ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ПРЕЗИДЕНТІНІҢ ЖАНЫНДАҒЫ
ҰЛТТЫҚ ҒЫЛЫМ
АКАДЕМИЯСЫ

№3 (октябрь-декабрь)
2025 год

НАУКА ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Научный журнал НАН РК при Президенте РК



ҚР Президентінің жанындағы ҚР ҰҒА
«Интеллектуалды жүйелер ғылымы» ғылыми журналы

Scientific Journal of the NAS RK under the President of the RK
«Science of intelligent systems»

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ПРЕЗИДЕНТІНІҢ ЖАНЫНДАҒЫ ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

№3 (октябрь-декабрь) 2025 год

«Интеллектуалды жүйелер ғылымы» ғылыми журнал

Научный журнал «Наука интеллектуальных систем»

Scientific Journal «Science of intelligent systems»

БАС РЕДАКТОР:

Тоқбергенов И.Т.
физика-математика ғылымдарының кандидаты, ҚР Президентінің жанындағы ҚР ҰҒА Бас ғалым хатшысы

ГЛАВНЫЙ РЕДАКТОР:

Токбергенов И.Т.,
кандидат физико-математических наук, Главный ученый секретарь НАН РК при Президенте РК

CHIEF EDITOR:

I.Tokbergenov
Candidate of Physical and Mathematical Sciences, Chief Scientific Secretary of the National Academy of Science of the Republic of Kazakhstan under the President of the Republic of Kazakhstan

РЕДАКЦИЯЛЫҚ АЛҚА МҮШЕЛЕРІ:

Бакенов Ж.Б., академик (Қазақстан);

Батырбеков Э.Г., академик (Қазақстан);

Қалтаев А.Ж., академик (Қазақстан);

Локшин В.Н., академик (Казахстан);

Миталипов Ш., профессор (АҚШ);

Сұраған Дурвудхан, академик (Қазақстан);

Телтаев Б.Б., академик (Қазақстан);

Чэнь Си, профессор (ҚХР);

Шеремет И. А., академик (Ресей).

ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ:

Бакенов Ж.Б., академик (Казахстан);

Батырбеков Э.Г., академик (Казахстан);

Калтаев А. Ж., академик (Казахстан);

Локшин В.Н., академик (Казахстан);

Миталипов Ш., профессор (США);

Сураган Дурвудхан, академик (Казахстан);

Телтаев Б.Б., академик (Казахстан);

Чэнь Си, профессор (КНР);

Шеремет И.А., академик РАН (Россия).

MEMBERS OF THE EDITORIAL BOARD:

Bakenov J.B Academic (Kazakhstan);

Batyrbekov E.G., Academic (Kazakhstan);

Kaltaev A. J., Academic (Kazakhstan);

Lokshin V.N., Academic (Kazakhstan);

Mitalipov Sh., Professor (USA);

Suragan Durvudhan, Academic (Kazakhstan);

Teltaev B.B., Academic (Kazakhstan);

Chen Xi, Professor (China);

Sheremet I. A., Academic (Russia).

УЧРЕДИТЕЛЬ:
Национальная академия наук Республики Казахстан при Президенте Республики Казахстан
Издается с 2025 года
Выходит 4 раза в год, ежеквартально
Свидетельство о постановке на учет периодического печатного издания, информационного агентства и сетевого издания № KZ90VPY00113742, выданное Министерством культуры и информации от 03.03.2025 г.
Адрес редакции: 050010, г. Алматы, ул. Шевченко, 28
Тел. канцелярии: +7 (727) 308 08 17
Тел. отв. редактора: +7 705 988 19 52
Сайт журнала: https://kazscience.kz

FOUNDER:
National Academy of Sciences of the Republic of Kazakhstan under the President of the Republic of Kazakhstan
Published since 2025 year.
Issued 4 times a year, quarterly
Certificate of registration of a periodical, news agency and online publication No. KZ90VPY00113742 issued by the Ministry of Culture and Information dated 03.03.2025.
Address of editorial offices: 050010, Almaty city, Shevchenko str., 28
Office phone number: +7 (727) 308 08 17
Tel. of the responsible editor: +7 705 988 19 52
Website: https://kazscience.kz

СОДЕРЖАНИЕ

Куришбаев А.К.
Закон об искусственном интеллекте: настоящее и будущее **4**

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Д. Тан, С. Адилжанова, Г. Амирханова, А. Ануарбек.
Предприятия на основе прогнозирования погоды и машинного обучения..... **7**

В.В. Жданов.
Искусственный интеллект и машинное обучение в изучении природных опасностей..... **18**

Темір Нұрмахан.
Использование алгоритма машинного обучения для выявления слабых мест
в процессе взаимодействия банка и клиентов..... **28**

CONTENT

Kurishbaev A.K.
The Law on Artificial Intelligence: Present and Future **4**

INTELLIGENT SYSTEMS

D. Tang, S. Adilzhanova, G. Amirkhanova, A. Anuarbek.
Research on enterprise carbon emission accounting system based on weather forecasting and machine lea..... **7**

V.Zhdanov.
Artificial intelligence and machine learning in the study of natural hazards..... **18**

Temir Nurmakhan.
Using a machine learning algorithm to identify weak points in the process of interaction between the bank and customers **28**



ЗАКОН РК «ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ»: НАСТОЯЩЕЕ И БУДУЩЕЕ

Влияние искусственного интеллекта на общество никогда не проявлялось так ярко и не развивалось такими быстрыми темпами, как в последние годы. В мире уже несколько лет ведется гонка технологий искусственного интеллекта, в которой есть два явных лидера – США и Китай, в основном благодаря значительным инвестициям в исследования и разработки в области искусственного интеллекта. Казахстан недавно включился в этот процесс, когда Президент Казахстана Касым-Жомарт Токаев в своем выступлении на Международном технологическом форуме «Digital Bridge 2023» в городе Астане подчеркнул важность искусственного интеллекта для экономического роста страны.

Казахстан стал одной из первых стран, принявших Закон об искусственном интеллекте (ИИ). Первый в мире закон об искусственном интеллекте – European Union Artificial Intelligence Act – был принят Европейским Союзом и начал действовать 1 августа 2024 года.

В целом, прогресс и усовершенствование искусственного интеллекта происходит стремительными темпами и в будущем значительно повлияет на облик всего мира. Поэтому активное участие нашей страны в развитии данной сферы является не только важным, но и единственно возможным. В данной связи Президент Республики Казахстан К.К.Токаев особо подчеркнул: «Без активного внедрения ИИ Казахстан не сможет выдержать глобальную конкуренцию».

В Казахстане уже сделаны определенные шаги по использованию ИИ. Так, были представлены Концепция развития искусственного интеллекта на 2024-2029 годы; образовано Министерство искусственного интеллекта и цифрового развития, разработана Концепция развития искусственного интеллекта на 2024-2029 годы; открыт Международный центр искусственного интеллекта Alemai в Астане, объединивший образование, науку, стартапы и цифровое госуправление; запущен самый мощный в Центральной Азии национальный суперкомпьютер Alemcloud, вошедший в TOP-500 мира; принят Цифровой кодекс, объединивший регулирование данных, платформ и государственных информационных систем; ИИ интегрирован в государственные услуги (электронное правительство egov), в банки (Kaspi, Halyk и др.).

Официальное (юридическое) определение искусственного интеллекта (ИИ) теперь звучит следующим образом:

«Искусственный интеллект – функциональная способность имитации когнитивных функций, характерных для человека, обеспечивающая результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их».

Это определение закреплено в Законе Республики Казахстан «Об искусственном интеллекте», подписанном Президентом РК 17 ноября 2025 года. Закон вступит в силу 18 января 2026 года.

Закон состоит из 31 статьи. В статье 1 Закона дается перечень ключевых понятий: библиотека данных, модель ИИ, текстовый запрос, синтетические результаты деятельности систем ИИ и другие. Особое место занимает понятие «модель искусственного интеллекта» – программный продукт, который: решает специализированные интеллектуальные задачи; способен обучаться и адаптироваться к условиям; оптимизирует процессы и результаты своей работы.

За последние годы Казахстан создал одну из наиболее развитых в регионе экосистем электронного правительства, включающую цифровые услуги, национальные платформы и интегрированные системы государственного управления.

В Казахстане искусственный интеллект уже активно внедряется в различные области научных исследований, что способствует развитию инновационных технологий и повышения эффективности решений. Среди ключевых отраслей:

Медицина и здравоохранение. Использование ИИ для диагностики заболеваний, анализа медицинских изображений, разработки персонализированного лечения, предсказания эпидемий и мониторинга здоровья населения.

Аграрные науки. Применение ИИ для совершенствования сельскохозяйственных технологий, прогнозирования урожаев, оптимизации использования ресурсов и обеспечения продовольственной безопасности.

Экология и охрана окружающей среды. Использование систем ИИ для мониторинга экологической ситуации, моделирования изменений климата, оценки воздействия промышленных предприятий и автоматического обнаружения загрязнений.

Геология и недропользование. Применение аналитических алгоритмов для поиска новых месторождений, оценки запасов и автоматизации разведочных работ.

Энергетика. Внедрение ИИ для оптимизации производства, распределения и потребления энергии, а также для прогнозирования аварийных ситуаций и повышения энергоэффективности.

Информационные технологии и кибербезопасность. Разработка систем защиты информации, предотвращения кибератак, автоматизированного анализа данных и создания интеллектуальных сетей.

Космические исследования. Обработка спутниковых данных, моделирование космических процессов, развитие технологий дистанционного зондирования и изучения планеты.

Образование и наука о данных. Внедрение ИИ для персонализации образовательных программ, автоматизации оценки знаний и обработки научных данных.

