МРНТИ 20.53.01

# **АНАЛИЗ И КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ЦЕН НА ФОНДОВОМ РЫНКЕ С ПОМОЩЬЮ ВАРИАЦИОННЫХ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ**

**А. Есенбек**[C:\Users\admin\Desktop\iD.png](https://orcid.org/0009-0006-3327-2696)**, А. Касымова**[C:\Users\admin\Desktop\iD.png](https://orcid.org/0000-0002-8612-4922)**, А. Картбаев**[C:\Users\admin\Desktop\iD.png](https://orcid.org/0000-0003-0592-5865)

*Казахстанско-Британский технический университет, Алматы, Казахстан*

Корреспондент-автор: a.kartbaev@kbtu.kz

Прогнозирование цен на финансовые активы остается жизненно важной областью финансовой аналитики, что стимулирует постоянные исследования более эффективных методов прогнозирования. В данном исследовании рассматривается вопрос о том, можно ли прогнозировать динамику цен на конкретную акцию на основе движения цен на другие связанные с ней акции. Наш подход направлен на выявление межфондовых связей и прогнозирование ценовых тенденций на основе анализа этих связей. Используя машинное обучение, были выявлены кластеры компаний, демонстрирующих схожие модели движения цен. Далее использована модель векторной авторегрессии (VAR) для создания функций импульсного отклика, прогнозирующих потенциальные колебания цен в последующие дни. Наши эксперименты направлены на повышение точности и вычислительной эффективности этих прогнозов. Кроме того, разработана интерактивная панель, которая наглядно отображает прогнозы изменения цен на акции, позволяя пользователям наблюдать за ожидаемыми тенденциями цен на акции компаний. Данная работа вносит вклад в развивающуюся область прогнозного моделирования, улучшая понимание взаимозависимости между активами и предоставляя доступные прогнозы для принятия финансовых решений.

**Ключевые слова**:машинное обучение, предиктивное моделирование, прогнозирование цен, нейронные сети, VAR, K-means, Big Data*.*

# **УАҚЫТ-ВАРИАЦИЯЛЫ СЕРИЯЛЫҚ МҮЛГІЛЕРДІ ПАЙДАЛАНУ МЕН ҚОР НАРЫҒЫНДАҒЫ БАҒАНЫ ТАЛДАУ ЖӘНЕ ҚЫСҚА МЕРЗІМДІ БОЛЖАУ**

**А. Есенбек, А. Касымова, А. Картбаев**

*Қазақ-Британ техникалық университеті, Алматы, Қазақстан,*

*e-mail: a.kartbaev@kbtu.kz*

Қаржы активтерінің бағасын болжау қаржылық аналитиканың маңызды саласы болып қала береді, бұл болжаудың тиімді әдістеріне үздіксіз зерттеулер жүргізуге түрткі болады. Бұл зерттеу белгілі бір акцияның бағасының қозғалысын басқа қатысты акциялардың баға қозғалысы негізінде болжауға болатынын зерттейді. Біздің көзқарасымыз қор аралық байланыстарды анықтауға және осы байланыстарды талдау негізінде баға үрдістерін болжауға бағытталған. Нейрондық желілерді пайдалана отырып, біз ұқсас баға қозғалысы үлгілерін көрсететін компаниялардың кластерлерін анықтаймыз. Содан кейін біз векторлық авторегрессия (VAR) моделін келесі күндердегі әлеуетті баға қозғалысын болжайтын импульстік жауап функцияларын жасау үшін қолданамыз. Біздің эксперименттеріміз осы болжамдардың дәлдігі мен есептеу тиімділігін арттыруға бағытталған. Сонымен қатар, біз пайдаланушыларға компания акцияларының бағамындағы күтілетін үрдістерді көруге мүмкіндік беретін акциялар бағасының болжамдарын көрнекі түрде көрсететін интерактивті бақылау тақтасын жасадық. Бұл жұмыс активтер арасындағы өзара тәуелділікті түсінуді жақсарту және қаржылық шешімдерді қабылдау үшін қолжетімді болжамдарды қамтамасыз ету арқылы көп акцияларды болжауды модельдеудің пайда болған саласына ықпал етеді.

**Түйін сөздер:**машиналық оқыту, болжамды модельдеу, бағаны болжау, нейрондық желілер, VAR, K-means, Big Data*.*

# **ANALYSIS AND SHORT-TERM PREDICTION OF STOCK MARKET PRICES USING VAR MODELS OF TIME SERIES**

**A. Yessenbek, A. Kassymova, A. Kartbayev**

*Kazakh-British Technical University, Almaty, Kazakhstan,*

*e-mail: a.kartbaev@kbtu.kz*

Forecasting stock assets prices remains a vital area in financial analytics, driving ongoing research into more effective predictive techniques. This study explores whether the price trends of a given stock can be forecasted based on the price movements of other related stocks. Our approach seeks to uncover inter-stock relationships and predict price trends by analyzing these connections. Leveraging neural networks, we identify clusters of companies exhibiting similar price movement patterns. Subsequently, we employ a Vector Auto Regression (VAR) model to generate impulse response functions, projecting potential price fluctuations over the following days. Our experiments focus on refining both the accuracy and computational efficiency of these predictions. Additionally, we developed an interactive dashboard that visually displays forecasts of stock price changes, enabling users to observe anticipated trends in companies’ stock prices. This work contributes to the growing field of predictive modeling by enhancing understanding of interdependencies among stocks and providing accessible predictive insights for financial decision-making.

**Keywords**:machine learning, predictive modeling, stock forecasting, neural networks, VAR, K-means, Big Data*.*

**Введение.** В настоящее время прогнозирование цен на акции и выявление связей между компаниями - две популярные темы в аналитике больших данных. Известно, что цены на различные акции коррелируют между собой; цены на акции иногда движутся вместе. Поэтому вполне естественно задаться вопросом, можно ли предсказать их цены, используя цены на другие акции. Цель работы - определить и доказать взаимосвязи между акциями и спрогнозировать динамику цен на основе изменений цен связанных с ними финансовых активов. Эта информация может иметь актуальность исследования для инвесторов, которые хотят принять решение о вложении средств в определенные виды активов. Понимание взаимосвязей между акциями может помочь инвесторам обнаружить больше скрытых возможностей, когда они видят изменения в определенных индексах.

В результате создана модель, которая поможет инвесторам получать информацию о связях между акциями и прогнозы будущих изменений цен. В качестве характеристик используются связи в сетях регистраторов фондовых бирж. Она имеет практическую ценность для организаций, как способ генерировать оценку из сети и использовать их для предварительного определения стоимости компании, так и теоретическую ценность для обучения финансовых моделей методами машинного обучения.

Некоторые другие модели, такие как модель ARIMA [1] и модель логистической регрессии [2], также используются для этой цели. Однако эти авторы используют отдельные данные для построения своей модели и не обращают внимания на связи между компаниями. Кроме того, как еще одно ограничение данных моделей, они не отбирают признаки перед прогнозированием новых цен на финансовые активы. Для дальнейшего развития принято решение использовать VAR-модели [3], основанные на исторических изменениях цен на акции, чтобы получить функции импульсного отклика, которые предсказывают будущие изменения. Более того, были использованы группы акций, которые ведут себя одинаково, в качестве признаков, а не просто используем исторические данные только целевой акции. Перед прогнозированием цен на акции проведена кластеризация компаний, чтобы выяснить, какие группы компаний наиболее связаны друг с другом. Далее эти связанные компаний были использованы в качестве характеристик VAR-моделей для прогнозирования изменений цен на акции компаний, выбранных пользователем. Была использована кластеризация K-means [4] и ANNOY [5], чтобы выяснить, какие группы компаний связаны между собой, и анализированы преимущества и недостатки наших методов. Затем результаты кластеризации и модель VAR использованы для расчета функций импульсного отклика и прогнозирования динамики цен на основе поиска по сетке для повышения точности и определение времени работы [6].

В последние время, после новаторской работы Симса (1980) по созданию моделей векторной авторегрессии (VAR) [7], эти модели получили широкое распространение при моделировании взаимосвязей между экономическими и финансовыми временными рядами и прогнозировании будущих величин. Поскольку количественные методологии все чаще применяются для прогнозирования цен на акции, VAR-модели не стали исключением. Во многих исследованиях VAR-модели использовались для прогнозирования индексов цен на акции и достигли хороших результатов [8 -10]. Однако они моделировали индексы цен акций, а не цены отдельных акций, в то время как инвесторы часто интересуются будущей доходностью отдельных акций, в которые они могли бы вложить средства. Ву и Чжоу (2014) рассмотрели применение моделей VAR для прогнозирования доходности акций, включая доходность отдельных акций, с относительно теоретической точки зрения и заложили основу для конкретной, прикладной работы в этой области [11].

В настоящее время в этой области существует не очень большое количество прикладных работ по данной теме. Например, в работе Лонг (2024) использовалась модель VAR для прогнозирования дневных максимумов и минимумов цен на 20 высокоторгуемых акций, и результаты оказались обнадеживающими: был сделан вывод, что "высокие и низкие цены на акции в значительной степени предсказуемы [12]". Однако данные охватывали только семь лет (2010-2017) и были статичными (не обновлялись постоянно). Кроме того, рассматривалось лишь небольшое количество акций.

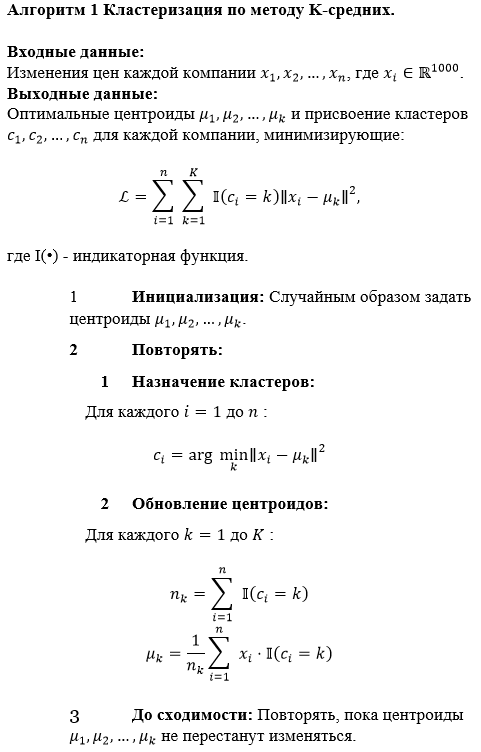
С точки зрения инвесторов, более ценным был бы инструмент прогнозирования, который можно использовать на большом количестве акций и который динамически обновляется на основе текущих цен на акции. В отличие от предыдущих работ, был использован ANNOY для поиска групп связанных акций, а затем рассчитана VAR-модель под каждую группу. Кроме того, в предыдущих работах по VAR обычно рассматривалось лишь небольшое количество акций, в данной работе были включены более 2500 акций. Поэтому использован динамический подход на основе больших данных для прогнозирования на основе ежедневно обновляемых цен на акции. Наши VAR-модели ежедневно обновлялимсь с использованием самых свежих данных о ценах на акции.Мы полагаем, что такие модели ранее не использовались с такой степенью широты и глубины для анализа определенных групп активов.

**Материалы и методы.** *Данные.* В начале исследования было собрано данные о более чем 2800 компаниях, входящих в NASDAQ с платформы Kaggle начиная с момента выхода на биржу, в зависимости от доступности информации в наиболее полном объеме. Для каждой акции и для каждого дня были включены данные о низкой цене, цене открытия, объеме, высокой цене, цене закрытия и скорректированной цене закрытия. Причем размер этих данных составил около 19 ГБ, а формат включает, в основном, *json* и *csv*.

Кроме того, наши данные обновлялись с помощью модуля yfinance в Python, который загружает цены на акций через Yahoo Finance.

В результате модель обновляла около 5-6 МБ дополнительных данных каждый будний день. Эти новые данные обрабатываются в формате процентных изменений и затем используются для обновления существующего набора данных, что гарантирует, что модели всегда будут соответствовать самой актуальной информации.

Эти данные были очищены с помощью библиотеки Pandas. Были отброшены акции, по которым имеется слишком мало данных, то есть если они были опубликованы менее четырех лет назад. Это связано с тем, что по таким акциям у нас нет достаточной информации для точной оценки их взаимосвязи с другими акциями и прогнозирования их будущих ценовых тенденций. Также были исключены акции, которые больше не торгуются на рынке. В результате всего насчитано 2511 акций. Затем данные были обработаны в формате, удобном для наших моделей. Далее данные переведены в формат процентных изменений. Также были добавлены к данным аддитивная обратная величина для каждой акции, чтобы выявить акции которые отрицательно связаны с другими акциями.



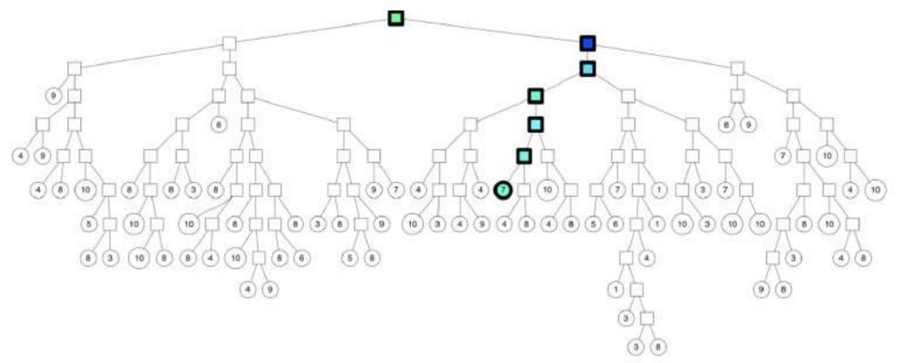
**Рис. 1- Алгоритм 1. Кластеризация по методу K-средних**

*Кластеризация по методу К-средних*. Было изучен метод кластеризации K-means clustering, где использовались К-средние для кластеризации компаний на основе изменений их биржевых цен. Поскольку целью кластеризации является получение связей между компаниями, группы, имеющие только одну компанию, не могут дать нам никакой информации. Поэтому алгоритм использован следующим образом, как на рисунке 1.

Для работы модели VAR необходимы кластеры акций. Естественно рассмотреть модели обучения без контроля, такие как k-means. Однако для того, чтобы неконтролируемое обучение давало хорошее представление данных, то есть кластеры с одинаковым размером, данные должны быть линейно разделимыми по своей сути. В связи с тем, что размер кластеров в данных по акциям сильно разбалансирован, был использован алгоритм ближайших соседей.

Был выбран ANNOY из-за высокой скорости и универсальности. Эта вариация метода широко используется в производстве функций поиска и ранжирования технологических компаний, таких как Netflix. Это библиотека для python, которая является самым быстрым способом ее практической реализации, где она случайным образом разрезает данные на небольшие фрагменты, пока все фрагменты не станут меньше n.

Затем он строит лес бинарных деревьев, в котором каждая режущая гиперплоскость является внутренним узлом, а каждый осколок - внешним узлом. Для поиска цели перебираются деревья от корня до соответствующего внешнего узла, содержащего цель и ближайшие осколки, если в узле меньше точек, чем нам нужно, как показано на рисунке 2. Результат поиска по всем деревьям используется для получения рейтинга сходства.



**Рис. 2 -Бинарное дерево ANNOY**

После очистки фактические цены были преобразованы в процентные изменения. В качестве признаков были взяты цены закрытия за последние 1000 дней. Чтобы понять отрицательную корреляцию между акциями, также добавлен к данным аддитивный обратный вектор. Затем было построено бинарное дерево ANNOY с векторным представлением каждой акции, используя евклидово расстояние как метрику сходства. Модель выводит 10 наиболее похожих акций и соответствующее сходство для каждого поиска. Названия акций с меткой "inversed" указывают на сильную отрицательную корреляцию. В нашей системе был построен ANNOY, которая принимает на вход коды акций и ищет по ним соответствующий вектор. Выходные данные также преобразуются в названия акций из вектора.

*Модель VAR.* Сразу после получения группы акций с помощью ANNOY, далее рассмотрена модель векторной авторегрессии (VAR) для акций этой группы. Модели VAR используются для моделирования отношений между несколькими временными рядами. Поскольку исторические цены каждой акции представляют собой временной ряд, VAR-модели хорошо подходят для моделирования взаимосвязи между ценами на акции. Она состоит из системы линейных уравнений, где каждый ряд рассматривается как зависимая переменная в одном уравнении, а независимыми переменными являются значения каждого временного ряда в предыдущие периоды. Например, простая VAR-модель двух временных рядов у 1 и у 2 может быть записана следующим образом:

(1)

далее можно записать в виде:

(2)

где и представляют собой значения у 1 и у 2 в момент времени и представляют значения у 1 и у 2 в момент времени n и коэффициенты, подлежащие оценке.

Было решено использовать VAR-модели, поскольку они имеют несколько важных преимуществ при моделировании отношений между акциями. Во-первых, в VAR-модели каждая переменная рассматривается как зависимая. Это позволяет полностью отразить динамические отношения между акциями; например, если цена акции X изменится, это может привести к изменению цены акции Y, которая затем может привести к другому изменению цены акции X. Модели VAR очень хорошо отражают этот тип сложной зависимости, в то время как другие регрессионные модели, такие как линейная регрессия, отражают зависимость только в одном направлении, поскольку в них некоторые акции рассматриваются как независимые переменные.

Таким образом, по сравнению с другими регрессионными моделями, VAR-модели наиболее подходят для полного и точного ответа на наш вопрос. Еще одно преимущество VAR-моделей заключается в том, что их легко оценивать, оценка производится с помощью обыкновенных наименьших квадратов для каждого уравнения. Кроме того, VAR-модели обладают хорошими возможностями предсказывать цены на много дней вперед, а не только на следующий день. Это важно, поскольку инвесторов часто интересует, как изменятся цены на акции в ближайшие несколько недель, а не только завтра. Наконец, четкая структура VAR-моделей позволяет легко интерпретировать результаты и прогнозы, что помогает инвесторам воплотить полученные результаты в реальные решения.

При каждом выполнении поиска акций в модели происходит оценка VAR-модель на группе из 10 акций, найденных с помощью ANNOY, поэтому VAR-модель состоит из 10 уравнений. Независимыми переменными являются цены каждой из 10 акций на день раньше (день ). Эта VAR-модель может быть записана как:

(3)

где цена акции і в день t , а все b - коэффициенты, подлежащие оценке. Эквивалентно, в векторной форме, эта модель VAR имеет вид:

(4)

где вектор 10 х 1 цен на акции в день t и и матрицы коэффициентов 10 x 1 и 10х10, подлежащих оценке. Модель VAR оценивается с помощью библиотеки statsmodels в Python. После оценки модели вычислены функции импульсного отклика, которые предсказывают, как изменение цены данной акции в настоящий день повлияет на цены других акций в последующие дни.

