относительно собственного валового регионального продукта. Вклад региона в ВВП страны здесь не измеряется.

В расчёт вошли 12 территорий с полными рядами, не затронутых выделением новых областей за рассматриваемый период: Акмолинская, Актюбинская, Атырауская, Западно-Казахстанская, Жамбылская, Костанайская, Кызылординская, Мангистауская, Павлодарская и Северо-Казахстанская области, города Астана и Алматы. Общереспубликанскую строку я исключила. Такой отбор уменьшает влияние изменений границ. При этом часть регионов остаётся за пределами исследования.

Одно наблюдение соответствует региону в определённом году. Прогнозируется доля ненаблюдаемой экономики в процентах; единственный признак — её значение годом ранее. Для обучения взяты 132 наблюдения с целевыми годами 2014–2024, для проверки — 12 наблюдений за 2025 год. Значения за 2025 год используются только при оценке ошибок.

Базовый прогноз равен прошлогоднему значению. Метод пяти ближайших соседей ищет в обучающей выборке пять наблюдений с наиболее близким значением признака и усредняет их последующие доли. Расстояние — абсолютная разность, веса соседей одинаковы. В поиске участвуют все выбранные регионы и годы; отдельного признака региона нет. При равных расстояниях сохраняется порядок регионов исходной таблицы, затем порядок лет по возрастанию. Число соседей задано заранее, без подбора по проверочным данным.

Ошибки сравниваются по MAE и RMSE. Первая величина — среднее абсолютных разностей прогноза и факта, вторая — корень из среднего квадратов этих разностей. Усреднение выполнено по 12 территориям, единица измерения — процентный пункт. Наблюдения одного региона за разные годы связаны между собой. Поэтому я ограничиваюсь описательным сравнением и не делаю вывода о статистической значимости различий.

5. Анализ и результаты

На данных за 2025 год ближайшие соседи проиграли базовому прогнозу по обеим метрикам (таблица 1). MAE составила 2,350 против 1,928 п. п., RMSE — 3,407 против 2,398 п. п. Ошибки оказались выше на 21,9% и 42,1% соответственно. Исходная гипотеза не подтвердилась. На этой выборке усложнять прогноз таким способом оснований нет.

Таблица 1 — Ошибки прогноза доли ненаблюдаемой экономики за 2025 год

Метод	MAE, п.п.	RMSE, п.п.
Значение предыдущего года	1,928	2,398
Пять ближайших соседей	2,350	3,407

Примечание — расчёты по данным [5]; проверочная выборка — 12 территорий.

Базовый прогноз точнее в семи территориях, ближайшие соседи — в пяти. Особенно сильно на итог повлияла Атырауская область. Её доля снизилась с 8,37% в 2024 году до 4,45% в 2025 году, тогда как прогноз соседей равен 12,786%. Абсолютная ошибка достигла 8,336 п. п. против 3,920 п. п. у базового метода. Это одно наблюдение даёт 49,9% суммы квадратов ошибок модели соседей. По одному прошлогоднему значению объяснить такое изменение нельзя.

Для Астаны картина обратная. При фактической доле 20,47% базовый метод дал 18,14%, а ближайшие соседи — 20,212%. Ошибки равны 2,330 и 0,258 п. п. Алгоритм здесь заметно точнее. Но удачный прогноз для одной территории ещё не делает его предпочтительным для всей выборки. Кроме того, близкие прошлогодние доли могут принадлежать регионам с разной экономической структурой.

Я трактую полученный результат узко: регрессия с одним лаговым признаком не улучшила прогноз в одном проверочном году. Данные не позволяют судить обо всех методах машинного обучения. Способность обнаруживать теневую деятельность в этом расчёте вообще не проверялась.

У такого расчёта есть ещё одно ограничение. Все годы взяты из одной сохранённой версии ряда, а исторические оценки могли пересматриваться. Значит, это ретроспективная проверка по доступному файлу. Она не воспроизводит набор сведений, которым исследователь располагал бы в каждый прошлый год. Результат также зависит от отбора территорий и не охватывает регионы с изменившимися границами.

6. Выводы и предложения

Для выбранных 12 территорий и проверки на 2025 год я считаю прогноз по прошлогоднему значению более обоснованным рабочим ориентиром. Он дал меньшую ошибку и проще по устройству. Поэтому в следующем расчёте я бы снова начала с него и уже по этому сравнению судила о пользе нового алгоритма.

Прежде чем пробовать более сложные алгоритмы, я бы расширила проверку на несколько лет. На каждом шаге модель должна обучаться только на предшествующих данных. Затем можно добавить показатели неформальной занятости, отраслевой структуры и предпринимательской активности, учитывая сроки их публикации. Параметры следует подбирать внутри обучающей выборки, сохраняя последний период для окончательной оценки.

Для оценки рисков конкретных нарушений нужны административные данные. Если известны подтверждённые результаты проверок, можно строить классификатор, учитывая избирательность меток. Если таких сведений нет, поиск аномалий можно использовать для отбора случаев на экспертный разбор. Его результативность придётся проверять по подтверждённым случаям. Число необычных наблюдений ещё не показывает, сколько нарушений удалось обнаружить.

В региональном мониторинге я бы использовала отклонение от прогноза как повод разобраться в данных и экономической ситуации. Сначала нужно проверить пересмотры оценок и сопоставимость ряда. Порог, после которого отклонение заслуживает внимания, пока не определён. Поэтому предложенный расчёт может

быть началом мониторинга, но вывод о правонарушении конкретного лица из него не следует.

СПИСОК ИСПОЛЬЗУЕМЫХ ЛИТЕРАТУРЫ

[1] Quirós-Romero G., Alexander T. F., Ribarsky J. Measuring the Informal Economy. IMF Policy Papers. 2021. No. 2021/002. DOI: 10.5089/9781513568249.007.

[2] Felix J., Alexandre M., Lima G. T. Applying Machine Learning Algorithms to Predict the Size of the Informal Economy. Computational Economics. 2025. Vol. 65. P. 1169–1189. DOI: 10.1007/s10614-024-10593-6.

[3] Battaglini M., Guiso L., Lacava C., Miller D. L., Patacchini E. Refining public policies with machine learning: The case of tax auditing. Journal of Econometrics. 2025. Vol. 249. Art. 105847. DOI: 10.1016/j.jeconom.2024.105847.

[4] Savić M., Atanasijević J., Jakovetić D., Krejić N. Tax evasion risk management using a Hybrid Unsupervised Outlier Detection method. Expert Systems with Applications. 2022. Vol. 193. Art. 116409. DOI: 10.1016/j.eswa.2021.116409.

[5] Бюро национальной статистики Республики Казахстан. Доля ненаблюдаемой экономики в ВВП по регионам. Динамический ряд, XLS, лист «доля ННЭ в ВРП». URL: https://stat.gov.kz/ru/industries/economy/national-accounts/dynamic-tables/?period=year (дата обращения: 09.11.2026).