Таким образом, современный этап развития общества характеризуется повышением темпов инновационного прогресса, и системы ИИ являются его основными движущими силами. Современные алгоритмы всё чаще избавляют нас от необходимости принимать решения – они выбирают маршруты, составляют тексты и даже формируют мнения. При этом важно развивать критическое мышление, учиться различать факты от сгенерированных данных и сохранять умение принимать самостоятельные решения.

Искусственный интеллект может помогать анализировать и структурировать информацию, но не заменяет человеческое суждение, мораль и эмоции. Закон прямо фиксирует, что ИИ не является субъектом права. За все решения, ошибки и последствия ответственности несут владельцы и пользователи системы. Полная ответственность за работу искусственного интеллекта в Казахстане лежит на человеке.

Приняв первый в регионе Закон «Об искусственном интеллекте», Казахстан делает важный шаг к безопасному и ответственному развитию цифровых технологий. Этот нормативный акт создает правовую основу для внедрения ИИ в экономику и государственные услуги, при этом защищая права, свободы и достоинство граждан. Человек имеет право знать, что в принятии решений участвовала система ИИ, может запросить объяснение

автоматического решения, если оно затронуло его интересы, а его данные обрабатываются только в рамках закона и с соблюдением надлежащих мер защиты.

Кроме того, Закон прямо запрещает создание и использование систем, направленных на манипуляцию поведением человека, эксплуатацию его уязвимостей, формирование социального рейтинга, определение эмоций без согласия и нелегальную сборку данных. Таким образом, государство закрывает дорогу опасным экспериментам и укрепляет доверие к цифровой среде.

Сегодня искусственный интеллект переживает настоящий технологический бум, становясь важнейшим фактором в сферах экономики, науки и политики. К 2025 году ИИ не только продемонстрировал впечатляющие результаты в решении сложных задач, но и все активнее интегрировался в повседневную жизнь, существенно повышая эффективность процессов и трансформируя рынок труда. Ресурсы ИИ уже превосходят человеческие возможности по оперативности поиска, сбора, обработки информации.

Итогом 2025 года является переход от цифровых сервисов к управляемой ИИ-экосистеме. Государство заложило правовую, вычислительную и кадровую базу, которая позволяет масштабировать искусственный интеллект как ключевой инструмент развития экономики и управления.

UDC 004.8:551.5:658.5
IRSTI 28.23.37
DOI https://doi.org/10.66941/3106-9037/2025/iss3pp7-17

D. Tang, S. Adilzhanova, G. Amirkhanova, A. Anuarbek*, 2025.
Al-Farabi Kazakh National University, Almaty, Kazakhstan.
E-mail: *aidosik165@gmail.com*

RESEARCH ON ENTERPRISE CARBON EMISSION ACCOUNTING SYSTEM BASED ON WEATHER FORECASTING AND MACHINE LEARNING

Tang Daoyun – 2 year PhD's student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: hfuttdy@163.com, https://orcid.org/ 0009-0000-7015-5276;
Saltanat Adilzhanova – PhD, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: asaltanat81@gmail.com, https://orcid.org/0000-0003-1768-064X;
Gulshat Amirkhanova - PhD, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: gulshat.aa@gmail.com, https://orcid.org/0000-0003-3933-5476;
Aidos Anuarbek – 2 year master's student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, E-mail: aidosik165@gmail.com, https://orcid.org/0009-0009-0669-1440.

Abstract. This study integrates weather forecasting and machine learning to propose a new method for corporate carbon emission accounting and management. First, set up a weather data acquisition system to obtain real-time and future 15-day weather forecast data; Then use the carbon emission calculation method to monitor and calculate the energy consumption and carbon dioxide emissions in the production process of the enterprise in real time; Then, with the help of machine learning algorithms such as neural networks, a prediction model is built based on historical data to analyze and predict the impact of different factors on carbon emissions. Finally, through data analysis and cost forecasting, we provide reliable carbon cost prediction and management services for enterprises, help enterprises adjust production plans, improve production efficiency, and formulate scientific and effective emission reduction strategies. This method significantly improves the accuracy and stability of carbon emission forecasting, and has important practical significance for enterprise carbon cost accounting and management, which can help enterprises reduce environmental risks and costs, and provide new ideas and methods for enterprise carbon emission management and sustainable development.

Keywords: weather forecasting; Machine learning; carbon emission accounting; Management; Cost forecasting.

Д. Тан, С. Адилжанова, Г. Амирханова, А. Ануарбек*, 2025.
Әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.
E-mail: *aidosik165@gmail.com*

АУА РАЙЫН БОЛЖАУ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУҒА НЕГІЗДЕЛГЕН КӘСІПОРЫННЫҢ КӨМІРТЕК ШЫҒАРЫНДЫЛАРЫН ЕСЕПКЕ АЛУ ЖҮЙЕСІН ЗЕРТТЕУ

Тан Даоюнь - 2 курс докторанты, Әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, E-mail: hfuttdy@163.com, https://orcid.org/ 0009-0000-7015-5276;
Адилжанова Салтант Альмуханбетовна – PhD, Әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, E-mail: asaltanat81@gmail.com, https://orcid.org/0000-0003-1768-064X;
Амирханова Гульшат Аманжоловна - PhD, Әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, E-mail: gulshat.aa@gmail.com, https://orcid.org/0000-0003-3933-5476;
Ануарбек Айдос Маратұлы - 2 курс магистранты, Әл – Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, E-mail: aidosik165@gmail.com, https://orcid.org/0009-0009-0669-1440.

Аннотация. Бұл зерттеуде ауа райын болжау мен машиналық оқыту біріктіріліп, корпоративтік көміртек шығарындыларын есепке алу және басқарудың жаңа тәсілі ұсынылады. Алдымен нақты уақыт режимінде және алдағы 15 тәулікке арналған ауа райы деректерін алуға мүмкіндік беретін метеодеректерді жинау жүйесі орнатылады. Кейін кәсіпорынның өндірістік үдерістеріндегі энергия тұтынуы мен көмірқышқыл газы (CO_2) шығарындылары нақты уақыт режимінде мониторинг жасалып, көміртек шығарындыларын есептеу әдістемесімен есептеледі. Бұдан соң нейрондық желілер сияқты машиналық оқыту алгоритмдерінің көмегімен тарихи деректер негізінде әртүрлі факторлардың

шығарындыларға әсерін талдап, болжайтын болжамдық модель құрылады. Соңында деректерді талдау және шығындарды болжау арқылы кәсіпорындарға көміртек құнын (carbon cost) болжау және басқару бойынша сенімді қызметтер көрсетіледі, бұл өндірістік жоспарларды түзетуге, өндіріс тиімділігін арттыруға және шығарындыларды төмендетудің ғылыми әрі тиімді стратегияларын әзірлеуге жәрдемдеседі. Ұсынылған әдіс көміртек шығарындыларын болжаудың дәлдігі мен тұрақтылығын айтарлықтай арттырады және кәсіпорынның көміртек құнын есепке алу мен басқару үшін маңызды практикалық мәнге ие: ол экологиялық тәуекелдер мен шығындарды азайтуға көмектеседі және көміртек шығарындыларын басқару мен орнықты дамуға жаңа идеялар мен тәсілдер ұсынады.

Түйін сөздер: ауа райын болжау; машиналық оқыту; көміртек шығарындыларын есепке алу; басқару; шығындарды болжау.

Д. Тан, С. Адилжанова, Г. Амирханова, А. Ануарбек*, 2025.
Казахский национальный университет имени аль – Фараби, Алматы, Казахстан.
E-mail: *aidosik165@gmail.com*

ИССЛЕДОВАНИЕ СИСТЕМЫ УЧЕТА УГЛЕРОДНЫХ ВЫБРОСОВ ПРЕДПРИЯТИЯ НА ОСНОВЕ ПРОГНОЗИРОВАНИЯ ПОГОДЫ И МАШИННОГО ОБУЧЕНИЯ

Тан Даоюнь – докторант 2 курса, Казахский национальный университет имени аль – Фараби, Алматы, Казахстан, E-mail: hfuttdy@163.com, https://orcid.org/ 0009-0000-7015-5276;
Адилжанова Салтант Альмуханбетовна – PhD, Казахский национальный университет имени аль – Фараби, Алматы, Казахстан, E-mail: asaltanat81@gmail.com, https://orcid.org/0000-0003-1768-064X;
Амирханова Гульшат Аманжоловна - PhD, Казахский национальный университет имени аль – Фараби, Алматы, Казахстан, E-mail: gulshat.aa@gmail.com, https://orcid.org/0000-0003-3933-5476;
Ануарбек Айдос Маратулы – магистрант 2 курса, Казахский национальный университет имени аль – Фараби, Алматы, Казахстан, E-mail: aidosik165@gmail.com, https://orcid.org/0009-0009-0669-1440.

Аннотация. В данном исследовании интегрируются прогнозирование погоды и методы машинного обучения для предложения нового подхода к корпоративному учёту и управлению углеродными выбросами. Сначала развёртывается система получения метеоданных для сбора погодной информации в реальном времени и 15-дневного прогноза. Затем с использованием методики расчёта углеродных выбросов в реальном времени осуществляются мониторинг и вычисление энергопотребления и выбросов диоксида углерода (CO_2) в производственных процессах предприятия. Далее при помощи алгоритмов машинного обучения, таких как нейронные сети, на основе исторических данных строится прогностическая модель для анализа и прогнозирования влияния различных факторов на объём выбросов. Наконец, с опорой на анализ данных и прогнозирование затрат предоставляются надёжные услуги по прогнозированию и управлению углеродными затратами, что помогает предприятиям корректировать производственные планы, повышать эффективность и разрабатывать научно обоснованные и эффективные стратегии снижения выбросов. Предложенный метод существенно повышает точность и устойчивость прогнозирования углеродных выбросов и имеет важное практическое значение для учёта и управления углеродными затратами предприятия: он помогает снижать экологические риски и издержки, а также предлагает новые идеи и методы управления углеродными выбросами и устойчивого развития.