Несмотря на множество очевидных преимуществ, которыми обладают VAR-модели, у них есть и недостаток. Было запланировано применение VAR-модели к группам связанных акций. Поскольку количество коэффициентов, которые необходимо оценить в модели, быстро растет с увеличением числа переменных (например, если переменных n, то существует коэффициентов, каждый из которых все сложнее вычислить при увеличении n), не представляется возможным за разумное время построить VAR-модели для групп из более чем 100 акций. Особенно учитывая, что модели оцениваются в реальном времени, когда пользователи ищут определенную акцию в приложении, малое время работы очень важно. Однако, когда была предпринята попытка создать группы связанных акций с помощью кластеризации K-means, некоторые группы оказались слишком большими ( акции), независимо от того, какое количество кластеров были использованы. Поэтому были рассмотрены несколько других методов поиска групп связанных акций, таких как ЕМ-алгоритмы и KNN, и в итоге решили создать группы с помощью приближенных ближайших соседей, благодаря существованию библиотеки ANNOY на языке python с быстрым временем выполнения.

**Результаты и обсуждение.** Задача заключалась в предсказании цен на акции на ближайшие несколько дней с использованием моделей VAR. Для оценки точности прогнозов случайным образом были выбраны 50 акций и 20 дней в качестве тестовых примеров.Затем для каждого из тестовых примеров был осуществлен поиск группы связанных акций с помощью ANNOY, после чего построили VAR-модель для прогнозирования цен на акции на следующие несколько дней.Прогнозы сравнивались с фактическими ценами в эти дни,чтобы вычислить среднюю абсолютную ошибку прогнозов. Также были рассчитаны средняя погрешность прогнозов, т. е. то, насколько прогнозы в среднем выше или ниже фактических цен, более полные данные приведены ниже в Таблице 1.

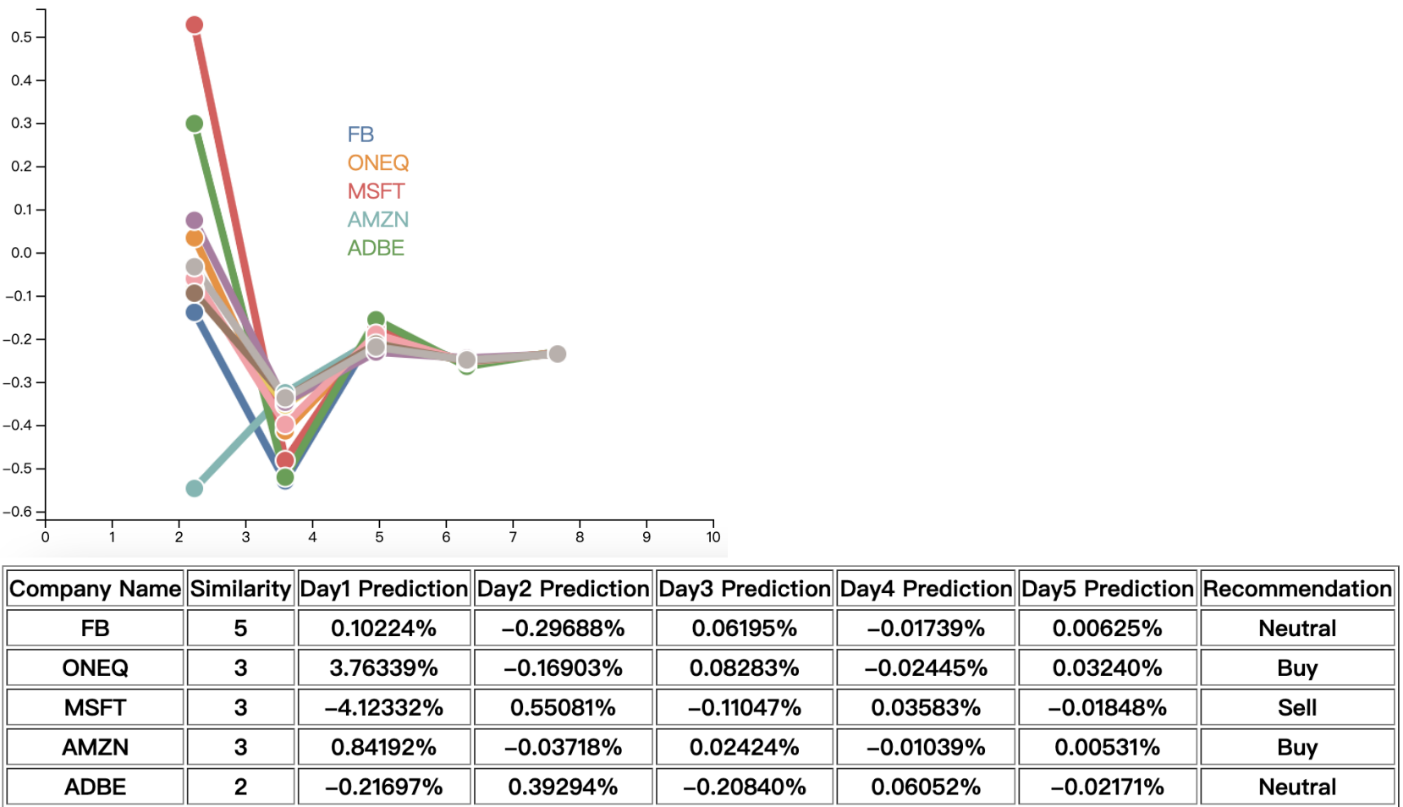
**Таблица 1 – Измерения точности работы модели**

|  |  |  |  |
| --- | --- | --- | --- |
| История\_Дней | Прогноз\_Дней | Кол\_Акций | Сред.кв. ошибка |
| 900 | 5 | 10 | 0.009192 |
| 900 | 5 | 9 | 0.009298 |
| 900 | 10 | 7 | 0.009582 |
| 900 | 3 | 7 | 0.010905 |
| 700 | 10 | 10 | 0.086412 |
| 700 | 10 | 7 | 0.086470 |
| 700 | 3 | 7 | 0.011312 |
| 900 | 10 | 9 | 0.011659 |
| 900 | 10 | 5 | 0.088065 |
| 300 | 3 | 10 | 0.012224 |
| 300 | 10 | 9 | 0.087300 |

Чтобы минимизировать среднюю абсолютную ошибку, был проведен сеточный поиск для оптимизации гиперпараметров модели, а именно продолжительности обучения, продолжительности прогнозирования и количества акций в группе. То есть для каждого тестового случая было использовано разное количество связанных акций и разное количество дней обучения, чтобы предсказать цены в разное количество будущих дней. Для этого был поиск с длиной обучения [1200, 900, 700, 500, 300, 150, 30], длительность прогнозирования [3, 5, 15, 30] и размер кластера [5, 12, 15]. Это показало, что оптимальными гиперпараметрами являются 600 дней обучения, 5-7 дней прогнозирования и до 10 акций на группу. Это позволило достичь средней абсолютной ошибки 0.089%, как указано в Таблице 1. Рассматривая каждый тестовый случай в отдельности, было обнаружено, что модель с оптимальными гиперпараметрами достигла абсолютной ошибки в диапазоне в широком диапазоне отраслей и во времени. Более того, смещение прогнозов было близким к нулю ( и ) для всех тестовых случаев, что говорит о том, что цены акций не переоцененные и не недооцененные.

Также была оценка времени работы программы для 20 случайно выбранных акций. Напомним, что всякий раз, когда запрашиваются результаты по определенной акции, используется ANNOY для поиска группы связанных акций, а затем подгоняют под эту группу модель, чтобы сгенерировать функции импульсного отклика, которые предсказывают будущие изменения цен. Было обнаружено, что общее время работы программы составляет от 1,2 до 1,6 секунды. Таким образом, пользователи приложения могут ожидать результатов в течение 1,5 секунд после выполнения поиска. Из общего времени работы построение и запрос ANN-модели занимает до 0,4 секунды, а оценка модели и вычисление функций импульсного отклика до 0,03 секунды.

Во время выполнения программы каждый день оператор python вызывает функцию, которая загружает цены на акции с помощью модуля yfinance, обрабатывает их в нужном нам формате и сохраняет в csv-файл. Когда пользователь ищет определенную акцию в нашем приложении, вызывается функция, которая с помощью ANNOY находит 10 наиболее тесно связанных акций, а затем оценивает модель VAR, используя исторические цены этих 10 акций. Затем рассчитываются функции импульсного отклика для каждой из 10 компаний по сравнению с искомой компанией, и результаты отображаются в виде таблицы. Кроме того, используется функция импульсного отклика для выработки инвестиционных рекомендаций. Для каждой из десяти компаний, если прогнозируемое кумулятивное изменение цены положительно по крайней мере в 4 из следующих 5 дней, рекомендуется "покупать"; если прогнозируемое кумулятивное изменение цены отрицательно по крайней мере в 4 из следующих 5 дней, рекомендуется "продавать"; в противном случае предлагают "нейтральную" рекомендацию. Эти рекомендации также отображаются в таблице. Ниже приведен рисунок 3, где показано прогноз изменения цены акций на следующие пять дней.



**Рис. 3 - Прогноз изменения цены акций на следующие пять дней по модели**

Эти рекомендации рассчитаны помочь пользователям более эффективно принимать решения на основе полученных прогнозов. После того как пользователь введет название компании и выберет цель поиска (отклик), полученные данные будут переданы в бэкенд. Функция get results в view.ру создаст модель ANNOY и VAR, используя csv-файл в главной директории, и вернет результаты, которые будут отправлены на фронт-енд для построения визуализации с помощью функции annoyvar в views.py, где будет отображен VAR. Во-первых, входные данные, подаваемые в таблицу А, будут отображать сходство каждой связанной акции с искомой, рекомендуемую операцию и фактические прогнозируемые процентные изменения в ближайшие 5 дней.

Чтобы улучшить подход, был построен динамический линейный график, который изначально отображает прогнозы по всем акциям. Поскольку у каждой компании есть различные связанные акции, и даже эта группа меняется каждый день, нет возможности просто обучить одну модель и вызывать ее каждый раз. Поэтому система мгновенно строит отдельную модель для искомой акции. Преимущество заключается в том, что модель всегда обновляется и не требует большого пространства для хранения.

**Выводы.** В результате была достигнута цель исследования по прогнозированию трендов цен на акции и отношений между акциями. Теперь на основе исследований получен инструмент, который может показать отношения между акциями, а также может предсказывать изменения цен на акции на следующие семь дней и предоставлять пользователям рекомендации. В процессе работы над данным проектом были выявлены многие ранее неизвестные проблемы. Изначально было решено использовать k-means для кластеризации, но позже обнаружили, что использование k-means метода приведет к множеству проблем для прогнозирования цены акций в паре с моделью VAR. Было решено использовать ANNOY для более полного решения этой проблемы. Когда исследование только началось было неизвестно, как динамически подключать результаты в модель из различных форматов данных. Позже было изучено, как другие модели решали эту проблему, и, наконец, проблема была решена самостоятельно с помощью методов из библиотеки на языке python. Также изначально не было известно, какой объем данных использовать для прогнозирования, а также неизвестно, через сколько дней после прогнозирования результаты будут наиболее точными. Но, для этого были использованы результаты поиска по сетке, чтобы выбрать параметры с наименьшей погрешностью.

В результате функции импульсного отклика рассчитываются непосредственно по коэффициентам модели VAR, с учетом всех разнонаправленных связей между ценами акций, определяемых каждым уравнением модели. Существует еще несколько будущих улучшений, которые можно сделать в рамках этой работы. Во первых, сейчас возможно простое предсказание изменения цен на акции на ближайшие несколько дней, поскольку результаты прогнозирования изменений цен на акции сходятся с моделью. Если нужно предсказывать цены на акции в течение очень долгого времени, то вероятно следует использовать другие более специализированные методы. Более того, была достигнута улучшение модели за счет уменьшения ошибок в процессе оценки изменений трендов.

**Литература**

1.George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung Time Series Analysis: Forecasting and Control,5th Edition. -Published by John Wiley and Sons Inc., Hoboken, New Jersey, 2015. –P. 712. ISBN 978-1-118-67502-1

2.David W Hosmer, S. L. Lemeshow Applied Logistic Regression. -John Wiley & Sons, New York, 2000. DOI 10.1002/0471722146

3.Agus Suharsono; Auliya Aziza; Wara Pramesti Comparison of vector autoregressive (VAR) and vector error correction models (VECM) for index of ASEAN stock price // AIP Conference Proceedings. - 2017. -Vol. 1913. DOI 10.1063/1.5016666

4.J. MacQueen. Some methods for classification and analysis of multivariate observations // Berkeley Symp. on Math. Statist. and Prob. -1967. -Vol. 5.1. -P. 281–297. URL: http://projecteuclid.org/euclid.bsmsp/1200512992

5.Erik BernhardssonANNOY: Approximate Nearest Neighbors in C++/Python // Github pages. -2014. URL: https://github.com/spotify/annoy

6.Bergstra J., Bengio Yo. Random search for hyper-parameter optimization// Journal of Machine Learning Research. -2012. -Vol. 13. -P. 281–305. DOI 10.5555/2188385.2188395

7.Christopher A. Sims Macroeconomics and Reality// Econometrica. -1980. -Vol. 48. -P. 1–48. DOI 10.2307/1912017

8.Helmut Lütkepohl New Introduction to Multiple Time Series Analysis: textbook. - 2005. ISBN 978-3-540-40172-8

9.Lutz Kilian,, Helmut Lütkepohl Structural Vector Autoregressive Analysis. -Cambridge University Press, 2017. DOI 10.1017/9781108164818

10.Ruey S. Tsay Analysis of Financial Time Series. – Wiley, 2010. -720 p. ISBN 9780470414354

11.David R. Rapach, Matthew C. Ringgenberg, , Guofu Zhou Short interest and aggregate stock returns // Journal of Financial Economics. -2016. -Vol. 121(1). -P. 46-65. DOI 10.1016/j.jfineco.2016.03.004

12.Wen Long & Jing Gao & Kehan Bai & Zhichen Lu A hybrid model for stock price prediction based on multi-view heterogeneous data // Financial Innovation, Springer;Southwestern University of Finance and Economics. -2024. -Vol. 10(1). –P. 1-50. DOI 10.1186/s40854-023-00519-w

***Cведения об авторах***

Есенбек А.- магистрант, Казахстанско-Британский технический университет, Алматы, Казахстан, e-mail:a\_yessenbek@kbtu.kz;

Касымова F.- магистрант, Казахстанско-Британский технический университет, Алматы, Казахстан, e-mail:a\_kassymova@kbtu.kz;

Картбаев А.-PhD, ассоциированный профессор, Казахстанско-Британский технический университет, Алматы, Казахстан, e-mail: a.kartbaev@kbtu.kz.

***Information about the authors***

Yessenbek A.- master student, Kazakh-British Technical University, Almaty, Kazakhstan, e – mail:a\_yessenbek@kbtu.kz;

Kassymova F.- master student, Kazakh-British Technical University, Almaty, Kazakhstan, e-mail:a\_kassymova@kbtu.kz;

Kartbayev A. – PhD, associate professor, Kazakh-British Technical University, Almaty, Kazakhstan, e-mail:a.kartbaev@kbtu.kz.

МРНТИ 28.17.19

**ПОСТРОЕНИЕ ПУТЕЙ ЭВАКУАЦИИ ЛЮДЕЙ ИЗ МНОГОЭТАЖНОГО**

**ЗДАНИЯ ПРИ УГРОЗЕ И ВОЗНИКНОВЕНИИ ПОЖАРА**

1А.Д. Бургегулов [D:\Desktop\иконка.png](https://orcid.org/0000-0003-3019-3352), 1Ш.А. Джомартова[D:\Desktop\иконка.png](https://orcid.org/0000-0002-5882-5588), 1А.Т. Мазакова[D:\Desktop\иконка.png](https://orcid.org/0000-0003-4904-3557), 2 М.С. Алиаскар[D:\Desktop\иконка.png](https://orcid.org/0000-0002-3013-6617),

1,2Т.Ж. Мазаков[D:\Desktop\иконка.png](https://orcid.org/%200000-0001-9345-5167), 3 А.Д. Майлыбаева[D:\Desktop\иконка.png](https://orcid.org/0009-0000-2958-1935)

*¹Казахский национальный университет имени аль-Фараби, Алматы, Казахстан,*

*2Международный инженерно-технологический университет, Алматы, Казахстан,*

*3Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан*

****Корреспондент-автор: [jomartova@mail.ru](mailto:jomartova@mail.ru)

Статья посвящена разработке алгоритма для построения эвакуационных маршрутов в случае чрезвычайных ситуаций, таких как пожар. Разработанный алгоритм используется в качестве математической модели интеллектуальной системы противопожарной безопасности и реализован на языке Python. Его эффективность подтверждается на примере модельной задачи, демонстрируя возможности мониторинга и реагирования на возникновение пожара.

Созданный алгоритм оптимизации маршрутов эвакуации основан на решении транспортных задач, обладает высокой универсальностью и легко адаптируется к практическим требованиям оперативного и безопасного выхода людей из здания. В статье представлены результаты численного моделирования, а также рассмотрены перспективы дальнейшего развития подобных систем.

Разработанные алгоритмы могут быть применены не только при пожаре, но и в случае других чрезвычайных ситуаций, таких как наводнения или землетрясения, что позволяет минимизировать риски и снизить ущерб от катастроф.

**Ключевые слова:** интеллектуальная система, противопожарная безопасность, транспортная задача, пожар, чрезвычайная ситуация, эвакуация.

ӨРТ ҚАУПІ ТӨНГЕН НЕМЕСЕ ПАЙДА БОЛҒАН ЖАҒДАЙДА КӨПҚАБАТТЫ ҮЙЛЕРДЕН АДАМДАРДЫ ЭВАКУАЦИЯЛАУ ЖОЛДАРЫН САЛУ

1А.Д. Бургегулов, 1Ш.А. Джомартова, 1Ә.Т. Мазақова, 2 М.С. Әлиасқар,

1,2Т.Ж. Мазақов 3А.Д. Майлыбаева

*¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,*

*,2Халықаралық инженерлік және технология университеті, Алматы, Қазақстан,*

*3 Х.Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан,*

*e-mail:* [*jomartova@mail.ru*](mailto:jomartova@mail.ru)

Мақала өрт сияқты төтенше жағдайлар кезінде эвакуациялау жолдарын құру алгоритмін әзірлеуге арналған. Әзірленген алгоритм өрт қауіпсіздігінің интеллектуалды жүйесінің математикалық моделі ретінде пайдаланылады және Python тілінде жүзеге асырылады. Оның тиімділігі өрттің пайда болуын бақылау және әрекет ету мүмкіндіктерін көрсете отырып, мысал ретінде үлгілік мәселені қолдану арқылы расталады.

Эвакуациялау жолдарын оңтайландырудың құрылған алгоритмі көлік мәселелерін шешуге негізделген, өте әмбебап және адамдардың ғимараттан тез және қауіпсіз шығуының практикалық талаптарына оңай бейімделуі мүмкін. Мақалада сандық модельдеу нәтижелері берілген, сондай-ақ мұндай жүйелерді одан әрі дамыту перспективалары қарастырылған.

Әзірленген алгоритмдер тек өрт кезінде ғана емес, сонымен қатар су тасқыны немесе жер сілкінісі сияқты басқа да төтенше жағдайлар кезінде де қолданылуы мүмкін, бұл тәуекелдерді барынша азайтуға және апаттардан келетін шығынды азайтуға мүмкіндік береді.