Ключевые слова: прогнозирование погоды; машинное обучение; учёт углеродных выбросов; управление; прогнозирование затрат.

Introduction. The impact of global warming is huge, and carbon emissions are one of the main causes. In order to achieve sustainable development, many countries have set carbon emission caps to encourage enterprises to reduce emissions. Corporate carbon emissions are related to energy consumption and type, and accurate calculation is of great significance for formulating emission reduction strategies. However, the traditional carbon emission calculation method has a large investment in manpower and material resources and is not accurate, which limits the implementation of emission reduction strategies.

Weather forecasting realizes weather prediction with the help of meteorological and geological information, which has been integrated into daily life, and the accuracy is constantly improving. Machine learning is increasingly widely used to mine the value of data and predict trends using computer algorithms and statistical theories.

Climate change is a serious global problem, and human activities are the main triggers, and reducing carbon emissions has become an international consensus. As a major energy consumer and greenhouse gas emitter, enterprises have made outstanding contributions to carbon emissions in large industrial countries. Therefore, it is important to scientifically calculate and manage the carbon emissions of enterprises, which is related to ecological and socio-economic development. At present, various fields are actively exploring the path of low-carbon economy, and the problem of corporate carbon emissions needs to be solved urgently.

The research on corporate carbon emission management at home and abroad focuses on carbon footprint calculation and management, the establishment and promotion of carbon trading market, and the openness and transparency of carbon emission data (Yang, et al, 2024). However, there are many problems with traditional methods, such as high energy consumption of carbon footprint calculation, difficulty in unified management of carbon trading markets, and easy concealment and falsification of carbon emission data.

At present, the measurement, management and reduction of corporate carbon emissions mainly use statistical algorithm-based methods, the former relies on a large amount of historical data and is susceptible to external interference, and model-based methods, the latter has high data requirements. The application of weather forecasting and machine learning technologies is expected to make up for the shortcomings of these two methods and improve the accuracy of measurement and management.

Theorical Framework

The traditional carbon emission accounting methods mainly include the combustion method, which estimates carbon dioxide emissions based on energy consumption, and the process analysis method, which calculates carbon dioxide emissions according to energy consumption, and the latter calculates according to the production process and raw material use (Wei, et al, 2024). Carbon reduction management strategies include energy substitution, process improvement, and resource optimization, etc., and can be evaluated by reducing energy consumption and optimizing production processes to reduce carbon emissions, and their effects can be evaluated with the help of an indicator system.

Weather forecast data is available through weather forecasting websites or API interfaces, covering key meteorological parameters, processed and cleaned for carbon accounting. Weather factors have a significant impact on corporate carbon emissions, such as temperature affecting energy consumption and humidity affecting chemical reactions. Machine learning algorithms, such as decision trees and support vector machines, are widely used in carbon emission management, which can build models to predict carbon emissions, and can also provide optimization and decision support for carbon emission reduction strategies, helping enterprises determine the best emission reduction plan. Combining weather forecast data with traditional accounting methods and using machine learning algorithms can more accurately predict carbon emissions and formulate effective emission reduction strategies for enterprises.

Carbon emission monitoring collects data such as energy consumption and emission sources with the help of sensors and monitoring equipment, and then analyzes, processes and interprets them. Enterprises should follow the requirements of scoping, accurate data, standardized format, and improved transparency in the preparation of carbon emission reports, and comprehensive and accurate reports can help enterprises manage carbon emissions (Zheyu, et al, 2024). Using data visualization tools to display carbon emission trends and compare data from different time periods and other enterprises can assist enterprises in scientific decision-making and efficient management of carbon emissions.

Clean energy technologies are the key to reducing carbon emissions, and solar, wind, and hydro can replace traditional fossil fuels, and large-scale promotion can significantly reduce carbon emissions. Energy efficiency and energy-saving technologies are widely used in buildings, transportation, and industry to reduce energy consumption without compromising service levels by improving thermal insulation, promoting electric vehicles, and optimizing production processes. Carbon capture and carbon storage technology captures carbon dioxide through chemical absorption, physical adsorption, etc., and sequesters it through underground storage or conversion and utilization (Chen, et al, 2024). New materials and low-carbon technology innovations bring new opportunities, new battery and energy storage technologies help the development of electric vehicles, and new materials improve solar energy efficiency, providing more options for reducing carbon emissions.

Materials and methods

Driven by the "dual carbon" goal, enterprises involved in carbon quotas and carbon tariffs are facing pressure on carbon management, and it is urgent to comprehensively verify the total amount of carbon emissions and establish a carbon cost management system. The power industry accounts for a high proportion of carbon emissions, and due to the influence of new energy power generation, the carbon emission factor fluctuates violently due to meteorological factors such as sunshine and wind speed, which brings uncertainty to carbon cost management (Li, et al, 2024).

This study combines weather forecast data with electricity carbon emission factors, and uses machine learning algorithms to calculate and predict real-time power carbon emission factors in the next 15 days, providing a scientific basis for enterprise carbon cost management and emission reduction.

In order to meet the needs of enterprises, a carbon emission monitoring information system is designed, and its design ideas and framework are shown in Figure 1. The system provides a visual representation of the past, current, and 15-day daily and hourly carbon emissions of electricity consumers. Enterprises can use this data to predict and analyze carbon emission costs on the same day and in the next 15 days, and obtain reliable carbon emission forecasting and management services in combination with production processes and plans. In addition, the system can help companies adjust production plans, increase or decrease production loads, and support decision-making (Li, et al, 2024).

The long-term operation of the monitoring system will help enterprises reduce costs, improve efficiency, and formulate more scientific and effective emission reduction plans. The research results are of great significance for enterprises to achieve carbon emission reduction, improve resource utilization efficiency, and formulate sustainable development strategies.

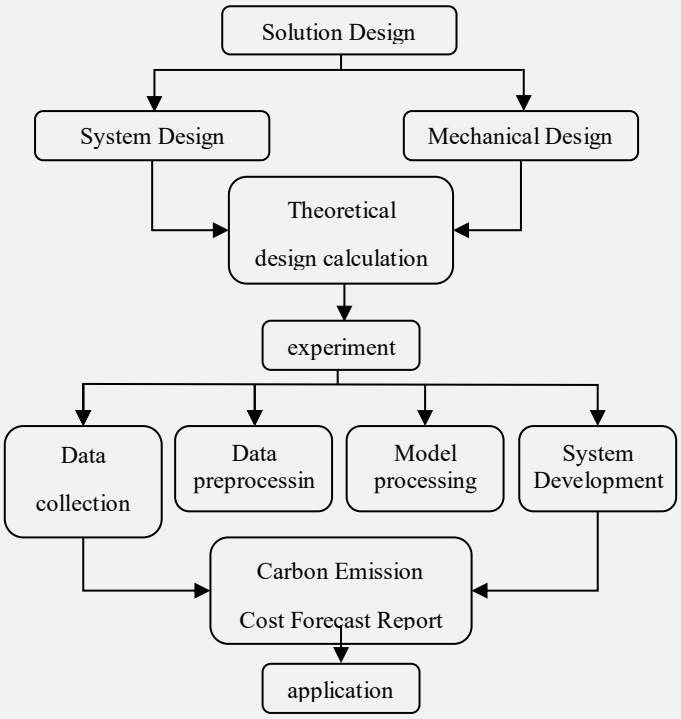


Figure 1: System design ideas and frameworks

The system collects data from multiple data sources such as meteorological bureaus, equipment sensors, and production sites, stores them in cloud databases, and cleans, converts, and standardizes pre-processing to meet the needs of subsequent model construction and analysis.

Meteorological data covers temperature, humidity, wind speed, precipitation and other parameters, analyzes their impact through historical data and machine learning algorithms, extracts key eigenvalues, and analyzes the correlation between meteorological factors and carbon emissions through statistical methods, so as to establish a prediction model suitable for enterprises to predict future meteorological conditions. Production data includes equipment usage, production process, energy consumption, pollutant emissions, etc., and analyzes these factors based on historical data and machine learning algorithms to predict future production processes and energy consumption. Real-time equipment data involves machine status, energy consumption details, flow meter and pressure gauge readings, etc., with the help of on-site sensors and equipment interfaces for real-time monitoring, combined with historical data to generate real-time operating charts and statistics.

The collected data needs to be cleaned, invalid and missing data removed, converted into a unified standard format, and then standardized to conform to specific distribution rules, such as normal distribution. This article uses daily weather forecast data, which can be obtained from meteorological websites or meteorological bureaus, where temperature, humidity, wind speed, rainfall and other information are essential for enterprise carbon emission accounting and management (Li, et al, 2024).

Corporate carbon emissions are generated by a combination of factors such as energy consumption, transportation and logistics, waste disposal, and chemical use. By measuring and tracking the carbon emissions of each link of the company's production activities, it identifies the peak period and emission sources of carbon emissions, and provides a scientific basis for enterprises to formulate emission reduction measures.

Based on the carbon emission factors that have been analyzed, we can build a carbon emission model based on weather forecasting and machine learning techniques (Mao, et al, 2024). The model can input weather forecast data and carbon emission factors in the production activities of the enterprise, and process and analyze the data in the model, and finally output the carbon emissions of the enterprise. The specific design steps of the model are as follows:

First, we need to preprocess weather forecast data and carbon emission factors. For weather forecast data, it is first necessary to standardize its format, such as unifying the date and time format. For carbon emission factors, we can use a standardized approach, that is, the values of multiple variables are scaled to the same range (Cui, et al, 2024).

Feature engineering is the process of transforming raw data into a collection of features that a model can recognize and utilize. In this paper, we feed weather forecast data and carbon emissions into the model as features. In terms of feature selection, we can use correlation coefficient analysis, principal component analysis and other techniques to screen out features with high correlation with carbon emissions (Chao, et al, 2024).

According to the results of feature engineering, we can choose machine learning algorithms suitable for enterprise carbon emission accounting, such as linear regression, decision trees, random forests, etc., for training. During the training process, we can use techniques such as cross-validation and grid search to improve the accuracy and generalization performance of the model.

The data used is the historical data of the city-level city where the enterprise is located, and the data collection interval is 5min, in which the weather indicators include irradiance, wind speed, humidity, temperature and other indicators, which are set as input indicators, and the power carbon emission factor at the corresponding time is the output index Using 80% of the local data for the whole year of 2022 as the initial dataset, the BP neural network model was trained. The other 20% of the data is used as the test set to test the accuracy of the model. The trend of carbon emission factors in Anhui Province in 2022 is shown in Figure 2.

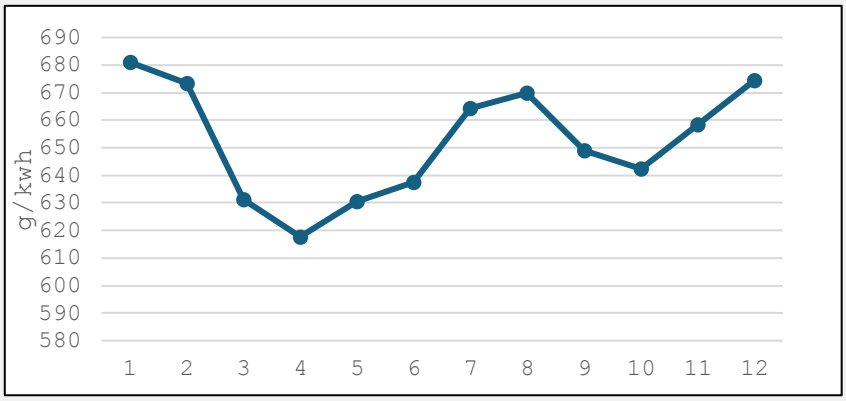


Figure 2: The trend of carbon emission factors in Anhui Province in 2022

The test dataset is fed into the trained model to evaluate the prediction performance of the model. Figure 3 shows the dynamic carbon emission factor for 24 hours in the future. The predicted dynamic carbon emission factors for the next 15 days are shown in Figure 4.

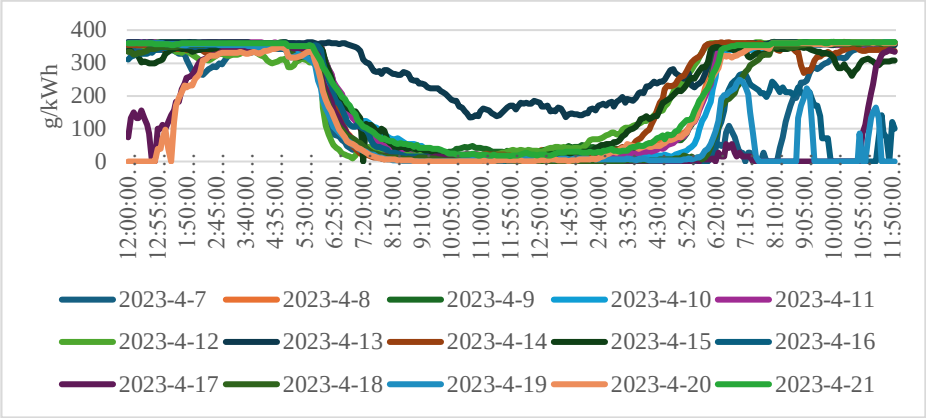


Fig. 3 The dynamic carbon emission factor for 24 hours in the future is predicted

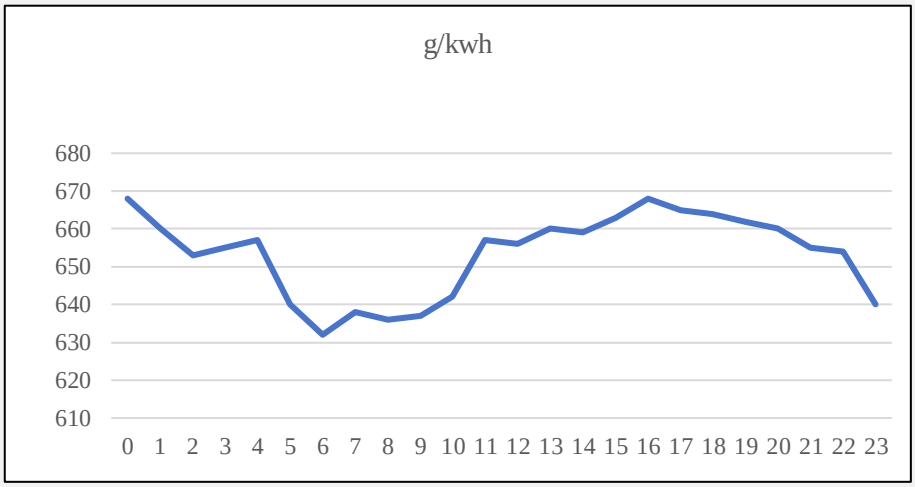


Figure 4: Predicted dynamic carbon emission factors for the next 15 days

In carbon emission accounting based on weather forecasting and machine learning, it is necessary to identify the correlation of model prediction results, identify the weather factors that mainly affect carbon emissions, and then design strategies to improve, control or reduce carbon emissions. Focus on model limitations and sources of error, such as the accuracy of weather forecast data.

This accounting method can help companies accurately measure carbon emissions, take effective control measures to reduce emissions, improve efficiency and reduce costs, and also help companies identify and track carbon footprints and set strategic goals that are in line with government regulations and market competition (Li, et al, 2024).

However, there are potential limitations to this approach. Weather forecast data may be inaccurate or updated late, resulting in increased error in the results; At the same time, it relies on large amounts of corporate energy use and weather condition data, which can affect the accuracy of the results if the data is incomplete or incorrect.

Weather forecasting and machine learning-based enterprise carbon cost forecasting needs to consider weather data acquisition, carbon emission calculation, machine learning modeling, and data analysis and cost forecasting. The working principle of the prediction system is shown in Figure 5.

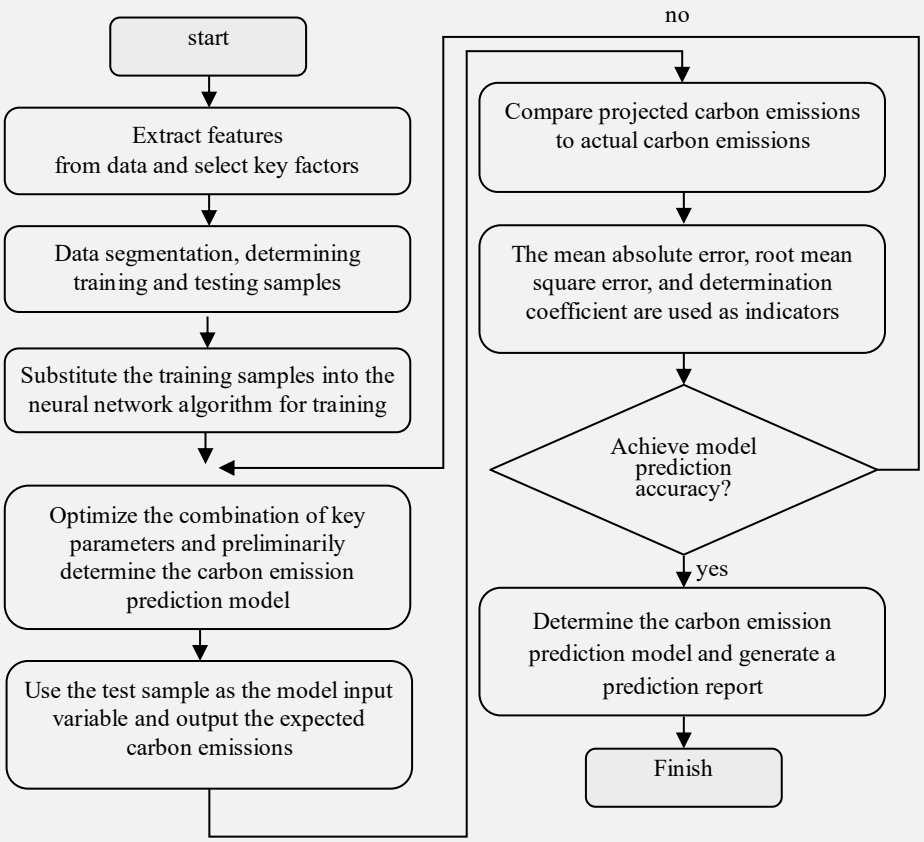


Figure 5: Schematic diagram of the working principle of the information system

Real-time weather data and weather forecast data for the next 15 days can be obtained from the Met website through the program. These data include indicators such as temperature, humidity, wind speed, and rainfall.

A company's carbon emissions can be calculated by monitoring energy consumption and CO_2 emissions from production processes. Methods such as direct monitoring of emissions from outlets, the use of fuel metering, or the use of carbon footprint tools can be employed.

Predictive models can be built based on existing historical data. In this process, the factors that affect carbon emissions need to be identified and used as input variables for the prediction model (Lou, et al, 2024). Machine learning algorithms such as BP neural networks can be used for modeling. The structure of the BP neural network is shown in Figure 6.