**Түйін сөздер:** интеллектуалды жүйе, өрт қауіпсіздігі, көлік тапсырмасы, өрт, төтенше жағдай, эвакуация.

**CONSTRUCTION OF EVACUATION ROUTES FOR PEOPLE FROM A MULTI-STORY BUILDING IN CASE OF A THREAT OR OCCURRENCE OF A FIRE**

**1A.D. Burgegulov, 1Sh.A. Jomartova, 1A.T.Mazakova, 2M.S. Aliaskar,**

**1,2T.Zh. Mazakov, 3Mailybayeva A.D.**

*¹Kazakh National University named after Al-Farabi, Almaty, Kazakhstan,*

*²International Engineering and Technology University, Almaty, Kazakhstan,*

*3Atyrau University named after Kh. Dosmukhambetov, Atyrau, Kazakhstan,*

*e-mail:* [*jomartova@mail.ru*](mailto:jomartova@mail.ru)

The article is devoted to the development of an algorithm for constructing evacuation routes in case of emergency situations, such as fire. The developed algorithm is used as a mathematical model of an intelligent fire safety system and is implemented in Python. Its effectiveness is confirmed by the example of a model problem, demonstrating the capabilities of monitoring and responding to a fire.

The created algorithm for optimizing evacuation routes is based on solving transport problems, is highly versatile and easily adapts to the practical requirements of prompt and safe exit of people from a building. The article presents the results of numerical modeling, and also considers the prospects for further development of such systems.

The developed algorithms can be applied not only in case of fire, but also in case of other emergency situations, such as floods or earthquakes, which allows to minimize risks and reduce damage from disasters.

Keywords: intelligent system, fire safety, transport task, fire, emergency, evacuation

**Введение.** Пожары представляют собой одни из наиболее разрушительных природных явлений, способных причинить значительный и зачастую необратимый вред экосистемам нашей планеты. В условиях быстрого научно-технического прогресса риски, связанные с пожарами, возрастают из-за усложнения технологических систем, развития электротехнической отрасли, широкого использования углеводородных материалов и влияния социально-экономических факторов. Все эти элементы в совокупности повышают вероятность возникновения аварий и катастроф, что приводит к серьезным экономическим и социальным последствиям.

Дополнительные трудности возникают из-за эксплуатации устаревших систем противопожарной безопасности, которые не соответствуют современным стандартам. Также важным аспектом является отсутствие комплексного подхода к решению данной проблемы на уровне государства.

В условиях крупных многоэтажных зданий и комплексов, где сосредоточено большое количество людей, быстрое и эффективное реагирование на угрозу пожара становится критически важным. Очевидно, что традиционные методы и технологии уже не способны обеспечить необходимый уровень безопасности.

**Материалы и методы.** Разработка специальных методов и вычислительных алгоритмов и программно-аппаратных комплексов, позволяющих в реальном времени предупреждать, строить оптимальные маршруты при возникновении ЧС являются актуальной проблемой [1-3]. Транспортная задача может быть применена для оптимального распределения людей по выходам в процессе эвакуации из здания, с целью сокращения времени эвакуации и предотвращения скопления людей.

*1. Постановка задачи*

Для разработки математической модели маршрутов эвакуации из здания введем следующие обозначения:

- количество этажей в здании;

- количество кабинетов на - м этаже, ;

- количество лестниц и выходов на - м этаже, ;

- количество людей в - м кабинете на - м этаже, , ;

- проходимость - й лестницы (выхода) на - м этаже, , ,

- количество людей из эвакуируемых из - го кабинета через - ю лестницу (выход) на - м этаже, , , ,

- время (или расстояние) необходимое для перемещения человека из - го кабинета через - ю лестницу (выход) на - м этаже, , , .

Создадим формализованную математическую модель, которая будет соответствовать определенным требованиям для организации эвакуации людей.

1. все люди должны покинуть каждый этаж здания

|  |  |
| --- | --- |
|  | (1) |

1. пропускная способность выходов не должна быть превышена (не создавалась давка):

|  |  |
| --- | --- |
|  | (2) |
|  |  |

1. условие неотрицательности переменных

|  |  |
| --- | --- |
| , , , , | (3) |
|  |  |

1. общее время эвакуации должно быть минимальным

|  |  |
| --- | --- |
| . | (4) |

1. *Теоретическая часть*

Существует множество методов для решения транспортной задачи (1)-(4). В дальнейшем мы рассмотрим метод потенциалов.

**Лемма 1.** Сложность метода потенциалов для задачи (1)-(4) может быть оценена следующим образом [4].

|  |  |
| --- | --- |
|  | (5) |

**Лемма 2.** Пусть имеются две числовые неотрицательные последовательности и , . Тогда справедливо следующее нераенство [5-7]

|  |  |
| --- | --- |
|  | (6) |

**Предлагается модификация транспортной задачи, адаптированная к проблеме эвакуации людей из многоэтажного здания.**

**На первом этапе решается транспортная задача для верхнего этажа, с целью минимизации функционала**

|  |  |
| --- | --- |
|  | (7) |

при ограничениях

|  |  |
| --- | --- |
| , , | (8) |

**Обозначим через s – номер шага итерации. Пусть s =1.**

**Шаг S. s=s+1. Если s=m то переходи к шагу N.**

**После выполнения предыдущего этапа становится известно, сколько людей спустилось с верхнего этажа по каждой лестнице. Введем фиктивные кабинеты, которые будут представлять собой лестницы, соединяющие исследуемый этаж с предыдущим. Количество людей в фиктивном кабинете будет определяться на основе предыдущего решения.**

**В результате передем к следующей транспортной задаче**

**минимизировать функционал**

|  |  |
| --- | --- |
|  | (9) |

при ограничениях

|  |  |
| --- | --- |
| , , . | (10) |

**Возвращаемся к шагу S.**

**Шаг N. Полученные на предыдущих шагах решения**  , , , **и составляют план эвакуации из многоэтажного здания.**

**Введем обозначение**

|  |  |
| --- | --- |
|  | (11) |

**Теорема.** Сложность предложенного метода для задачи (1)-(4) имеет следующую оценку

|  |  |
| --- | --- |
|  | (12) |

**Доказательство.**

1. ***Программная реализация***

Исходные данные о здании извлекаются из текстового файла Fzdan.txt, первая строка которого содержит целое число, обозначающее:

**-** Ne - количество этажей.

В последующих строках представлена информация о количестве кабинетов и лестниц (выходов), ведущих вниз (или из здания):

**-** Nk - количество кабинетов,

**-** Nl - количество лестниц (выходов).

Данные о кабинетах здания берутся из текстового файла Fkabn.txt, где построчно указаны:

* этаж, на котором расположен кабинет,
* номер кабинета,
* максимальное количество людей, которые могут находиться в кабинете.

Информация о лестницах здания содержится в текстовом файле Flest.txt, в котором построчно указаны:

* этаж, на котором он находится,
* номер лестницы,
* пропускная способность лестницы.

В текстовом файле Fsvja.txt описаны взаимосвязи между кабинетами, лестницами и выходами. Каждая строка включает следующие данные:

* первый объект,
* второй объект, с которым он связан,
* тип связи (1 — между кабинетом и лестницей, 2 - между лестницами),
* расстояние между объектами в метрах или время, необходимое для преодоления расстояния.

1. ***Экспериментальная часть***

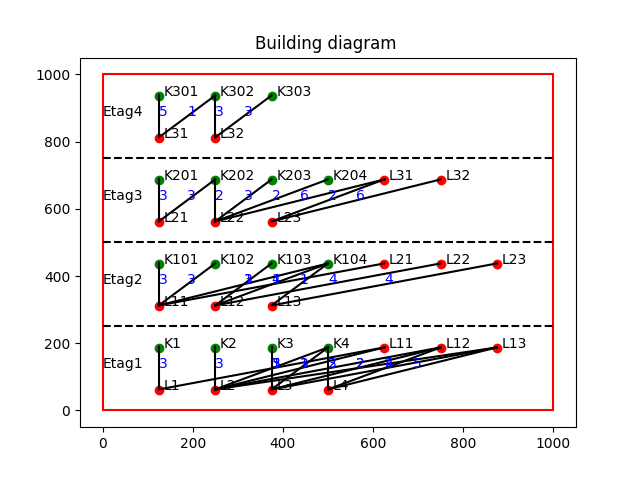
Для иллюстрации возможностей разработанного алгоритма были выполнены расчеты в рамках следующей модельной задачи.

Здание состоит из четырех этажей. На первом этаже размещены четыре кабинета и предусмотрены четыре выхода на улицу. На втором и третьем этажах также находятся по четыре кабинета, а количество лестниц, ведущих на нижний этаж, составляет три. Четвертый этаж включает три кабинета и две лестницы, ведущие на третий этаж.

Для определения оптимальных маршрутов эвакуации использованы методы решения транспортных задач [8-9].

Программа сохраняет результаты численных расчетов в текстовом файле и визуализирует их в виде графа. Это позволяет наглядно представить оптимальные маршруты эвакуации и структуру взаимосвязей между элементами здания.

На основе предоставленных данных создается граф, где вершины Ki обозначают кабинеты, а вершины Li – лестницы и выходы.



**Рис.1 - Экран с визуализацией полученного графа**

На рисунке 1 кабинеты обозначены зеленым кругом, лестницы и выходы – красным.

Связи между объектами представлены стрелками: кабинет-лестница, лестница-лестница. Количество людей, эвакуируемых через соответствующую лестницу, отображается синим цветом.

**Выводы.** В данной статье были проанализированы основные аспекты проектирования и создания интеллектуальной централизованной противопожарной системы безопасности для умных городов. После изучения существующих традиционных систем и выявления их недостатков, мы перешли к рассмотрению инновационных решений, включая применение искусственного интеллекта и машинного обучения, а также их интеграцию с другими компонентами умного города.

Современная версия системы успешно справляется с задачей определения и оптимизации маршрутов эвакуации, принимая во внимание характеристики здания и распределение людей. Тем не менее, существуют определенные ограничения, связанные с фиксированностью входных данных. В частности, система не учитывает динамические изменения в конфигурации здания, такие как временное закрытие или блокировка отдельных выходов. Это сказывается на актуальности некоторых эвакуационных сценариев в реальных условиях. Для преодоления этих ограничений и расширения функциональности системы предлагается интеграция датчиков, способных передавать информацию о состоянии выходов и других ключевых элементов здания в режиме реального времени.

В заключение следует подчеркнуть, что интеллектуальные системы противопожарной безопасности, развиваясь и интегрируясь в инфраструктуру умных городов, открывают новые горизонты для повышения безопасности и снижения рисков для жизни и здоровья граждан. Однако их успешное внедрение требует глубокого анализа и учета множества факторов, а также тесного сотрудничества между технологическими компаниями, научными учреждениями и городскими властями. Кроме того, такие системы должны функционировать в рамках единой городской управленческой структуры, объединяя данные и функции с другими подсистемами, такими как экологический мониторинг, транспортные системы и управление зданиями. Использование отдельных сенсоров и актуальных данных позволит создать более гибкую и адаптивную систему, способную оперативно реагировать на изменения ситуации и определять наиболее безопасные и эффективные маршруты эвакуации.

Искусственный интеллект и машинное обучение могут быть применены для постоянного совершенствования алгоритмов и моделей, основанных на собственном опыте компании и данных из реальной жизни. Это позволит разрабатывать более точные прогнозы и предлагать рекомендации по повышению общей безопасности зданий. Система может расширить свои возможности, создав мобильное приложение для информирования граждан и координации их действий. Это будет способствовать формированию единого информационного пространства, в котором все участники процесса, от спасателей до обычных граждан, смогут получать актуальную информацию и действовать более эффективно. Технология дополненной реальности может быть использована для моделирования следов на месте происшествия и других визуальных данных [10].

Предложенные методы и подходы имеют значительный потенциал в области защиты от природных катастроф, управления массовыми мероприятиями и в других сферах, где требуется быстрое и эффективное реагирование на изменения обстановки. [11-13].

***Финансирование.****Работа выполнена за счет средств НИИ математики и механики при КазНУ имени аль-Фараби и грантового финансирования научных исследований на 2023–2025 годы по проекту AP19678157.*

**Литература**

1.Пирогов Н.С. Анализ причин пожаров в зданиях культурно-досугового значения. // Пожарный надзор. – 2023.- №1. - С.67-73.

2.Смирнов В.П. Прогнозирование путей эвакуации при пожаре в общественных зданиях // Авт.доктор.дисс. спец. 05.26.03, 2018. - 50 с.

3.Иванов П.Н. Исследование времени блокирования путей эвакуации токсичными продуктами горения на промышленных объектах //Авт.канд.дисс. спец. 05.26.03, 2019.30 с.

4.Кормен, Томас X. и др. А45 Алгоритмы: Построение и анализ, 3-е изд.: Пер. с англ. -М.: ООО “И. Д. Вильямс”, 2013.-1328 с. ISBN 978-5-8459-1794-2

5.[Derek John Clements-Croome](https://www.researchgate.net/profile/Derek-Clements-Croome) Intelligent Buildings: Design Management and Operation- ICE Publishing,2004.-408 p. ISBN 0 7277 3266 8

6. Бургегулов А.Д., Мазаков Т.Ж., Зиятбекова Г.З., Саметова А.А., Джолдасова Б.У. Применение интеллектуальных систем пожарной безопасности в умных городах // Вестник КазУТБ.- 2023. - № 2(19). - С. 7-20. DOI 10.58805/kazutb.v.2.19-84

7. Жадан В. Г. Методы оптимизации. Часть II. Численные алгоритмы: учебное пособие / В. Г. Жадан. - М. : МФТИ, 2015. - 320 с. ISBN 978-5-7417 (Ч. II)

8. Рафгарден Т[.](https://www.meloman.kz/other/rafgarden-t-sovershennyj-algoritm-grafovye-algoritmy-i-struktury-dannyh.html?srsltid=AfmBOoonzl3DLF5EZHXaStUKbeOPMt0iOX2nrgpmPI-HIND3-7hOvRYs): Совершенный алгоритм. Графовые алгоритмы и структуры данных. – Питер: Питер-Трейд, 2020. - 340 с. ISBN 978-5-4461-1272-2

9. Сесекин А. Н., Ченцов А. А., Ченцов А. Г. Задачи маршрутизации перемещений. – Мз-во “Лань”, 2022. - 240 с. ISBN 978-5-8114-9999-1

10. Paes D., Feng Zh., King M., Shad H.K., Sasikumar P., Pujoni D., Lovreglio R. Optical see-through augmented reality fire safety training for building occupants //Automation in Construction. - 2024.-Vol.162. - ISSN 0926-5805. DOI [10.1016/j.autcon.2024.105371](https://doi.org/10.1016/j.autcon.2024.105371)

11. Krichen M., Abdalzaher M.S., Elwekeil M., Fouda M.M. Managing natural disasters: An analysis of technological advancements, opportunities, and challenges //Internet of Things and Cyber-Physical Systems.- 2024.-Vol.4.- P.99-109, ISSN 2667-3452.

[DOI 10.1016/j.iotcps.2023.09.002](https://doi.org/10.1016/j.iotcps.2023.09.002)

12.Мазаков Т.Ж., Бургегулов А.Д., Джомартова Ш.А., Мазакова А.Т., Саметова А.А., Досаналиева А.Т. Построение маршрутов для эвакуации сотрудников из здания при возникновении чрезвычайной ситуации // Вестник КазУТБ.- 2024.- № 1(22).- С.68-74

DOI 10.58805/kazutb.v.1.22-292

13.Бургегулов А., Зиятбекова Г., Мазаков Т., Саметова А., Алиаскар М. Использование технологий Интернета вещей для предотвращения пожаров // Международная конференция по электротехнике и информационным технологиям (ICEF 2023), Алматы, 2023. С.349-355

**References**

1.Pirogov N.S. Analiz prichin pozharov v zdanijah kul'turno-dosugovogo znachenija. // Pozharnyj nadzor. -2023.- №1. - S.67-73.[in Russian]

2.Smirnov V.P. Prognozirovanie putej jevakuacii pri pozhare v obshhestvennyh zdanijah // Avt.doktor.diss. spec. 05.26.03, 2018. - 50 s. [in Russian]

3.Ivanov P.N. Issledovanie vremeni blokirovanija putej jevakuacii toksichnymi produktami gorenija na promyshlennyh ob#ektah //Avt.kand.diss. spec. 05.26.03, 2019.30 s. [in Russian]

4.Kormen, Tomas X. i dr. A45 Algoritmy: Postroenie i analiz, 3-e izd.: Per. s angl. -M.: OOO “I. D. Vil'jams”, 2013.-1328 s. ISBN 978-5-8459-1794-2[in Russian]

5.[Derek John Clements-Croome](https://www.researchgate.net/profile/Derek-Clements-Croome) Intelligent Buildings: Design Management and Operation- ICE Publishing,2004.-408 p. ISBN 0 7277 3266 8

6.Burgegulov A.D., Mazakov T.Zh., Zijatbekova G.Z., Sametova A.A., Dzholdasova B.U. Primenenie intellektual'nyh sistem pozharnoj bezopasnosti v umnyh gorodah // Vestnik KazUTB.- 2023. - № 2(19). - S. 7-20. DOI 10.58805/kazutb.v.2.19-84. [in Russian]

7. Zhadan V. G. Metody optimizacii. Chast' II. Chislennye algoritmy: uchebnoe posobie / V. G. Zhadan. - M. : MFTI, 2015. - 320 s. ISBN 978-5-7417 (Ch. II). [in Russian]

8.Rafgarden T.: Sovershennyj algoritm. Grafovye algoritmy i struktury dannyh. – Piter: Piter-Trejd, 2020. - 340 s. ISBN 978-5-4461-1272-2. [in Russian]

9. Sesekin A. N., Chencov A. A., Chencov A. G. Zadachi marshrutizacii peremeshhenij. – Mz-vo “Lan'”, 2022. - 240 s. ISBN 978-5-8114-9999-1. [in Russian]

10. Paes D., Feng Zh., King M., Shad H.K., Sasikumar P., Pujoni D., Lovreglio R. Optical see-through augmented reality fire safety training for building occupants //Automation in Construction. - 2024.-Vol.162. - ISSN 0926-5805. DOI [10.1016/j.autcon.2024.105371](https://doi.org/10.1016/j.autcon.2024.105371)

11. Krichen M., Abdalzaher M.S., Elwekeil M., Fouda M.M. Managing natural disasters: An analysis of technological advancements, opportunities, and challenges //Internet of Things and Cyber-Physical Systems.- 2024.-Vol.4.- P.99-109, ISSN 2667-3452.