BP neural network is a powerful feedforward neural network model, which is efficient and accurate in achieving complex nonlinear mapping. The model trains the neural network by minimizing the error function, and uses the error backpropagation algorithm to continuously correct the weights, and solves complex problems by optimizing the network structure and model parameters.

The structure of the BP neural network includes an input layer, a hidden layer and an output layer, and the training is completed by minimizing the mean square deviation between the actual output and the desired output through gradient search technology. The BP algorithm includes forward propagation of signals and backpropagation of errors. When calculating the actual output, the signal is passed from the input layer through the hidden layer to the output layer; When correcting weights and thresholds, errors are backpropagated from the output layer to the input layer.

By combining weather forecast data and machine learning modeling, data analysis and cost prediction can be carried out to provide enterprises with reliable carbon cost prediction and management services. These information technology tools can help enterprises adjust production plans, improve production efficiency, and formulate more scientific and effective emission reduction strategies.

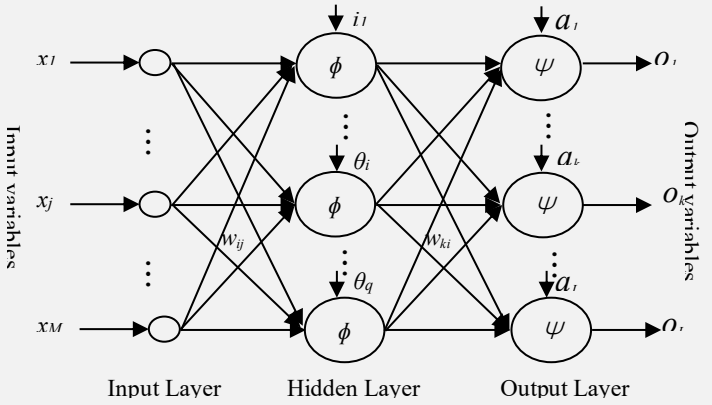


Figure 6: Schematic diagram of BP neural network structure

Where x_j represents the input of the jth node in the input layer, $j = 1, \ldots, M$; w_{ij} represents the weight between the i-th node in the hidden layer and the j-th node in the input layer; θ_i represents the threshold of the i-th node in the hidden layer; Φ represents the activation function of the hidden layer; w_{ki} represents the weight between the k-th node in the output layer and the i-th node in the hidden layer, $i=1,2,\ldots,q$; a_k represents the threshold of the k-th node in the output layer, $k=1,\ldots,L$; $\phi(x)$ represents the activation function of the output layer; O_k represents the output of the k-th node in the output layer.

Although the traditional BP neural network has good results in practical applications, there are still some limitations. The BP algorithm is easily disturbed by noise during the training process, resulting in the over-complexity of the parameters of the neural network and the over-fitting situation, which can only ensure the smallest error on the training set, and the actual error on the test set may be large, resulting in the reduction of prediction accuracy. Secondly, the BP algorithm is an optimization method based on gradient descent, which can only guarantee to find the optimal solution of the current point, and it is easy to fall into the local optimum, but cannot guarantee the global optimal solution. Figure 7 shows the BP algorithm program flow.

In order to improve the accuracy of the traditional BP neural network model, this work improves the traditional BP neural network model by increasing the adaptive adjustment learning step size and increasing the momentum term, so as to further improve the optimization ability of the neural network, so as to make more accurate predictions of power carbon emission factors in the future.

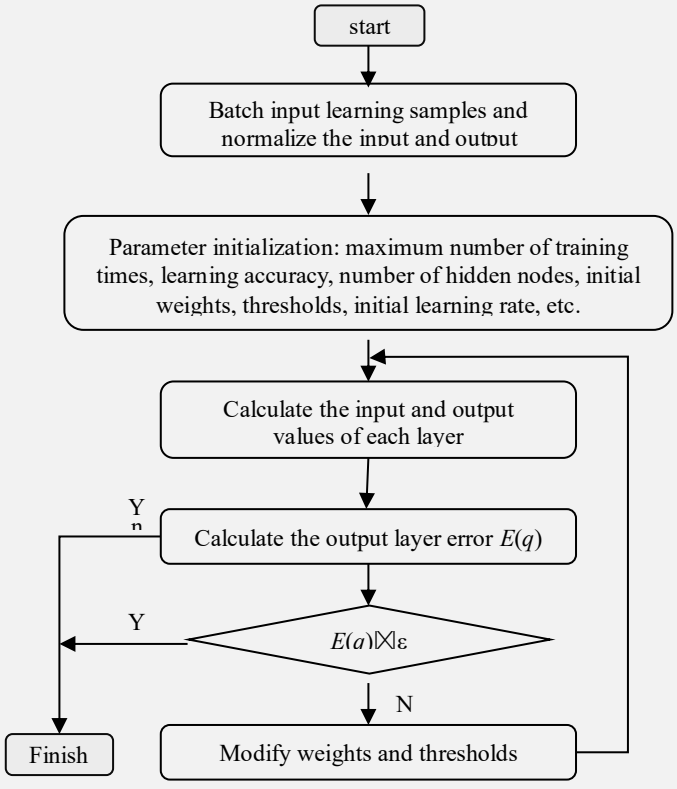


Figure 7: BP algorithm program flow chart

In order to solve the problems of low prediction accuracy and possible overfitting in the training of traditional BP neural network models, this paper proposes to increase the adaptive adjustment learning step size on the traditional model. The learning step size is a constant η, and if the value is too large or too small, it will affect the training effect of the model. When the η error surface is relatively flat, the learning step size is too small, which will cause the increase of learning times, and it should be increased η appropriately. When around the minimum, the learning step size is too large to cause oscillations, which should be reduced η. Based on this, the rules for the adaptive learning step size are set as follows. where is the $E(k)$ sum of squares of the error of the first step k.

$$\eta(k+1)=\begin{cases}1.05\eta(k) & E(k+1)<E(k)\\ 0.7\eta(k) & E(k+1)>1.04E(k)\\ \eta(k) & Other\end{cases} \quad (1)$$

The idea of overall adjustment is to increase the adaptive increase η when the learning converges to shorten the learning time. When it is too large to converge, η it should be reduced immediately η until it converges.

In order to solve the problem that the traditional BP neural network model is easy to fall into the local minimum value and cannot get the optimal value in the training, this paper proposes a method to increase the momentum term, and uses the momentum factor to transmit the last weight change. When the momentum factor is 0, the last weight is generated by the gradient descent method. When the momentum factor is 1, the new weight generated by the momentum factor method is the last weight change. Increasing the momentum method is essentially equivalent to adding a damping term in the training process, which reduces the sensitivity of the network to the local details of the error surface and slows down the oscillation trend of the learning process, so as to avoid the network falling into the local minima Improve its astringency. The modified formula for the momentum term is as follows:

$$\Delta w_{ij}(k+1)=(1-mc)\eta\delta_i p_j+mc\Delta w_{ij}(k)$$

$$\Delta b_i(k+1)=(1-mc)\eta\delta_i+mc\Delta b_i(k) \quad (2)$$

Among them, k is the number of training times, mc is the momentum factor, which is generally 0.95. By adding the momentum term, the network optimization can be helped to break away from the local minimum of the error surface, effectively preventing the situation of falling into the local optimum.

Therefore, in order to ensure the generalization ability and reliability of the model, the model is improved to a certain extent when selecting the model.

Based on the established machine learning model, combined with historical data and current weather forecasts, the future carbon cost can be predicted, and management measures can be formulated accordingly to reduce the carbon emissions and carbon costs of enterprises. With the help of real-time sensor monitoring and data analysis, the system can track the actual carbon emissions in real time, compare them with the predicted values, and identify potential problems in time and take countermeasures. Once the actual carbon emissions exceed the predicted value, the system will automatically alarm and remind the user to deal with it (Liu, et al, 2024).

The system displays data in various forms such as charts and maps to facilitate users to grasp data trends. At the same time, it supports custom reports and indicator cards to meet the needs of users to view data personally. Users can use the reporting feature to export data as a table or PDF file for in-depth analysis. The system can also provide targeted recommendations, such as optimizing production processes or equipment, to help companies reduce their carbon footprint.

In addition, the system has strict security measures, including user authentication, data encryption, etc., to ensure data security, prevent leakage and tampering, and strictly comply with relevant data protection regulations (Qiu, et al, 2024).

Results and discussion

Studies have shown that weather factors have a significant impact on carbon costs. Temperature, humidity, wind speed, etc. are closely related to carbon cost, and when the temperature increases, the humidity increases, and the wind speed increases, the carbon cost of enterprises often decreases, which is closely related to energy consumption. For example, the use of air conditioners by enterprises at high temperatures increases

carbon emissions; The use of heating equipment increases at low temperatures, and carbon emissions are reduced; The increase in wind speed promotes natural ventilation of the atmosphere, reduces the accumulation of indoor pollutants, and lowers carbon costs. Different industries respond differently to weather factors, with manufacturing and construction industries responding significantly to temperature and humidity and less to wind speed, which is related to their energy consumption patterns and utilization efficiency. The impact of weather factors on carbon costs is not direct or linear, but through various approaches such as production efficiency, employee behavior, and energy efficiency, and subsequent studies need to consider these intermediate variables to more accurately simulate and predict their impacts (Shang, et al, 2024).

In this study, a neural network-based machine learning algorithm is used to predict the carbon cost of enterprises, which outperforms the traditional linear regression method and can predict more accurately and sensitively. With the help of deep learning algorithms, the neural network learns complex patterns and relationships in historical data through multi-layer neurons, uses backpropagation algorithms to optimize parameters during training, and uses cross-validation to evaluate prediction accuracy and generalization ability during testing. The results show that the model has high accuracy and sensitivity in predicting corporate carbon costs and can be effectively applied to corporate carbon management. At the same time, the model copes with data noise, missing or outliers through data cleaning and preprocessing, and has high stability and reliability.