[DOI 10.1016/j.iotcps.2023.09.002](https://doi.org/10.1016/j.iotcps.2023.09.002)

12.Mazakov T.Zh., Burgegulov A.D., Dzhomartova Sh.A., Mazakova A.T., Sametova A.A., Dosanalieva A.T. Postroenie marshrutov dlja jevakuacii sotrudnikov iz zdanija pri vozniknovenii chrezvychajnoj situacii // Vestnik KazUTB.- 2024.- № 1(22).- S.68-74

DOI 10.58805/kazutb.v.1.22-292. [in Russian]

13.Burgegulov A., Zijatbekova G., Mazakov T., Sametova A., Aliaskar M. Ispol'zovanie tehnologij Interneta veshhej dlja predotvrashhenija pozharov // Mezhdunarodnaja konferencija po jelektrotehnike i informacionnym tehnologijam (ICEF 2023), Almaty, 2023. S.349-355. [in Russian]

***Сведения об авторах***

Бургегулов А.Д.- докторант КазНУ им.аль-Фараби, Алматы, Казахстан, e-mail: [dizel\_kz@bk.ru](mailto:dizel_kz@bk.ru);

Джомартова Ш.А. - доктор технических наук, доцент, КазНУ им. аль-Фараби, Алматы, Казахстан, e-mail: [jomartova@mail.ru](mailto:jomartova@mail.ru);

Мазақова Ә.Т.- докторант КазНУ им.аль-Фараби, Алматы, Казахстан, e-mail: [aigerym97@mail.ru](mailto:aigerym97@mail.ru);

Әлиасқар М.С. - старший преподаватель МИТУ, Алматы, Казахстан, e-mail: [m.alyasqar@gmail.ru](mailto:m.alyasqar@gmail.ru);

Мазаков Т.Ж. - доктор физико-математических наук, профессор, КазНУ им. аль-Фараби, Алматы, Казахстан, e-mail: [tmazakov@mail.ru](mailto:tmazakov@mail.ru);

Майлыбаева А.Д. – к.ф.-м.н., доцент кафедры «Информатика» Атырауского универститета имени Х. Досмухамедова, Атырау, Казахстан, e-mail: [a.maylibayeva@asu.edu.kz](mailto:a.maylibayeva@asu.edu.kz)

***Information about the author***

Burgegulov A.D. - Doctoral student of Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: [dizel\_kz@bk.ru](mailto:dizel_kz@bk.ru);

Jomartova Sh.A. - Doctor of Technical Sciences, Associate Professor, Al-Farabi KazNU, Almaty, Kazakhstan, e-mail: [jomartova@mail.ru](mailto:jomartova@mail.ru);

Mazakova A.T.- Doctoral student of Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: [aigerym97@mail.ru](mailto:aigerym97@mail.ru);

Aliaskar M.S. - Senior Lecturer, MITU, Almaty, Kazakhstan, e-mail: [m.alyasqar@gmail.ru](mailto:m.alyasqar@gmail.ru);

Mazakov T.Zh. - Doctor of Physical and Mathematical Sciences, Professor, Al-Farabi KazNU, Almaty, Kazakhstan, e-mail: [tmazakov@mail.ru](mailto:tmazakov@mail.ru);

Mailybayeva A.D. - PhD, Associate Professor, Department of Computer Science, Atyrau University named after Kh. Dosmukhambetov, Atyrau, Kazakhstan, e-mail: [a.maylibayeva@asu.edu.kz](mailto:a.maylibayeva@asu.edu.kz)

ҒТАМР 81.93.29

**ЖЕЛІЛЕРДІҢ ОСАЛДЫҚТАРЫН АНЫҚТАУ ҮШІН КӨП ФАКТОРЛЫ БОЛЖАУ АЛГОРИТМІН БЕЙІМДЕУ**

**1А.К.Шегетаева**[C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0000-0002-9212-251X)**, 1А.Б.Оспанова**[C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0000-0003-2811-9997)**, 2 Н.С.Смакова**[C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0009-0002-5754-4001)**, 3 Б.Рысбекқызы**[C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0000-0003-4353-1728)**,**

**2 С.А.Алтынбек** [C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0000-0002-8435-7773)**,4Л.Н. Кулбаева**[C:\Users\admin\AppData\Local\Microsoft\Windows\INetCache\Content.Word\iD.PNG](https://orcid.org/0000-0001-9029-5102)

*1Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан,*

*2Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,*

*3Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағнды, Қазақстан,*

*4 Astana IT University, Астана, Қазақстан*

Корреспондент – автор: aizhanshegetaeva@mail.ru

Желілік қауіпсіздік қазіргі цифрлық әлемдегі ең жылдам өзгеретін салалардың бірі болып қала береді. Бұл әсіресе жаңа технологиялардың дамуына және кибершабуылдар санының артуына байланысты өзекті болып табылады. Уақыт өте келе қорғау әдістері мен технологиялары ескіреді және жаңартуды қажет етеді. Бүгінгі таңда өнеркәсіптік желілер мен компьютерлік қауіпсіздік кибершабуылдар санының өсуіне және желінің осалдықтарына байланысты маңыздырақ болып отыр. Сондықтан жыл сайын анықталатын желілік осалдықтарды болжау және автоматтандыру өте маңызды. Бұл зерттеудің негізгі мақсаты - көп факторлы болжау алгоритмі арқылы желідегі осалдықтарды және кибершабуылдарды кешенді зерттеу. Көп факторлы болжау алгоритмі желі осалдықтарының жиілігін, желі трафигінің белсенділігін және қауіпке жауап беру уақытын ескере отырып, шабуылды болжау дәлдігін арттыруға мүмкіндік береді. Зерттеу барысында, Scopus және Web of Science сайттарында жарияланған зерттеу мақалаларына әдебиеттік шолу жүргізу және нейрондық желілер, логистикалық регрессия және кездейсоқ ормандар сияқты машиналық оқыту әдістерін қолдана отырып, деректерді қалыпқа келтіру және талдау арқылы әртүрлі көздерден деректерді жинауды қамтыды. Зерттеудің практикалық маңыздылығы әзірленген алгоритмді желі осалдықтарын бақылау жүйелерін құру үшін пайдалануға болады. Бұл корпоративтік желілердің қауіпсіздігін айтарлықтай жақсартады, кибершабуылдардан болатын зиянды азайтады және жалпы киберқауіпсіздікті жақсартады.

**Түйін сөздер:** киберқауіпсіздік, көп факторлы болжау, машиналық оқыту, нейрондық желілер , желі осалдықтар, эксплойттар

**АДАПТАЦИЯ АЛГОРИТМА МНОГОФАКТОРНОГО ПРОГНОЗИРОВАНИЯ ДЛЯ ОБНАРУЖЕНИЯ СЕТЕВЫХ УЯЗВИМОСТЕЙ**

**1А К.Шегетаева, 1А.Б.Оспанова, 2 Н.С.Смакова, 3 Б.Рысбекқызы,**

**2 С.А.Алтынбек, 4 Л.Н Кулбаева**

*1Евразийский национальный университет им. Л.Н.Гумилева, Астана, Казахстан,*

*2Казахский университет технологий и бизнеса имени К.Кулажанова, Астана, Казахстан,*

*3Карагандинский технический университет имени Абылкас Сагинова, Караганда, Казахстан,*

*4Astana IT University, Астана, Казахстан,*

*e-mail: aizhanshegetaeva@mail.ru*

Сетевая безопасность по понятным причинам остается одной из самых быстро меняющихся областей в цифровом мире. Кроме того, в связи с развитием новых технологий и появлением новых кибератак устаревание методов и технологий защиты происходит быстрыми темпами. В настоящее время сетевая и компьютерная безопасность в производственных средах становится все более важной из-за растущего числа кибератак и сетевых уязвимостей. Из-за увеличения количества удаленно подключенных устройств, увеличения объема данных и сложности сетевых технологий. Учитывая большое количество уязвимостей, выявляемых каждый год, автоматизация их прогнозирования имеет решающее значение. Целью данного исследования является комплексный анализ уязвимостей и атак в сети с использованием многофакторного алгоритма прогнозирования. Алгоритм многофакторного прогнозирования повышает точность прогнозирования атак за счет учета таких параметров, как частота сбоев в работе сети, активность сетевого трафика и время реакции на угрозы. Исследование включало проведение обзора литературы статей, опубликованных в Scopus и Web of Science, сбор данных из различных источников, их нормализацию и анализ с использованием методов машинного обучения, таких как нейронные сети, логистическая регрессия и случайные леса. Практическая ценность исследования заключается в том, что разработанный алгоритм может быть использован для создания систем мониторинга уязвимостей сетей и оценки эффективности эксплуатации в режиме реального времени. Это значительно улучшает защиту корпоративных и государственных сетей, снижает ущерб от кибератак и повышает общую кибербезопасность.

**Ключевые слова:** многофакторное прогнозирование, машинное обучение, нейронные сети, уязвимости сетей, эксплойты

**ADAPTATION OF MULTI-FACTOR FORECASTING ALGORITHM FOR DETECTING NETWORK VULNERABILITIES**

**1A.Shegetayeva, 1A.Ospanova, 2 N Smakova, 3 B.Rysbekkyzy,**

**2 S. Altynbek, 4L.Kulbayeva**

*1 L.N. Gumilyov Eurasian National University, Astana, Kazakhstan,*

*2 K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan,*

*3Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan,*

*4Astana IT University, Astana, Kazakhstan,*

*e-mail: aizhanshegetaeva@mail.ru*

Network security understandably remains one of the fastest changing areas in the digital world. In addition, with the development of new technologies and the emergence of new cyber attacks, the obsolescence of security methods and technologies occurs at a rapid pace. Nowadays, network and computer security in production environments is becoming increasingly important due to the growing number of cyber attacks and network vulnerabilities. Due to the increase in the number of remotely connected devices, the increase in the volume of data and the complexity of network technologies. Given the large number of vulnerabilities identified each year, automating their prediction is crucial. The objective of this study is to comprehensively analyze network vulnerabilities and attacks using a multi-factor forecasting algorithm. The multi-factor forecasting algorithm improves the accuracy of attack forecasting by taking into account parameters such as network failure rate, network traffic activity and threat response time. The study included a literature review of articles published in Scopus and Web of Science, data collection from various sources, their normalization and analysis using machine learning methods such as neural networks, logistic regression and random forests. The practical value of the study is that the developed algorithm can be used to create systems for monitoring network vulnerabilities and assessing the effectiveness of operation in real time. This significantly improves the protection of corporate and government networks, reduces damage from cyberattacks and improves overall cybersecurity.

**Keywords:** multivariate forecasting, machine learning, neural networks, network vulnerabilities, exploits

**Кіріспе.** Соңғы жылдары кибершабуылдар мен желінің осалдығы күрделі жаһандық проблемаға айналды. Қосылған құрылғылардың санының өсуі, деректер көлемінің өсуі және желілік технологиялардың күрделілігі бұл мәселені айқын көрсетеді. Көптеген компаниялар мен ұйымдар қашықтан жұмыс істейтіндіктен, қылмыскерлердің жаңа мүмкіндіктері бар. Мысалы, фишинг, DNS жазбаларын ұрлау, DDoS шабуылдары және құпия сөзді бұзу сияқты қауіптер бүгінгі күні өте өзекті. Сонымен қатар, бағдарламалық қамтамасыз ету қателері, желілік хаттаманың осалдықтары және маңызды жүйелердегі қауіпсіздік мәселелері мен интернеттегі пайдаланушылардың абайсыз әрекеті желінің осалдықтары айтарлықтай қауіп төндіреді.

Кибершабуылдармен күресу әдістері, олардың өсуіне қарамастан айтарлықтай өзгерді. Автоматтандырылған шешімдерді және көп факторлы болжау алгоритмдерін пайдалану желінің осалдықтар мен шабуылдардың алдын алудың маңызды шаралары болып есептелінеді [1]. Интрузияны анықтау жүйелері (IDS), желіаралық қалқандар және антивирустық бағдарламалық қамтамасыз ету көп деңгейлі қорғаныстың негізін құрайды. Бұл шаралар деректерге рұқсатсыз қол жеткізуді болдырмайды, сонымен қатар фишинг және DDoS сияқты шабуылдардан болатын зиянды азайтады [2].

Брандмауэр және шабуылды анықтау жүйелері (IDS) сияқты киберқауіпсіздік жүйелері заманауи шабуылдардың алдын алу үшін жеткілікті тиімді емес. Мысалы, эксплойттар қазір күрделі «камуфляж» әдістерін жиі пайдаланады, бұл анықтау мен бейтараптандыруды қиындатады. Нәтижесінде, қазіргі заманғы машиналық оқыту әдістерін қолданатын және бірнеше факторларды ескеретін кибершабуылдарды болжау және алдын алудың жаңа, дәлірек әдістерін жасау қажеттілігі туындады.

Зерттеудің мақсаты – көп факторлы болжау алгоритмін желінің осалдықтары мен эксплойттарына бейімдеу. Бұл кибершабуылдарды жақсы түсіну және алдын алу, сондай-ақ олар келтіруі мүмкін зияндардан қорғау және азайту стратегияларын әзірлеу үшін қажет. Зерттеудің негізгі міндеттері: 1) жүйелердің осалдығына әсер ететін негізгі факторларды және олардың эксплуатацияларын анықтау; 2) Желілік трафик, шабуыл жиілігі және қауіптерге жауап беру уақыты туралы деректер негізінде көп факторлы шабуылды болжау алгоритмін құру; және 3) желілік трафик деректеріне негізделген алгоритмді әзірлеу.

**Материалдар мен әдістер.**Заманауи осалдықты бағалау әдістері киберқауіпсіздік тәуекелдерін талдау және басқару үшін әртүрлі әдістер мен алгоритмдерді қолдайды. Графикалық шабуылдарды желілер мен жүйелердегі ықтимал шабуыл жолдарын модельдеу үшін пайдалануға болады. Карталар әлсіз жақтарды анықтауға және шабуылдарды болжауға көмектеседі. Мысалы, графикалық өндіру әдістері IoT желілеріндегі осалдықтарды бағалау үшін Марковтың шешім қабылдау процестері және машиналық оқыту алгоритмдері сияқты модельдерді пайдаланады [3].

Қауіпсіздік мәселелері архитектурада, жобалауда, кодта немесе бағдарламалық қамтамасыз етуді енгізуде пайда болады, жалпы қауіпсіздік ақауларының ресми тізілімі немесе сөздігі бар. Бұл тізім бағдарламалық қамтамасыз етудің қауіпсіздік ақауларын сипаттайтын, сондай-ақ осы ақауларды тануға, жоюға және болдырмауға арналған әмбебап ресми тіл ретінде әрекет етеді. CWE кибершабуылдар тізімі мен осалдықтар деректер қорына NDV, CVE, VulnDB, X-Force және т.б. қоса, осалдықтарды жіктеу мен бағалаудың әртүрлі жүйелері NIPC, SANC, nCirle, CVSS кеңінен қолданылады.

CVSS (Жалпы осалдықты бағалау жүйесі) – осалдықтарды пайдалану ықтималдығы мен желіге әсер ету негізінде бағалаудың жалпы қабылданған классификациясы. Соңғы зерттеулер әрбір инфрақұрылымның сипаттамаларын ескере отырып, нақты желілік орталарда тәуекелді дәлірек бағалау үшін жетілдірілген алгоритмдерді қолдануды көрсетеді [4].

Осалдықтар – жүйенің әлсіз жері, шабуылдаушылар оның қауіпсіздігі мен тұтастығын бұзу үшін пайдалана алады. Осалдықты талдау жүйелердегі ақпараттық қауіпсіздікті қамтамасыз етудің маңызды құралы болып табылады. Ол рұқсат етілмеген қол жеткізу, ақпаратты өзгерту немесе жою үшін шабуылдаушылар пайдаланатын деректер қауіпсіздігін бұзудың ықтимал осалдықтары мен тәуекелдерін анықтауға мүмкіндік береді [5].

Бұл құнды деректер мен активтердің тікелей ағып кетуі (әртүрлі ішкі жүйелерді басқаруға қашықтан қол жеткізу, шоттарды бұзу), соның ішінде бизнес-процестердің үзілуі, компанияның жұмысын тоқтату (қажет етпейтін шабуылдар) түріндегі елеулі қаржылық шығындар болуы мүмкін. желіге ену де жиі кездеседі (DoS, DDoS). Сондай-ақ, компанияның беделі мен тұтынушылардың сеніміне нұқсан келтіретін бопсалау немесе жариялаумен жалғасатын корпоративтік және құпия ақпараттың (деректер базасын бұзу) шығуы экономикалық шығындарға әкеледі. Жаңа желілік және бұлттық технологиялар, корпоративтік желілердің әртүрлі деңгейлеріндегі инфрақұрылым сөзсіз дерлік белгілі бір осалдықтарға ие, олардың көпшілігін жойылғанға дейін жеке хакерлер немесе тұтас топтар әдейі пайдаланады [6].

Эксплойттар SQL инъекциясы, буфердің толып кетуі, қол жеткізуді басқару осалдықтары және сайтаралық сценарийлер (XSS) осалдықтары сияқты осалдықтардың әртүрлі түрлеріне қарай жіктеледі. Эксплойттар прокси серверлер және қауіпсіздік механизмдерін айналып өту сияқты күрделі әдістерін пайдаланады, бұл шабуылды анықтау мен ізді қалпына келтіруді қиындатады.

**Нәтижелер мен талқылау.** Болжаудың дәлдігі қазіргі заманғы машиналық оқыту әдістері мен тарихи деректерді пайдалану арқылы артады. Бұл жүйелерді тиімдірек қорғауға және шабуылдардың алдын алуға көмектеседі.

Киберқауіпсіздікте көп факторлы болжау алгоритмі қауіпті болжау мен тәуекелдерді басқарудың қуатты құралы болып табылады. Болжаудың дәлдігі қазіргі заманғы машиналық оқыту әдістері мен тарихи деректерді пайдалану арқылы артады. Бұл жүйелерді тиімдірек қорғауға және шабуылдардың алдын алуға көмектеседі. Көп факторлы болжау алгоритмін құрудың блок-схемасы 1-суретте көрсетілген. Бұл қадамдар деректерді жинауды, деректерді тазалауды және нормалауды, корреляциялық талдауды және факторларды таңдауды, модельді және алгоритмді таңдауды, оқыту мен тестілеуді, болжау және қорытынды жасауды қамтиды.

**1- сурет. Көп факторлы болжауды дамытудың блок-схемасы**

Көп факторлы болжау алгоритмін жасау математикалық әдістерді қолдануды және деректерді дұрыс бөлуді талап етеді, толығырақ процесс төменде берілген. Ол деректерді тазалау және қалыпқа келтіру формулаларын, корреляциялық талдауды және үлгілерді оқыту мен сынау әдістерін қамтиды.

*1 . Дереккөздерден мәліметтер жинау*

Кез келген болжау алгоритмінің негізі - деректер болып табылады. Бұл оқиғалар журналдары, кибершабуылдар мен шабуылдарды анықтау жүйесінің (IDS) дерекқорлары және сыртқы осалдық репозиторийлері сияқты әртүрлі көздерден алынуы мүмкін. EXPLOIT - DB ([https://www.exploit-db.com/](https://www.exploit-db.com/%20)). Зерттеу үшін Exploit-DB дерекқорынан алынған 500 желінің осал тұстары мен қатысты эксплойттар үлгісі құрастырылды. Үлгі буфердің толып кетуі, веб-бағдарламаның осалдықтары (SQL инъекциясы, XSS), аутентификация және артықшылық қателері, сондай-ақ желі құрылғысы мен қызмет конфигурацияларындағы әртүрлі типтегі осалдықтарды қамтиды. Әрбір эксплуатация бірнеше критерийлер негізінде бағаланады, соның ішінде пайдалану қиындығы, сәттілік ықтималдығы және ықтимал зиян. Деректерді келесі критерийлер бойынша таңдалды, олардың келесі айнымалылары бар: – осалдықтардың жиілігі (белгілі бір кезеңде жүйеде анықталған осалдықтардың саны); – желілік трафик белсенділігі (кіріс және шығыс деректер пакеттерінің көлемі); – шабуыл жиілігі (желіге жазылған шабуылдар саны); – қауіптерге жауап беру уақыты (қауіптерді анықтау және жоюдың орташа уақыты); – осалдықтардың түрі (CVSS стандартына сәйкес осалдықтардың жіктелуі).

*2. Деректерді тазалау және қалыпқа келтіру*

Деректерді тазалау көшірмелерді жоюды, жетіспейтін мәндерді жоюды және қателерді түзетуді қамтиды. Аномалияларды жою және деректерді қалыпқа келтіру үшін келесі әдістер қолданылады.

Аномалияларды жою формуласы

(1)

мұндағы - айнымалының бастапқы мәні, - орташа мәні және стандартты ауытқуы. (1) формула шектен шығуды жоюға көмектеседі және қалыпты таралуға әкеледі.

Нормалау:

, (2)

мұндағы және әрбір айнымалы үшін ең төменгі және ең үлкен мәндер. Бұл процесс деректерді 0-ден 1-ге дейінгі диапазонға қалыпқа келтіріп, машиналық оқыту үлгілерінің жұмысын жеңілдетеді.

*3. Корреляциялық талдау және факторларды таңдау*

Корреляциялық талдау әртүрлі айнымалылар мен мақсатты айнымалылар арасындағы байланысты анықтауға көмектеседі (мысалы, шабуылдың ықтималдығы). Корреляцияны өлшеудің бір жолы Пирсон коэффициентін қолдану болып табылады:

Пирсон корреляция коэффициентінің формуласы

(3)

мұндағы тәуелсіз және тәуелді айнымалылар, және олардың орташа мәндері.

Бұл коэффициент факторлар мен нәтиже арасындағы сызықтық байланыстың дәрежесін көрсетеді.

*4. Алгоритм мен модельді таңдау*

Киберқауіпсіздікте көп факторлы болжау үшін әртүрлі алгоритмдерді қолдануға болады, олардың арасында:

4.1 *Логистикалық регрессия* оқиғаның ықтималдығы болжанатын екілік жіктеу үшін қолданылады (мысалы, шабуыл/шабуылсыз):

(4)

мұндағы , а - үлгілік оқыту нәтижесінде алынған коэффициенттер.

– оқиғаның орын алу ықтималдығы (мысалы, шабуылдың ықтималдығы).

– тәуелсіз айнымалылар (факторлар).

– оқытылатын модельдің параметрлері.

– көрсеткіш.

4.2 *Шешім ағаштары (Decision Trees)*

Шешім ағаштары шешімдер қабылданатын шарттар (тармақтар) тізбегін құрайды. Тармақ айнымалылардағы шарттарды білдіреді, ал жапырақтарда сынып белгілері бар (мысалы, шабуыл). Алгоритм деректерді энтропия немесе Дженни индексі сияқты критерий негізінде бөледі.

Энтропия:

(5)

мұндағы сыныпқа жататын объектілердің үлесі, және сыныптар саны.

Джини индексі:

(6)

Алгоритм бөлу үшін ең тиімді сипатты таңдау үшін осы көрсеткіштердің бірін азайтуға тырысады.

4.3 *Кездейсоқ орман* *(Random Forest)*.

Кездейсоқ орман - дәлдікті жақсарту және артық орнатуды болдырмау үшін бірнеше шешім ағаштарын біріктіретін ансамбль әдісі. Алгоритм деректердің әртүрлі жиынтықтары бойынша көптеген шешім ағаштарын құрастырады және дауыс беру арқылы соңғы шешімді шығарады:

(7)

мұндағы ағаштар саны, сынып үшін ағаштың болжаған ықтималдығы .

4.4 *Нейрондық желілер (Neural Networks).*

Нейрондық желілер қабаттарға біріктірілген есептеу түйіндерін (нейрондарды) көрсету арқылы мидың жұмысын модельдейді. Ең көп тараған түрі – көпқабатты перцептрондар (MLP).

Бір нейронның формуласы**:**

(8)

мұндағы нейронның шығысы, кіріс пен нейрон арасындағы байланыстың салмағы, кіріс деректері, ығысу, белсендіру функциясы (мысалы, сигма тәрізді немесе ReLU).

Сигмоидты белсендіру функциясы:

(9)

ReLU (түзетілген сызықтық бірлік)

(10)

Нейрондық желілер қателіктердің таралуына кері әсер ететін логистикалық жоғалтуды минимизациялайды (мысалы, орташа квадраттық қателіктер).

Бұл алгоритмдер көбінесе шабуыл және осалдықты болжау сияқты киберқауіпсіздік тапсырмалар үшін пайдаланылады.

*5. Оқыту және тестілеу*

Деректер оқу және сынақ үлгілеріне бөлінуі, әдеттегі деректер 70/30 қатынасын құрайды, мұнда 70% оқыту үшін және 30% тестілеу үшін пайдаланылады. Бұл модельге деректердің бір бөлігінен «үйренуге» және шамадан тыс өзін жаңа ақпаратта сынауға мүмкіндік береді.

Модельді үйрету үшін бинарлық есептеулер үшін логистикалық жоғалту функциясы пайдаланылады:

мұндағы ақиқат мән, болжамды мән және бақылаулар саны.

Дәлдік, толықтық және F1 өлшемі, классификация моделінің өнімділігін бағалауға арналған негізгі көрсеткіштер болып табылады, әсіресе шабуылдар мен осалдықтарды дұрыс анықтау келесі формулалар арқылы есептеледі:

Дәлдік (Precision):

(11)

мұндағы (True Positive) – дұрыс болжанған оң мысалдар саны (мысалы, анықталған шабуылдар); (False Positives) – модель теріс мысалды оң деп қате жіктеген кездегі жалған оң мәндердің саны (мысалы, «жалған дабыл»).

Толықтық (Аccuracy):

мұндағы (False Negatives) – модель оң мысалды анықтай алмаған мысалдар саны (мысалы, өткізіп алған шабуылдар саны).

F1- өлшем (F1-балл):

(12)

Жоғарыдағы (12)-ші формулада F1-өлшемі дәлдік пен толықтық арасындағы гармоникалық орташа мән болып табылады, ол модельді неғұрлым теңгерімді бағалау үшін қолданылады, әсіресе жалған позитивтерді де, өткізіп алған шабуылдарды да есепке алу маңызды.

*6. Болжау және қорытындылар*

Модель оқытылып, тексерілгеннен кейін, оны болжау үшін пайдалануға болады. Болжау кіріс ретінде қабылдайтын және оқиғалардың ықтималдығын шығаратын болжау функциясына негізделген маңызды фактор.

500 желі осалдықтарының үлгісі мен Exploit-DB дерекқорынан эксплойттардың тиімділігін - көрсеткіштерге сәйкес сипаттау үшін 1 - кестені әрбір индикатор үшін максималды, ең төменгі және орташа мәндермен ұсынылды.

**1- кесте. Зерттелетін көрсеткіштердің максималды, ең төменгі және орташа мәндері**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Көрсеткіш | Өлшем бірліктері | Максималды мән | Ең төменгі мән | Орташа мән () |
|  | тәулігіне жазбалар саны (бірлік/күн) | 200 | 1 | 50 |
|  | желілік трафик көлемі (ГБ/күн) | 800 | 10 | 250 |
|  | күніне шабуылдар саны (бірлік/күн) | 500 | 20 | 150 |
|  | қауіптерге жауап беру уақыты (минут) | 600 | 5 | 180 |
|  | CVSS осалдық көрсеткіші (баллдар) | 10 | 3 | 6.5 |

**2 - сурет. 1 тәулікте болатын осалдықтар мен шабуылдардың максималды және минималды көрсеткіштері**

1-ші кесте жаңа өлшем бірліктерін ескере отырып, осалдықтардың пайда болу жиілігін, шабуылдар жиілігін және қауіптерге жауап беру уақытын түсіндірудегі өзгерістерді көрсетеді (2 сурет). 2-ші кесте деректерді тазалау (аномалияларды жою) мысалын көрсетеді.

**2-кесте. Деректерді тазалау (аномалияларды жою)**

|  |  |  |  |
| --- | --- | --- | --- |
| Көрсеткіш | Орташа ( ) | Стандартты ауытқу () | (аномалияларды жойғаннан кейін) |
|  | 50 | 70 | ( -50) / 70 |
|  | 250 | 200 | ( -250) / 200 |
|  | 150 | 120 | ( -150) / 120 |
|  | 180 | 160 | ( -180) / 160 |
|  | 6.5 | 2 | ( -6,5) / 2 |

Деректерді қалыпқа келтіру үшін (2) формула пайдаланылды. Деректерді қалыпқа келтіру нәтижелері 3-ші кестеде көрсетіледі.

**3-кесте. Деректерді қалыпқа келтіру**

|  |  |  |  |
| --- | --- | --- | --- |
| Көрсеткіш | Ең аз () | Максимум  () | Шамамен нормаланған мән () |
|  | 1 | 200 | ( -1) / (200-1) |
|  | 10 | 800 | ( -10) / (800-10) |
|  | 20 | 500 | ( -20) / (500-20) |
|  | 5 | 600 | ( -5) / (600-5) |
|  | 3 | 10 | ( -3) / (10-3) |

- көрсеткіштер және мақсатты айнымалылар арасындағы корреляцияны талдау жүргізу үшін талдау Пирсон корреляция коэффициенті (3) формуласы пайдаланылды. Корреляциялық талдаудың нәтижелері 3-ші суретте көрсетіледі.

**3 - сурет. Пирсон корреляция коэффициенттері**

Ең жоғары оң корреляция () (осалдықтардың пайда болу жиілігі) және (шабуылдар жиілігі) арасында байқалады, бұл осалдықтар саны мен шабуылдар саны арасындағы өзара байланысты көрсетеді.

Орташа корреляция () (желі трафик белсенділігі) және (қауіпке жауап беру уақыты) арасындағы трафик көлемі мен қауіпке жауап беру уақыты арасындағы тәуелділікті көрсете алады.

Төмендеген корреляция (күніне шабуылдар саны) және (CVSS бойынша осалдықты бағалау) көрсеткіштері арасында және (осалдықтардың жиілігі) және (CVSS бойынша осалдық көрсеткіші) индикаторлар арасында төмендеген корреляция бар екенін де атап өтуге болады. (желінің трафик белсенділігі) және (шабуыл жиілігі) арасындағы корреляция коэффициенті жеткілікті жоғары екенін белгілеп, айта кеткен жөн. Бұл нәтижелер көп факторлы болжау үшін негізгі факторларды таңдауға мүмкіндік береді.

3 және 4-суреттер шабуылдың ықтималдығын болжау үшін төрт машиналық оқыту алгоритмін пайдалану нәтижелерін көрсетеді. Дәлдік (Precision), толықтық (Аccuracy) және F1-өлшемі (F1-балл) көрсеткіштері оқыту және сынақ үлгілері үшін есептелінді.

**4-сурет. Зерттеу үлгісінде алгоритмдерін қолдану нәтижелері**

4 cуретте көрсетілген деректерді талдау арқылы, Логистикалық регрессия 0,85 дәлдігін көрсетеді, яғни болжанған шабуылдардың 85% ақиқат мәнінде шабуылдар болды. Толықтығы 0,78 деңгейінде барлық нақты шабуылдардың 78% дұрыс болжағанын көрсетеді. F1-өлшемі 0,81 деңгейі дәлдік пен толықтық арасындағы жақсы теңгерімді көрсетеді.

Шешім ағаштары дәлдігі 0,82 және толықтық 0,80 сәл теңдестірілген нәтижелерді көрсетеді, бұл 0,81 F1-өлшемінде де көрінеді. Бұл модель шабуылдарды болжау міндетін тиімді екенін білдіреді.

Кездейсоқ орман барлық алгоритмдер арасында ең жоғары 0,88 дәлдікті көрсетті, бұл дұрыс болжамдардың жоғары деңгейін көрсетеді. Алайда, толықтығы 0,74 болды, бұл модель шабуылдарды шамамен 26% өткізіп алғанын көрсетеді. F1-өлшемі 0,80-ге тең, жоғары дәлдікке қарамастан, дәлдік пен толықтық арасындағы тепе-теңдік идеалды емес екенін көрсетеді.

Нейрондық желілер 0,87 дәлдігін көрсетеді, бұл кездейсоқ орман нәтижелеріне жуық, ал еске түсіру 0,76. Бұл нейрондық желілер барлық шабуылдардың 76% дұрыс анықтағанын білдіреді және F1-өлшемі 0,81-ге тең дәлдік пен толықтық арасындағы теңгерімді көрсетеді.

**5 -сурет. Сынақ үлгісінде алгоритмдерді қолдану нәтижелері**

5 суретінде көрсеткендей, Сынақ үлгісінде алгоритмдерді қолдану нәтижелері бойынша Логистикалық регрессия 0,84 дәлдікпен 0,76 толықтық берді, бұл Оқыту үлгісінің нәтижесінен сәл төмен, бірақ F1-өлшемі 0,80 модель жаңа деректерге шабуылдарды болжауда әлі де тиімді.

Сынақ үлгісінде Шешім ағаштары 0,80 дәлдігін және 0,78 толықтықты көрсетті, бұл олардың оқыту үлгісінің нәтижелеріне ұқсас. F1 өлшемі 0,79 мәнді алгоритмнің әртүрлі деректер жиындарында тұрақты орындалатынын растайды, бұл жеткілікті теңдестірілген нәтижелер береді.

Кездейсоқ орман 0,85 дәлдікті көрсетеді, бұл жаттығу жиынындағы нәтижеден небәрі 0,03 төмен. Дегенмен, толықтық 0,72-ге дейін төмендейді, бұл модельдің шабуылдарды анықтау сезімталдығының төмендеуін көрсетеді. F1-өлшемі 0,78-ге тең жаңа деректермен жұмыс істеу кезінде модельдің жалпы балансының төмендеуін көрсетеді.

Нейрондық желілер сынақ үлгісінде 0,86 дәлдікпен тұрақтылықты көрсетті, бұл олардың жаңа деректерге шабуылдарды болжаудың жоғары қабілетін растайды. Толықтығы 0,74-ке дейін төмендеді, бірақ F1-өлшемі 0,80-ге тең, модель әлі де болжамдылықтың жоғары деңгейін және дәлдік пен толықтық арасындағы тепе-теңдікті сақтайтынын растайды.

4 және 5 суреттегі салыстырулар нәтижелерінен, нейрондық желілер шабуылды болжаудың жоғары дәлдігін және теңдестірілген F1-өлшемдерін көрсете отырып, екі үлгіде де ең жақсы нәтижелерді көрсетеді деп қорытынды жасауға болады. Кездейсоқ орман оқыту үлгісінде ең жоғары дәлдікті көрсетті, бірақ оның сынақ үлгісіндегі деректерді жалпылау қабілеті нейрондық желілерге қарағанда төмен болды. Логистикалық регрессия және шешім ағаштары нейрондық желілерге қарағанда тұрақтырақ, бірақ дәлірек нәтижелерді береді.

Машиналық оқытуды пайдаланатын көп факторлы болжау алгоритмі желінің осал тұстарын бағалауда жоғары дәлдікті көрсетті. Бұл осалдықтардың жиілігі, желілік трафик белсенділігі, қауіптерге жауап беру уақыты және басқа да факторлар сияқты әртүрлі параметрлерді ескеру мүмкіндігіне байланысты. Бұл айнымалылар арасындағы күрделі қатынастарды жиі есепке алмайтын дәстүрлі бір факторлы әдістерден оның артықшылығын растайды [7].

Зерттеулер көрсеткендей, осалдықтардың жоғары жиілігі (мысалы, тәулігіне осалдықтардың саны) жүйелерге жасалған шабуылдардың жиілігімен сәйкес келеді. Сонымен қатар, біздің нәтижелеріміз сәтті шабуылдардың осал желілерді, әсіресе желілік белсенділік жағдайында көбірек нысанаға алатынын көрсетті. Басқа авторлар инфрақұрылымға шабуылдардың жиілігі мен желілік трафик белсенділігі арасындағы байланысты көрсетті, бұл нәтижелерді растайды [8].

Нейрондық желілер және шешім ағаштары сияқты заманауи машиналық оқыту әдістерін пайдалану арқылы көп факторлы болжамды модельдер ықтимал қауіптерді дәлірек анықтай алады. Бұл алгоритмдер үлкен көлемдегі деректерді өңдей алады және дәстүрлі талдау әдістері анықтай алмайтын жасырын үлгілерді таба алады [9].

Көп факторлы болжау алгоритмінің үлкен артықшылығы, оның жоғары дәлдігі және әртүрлі желі жағдайларына бейімделу мүмкіндігі болып табылады. Дегенмен, оның кемшіліктері бар, оларды одан әрі оңтайландыру үшін ескеру қажет.   
Шабуыл жиілігі мен осалдықтар сияқты әртүрлі факторлар қосылғанда, әсіресе нейрондық желілер мен кездейсоқ ормандарды пайдаланғанда, болжау дәлдігі айтарлықтай артады. Біздің экспериментіміз модель болашақ қауіптерді дәл болжай алатынын көрсетті, өйткені сынақ деректері F1-өлшемін 0,80 көрсетті. Алгоритм бірнеше түрлі желілік инфрақұрылымдар мен кибер орталарда жұмыс істейді. Мысалы, бірнеше зерттеулер оның шағын желілерді де, кәсіпорын инфрақұрылымдарын да сәтті модельдейтінін көрсетті. Кибершабуылдарды болжаудағы тиімділігін растай отырып, нейрондық желілер дәлдікте де, F1-өлшемінде де ең жақсы нәтиже көрсетті. Нейрондық желілер күрделі желілік деректерді талдау кезінде жоғары дәлдік көрсетті, бұл осы нәтижелерді растайды [10].

Көп факторлы модельдерді, әсіресе нейрондық желілерге негізделген модельдерді пайдалану үлкен есептеу ресурстарын қажет етеді, бұл оларды нақты уақытта пайдалануды қиындатады. Олардың қаншалықты дәл екендігіне қарамастан, кездейсоқ орман және шешім ағаштары сияқты машиналық оқыту үлгілері, әсіресе шағын деректермен жұмыс істегенде, шамадан тыс орнатуға бейім. Бұл олардың жаттығу деректерінде анықталмаған жаңа шабуылдарды дәл болжау қабілетін төмендетуі мүмкін [11].