In practical applications, the model has significant advantages. The deep learning algorithm enables it to better capture the nonlinear relationship of data and provide more accurate carbon cost prediction. The cross-validation method ensures the stability of the prediction results of the model on different datasets. Data cleaning and preprocessing ensure model stability and reliability. However, the model also has shortcomings. On the one hand, when a large amount of historical data is required to make the amount of data insufficient or inaccurate, it will affect the prediction results. Deep learning algorithms, on the other hand, involve a large number of matrix operations and backpropagation calculations, relying on high-performance computers or cloud computing support.

Conclusion. Based on weather forecasting and machine learning techniques, this paper proposes a new enterprise carbon cost prediction and management model and verifies its practical effect through a case study. The main contributions of this paper include providing a new way of thinking and methodology, identifying the main factors affecting corporate carbon emissions, and formulating carbon management strategies. In the future, the model can be further improved and optimized to better cope with practical application scenarios, including adding more feature variables and improving prediction algorithms. In addition, the relationship between carbon costs and corporate profits can be studied, and corresponding strategies can be developed to balance the relationship between carbon emissions and corporate profits. In short, the model has a wide range of application prospects in the prediction and management of corporate carbon costs, which can help enterprises effectively control carbon emissions, reduce carbon costs, and promote the process of sustainable development.

Әдебиет

Chao Zhang, Yuxue Zhang, Qingjia Li, Rui Wang & Xiaoyu Liu. (2024) Research on the differential distribution characteristics and influencing factors of carbon emissions in town-level governance units: a case study of Guangzhou. Journal of Huazhong Agricultural University (06). – P. 75-86 (in Chinese)

Chen Nuo, Liu Lu, Yao Jiudi & Gao Pin. (2024) Construction and application of carbon emission accounting model for enterprises in paper and paper products industry. Journal of Donghua University (Natural Science Edition). – P. 1-12 (in Chinese)

Cui Qian, Zhou Zhixiang, Guan Dongjie & Xue Yuqian. (2024) Research Progress on Accounting, Simulation and Influencing Factors of Transportation Carbon Emissions. Environmental Engineering. – P. 1-12 (in Chinese)

Li Guoqing, Bai Bin, Li Guiqi, Guan Kai, Hou Jie & Lin Chunping. (2024) Research on carbon emission accounting model of open-pit mine in alpine ecologically fragile area. Chinese Journal of Engineering. – P. 1-13 (in Chinese)

Li Jie, Zhou Juan, Li Qinyuan, Li Songzhi, Ming Zhijun & Yu Xueting. (2024) Evolution and Application Analysis of Carbon Accounting Methods for Urban Ecosystems. Journal of Environmental Engineering and Technology. – P. 1-14 (in Chinese)

Li Xia & Lan Xiaoji. (2024). Analysis and multi-scenario prediction of land use carbon emissions in the urban agglomeration around Poyang Lake. Environmental Science and Technology (11). – P. 11-21 (in Chinese)

Li Xiaojiang, Lv Xiaobei, Shao Ling, Yu Miao, Xiao Yinghe, Hu Jingjing... & Shi Shuai. (2024). "Community Carbon Reduction": Measurement Method, Emission Characteristics and Realization Path. Urban Planning Journal (05). – P. 40-46 (in Chinese)

Liu Hongye, Wang Min & Wang Zhen. (2024). Research on the construction of carbon emission reduction accounting system for land reclamation project in coal mining area. China Coal (11). – P. 123-132 (in Chinese)

Lou Mingyue, Liu Guangbing, Liu Weijing, Meng Xi, Shi Mengqi & Guo Mingchen. (2024). Research on carbon emission accounting method of typical unit of sewage collection based on anaerobic carbon cycle theory. Environmental Engineering (11). – P. 61-71 (in Chinese)

Mao Geping, Meng Ting, Wang Na & Li Xiang. (2024) Village-scale carbon emissions based on land use: accounting methods and micro-empirical evidence. Chinese Journal of Eco-Agriculture. – P. 1-14 (in Chinese)

Qiu Yong, Sun Ping, Liu Xuejie, Sun Wenjun, Ma Xueyan, Shi Peipei... & Li Bing. (2024). Research on the localization of enterprise production carbon emission factors based on time series monitoring data. Environmental Pollution and Prevention (10). – P. 1396-1401 (in Chinese)

Shang Min, Liu Yi, Deng Yi, Chen Jing, Xie Na, Tan Gang... & Chen Mengjun. (2024). Construction and case study of carbon emission reduction accounting model for solid waste-derived fuels. Environmental Pollution and Prevention (10). – P. 1402-1407+1413 (in Chinese)

Wei Yuanyuan, Ding Chao & Feng Qian. (2024) Research on the influencing factors of carbon emissions during the operation and maintenance period of gray rainwater drainage system in typical towns in the Yangtze River Delta. Water conservation. – P. 1-15 (in Chinese)

Yang Shixuan, Yu Yang, Li Yang, Wu Zaoliang, Yilierjiang Alimujiang & Ge Suan. (2024) Verification of applicability and optimization path of carbon emission accounting method for peak shaving gas generator set. Chinese Journal of Environmental Engineering. – P. 1-14 (in Chinese)

Zheyu Wang, Ping Cao, Li Han & Xiangnan Wang. (2024) Temporal and spatial variation characteristics and influencing factors of land use carbon emissions at the county scale: A case study of Shaanxi Province. Journal of Xi'an University of Technology. – P. 1-11 (in Chinese)

References

Chao Zhang, Yuxue Zhang, Qingjia Li, Rui Wang & Xiaoyu Liu. (2024) Research on the differential distribution characteristics and influencing factors of carbon emissions in town-level governance units: a case study of Guangzhou. Journal of Huazhong Agricultural University (06). – P. 75-86 (in Chinese)

Chen Nuo, Liu Lu, Yao Jiudi & Gao Pin. (2024) Construction and application of carbon emission accounting model for enterprises in paper and paper products industry. Journal of Donghua University (Natural Science Edition). – P. 1-12 (in Chinese)

Cui Qian, Zhou Zhixiang, Guan Dongjie & Xue Yuqian. (2024) Research Progress on Accounting, Simulation and Influencing Factors of Transportation Carbon Emissions. Environmental Engineering. – P. 1-12 (in Chinese)

Li Guoqing, Bai Bin, Li Guiqi, Guan Kai, Hou Jie & Lin Chunping. (2024) Research on carbon emission accounting model of open-pit mine in alpine ecologically fragile area. Chinese Journal of Engineering. – P. 1-13 (in Chinese)

Li Jie, Zhou Juan, Li Qinyuan, Li Songzhi, Ming Zhijun & Yu Xueting. (2024) Evolution and Application Analysis of Carbon Accounting Methods for Urban Ecosystems. Journal of Environmental Engineering and Technology. – P. 1-14 (in Chinese)

Li Xia & Lan Xiaoji. (2024). Analysis and multi-scenario prediction of land use carbon emissions in the urban agglomeration around Poyang Lake. Environmental Science and Technology (11). – P. 11-21 (in Chinese)

Li Xiaojiang, Lv Xiaobei, Shao Ling, Yu Miao, Xiao Yinghe, Hu Jingjing... & Shi Shuai. (2024). "Community Carbon Reduction": Measurement Method, Emission Characteristics and Realization Path. Urban Planning Journal (05). – P. 40-46 (in Chinese)

Liu Hongye, Wang Min & Wang Zhen. (2024). Research on the construction of carbon emission reduction accounting system for land reclamation project in coal mining area. China Coal (11). – P. 123-132 (in Chinese)

Lou Mingyue, Liu Guangbing, Liu Weijing, Meng Xi, Shi Mengqi & Guo Mingchen. (2024). Research on carbon emission accounting method of typical unit of sewage collection based on anaerobic carbon cycle theory. Environmental Engineering (11). – P. 61-71 (in Chinese)

Mao Geping, Meng Ting, Wang Na & Li Xiang. (2024) Village-scale carbon emissions based on land use: accounting methods and micro-empirical evidence. Chinese Journal of Eco-Agriculture. – P. 1-14 (in Chinese)

Qiu Yong, Sun Ping, Liu Xuejie, Sun Wenjun, Ma Xueyan, Shi Peipei... & Li Bing. (2024). Research on the localization of enterprise production carbon emission factors based on time series monitoring data. Environmental Pollution and Prevention (10). – P. 1396-1401 (in Chinese)

Shang Min, Liu Yi, Deng Yi, Chen Jing, Xie Na, Tan Gang... & Chen Mengjun. (2024). Construction and case study of carbon emission reduction accounting model for solid waste-derived fuels. Environmental Pollution and Prevention (10). – P. 1402-1407+1413 (in Chinese)

Wei Yuanyuan, Ding Chao & Feng Qian. (2024) Research on the influencing factors of carbon emissions during the operation and maintenance period of gray rainwater drainage system in typical towns in the Yangtze River Delta. Water conservation. – P. 1-15 (in Chinese)

Yang Shixuan, Yu Yang, Li Yang, Wu Zaoliang, Yilierjiang Alimujiang & Ge Suan. (2024) Verification of applicability and optimization path of carbon emission accounting method for peak shaving gas generator set. Chinese Journal of Environmental Engineering. – P. 1-14 (in Chinese)

Zheyu Wang, Ping Cao, Li Han & Xiangnan Wang. (2024) Temporal and spatial variation characteristics and influencing factors of land use carbon emissions at the county scale: A case study of Shaanxi Province. Journal of Xi'an University of Technology. – P. 1-11 (in Chinese)

УДК 551.578.48
МРНТИ 37.29.25
DOI https://doi.org/10.66941/3106-9037/2025/iss3pp18-27