**Қорытынды.** Бұл зерттеудің мақсаты көп факторлы болжау алгоритмі арқылы желінің осалдықтары мен эксплуатацияларының өнімділігін бағалау болды. Осы мақсатқа жету үшін желі белсенділігінің әртүрлі параметрлерін, шабуылдардың жиілігін және қауіпке реакция уақытын ескере отырып, болашақ кибершабуылдарды дәлірек болжауға мүмкіндік беретін алгоритм әзірленді және қолданылды [12].

Зерттеуге арналған деректер оқиғалар журналдары және осалдық дерекқорлары сияқты әртүрлі көздерден жиналды. Деректерді қалыпқа келтіру және тазалаудан кейін логистикалық регрессия, нейрондық желілер, кездейсоқ ормандар және шешім ағаштары сияқты машиналық оқыту әдістері қолданылды. Осалдықтардың түрі, трафик белсенділігі және осалдықтардың жиілігі осы алгоритмдер шабуылдардың ықтималдығын болжау үшін пайдаланатын факторлар болды.

Негізгі зерттеу нәтижелері нейрондық желілер шабуылдарды болжауда тамаша екенін көрсетті. Нейрондық желілер шабуылдарды дәл болжау және сынақ жиынындағы жалған позитивтердің санын азайту қабілетін дәлелдеді, дәлдік 86% және F1-өлшемі 0,80. Екінші жағынан, шешім ағаштары 80% дәлдік пен F1-өлшемі 0,79, ал кездейсоқ ормандар 85% және F1- өлшемі 0,78 дәлдік көрсетті . Бұл осалдықтардың жиілігі мен шабуылдардың жиілігі бір-бірімен тығыз байланысты екенін көрсетті және олардың арасындағы ең жоғары корреляция r = 0,75 болды.

Алынған нәтижелер деректердің ағып кетуін және басқа кибершабуылдарды болдырмау үшін шабуылдарды дәл болжау қажет болатын корпоративтік және мемлекеттік желілерді қорғау саласында қолданылуы мүмкін. Алгоритмді нақты уақыт режимінде желі белсенділігін бақылау жүйелерін құру үшін пайдалануға болады, бұл қауіптерге жауап беру уақытын қысқартуға және инфрақұрылымның жалпы қауіпсіздігін арттыруға мүмкіндік береді.

**Әдебиеттер**

1.Сасенов А.Б. Прогнозирование развития криминальной ситуации в Республике Казахстан: монография/КариповаА.Т. Касимова М.О., Абулгазина А.Ж., Казбекова А.Б., Тулеуов Б.И., Сыздыков А.Ж., Муканов Д.Ж. Под общей редакцией Ахметзакирова Н.Р. - Астана: Академия правоохранительных органов при Генеральной прокуратуре Республики Казахстан, 2017.- 172 с. ISBN 978-601-06-4184-6

2.Falkevych, V., Lisnyak, A. Internal and External Threats in Cyber Security and Methods for Their Prevention // In 2023 13th International Conference on Advanced Computer Information Technologies (ACIT). -2023. -P.414 - 419. DOI 10.1109/ACIT58437.2023.10275516

3.Almazrouei, O., Magalingam , P., Hasan, M., & Shanmugam, M. A Review on Attack Graph Analysis for IoT Vulnerability Assessment: Challenges, Open Issues, and Future Directions // IEEE Access. -2023. -Vol.11. -Р. 44350-44376. DOI 10.1109/ACCESS.2023.3272053

4.Wang, W., Shi, F., Zhang, M., Xu, C., & Zheng, J. A Vulnerability Risk Assessment Method Based on Heterogeneous Information Network // IEEE Access. -2020. -Vol. 8. *-*P*.*148315–148330. [DOI 10.1109/ACCESS.2020.3015551](https://doi.org/10.1109/ACCESS.2020.3015551)

5.Хромова А.Р., Петросян Л.Э. Анализ уязвимостей в системах безопасности данных // Инженерный вестник Дона. -2023. -№ 6.(102)-C.67-76

6.Оспанова А.Б., Шегетаева А.К., Түсіпханов А. Т., Жалгасбаев А. Б., Кадринов Д. М. Прогнозирование сетевых уязвимостей и эксплойтов // Международная научно-практическая конференции «XVI Сагиновские чтения. Интеграция образования, науки и производства». - Караганда: Изд-во КарТУ им. А.Сагинова, 2024. -Ч.2. -С. 287-289.

7.Carriegos, M., Castañeda, Á., Trobajo , M., & Zaballa , D. On Aggregation and Prediction of Cybersecurity Incident Reports //IEEE Access.-2021.-Vol.9.-P.102636-102648.

[DOI 10.1109/ACCESS.2021.3097834](https://doi.org/10.1109/ACCESS.2021.3097834)

8.Demidova, L., & Stepanov, M. Development of Multifactor Forecasting Model based on Fuzzy Time Series // 2021 International Conference on Information Technologies (InfoTech). -2021. [DOI 10.1109/InfoTech52438.2021.9548389](https://doi.org/10.1109/InfoTech52438.2021.9548389)

9.Injadat, M., Moubayed, A., & Shami, A. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach // 202032nd International Conference on Microelectronics (ICM*)*. -2020. DOI [10.1109/ICM50269.2020.9331794](https://doi.org/10.1109/ICM50269.2020.9331794)

10.Pavlov, A. Analysis of Network Interaction of Modern Exploits // Informacionnye Tehnologii. -2022. -Vol. 28(2). [DOI 10.17587/it.28.75-80](https://doi.org/10.17587/it.28.75-80)

11.Aota, М., Kanehara, H., Кubо, М., Мurata, N., Sun, B., Takahashi, Т. Automation of Vulnerability Classification from its Description using Machine Learning //2020 IEEE Symposium on Computers and Communications (ISCC). -2020.

DOI 10.1109/ISCC50000.2020.9219568

12.Шегетаева А.К. Всесторонний обзор многофакторного прогнозирования сетевых уязвимостей // International Scientific Symposium Karabakh and West Azerbaijan: Triumph of Victory. The 26th of October 2024. - P. 553-560. ISBN 978-625-98125-4-0

**References**

1.Sasenov A.B. Prognozirovanie razvitija kriminal'noj situacii v Respublike Kazahstan: monografija/KaripovaA.T. Kasimova M.O., Abulgazina A.Zh., Kazbekova A.B., Tuleuov B.I., Syzdykov A.Zh., Mukanov D.Zh. Pod obshhej redakciej Ahmetzakirova N.R. - Astana: Akademija pravoohranitel'nyh organov pri General'noj prokurature Respubliki Kazahstan, 2017.- 172 s. ISBN 978-601-06-4184-6. [in Russian]

2.Falkevych, V., Lisnyak, A. Internal and External Threats in Cyber Security and Methods for Their Prevention // In 2023 13th International Conference on Advanced Computer Information Technologies (ACIT). -2023. -P.414 - 419. DOI 10.1109/ACIT58437.2023.10275516

3.Almazrouei, O., Magalingam , P., Hasan, M., & Shanmugam, M. A Review on Attack Graph Analysis for IoT Vulnerability Assessment: Challenges, Open Issues, and Future Directions // IEEE Access. -2023. -Vol.11. -Р. 44350-44376. DOI 10.1109/ACCESS.2023.3272053

4.Wang, W., Shi, F., Zhang, M., Xu, C., & Zheng, J. A Vulnerability Risk Assessment Method Based on Heterogeneous Information Network // IEEE Access. -2020. -Vol. 8. *-*P*.*148315–148330. [DOI 10.1109/ACCESS.2020.3015551](https://doi.org/10.1109/ACCESS.2020.3015551)

5.Hromova A.R., Petrosjan L.Je. Analiz ujazvimostej v sistemah bezopasnosti dannyh // Inzhenernyj vestnik Dona. -2023. -№ 6.(102)-C.67-76. [in Russian

6.Ospanova A.B., Shegetaeva A.K., Tүsіphanov A. T., Zhalgasbaev A. B., Kadrinov D. M. Prognozirovanie setevyh ujazvimostej i jeksplojtov // Mezhdunarodnaja nauchno-prakticheskaja konferencii «XVI Saginovskie chtenija. Integracija obrazovanija, nauki i proizvodstva». - Karaganda: Izd-vo KarTU im. A.Saginova, 2024. -Ch.2. -S. 287-289.[in Russian]

7.Carriegos, M., Castañeda, Á., Trobajo , M., & Zaballa , D. On Aggregation and Prediction of Cybersecurity Incident Reports //IEEE Access.-2021.-Vol.9.-P.102636-102648.

[DOI 10.1109/ACCESS.2021.3097834](https://doi.org/10.1109/ACCESS.2021.3097834)

8.Demidova, L., & Stepanov, M. Development of Multifactor Forecasting Model based on Fuzzy Time Series // 2021 International Conference on Information Technologies (InfoTech). -2021. [DOI 10.1109/InfoTech52438.2021.9548389](https://doi.org/10.1109/InfoTech52438.2021.9548389)

9.Injadat, M., Moubayed, A., & Shami, A. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach // 202032nd International Conference on Microelectronics (ICM*)*. -2020. DOI [10.1109/ICM50269.2020.9331794](https://doi.org/10.1109/ICM50269.2020.9331794)

10.Pavlov, A. Analysis of Network Interaction of Modern Exploits // Informacionnye Tehnologii. -2022. -Vol. 28(2). [DOI 10.17587/it.28.75-80](https://doi.org/10.17587/it.28.75-80)

11.Aota, М., Kanehara, H., Кubо, М., Мurata, N., Sun, B., Takahashi, Т. Automation of Vulnerability Classification from its Description using Machine Learning //2020 IEEE Symposium on Computers and Communications (ISCC). -2020.

DOI 10.1109/ISCC50000.2020.9219568

12.Shegetaeva A.K. Vsestoronnij obzor mnogofaktornogo prognozirovanija setevyh ujazvimostej

// International Scientific Symposium Karabakh and West Azerbaijan: Triumph of Victory. The 26th of October 2024. - P. 553-560. ISBN 978-625-98125-4-0.[in Russian]

***Авторлар туралы мәліметтер***

Шегетаева А. К. - докторант, Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан, e-mail: [aizhanshegetaeva@mail.ru](mailto:aizhanshegetaeva@mail.ru);

Оспанова А. Б. - ф. -м.ғ.к., Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан, e-mail: [o.ademi111@gmail.com](mailto:o.ademi111@gmail.com);

Смакова Н.С.-PhD, ассоц. профессор, Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,, Астана, Қазақстан, e-mail: [nuri\_5@mail.ru](mailto:nuri_5@mail.ru);

Рысбекқызы Б. - PhD, аға оқытушысы, Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан, e-mail: [Bakhytgulz@mail.ru](mailto:Bakhytgulz@mail.ru);

Алтынбек С. А. - PhD, ассоц. профессор, Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,, Астана, Қазақстан, e-mail: [serik\_aa@bk.ru](mailto:serik_aa@bk.ru);

Кулбаева Л.Н. - магистр, Astana IT University, Астана, Қазақстан, e-mail: [laukakn@mail.ru](mailto:laukakn@mail.ru)

***Information about authors***

Shegetayeva A. - doctoral student, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: [aizhanshegetaeva@mail.ru](mailto:aizhanshegetaeva@mail.ru);

Ospanova A. - candidate of physical and mathematical sciences, senior lecturer, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: [o.ademi111@gmail.com](mailto:o.ademi111@gmail.com);

Smakova N.- PhD, K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: [nuri\_5@mail.ru](mailto:nuri_5@mail.ru);

Rysbekkyzy B. - PhD, senior lecturer, Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan, e-mail: [Bakhytgulz@mail.ru](mailto:Bakhytgulz@mail.ru);

Altynbek S.- PhD, K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: [serik\_aa@bk.ru](mailto:serik_aa@bk.ru);

Кулбаева Л.- магистр, Astana IT University, Astana, Kazakhstan, e-mail: [laukakn@mail.ru](mailto:laukakn@mail.ru)

IRSTI 28.23.29

**OPTIMIZING PROJECT DEVELOPMENT RISKS AND MARKET VOLATILITY USING**

**DEEP LEARNING METHODS**

**A. Bolatova**[](https://orcid.org/0009-0007-8674-7306)**, U. Kylyshbek**[](https://orcid.org/0009-0008-1386-2984)**, A. Baktygaliyev**[](https://orcid.org/0000-0001-6163-4451)**, A. Kartbayev**[C:\Users\admin\Desktop\iD.png](https://orcid.org/0000-0003-0592-5865)

*Kazakh-British Technical University, Almaty, Kazakhstan*

Correspondent-author: a.kartbaev@kbtu.kz

This study addresses computer modeling challenges by focusing on the risks in IT projects, with particular emphasis on managing investment processes under conditions of uncertainty and incomplete information. The growing number of IT projects in recent years has brought new challenges to assessing and managing associated risks. As technology advances and IT initiatives expand in scale, uncertainties in investment processes have intensified, requiring more sophisticated evaluation methods. The study introduces a RIC methodology for calculating the risk function of investment projects, incorporating fluctuations in projected cash flows. Investment project development is often characterized by uncertainty and a lack of robust statistical data, necessitating advanced analytical approaches for sound decision-making. This research applies modern scientific techniques, including machine learning and convolutional neural networks, to develop an algorithm for risk assessment in investment projects. The proposed algorithm provides practical recommendations to improve the evaluation and management of investment-related risks. The findings of this study offer valuable tools for planning and risk analysis, making them applicable to various stakeholders engaged in investment activities.

**Keywords:** investment risk, IT projects, fuzzy fields, information uncertainty, Big Data, CNN models, machine learning.

**ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНУ АРҚЫЛЫ ЖОБАЛАРДЫ ДАМЫТУ ТӘУЕКЕЛДЕРІН ЖӘНЕ НАРЫҚТЫҢ ҚҰБЫЛМАЛЫЛЫҒЫН ОҢТАЙЛАНДЫРУ**

**А. Болатова, У. Кылышбек, А. Бақтығалиев, А.Картбаев**

*Қазақстан-Британ техникалық университеті, Алматы, Қазақстан,*

*e-mail: a.kartbaev@kbtu.kz*

Бұл зерттеу компьютерлік модельдеу мәселелерін шешуге арналған, әсіресе ақпараттық технологиялар (АТ) жобаларындағы тәуекелдерге, сондай-ақ белгісіздік және толық емес ақпарат жағдайында инвестициялық процестерді басқаруға ерекше назар аударады. Соңғы жылдары АТ жобаларының санының өсуі тәуекелдерді бағалау және басқару бойынша жаңа қиындықтарды тудырды. Технологиялар дамып, АТ бастамаларының ауқымы кеңейген сайын, инвестициялық процестердегі белгісіздіктер күшейіп, неғұрлым күрделі бағалау әдістерін қажет етеді. Зерттеуде инвестициялық жобалардың тәуекел функциясын есептеуге арналған RIC әдістемесі ұсынылған, ол болжанған ақша ағындарының ауытқуларын ескереді. Инвестициялық жобаларды әзірлеу жиі белгісіздікпен және сенімді статистикалық деректердің жетіспеушілігімен сипатталады, бұл негізделген шешімдер қабылдау үшін заманауи аналитикалық тәсілдерді қолдануды талап етеді. Бұл зерттеуде инвестициялық жобалардың тәуекелдерін бағалау алгоритмін әзірлеу үшін машиналық оқыту және конволюциялық нейрондық желілер сияқты заманауи ғылыми әдістер қолданылады. Ұсынылған алгоритм инвестициялық тәуекелдерді бағалау және басқаруды жақсарту бойынша практикалық ұсыныстар береді. Зерттеу нәтижелері жоспарлау және тәуекелдерді талдау үшін құнды құралдарды ұсынады және инвестициялық қызметке қатысатын түрлі мүдделі тараптарға қолданыла алады.

**Tүйін сөздер:** инвестициялық тәуекел, АТ жобалары, анық емес өрістер, ақпараттық белгісіздік, үлкен деректер, CNN үлгілері, машиналық оқыту.

**ОПТИМИЗАЦИЯ РИСКОВ РАЗРАБОТКИ ПРОЕКТА И ВОЛАТИЛЬНОСТИ РЫНКА**

**С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ**

**А. Болатова, У. Кылышбек, А. Бактыгалиев, А.Картбаев**

*Казахстанско-Британский технический университет, Алматы, Казахстан,*

*e-mail: a.kartbaev@kbtu.kz*

Данное исследование посвящено решению задач компьютерного моделирования, сосредотачивая внимание на рисках в ИТ-проектах, с особым акцентом на управление инвестиционными процессами в условиях неопределенности и неполной информации. Рост числа ИТ-проектов в последние годы привел к появлению новых вызовов, связанных с оценкой и управлением сопутствующими рисками. По мере развития технологий и расширения масштабов ИТ-инициатив неопределенность в инвестиционных процессах усиливается, требуя более сложных методов оценки. В исследовании предлагается методология RIC для расчета функции риска инвестиционных проектов с учетом колебаний прогнозируемых денежных потоков. Разработка инвестиционных проектов часто характеризуется неопределенностью и недостатком надежных статистических данных, что требует применения современных аналитических подходов для принятия обоснованных решений. В данном исследовании применяются современные научные методы, включая машинное обучение и сверточные нейронные сети, для разработки алгоритма оценки рисков в инвестиционных проектах. Предлагаемый алгоритм содержит практические рекомендации по улучшению оценки и управления инвестиционными рисками. Результаты исследования предлагают ценные инструменты для планирования и анализа рисков, которые могут быть применимы различными заинтересованными сторонами, участвующими в инвестиционной деятельности.

**Ключевые слова:** инвестиционный риск, ИТ-проекты, нечеткая логика, неопределенность, Большие данные, модели CNN, машинное обучение.

**Introduction.** In recent years, the increasing number and complexity of Information Technology (IT) projects have posed significant challenges in assessing and managing associated risks. The rapid advancements in technology and the growing scale of IT initiatives have amplified uncertainties, particularly in investment processes, requiring more sophisticated and reliable evaluation methods. In today's dynamic business environment, IT infrastructure and architecture are critical investments that command significant financial resources. These investments are crucial for constructing advanced production environments and operational IT architectures. The investments cover a wide spectrum, including projects, computers, telecommunications, and services. These are integral to enhancing a company's productivity and operational efficiency. Yet, assessing the return on these investments poses a challenge, often taking years to realize their full benefits.

Recent scholarly and professional investigations have focused on evaluating the impact of IT investments on business performance. Various approaches have been suggested to assess these investments effectively and efficiently. *McKinsey* reports that IT infrastructure has undergone significant evolution-from small setups with few servers to massive data centers with thousands of servers. This transformation is driven by the need for robust systems capable of supporting essential business functions like transaction processing, customer data management, and complex decision-making processes. Investing in IT infrastructure brings substantial benefits. Well-integrated IT systems can improve real-time data collection, support extensive analytics, and enhance market responsiveness, thereby providing a competitive edge. For example, sophisticated infrastructure allows quicker establishment of sales offices in emerging markets and improved customer support.

Despite these benefits, the high costs and long commitments required for developing IT infrastructure present considerable challenges. The complexity of managing and integrating diverse technologies demands careful planning and strategic investments. Beyond individual companies, the strategic importance of infrastructure investments impacts broader economic activities. Enhancements in digital connectivity and transport networks facilitate business operations and can significantly boost economic growth by reducing costs, improving mobility, and bolstering competitiveness.

From the perspective of Ilin et al., the development and implementation of Enterprise Architecture (EA), including its IT components, significantly enhance transparency in business operations and agility in business re-engineering. They advocate for investment and assessment models that provide numerous advantages, such as enabling integrated comparisons of the impact of adopting IT solutions versus fragmented deployment, more precise calculations of investment project costs, reduced investment cycles for both physical and IT components, and the application of international software standards like COSMIC-ISO 19761 for practical implementation [1]. As well, a research by Purwita and Subriadi shows how IT investment valuations are influenced by both tangible and intangible benefits [2].

In another study, researchers analyzed Health Information Technology (HIT) investments in relation to hospital financial outcomes using econometric and microeconomic techniques. They optimized investment distribution based on productivity associated with each input, determining the appropriate investment levels for a global portfolio to achieve a desired confidence level [3]. Ali et al. underscores the significance of Information Technology Investment Governance (ITIG) as a vital organizational competency, highlighting its role in enhancing the relationship between IT investments and business performance, based on resource-based theory [4]. Berghout's research emphasizes the benefits of developing detailed business cases for IT projects, arguing that while resource-intensive, these cases are critical for organizations engaging in unfamiliar IT projects by providing a deeper understanding of the project's business value and supporting informed decision-making [5].

Further, Chen et al. proposes the development of an effective IT investment decision model for global organizations, aiming to demonstrate the framework's utility in supporting IT investment decisions [6]. Additionally, Witra et al. study on gender and IT investment decision-making reveals that women’s risk-averse behavior has significant implications for investment efficiency, especially in complex calculations [7]. Mirza et al. research suggests that female managers enhance asset efficiency and help organizations reduce unnecessary expenditures, particularly in developing economies [8]. Shin’s research further supports the idea that female directors strengthen board monitoring, impacting decision-making processes significantly [9]. Lee's research focuses on the critical need for businesses to understand how technology investments contribute to business outcomes, rather than just the technological aspects themselves. The study links IT spending to business growth and stresses the importance of prioritizing investments that produce tangible bottom-line results, navigating the so-called IT paradox effectively [10].

**Materials and Methods.** This research aims to establish a robust quantitative model to evaluate the effectiveness of IT investments in promoting business growth, using real-world data. The core proposition is that by strategically focusing IT investments on areas that generate substantial bottom-line results, businesses can optimize their expansion and performance. To achieve this, the study introduces a novel evaluation method named "RIC" which incorporates three critical criteria:

1. **Risk Score:** This metric assesses the attractiveness of an investment based on its risk level. The score ranges from 0% (indicating very high risk) to 100% (indicating very low risk), providing a straightforward, quantifiable measure to gauge potential risk associated with each IT investment.
2. **Return on Investment (ROI):** ROI is used to measure the profitability of an investment by comparing the return or profit generated against the initial investment cost. This criterion is essential for assessing the potential financial gains or losses from IT investments and supports decision-making by highlighting the efficiency and effectiveness of the expenditure.
3. **Criteria yet unspecified:** The third criterion, which will be detailed further in the study, complements the Risk Score and ROI to provide a comprehensive view of the investment’s value.

The RIC method's integration of these criteria aims to provide a multidimensional analysis of IT investments, enabling organizations to make informed, data-driven decisions that align IT spending with strategic business objectives. This approach not only seeks to clarify the direct impact of IT investments but also to guide companies in prioritizing investments that promise the most significant returns, thereby enhancing their competitive edge and operational efficiency.

This research commenced with the use of Python libraries to systematically gather and process investment-related datasets from a variety of sources, aiming to build a comprehensive foundation for the analysis. The primary dataset utilized was the "Twitter Dataset" obtained from *Kaggle*, which encompasses approximately 1.5 million tweets and 1.2 GB. These tweets were analyzed to extract insights related to financial markets and investment trends.

Additionally, it’s been integrated a substantial dataset provided by *McKinsey*, which includes detailed records of around 500 million interactions from regular users engaging with retail services. This dataset was pivotal in understanding consumer behavior and its impact on retail investment trends. The data collection also extended to financial markets, specifically through the use of data from *Yahoo Finance*. This source provided daily updates on a selected range of US-listed financial instruments. Although this dataset is comprehensive, it is important to note that it may contain some gaps, possibly due to the selective criteria used in data compilation, which could influence the availability of complete historical data.

To ensure a robust analytical framework, these datasets were meticulously cleaned and pre-processed to address inconsistencies and prepare them for integrated analysis. The aim was to correlate the risks derived from social media and consumer behavior with actual market movements, thereby providing a multi-dimensional perspective on investment strategies and market dynamics. This approach allowed us to harness big data analytics to derive actionable insights that could potentially guide investment decisions and strategy formulation.



**Fig. 1 – RIC methodology**

It has been considered quantitative methods as a cornerstone of the research, as a way to explore something better. In order to solve the investment problem, it has been created a method called RIC and uses three criteria (See Fig.1).

1. R stands for Risk (Risk Assessment)
2. I stands for Investment (Return On Investment)
3. C stands for Customer (Customer Satisfaction)

(1)

Each criterion is responsible for the main investment parameters: 1) both tangible and intangible; 2) tangible; 3) intangible, so that the method produces the best predictable result. In the realm of investment, accurately gauging the risk associated with stocks is paramount for determining their attractiveness and potential returns. The "Risk Score" is a fundamental metric developed to quantify this aspect, ranging from 0% to 100%. A score of 0% indicates a very high risk, suggesting that the investment is highly volatile or uncertain. Conversely, a score of 100% represents a very low risk, pointing to a stable and secure investment.

Investors rely on the Risk Score to make informed decisions by evaluating various factors that contribute to the risk profile of an investment. These factors include market volatility, which reflects the frequency and magnitude of price fluctuations; economic conditions, such as inflation rates, employment levels, and GDP growth, which can affect the overall investment climate; regulatory risks, involving changes in laws and regulations that could impact business operations; technological risks, particularly relevant in sectors where rapid innovation can render existing technologies obsolete; and operational risks, which encompass issues related to internal processes, systems, and people.

Understanding these risks is crucial as they directly influence the potential returns from an investment. By integrating the Risk Score into their analysis, investors can align their investment choices with their risk tolerance and investment objectives, aiming to optimize their portfolios for both risk and return. This approach to risk assessment not only aids in identifying potentially lucrative investments but also helps in mitigating potential losses, making it an indispensable tool in the financial decision-making process. ROI measures the profitability of an investment by comparing the return or profit generated to the initial investment cost. It helps assess the potential financial gains or losses associated with an investment.

(2)

Measuring customer satisfaction is a critical aspect of evaluating business performance and understanding the effectiveness of various investments. One common method to gauge this is through customer surveys or feedback mechanisms, which can collect detailed insights from customers about their experiences and satisfaction levels. These results are often quantified and reported as a percentage of satisfied customers. This percentage provides a direct indicator of how well a company is meeting customer expectations and needs. It is a valuable metric because it offers a clear, numerical benchmark that businesses can track over time and use to implement improvements. For instance, a high percentage of satisfied customers generally correlates with better customer loyalty, repeat business, and positive word-of-mouth, all of which are crucial for long-term success.

Businesses might utilize various tools for this purpose, including electronic surveys, feedback forms, social media interactions, and review platforms. By analyzing the data collected from these sources, companies can identify strengths and weaknesses in their products or services and make informed decisions to enhance customer satisfaction. This process not only helps in retaining existing customers but also attracts new ones by showcasing the company’s commitment to meeting their needs and expectations.

Further it’s been possible to conceptualize a model where the change in customer satisfaction over time is a function of various factors. Let represent the customer satisfaction level at time , measured as the percentage of satisfied customers. To model the change in satisfaction over time as a function of factors such as improvements in service quality , responsiveness to feedback , and changes in customer expectations was developed an equation that can be:

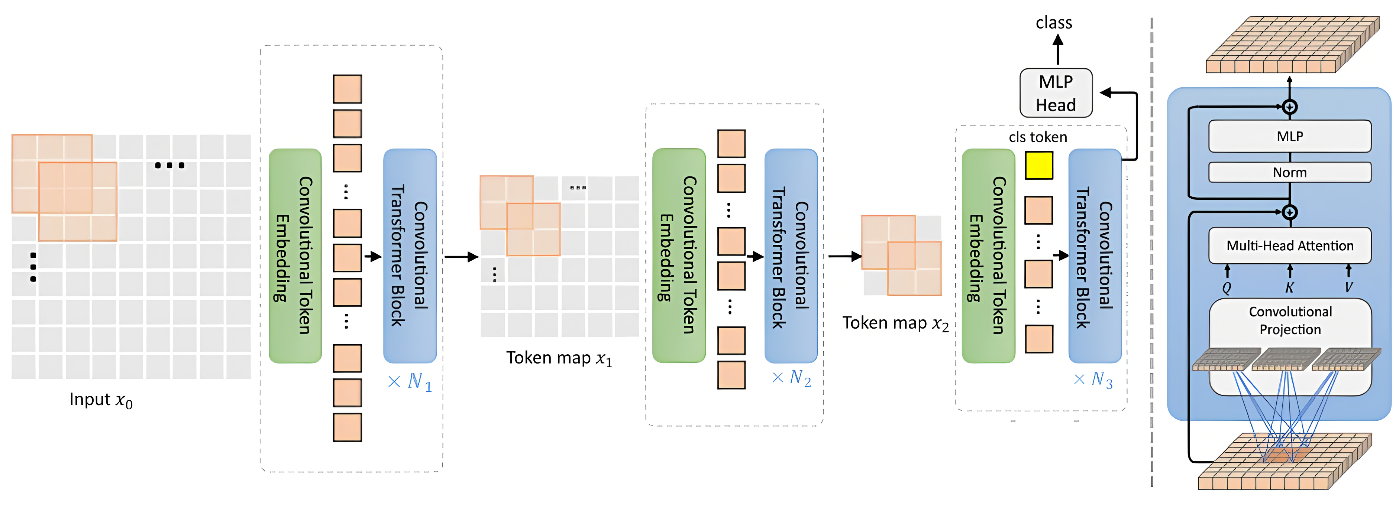
, (3)

where:

* is the rate of change of customer satisfaction.
* , and represent the rates of change in service quality, responsiveness, and customer expectations, respectively.
* , and are constants that determine the sensitivity of customer satisfaction to changes in each of these areas.

This model assumes that improvements in service quality and responsiveness directly contribute to increasing satisfaction, whereas rising customer expectations might decrease it. The constants , , and would need to be empirically determined based on data specific to a company or industry.

One of the RIC framework's key advantages is its adaptability to integrate advanced machine learning models, such as convolutional neural networks (CNNs). While it’s traditionally associated with image recognition, their utility in analyzing financial data lies in their ability to process sequential and structured datasets with remarkable precision. CNNs can be adapted to handle financial time-series data by treating the temporal progression of market events as layers of interconnected features. This approach enables the extraction of patterns and trends that might otherwise remain obscured. For example, in the context of the RIC framework, it could analyze multi-dimensional data inputs such as historical price movements, trading volume, sentiment scores, and macroeconomic indicators. The architecture’s convolutional layers can identify relationships between these variables, while pooling layers reduce dimensionality, ensuring efficient computation. Dilated convolutions could further expand the receptive field, capturing broader market contexts without increasing computational overhead. See in Fig.2 more details of the model used for risk analysis.

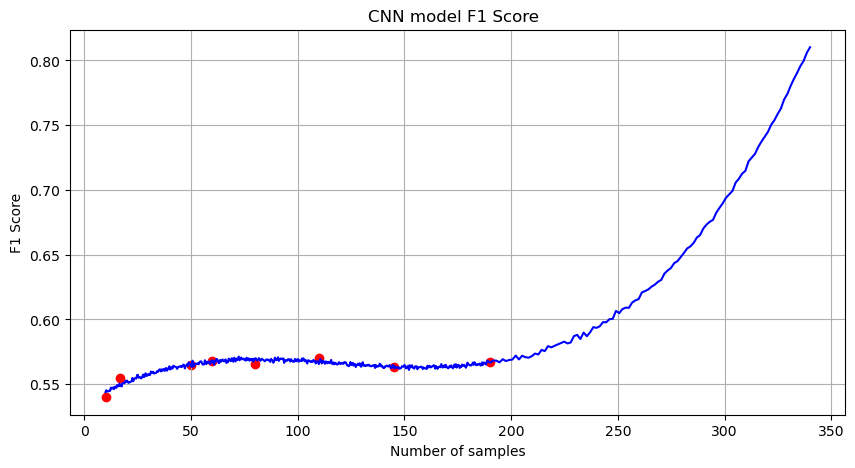


**Fig. 2 – Design of the CNN model for this study**

The CNN architecture was originally designed to address image generation problems. A key component of CNNs is the use of convolutions, which play a central role in feature extraction. Causal convolutions, in particular, are employed to preserve the temporal ordering of data, ensuring that the model’s output at timestep *t* depends only on current and past timesteps, and not on any future information. Additionally, the architecture incorporates dilated convolutions, which skip input values at regular intervals. This design enables the receptive field to expand exponentially with depth, effectively capturing broader context in the data. CNNs have demonstrated remarkable success in tasks such as music audio modeling and speech recognition. Specifically, the dilated causal convolution layers in the architecture are instrumental in capturing long-term dependencies, making them well-suited for sequential data modeling.

Due to the nature of the model, which greedily selects the highest and lowest risk points within a range, a few challenges arise in identifying risks. Firstly, selecting CNN parameters involves balancing complexity and efficiency. Typical architectures use 3–5 convolutional layers with 3×33 times 33×3 or 5×55 times 55×5 kernels, a stride of 1 or 2, and ReLU activation Pooling layers, typically max pooling with a 2×22 times 22×2 window, are used to reduce spatial dimensions while retaining important features. The number of filters often increases in deeper layers, starting from 32 or 64 in initial layers and scaling up to 256 or 512 in later layers to capture complex patterns. Learning rates between 0.001 and 0.0001, tuned via optimizers like Adam or SGD with momentum, ensure stable and effective training. These parameter choices help build a CNN that generalizes well across different datasets, as shown in Table 1.

Specifically, the sizes of buying points (peaks) and selling points (valleys) are relatively smaller compared to holding points. In the dataset, buying and selling risks account for only 3% of the total data, highlighting a significant class imbalance. To address this imbalance, it’s been proposed a sampling method that adjusts the data distribution based on the rate of rare events, ensuring a more balanced representation for effective model training, as shown in Fig.3.



**Fig. 3 – Evaluation of the CNN model after sampling**

In the Table 1, the study highlights the superior performance of the proposed CNN model, which outperforms traditional machine learning models and a previous hybrid model in terms of accuracy, precision, recall, and F1-score. The improvement in performance demonstrates the effectiveness of the updated CNN architecture, likely due to better hyper parameter tuning, deeper layers, and optimized feature extraction techniques.

**Table 1 - Comparison of CNN and machine learning models over 350 samples**

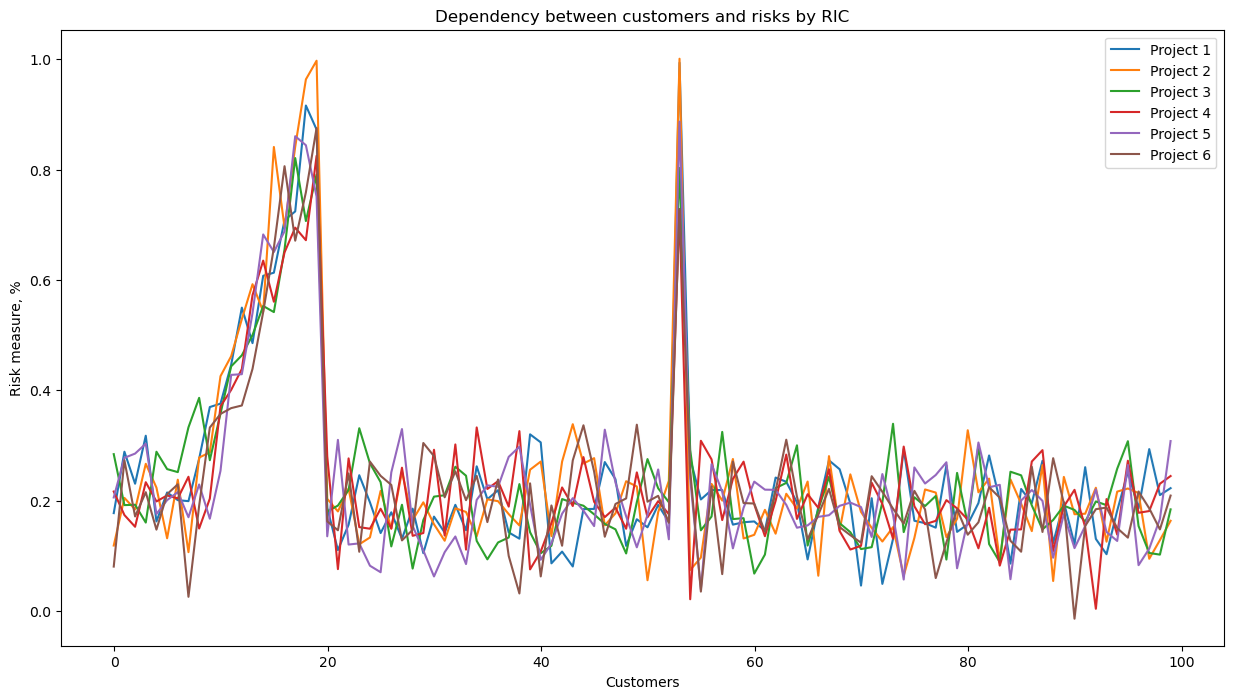
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| Proposed CNN Model | 92.5 | 93.1 | 91.8 | 92.4 |
| The Hybrid Model [14] | 88.3 | 87.9 | 88.0 | 87.9 |
| Random Forest [16] | 85.7 | 84.5 | 86.2 | 85.3 |
| Regression model [15] | 84.2 | 83.0 | 85.0 | 84.0 |
| Autoregressive (AR) model [19] | 89.1 | 88.7 | 89.5 | 89.1 |

**Results and discussion.** To validate the effectiveness of the Risk, Investment, and Compliance (RIC) method in assessing investment opportunities, the study utilized historical data from several authoritative financial analytics sources. Specifically, data spanning the past five years from Macrotrends [11], Infront Analytics [12] and Comparably [13] were employed to analyze the investment potential of ten selected companies across various industries.

This approach involved a systematic evaluation of each company's performance and market behavior using the RIC. This formula integrates multiple financial indicators to produce a composite score reflecting the investment reliability of a company. The scoring system was categorized into three distinct risk levels:

* **30% and below:** Investments in companies scoring within this range are considered high risk, and thus not recommended;
* **31% to 60%:** Companies falling within this middle range are deemed moderately safe investments;
* **Above 60%:** A score above 60% indicates a high level of investment safety and is strongly recommended for investors.