В.В. Жданов
АО «Институт географии и водной безопасности» МНВО РК, Алматы, Казахстан

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ В ИЗУЧЕНИИ ПРИРОДНЫХ ОПАСНОСТЕЙ

Жданов Виталий Владимирович, к.т.н., старший научный сотрудник АО «Институт географии и водной безопасности» МНВО РК, Алматы, Казахстан.
E-mail: Zhdanovitaliy@yandex.kz; https://orcid.org/000-0001-7678-1374

Аннотация. Статья посвящена применению искусственного интеллекта и машинного обучения при прогнозировании снежных лавин и селевых потоков. В течение многих лет в Институте географии и водной безопасности проводятся работы по оценке и прогнозу природного риска. Для этого был изучен передовой мировой опыт в этой области и принято решение применить технологии интеллектуального анализа данных. В работе использовано программное обеспечение от мировых лидеров – «StatSoft». Симулятор нейронной сети был обучен на архивных данных о погодных условиях за период 1950-2023 гг. Качество нового метода было протестировано на практике и составило 85-95 %, что является хорошим показателем на уровне мировых прогностических центров. Разработанная автоматическая экспертная система упрощает работу инженера прогнозиста и улучшает качество прогнозов стихийных явлений. А это очень важно для обеспечения безопасности населения и объектов в горах. Была разработана концепция региональной системы оценки и прогноза снежных лавин и селевых потоков. Полученный нами опыт может быть полезен в других областях науки, где необходимо обрабатывать большие массивы входящей информации.

Ключевые слова: искусственный интеллект, машинное обучение, оценка, риска прогноз, природные опасности, снежные лавины, селевые потоки

В.В. Жданов
ҚР ҰҚМ «География және су қауіпсіздігі институты» АҚ, Алматы, Қазақстан

ТАБИҒИ ҚАУІПТЕРДІ ЗЕРТТЕУДЕГІ ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ

Жданов Виталий Владимирович, т.ғ.к., аға ғылыми қызметкер, ҚР ҰҚМ «География және су қауіпсіздігі институты» АҚ, Алматы, Қазақстан.
E-mail: Zhdanovitaliy@yandex.kz; https://orcid.org/000-0001-7678-1374

Аннотация. Мақалада жасанды интеллект пен машиналық оқытуды қар көшкіні мен сел тасқындарын болжауда қолдану мәселесі қарастырылған. Көп жылдар бойы География және су қауіпсіздігі институтында табиғи қатерлерді бағалау және болжау бойынша зерттеулер жүргізіліп келеді. Осы мақсатта аталған саладағы озық әлемдік тәжірибе зерттеліп, деректерді интеллектуалды талдау технологияларын қолдану туралы шешім қабылданды. Жұмыста әлемдік көшбасшылардың бағдарламалық қамтамасыз етуі – «StatSoft» пайдаланылды. Нейрондық желі симуляторы 1950–2023 жж. аралығындағы ауа райы жағдайларының архивтік деректері негізінде оқытылды. Жаңа әдістің сапасы тәжірибеде сыналып, 85–95% аралығында нәтиже көрсетті, бұл әлемдік болжамдық орталықтар деңгейіндегі жақсы көрсеткіш болып табылады. Дамытылған автоматтандырылған эксперттік жүйе инженер-болжамшының жұмысын жеңілдетіп, стихиялық құбылыстарды болжау сапасын арттырады. Бұл өз кезегінде таулы аймақтардағы халық пен нысандардың қауіпсіздігін қамтамасыз ету үшін аса маңызды. Аймақтық қар көшкіні мен сел тасқындарын бағалау және болжаудың тұжырымдамасы әзірленді. Алынған тәжірибеміз ауқымды ақпарат ағындарын өңдеуді қажет ететін ғылымның басқа да салаларында пайдалы болуы мүмкін.

Түйін сөздер: жасанды интеллект, машиналық оқыту, бағалау, қатер болжамы, табиғи қатерлер, қар көшкіні, сел тасқындары.

V.Zhdanov
JSC Institute of Geography and Water Security, Almaty, Kazakhstan

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN THE STUDY OF NATURAL HAZARDS

Zhdanov Vitaly Vladimirovich, PhD., Senior Researcher, JSC Institute of Geography and Water Security of the Ministry of National Defense of the Republic of Kazakhstan, Almaty, Kazakhstan.
E-mail: Zhdanovitaliy@yandex.kz; https://orcid.org/000-0001-7678-1374

Abstract. The article is devoted to the application of artificial intelligence and machine learning in forecasting snow avalanches and mudflows. For many years, the Institute of Geography and Water Security has been working on assessing and forecasting natural risks. For this purpose, the best world practices in this field were studied and a decision was made to apply data mining technologies. The work used software from world leaders - StatSoft. The neural network simulator was trained on archived weather data for the period 1950-2023. The quality of the new method was tested in practice and amounted to 85-95%, which is a good indicator at the level of world forecasting centers. The developed automatic expert system simplifies the work of the forecast engineer and improves the quality of natural phenomena forecasts. And this is very important for ensuring the safety of the population and objects in the mountains. The concept of a regional system for assessing and forecasting snow avalanches and mudflows was developed. The experience we have gained can be useful in other areas of science where it is necessary to process large amounts of incoming information.

Keywords. Artificial intelligence, machine learning, assessment, risk forecast, natural hazards, snow avalanches, mudflows

Введение. Проблема снежных лавин и селевых потоков актуальна во всём мире, поскольку данные природные явления способны вызывать значительный материальный ущерб и приводить к многочисленным человеческим жертвам. В связи с этим защита от них имеет важное государственное и социальное значение. На основе своевременных предупреждений организуется работа спасательных служб, осуществляется эвакуация населения и принимаются меры по минимизации последствий чрезвычайных ситуаций (Баймолдаев, 2007:284).

В последние годы разработка новейших методов прогнозирования селевых потоков и лавин активно ведётся в ведущих научно-исследовательских центрах мира. Лидирующие позиции занимают государства, где горные районы занимают значительные территории – США, Китай, Индия, Россия, Швейцария, Япония и Южная Корея (Yang, 2024:623; Tang, 2021:51-67; Zhao, 2024:397-412; Ponzani, 2023:143-162). Применяются два подхода к прогнозу – детерминистский и стохастический. В первом используют расчетные формулы, учитывая физические и гидравлические законы возникновения потоков. Во втором применяют статистические зависимости между опасными явлениями погоды и возникновением селевого потока.

В прогнозировании лавинного и селевого риска существуют несколько проблем – редкость мониторинговой сети в горах, влияние малоизученных погодных факторов, короткий период архивных данных. Поэтому некоторые прогнозные модели имеют невысокую точность. В среднем точность предупреждений 80 % для редкого опасного процесса является успешным (Черников, 1990:198). Большинство мировых методов прогноза являются сверхкраткосрочными, относятся к классу «Нау-кастинг» прогнозов. В них учитывается статистическая зависимость между текущими гидрометеорологическими условиями и прохождением катастрофического потока (Таланов, 2001: 410). Для увеличения заблаговременности предупреждений повсеместно применяются данные численных моделей погоды. Однако с увеличением заблаговременности падает общая точность модели. Часто для моделирования потока применяются формулы гидравлики и GIS-технологии. Подобные методы позволяют оценить территории, подверженные селевому потоку, и оценить время добегания паводочной волны.

В Казахстане снежные лавины и селевые потоки широко распространены в предгорных районах на востоке, юго-востоке и юге республики. Масштабы явления могут меняться от локальных явлений до гигантских с многочисленными жертвами. Их объемы могут достигать миллионов кубомеров. Крупнейшие селевые катастрофы в 1921, 1963, 1973 г. приводили к огромным жертвам и ущербу в миллиарды долларов (Виноградов, 1976:60-72). Крупнейшая лавинная катастрофа произошла в марте 1966 г., когда был нанесен огромный ущерб и разрушены старые здания альпинистских и

туристических баз, автодорог и линий электропередач в ущельях рек Улкен и Киши Алматы, Талгар и Турген (Северский, 2006:184). Опасность этих природных явлений вызывает необходимость в разработке разных методов защиты. Долговременная постоянная защита в виде дамб и отводящих каналов стоит очень дорого, но значительно уменьшает риск для населения и объектов. Прогнозирование опасности прохождения селя или лавины стоит значительно меньше и позволяет организовать предупреждение для населения. Поэтому оно очень широко применяется в стране (Степанов, 2023:113-123).

Цель и задачи исследований. Целью выполненной работы было внедрение технологий машинного обучения в методы оценки и прогноза природного риска. Поскольку качественные прогнозы снежных лавин и селевых потоков являются важным элементом государственной системы предупреждения населения, то совершенствование методов оценки очень важны для обеспечения безопасности населения. В ходе работы выполнялись следующие задачи: изучение передового мирового опыта, сбор и анализ архивных данных, поиск подходящего программного обеспечения, оценка качества работы полученной модели.

Искусственный интеллект (ИИ) – это область современной информатики и программирования, в рамках которой разрабатываются алгоритмы, имитирующие свойства человеческого мозга, включая способность к обучению (Бишоп, 2020:960). Часто в этом контексте используется синоним – (машинное обучение). ИИ находит широкое применение в различных областях науки и техники, выполняя задачи, связанные с математическими расчётами, статистической обработкой данных, анализом текстов, изображений и других видов информации. Благодаря своей универсальности и эффективности, технологии искусственного интеллекта активно внедряются в научные исследования по всему миру.

Одним из ключевых направлений машинного обучения является «аналитика данных» (также известная как «наука о данных» или «добыча данных»). Это современный инструмент обработки и анализа больших объёмов информации, особенно актуальный в таких сферах, как климатология, археология и банковское дело, где исследователи работают с масштабными статистическими архивами (Бурова, 2020;60-72; Lutsenko, 2022:96-177).