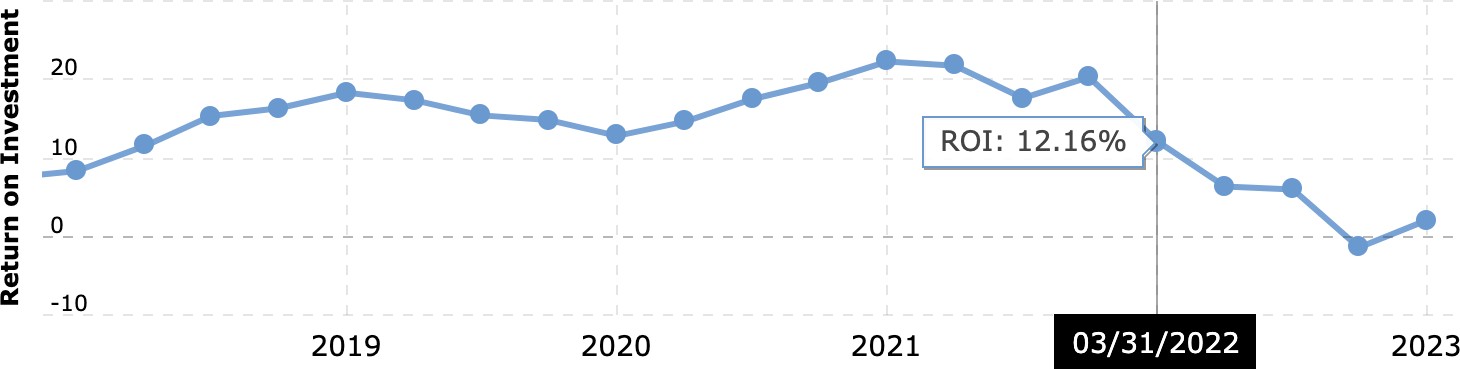
For each of the ten companies, the RIC scores were calculated based on their financial data, market trends, and other relevant economic indicators provided by the chosen data sources. This approach allowed us to map each company onto the risk assessment scale effectively, as shown in Fig.4.



**Fig. 4 – Risk measurement by customers**

The machine learning (ML) model further enhances the analysis by leveraging these RIC metrics alongside other key data points to refine investment predictions. The ML model integrates both time-series analysis and sentiment analysis to capture trends and market sentiment, offering a more comprehensive assessment of investment opportunities. By utilizing algorithms and back testing procedures, the model incorporates confidence levels and simulates real-world scenarios, effectively identifying buying and selling points to optimize investment strategies. The combination of RIC and ML-driven insights provides a robust framework for evaluating investment risks.

The application of the RIC yielded varied results across the board, reflecting a broad spectrum of investment reliability among the companies analyzed. Notably, the formula demonstrated its utility in distinguishing between high-risk and secure investment opportunities based on quantifiable metrics. Several companies scored above 60%, indicating strong financial health and market position, thus making them highly recommended for investment. Conversely, a few companies scored below 30%, suggesting significant risks that might outweigh potential returns, as shown in Fig. 5.



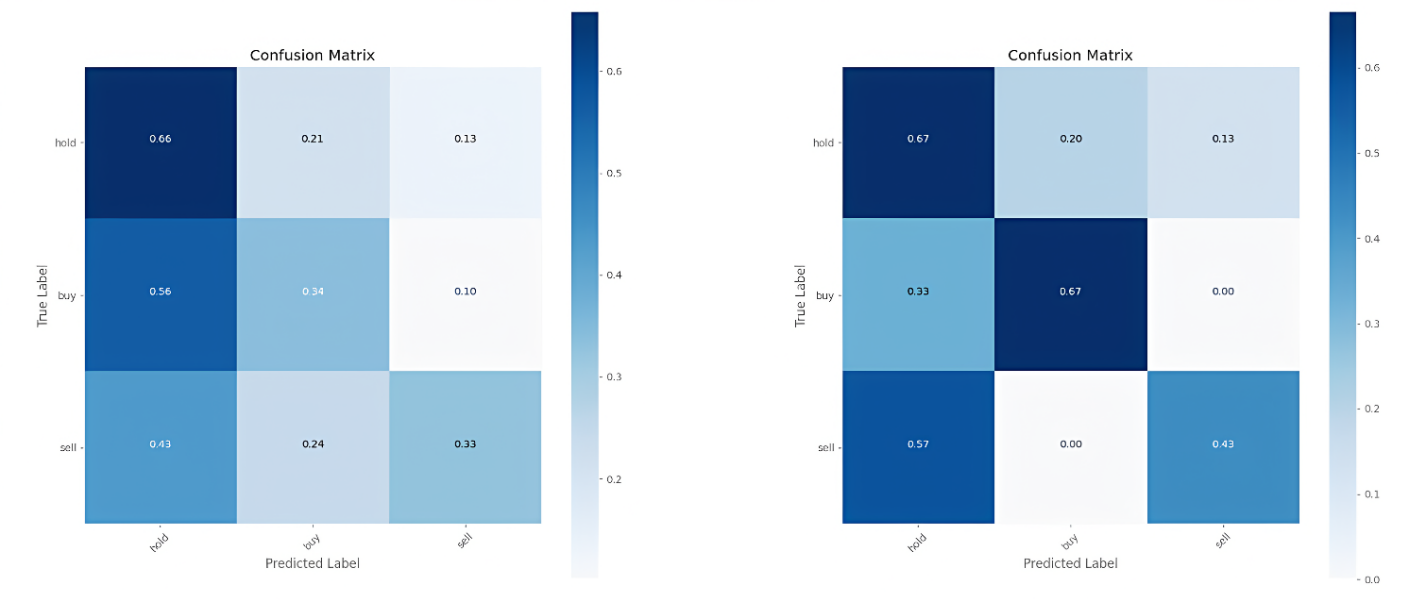
**Fig. 5 – ROI data for a 5-year period**

The results from this empirical investigation confirm that the RIC provides a robust framework for evaluating investment opportunities. By quantifying risk and correlating it with market and financial data, the formula helps investors make informed decisions grounded in comprehensive analytics [14]. Moving forward, there are obvious recommendations to refine the RIC parameters to enhance its predictive accuracy and applicability across different economic cycles and industry sectors. This ongoing validation process will ensure that the RIC remains a reliable tool for investment assessment in the dynamic global market (See Table 2). The machine learning model well complements the RIC framework by incorporating advanced analytical techniques to refine predictions and improve investment decisions. The model combines time-series analysis and sentiment analysis to identify patterns and gauge market sentiment, enriching the insights provided by the RIC.

**Table 2 - Investment assessment of the global market companies (NASDAQ)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | R (%) | I (%) | C (%) | Result |
| Amazon | 80 | 13.53 | 79 | 57.51 |
| Microsoft | 80 | 29.19 | 79 | 62.73 |
| AMD | 60 | 19.89 | 78 | 52.63 |
| Intel | 80 | 15.35 | 79 | 58.11 |
| Nokia | 80 | 2.54 | 67 | 49.85 |
| IBM | 80 | 9.13 | 68 | 52.37 |
| Netflix | 80 | 12.48 | 79 | 57.16 |
| NVIDIA | 70 | 24.73 | 85 | 59.91 |
| SAP | 90 | 8.87 | 83 | 60.62 |
| Oracle | 80 | 13.26 | 69 | 54.08 |

The machine learning module collects data from various sources to train predictive models. To preliminarily evaluate the usefulness of the data, the employment of time-series analysis and sentiment analysis provides a quick indication of whether valuable information is present. At this stage, the module is dedicated solely to Buy/Sell prediction tasks. To label the data, the authors devised a custom algorithm that greedily assigns points into three categories by identifying buying and selling points at the lowest and highest prices within a defined range. However, this labeling approach introduces significant class imbalance, leading the model to predominantly predict "hold," thereby failing to learn meaningful patterns. The results of this approach are illustrated in Fig. 6.



**Fig. 6 - Predicted investment decisions for the companies**

Given the inherent difficulty of the task, there’s been a question whether the model could effectively identify upward and downward trends necessary for accurate predictions, particularly as the confusion matrix alone does not adequately capture its performance. To address this, the study adopted a more practical evaluation metric by simulating real-world conditions-investing funds and measuring potential returns. It’s been considered to implement a back testing strategy that incorporates the model’s confidence levels.

The application of the RIC across various companies and industries has offered significant insights into the complexities of financial risk assessment. Leveraging data from sources such as Macrotrends, Infront Analytics, and Comparably, this study demonstrates the formula's ability to categorize companies into low, medium, and high investment recommendation tiers based on their scores. This stratification serves as a vital tool for investors aiming to make data-driven decisions in a competitive market environment [15]. Another key finding of this study is the RIC's reliability in delivering consistent risk assessments under diverse economic conditions [16-17]. The formula's robustness lies in its capacity to integrate immediate financial metrics with broader economic indicators, making it an adaptable framework even in fluctuating markets.

One of the primary limitations of the RIC method lies in its reliance on historical financial data and predefined risk thresholds, which may not fully capture the dynamic processes of financial markets. The fixed risk categories (30%, 60%) are based on past trends and economic theories, making them potentially outdated in rapidly evolving conditions. Additionally, while the RIC effectively quantifies investment risk using financial indicators, it lacks the ability to incorporate qualitative factors such as market innovation, or industry disruptions, elements that can impact a company's long-term performance. The RIC method operates under the assumption that past performance is indicative of future results, which may not always hold true in speculative markets [18]. While the integration of machine learning generally improves predictive capabilities, issues such as data imbalance and overfitting can limit the accuracy of uncertain forecasts.

To address these limitations, this study introduces machine learning (ML) techniques as a complementary approach to refine the framework. By integrating ML models, such as time-series forecasting and sentiment analysis, the RIC can incorporate a more dynamic and adaptive methodology. For instance, the ML models analyze historical and real-time market data to identify patterns, predict future trends, and enhance the precision of RIC scores. Furthermore, advanced ML-driven algorithms allow for the consideration of qualitative factors by processing textual and sentiment data from news, reports, and social media [19-20]. This holistic integration of quantitative and qualitative metrics improves the RIC's predictive capabilities, enabling it to provide a more comprehensive assessment of a company's investment potential.

Future research should focus on optimizing the framework by incorporating these ML advancements and exploring additional data sources. By employing techniques such as neural networks, clustering, and reinforcement learning, the RIC can evolve into a model that adapts to market changes and investor behaviors in real time [21]. This evolution will enhance the tool’s utility for investors seeking actionable insights in increasingly complex financial landscapes. The RIC method provides a foundational yet flexible framework for assessing and comparing the investment reliability of companies systematically. While this study affirms its utility and relevance, the integration of ML models and continuous validation with diverse datasets will be essential for maintaining its effectiveness. As financial markets become more interconnected and data-rich, the combination of the RIC framework with advanced ML techniques will pave the way for a new generation of investment assessment tools, capable of delivering better support.

**Conclusion.** The application of the RIC method in this study has demonstrated its efficacy in evaluating investment opportunities with above 80% risk score. By incorporating both tangible and intangible factors, the formula provides a comprehensive tool for assessing the attractiveness of investments. This practical approach is crucial for capturing the full spectrum of influences that can impact investment decisions. The findings indicate that the method enables quick and accurate determination of investment viability, supporting its use as a reliable decision-making tool in financial analysis. The success of this initial application suggests that the formula performs well across a diverse range of company types and sizes, from emerging businesses to large global corporations in the IT sector.

Given the positive results, future research is planned to refine and potentially expand the criteria used in the RIC method. Considerations such as Cost Benefit Analysis and Return on Assets could be integrated to enhance the formula's accuracy and relevance. Such developments could lead to a more precise method for assessing investments, tailored to the specific needs and contexts of various companies, thereby supporting investors in making even more informed decisions. The continual improvement and testing of the method will be essential to adapt to the evolving economic landscapes and investment scenarios, ensuring that it remains a robust tool in the arsenal of financial analysts and investors worldwide.

**References**

1.Ilin V., Levina A., Dubgorn A., Abran A. Investment Models for Enterprise Architecture and IT Architecture Projects within the Open Innovation Concept // Journal of Open Innovation: Tech, Market, and Complex. -2021. -Vol. 7. -P. 1-18. DOI [10.3390/joitmc7010069](https://doi.org/10.3390/joitmc7010069).

2.Purwita W., Subriadi A. P. Information Technology Investment: In Search of the Closest Accurate Method // Procedia Computer Science. -2019.-Vol.161.-P.300 -307.

DOI [10.1016/j.procs.2019.11.127](https://doi.org/10.1016/j.procs.2019.11.127).

3.Meyer R., Degoulet P. Choosing the Right Amount of Healthcare Information Technologies Investments // International Journal of Medical Informatics.-2010.- Vol. 79.-P. 225–231. DOI: [10.1016/j.ijmedinf.2010.01.001](https://doi.org/10.1016/j.ijmedinf.2010.01.001).

4.Ali S., Green P., Robb A. Information Technology Investment Governance: What Is It and Does It Matter? // International Journal of Accounting Information Systems. -2015. -Vol.18. -P.1-25. DOI [10.1016/j.accinf.2015.04.002](https://doi.org/10.1016/j.accinf.2015.04.002).

5.Berghout E., Tan C. W. Understanding the Impact of Business Cases on IT Investment Decisions: An Analysis of Municipal E-Government Projects // Information and Management. - 2013.- Vol.50. - P. 489-506. - DOI [10.1016/j.im.2013.07.010](https://doi.org/10.1016/j.im.2013.07.010).

6.Chen P. S., Yen D. C., Lin S. C., Chou C. S. Toward an IT Investment Decision Support Model for Global Enterprises//Computer Standards and Interfaces. -2018.-Vol. 59. -P.130-140.

DOI [10.1016/j.csi.2018.04.001](http://dx.doi.org/10.1016/j.csi.2018.04.001).

7.Witra W. P. P., Subriadi A. P. Gender and Information Technology (IT) Investment Decision-Making//Procedia Computer Science.-2021.-Vol.197.-P. 583-590.-

DOI [10.1016/j.procs.2021.12.176](https://doi.org/10.1016/j.procs.2021.12.176).

8.Mirza S. S., Majeed M. A., Ahsan T. Board Gender Diversity, Competitive Pressure and Investment Efficiency in Chinese Private Firms // Eurasian Business Review.-2020.-Vol.10(3).- P. 417-440. DOI: [10.1007/s40821-019-00138-5](https://doi.org/10.1007/s40821-019-00138-5).

9.Shin Y. Z., Chang J. Y., Jeon K. et al. Female Directors on the Board and Investment Efficiency: Evidence from Korea // Asian Business & Management.- 2020.-Vol.19.-P. 438-479.

DOI [10.1057/s41291-019-00066-2](https://doi.org/10.1057/s41291-019-00066-2).

10.Lee H., Choi H., Lee J., Min J., Lee H. Impact of IT Investment on Firm Performance Based on Technology IT Architecture // Telematics and Informatics.-2016.-Vol.91. -P. 652-661.

DOI [10.1016/j.procs.2016.07.164](https://doi.org/10.1016/j.procs.2016.07.164).

11.Macrotrends -Website with Financial Data. URL: https://www.macrotrends.net.Accessed: 18.02.2025.

12.Infront Analytics -Website with Financial Data. URL: https://infrontanalytics.com. Accessed: 18.02.2025.

13.Comparably-Website with Financial Data.URL:https://www.comparably.com/brands/ Accessed: 18.02.2025).

14.Suchkov M., Bekturgan B., Kartbayev A. Gamification Effects on Employee Engagement and Business Process Risk Evaluation // 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST).-Astana, Kazakhstan, 2024.- P. 594–599.

DOI [10.1109/SIST61555.2024.10629629](https://doi.org/10.1109/SIST61555.2024.10629629).

15.Gao Y., Zhao C., Sun B., Zhao W. Effects of Investor on Stock Volatility: New Evidences from Multi-Source Data in China's Green Stock Markets // Financial Innovation.-2022.- Vol. 8(1).- Art. 77. - P. 1-30. DOI [10.1186/s40854-022-00381-2](https://doi.org/10.1186/s40854-022-00381-2)

16.Zhang F., Ding Y., Liao Y. Financial Data Collection Based on Big Data Intelligent Processing // International Journal of Information Technology and Systems Approach. -2023.- Vol.16( 3). - P. 1–13. DOI 10.4018/IJITSA.320514.

17.Sonkavde G., Dharrao D. S., Bongale A. M., Deokate S. T., Doreswamy D., Bhat S. K. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis, and Discussion of Implications // International Journal of Financial Studies.-2023.-Vol. 11. - Art. 94. DOI [10.3390/ijfs11030094](https://doi.org/10.3390/ijfs11030094).

18.Bhowmik R., Wang S. Stock Market Volatility and Return Analysis: A Systematic Literature Review // Entropy. -2020. - Vol. 22(5). -Art. 522. DOI [10.3390/e22050522](https://doi.org/10.3390/e22050522).

19.Gong X., Wang C., Zhang W., Wang J. Investor Sentiment and Stock Volatility: New Evidence // International Review of Financial Analysis.-2022.-Vol.81.-Art.102008.

DOI:[10.1016/j.irfa.2022.102028](http://dx.doi.org/10.1016/j.irfa.2022.102028).

20.Rouatbi W., Demir E., Kizys R., Zaremba A. Immunizing Markets Against the Pandemic: COVID-19 Vaccinations and Stock Volatility Around the World // International Review of Financial Analysis. - 2021.-Vol.76.- Art. 101785. DOI [10.1016/j.irfa.2021.101819](https://doi.org/10.1016/j.irfa.2021.101819).

21.Chevallier J., Ma F., Guo Y., Huang D. Macroeconomic Attention, Economic Policy Uncertainty, and Stock Volatility Predictability // International Review of Financial Analysis. - 2022. -Vol.81. -Art. 102037. DOI [10.1016/j.irfa.2022.102339](https://doi.org/10.1016/j.irfa.2022.102339).

***Information about the authors***

Bolatova A.- master student, Kazakh-British Technical University, Almaty, Kazakhstan, aru\_bolatova@kbtu.kz;

Kylyshbek U.- master student, Kazakh-British Technical University, Almaty, Kazakhstan, u\_kylyshbek@kbtu.kz;

Baktygaliyev A.- master student, Kazakh-British Technical University, Almaty, Kazakhstan, a\_baktygaliyev@kbtu.kz;

Kartbayev A.- PhD, associate professor, Kazakh-British Technical University, Almaty, Kazakhstan, a.kartbaev@kbtu.kz

***Сведения об авторах***

Болатова А.- магистрант, Казахстанско-Британский технический университет, Алматы, Казахстан, aru\_bolatova@kbtu.kz;

Кылышбек У.- магистрант, Казахстанско-Британский технический университет, Алматы, Казахстан, u\_kylyshbek@kbtu.kz;

Бактыгалиев А.- магистрант, Казахстанско-Британский технический университет,

Алматы, Казахстан, a\_baktygaliyev@kbtu.kz;

Картбаев А. PhD, ассоциированный профессор, Казахстанско-Британский технический университет, Алматы, Казахстан, a.kartbaev@kbtu.kz