Применение ИИ особенно важно в ситуациях, когда традиционные методы анализа дают слабые результаты из-за слабовыраженных статистических зависимостей. В таких случаях машинное обучение позволяет выявлять скрытые закономерности и делать более точные прогнозы. Кроме того, технологии искусственного интеллекта применяются для прогнозирования и оценки опасных природных процессов, таких как снежные лавины, селевые потоки, оползни и обвалы.

Существуют два способа внедрения технологий ИИ в работу ученых, анализирующих природные явления. Первый из них – это написание собственной программы нейросимулятора. Самый распространенный язык программирования – PYTON с открытым кодом и использованием бесплатных библиотек. Но этот способ требует соответствующей квалификации программиста и больших затрат времени на разработку собственной нейронной сети (ИНС). Другой способ создания ИНС – это использование готового программного обеспечения от известных производителей, таких как “StatSoft”, “MathCad”, “SPSS” (Ada, 2018: 237-263; Akgum, 2012: 23-34).

Мы применили современные ИИ-методы обработки данных для изучения природных опасностей. Это позволило выявить новые зависимости и разработать методы оценки лавинной и селевой опасности. Накопленный опыт может использоваться и в других географических исследованиях, требующих анализа больших массивов данных.

Материалы и методы исследований. Аналитика и добыча данных – это современное направление математической статистики, позволяющее решать задачи классификации, кластеризации, регрессии, анализа временных рядов и прогнозирования. Оно упрощает работу с большими массивами информации и автоматизирует вычисления статистических коэффициентов. Машинное обучение снижает необходимость ручных расчетов, однако его точность зависит от качества обучающей выборки. При отсутствии надежных статистических зависимостей или при больших погрешностях модель может выдавать ложные результаты, поэтому необходима калибровка на реальных данных. Поскольку все природные явления подчиняются законам распределения, важным этапом исследования является выбор подходящего закона, от которого зависят вычисления средних характеристик и ошибок (Чернышова, 2009:88; Аргучинцева, 2007:105).

Параметрическая статистика и непараметрическая статистика. В точных науках и фундаментальных исследованиях переменные подчиняются закону нормального распределение (Гаусса, Стьюдента). При этом среднее значение переменной (математическое ожидание) и

стандартное отклонение описывают большинство случаев. При этом отсутствуют редкие крупные события (выбросы), которые искажают величину математического ожидания. «Золотое сечение» или «Правило трех сигм» означает, что в пределах трех стандартных отклонений (σ) расположено 99 % случаев; двух (σ) – 95 % случаев. Это называется доверительным интервалом. Соответственно в точных науках используют стандартные 1 и 5 % уровни значимости – вероятности ошибок.

В природе, экономике и социологии встречаются непараметрические распределения случайных величин (Парето, Экспоненциальное, Вейбулла). Степенное распределение характеризуется резким уменьшением вероятности больших значений переменной (например, количества больших звезд или миллионеров). Величина математического ожидания сильно искажается из-за одного редкого, но крупного события (выброса). Также характерно «Золотое сечение» 80×20. Это означает, что 20 % миллионеров имеют 80 % богатств; 20 % крупных звезд имеют 80 % массы вещества галактики. Для оценки социологических и природных явлений используют непараметрические статистические величины – медиану (центральную квартиль) и межквартильный размах. Параметры, показывающие срединное значение и процентное соотношение – 25, 50, 75 % от всей выборки. Виды распределений показаны на рисунке 1.

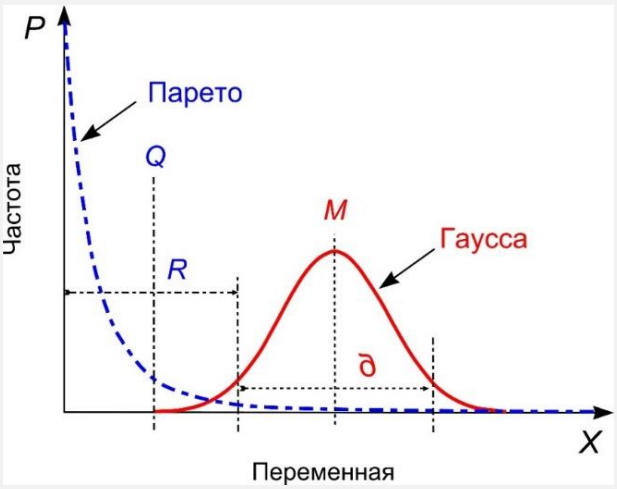


Рисунок 1 – Виды распределений случайной величины и их основные параметры.
X – изучаемая переменная; *P* – вероятность ее появления; *M* – математическое ожидание;
σ – стандартное отклонение; *Q* – квартиль; *R* – квартильный размах
Figure 2 – Types of distributions of random variables and their main parameters.
X – variable under study; P – probability of its occurrence; M – mathematical expectation;
□ – standard deviation; Q – quartile; R – quartile range

Нейросетевое программное обеспечение. Для практического применения искусственного интеллекта в математической статистике были разработаны нейросимуляторы от различных производителей программного обеспечения (Черепанов, 2013; Боровиков, 2018:354). Приведем несколько из них:

1. Надстройки для программ прикладных вычислений.

– Matlab_Neural_Network – набор нейросетевых расширений для пакета прикладных вычислений Matlab.

– Statistica_Neural_Networks – набор нейросетевых расширений для пакета прикладной статистики Statistica.

– Excel_Neural_Package – набор библиотек и скриптов для электронных таблиц Excel, реализуют некоторые возможности нейросетевой обработки данных.

2. Универсальные нейросетевые пакеты.

– NeuroSolutions – нейропакет предназначен для моделирования широкого круга искусственных нейронных сетей.

– NeuroPro – менеджер обучаемых искусственных нейронных сетей.

– NeuralWorks – нейропакет, в котором основной упор сделан на применение стандартных нейронных парадигм и алгоритмов обучения.

3. Специализированные.

–Neuroshell Trader – одна из наиболее известных программ создания нейронных сетей для анализа рынков.

–Глаз – используется для обработки аэрокосмической информации.

Нейросимулятор Ясницкого и Черепанова (версия 1,2,3,4). Был разработан авторами Ясницким Л.Н. и Черепановым Ф.М. из Пермской школы искусственного интеллекта при Пермском Государственном Гуманитарно-педагогическом университете в учебных целях и распространяется совершенно бесплатно. Однако он включает в себя современные алгоритмы обучения ИНС. Симулятор способен обучатся на вложенных архивных данных, которые можно загрузить из электронных таблиц или вбивать вручную. Также включена функция проверки полученной сети на тестовой выборке. Это делает программу универсальной и пригодной для научно-исследовательской работы. С применением нейросимулятора (версия 1,2,3,4) проводились научные исследования и опубликованы серии статей, что доказывает его практическую пользу.

Нейросетевое приложение "Statistica neural networks" от компании "StatSoft". Пакет профессиональной статистики "StatSoft" от компании "Dell". Русскоязычная версия адаптирована для рынка СНГ компанией "StatSoft Russia" и широко применяется многими научными организациями (Сайт StatSoft, 2022). Пакет представляет электронные таблицы данных и блоки статистического анализа. Программа включает интеллектуальный анализ данных использующих различные алгоритмы контролируемого и неконтролируемого обучения (Боровиков, 2018:354). Присутствуют режимы обучения, тестирования и прогнозирования с помощью полученной ИНС.

В симуляторе нейронной сети "StatSoft" реализованы следующие параметры:

– Тип сети: многослойный персептрон (MPL). Математическая модель восприятия информации, подражающая живому мозгу. Логический элемент электроники – включает серию входных, решающих и выходных элементов. При поступлении сигнала со входного элемента происходит обработка сигнала в решающем элементе. При превышении пороговых значений он активируется и передает сигнал на выходные элементы.

– Алгоритм обучения: итерационный метод численной оптимизации (BFGS), разновидность метода обратного распространения ошибки. Метод обучения персептронов. Многократное повторение для получения правильной реакции персептрона на входные сигналы. Метод BFGS названный в честь авторов в настоящее время считается очень эффективным и используется во многих симуляторах ИНС. Относится к классу методов контролируемого обучения с известным входным сигналом.

– Функция активации: Тождественная, Гиперболическая, Логистическая, Софтмакс. Функция активации или передаточная функция – это способность модели нейрона реагировать на изменение входных сигналов. Функция Софтмакс – частный случай логистической функции, эффективный для анализа многомерных входных данных.

– Эпохи обучения. Полный цикл обработки всего обучающего массива. Реализация от 1 до нескольких тысяч.

– Функция ошибки обучения: сумма квадратов и кросс-энтропия. Математический метод выражения суммарной ошибки расчетов. Сумма квадратов – известный статистический метод учета средней квадратической ошибки. Кросс-энтропия – новый метод идентификации различий для двух распределений вероятностей. Распространен в современном программировании.

– Режим хранения и прогнозирования. Позволяет сохранить готовую нейронную сеть в формате PMML и использовать для прогнозирования. Язык разметки прогнозного моделирования. Формат, используемый в совместимых программах «добычи данных» и «аналитики данных».

Данный интеллектуальный пакет программ позволяет решать большинство задач современной математической статистики и анализа больших массивов данных. Это подходит для создания собственной модели оценки и прогноза опасных природных явлений.

Результаты исследований. В течение многих лет в лаборатории природных опасностей проводились исследования, направленные на разработку и внедрение новых экспериментальных методов оценки и прогнозирования снежных лавин и селевых потоков. В ходе работы были достигнуты следующие результаты:

– Разработаны и протестированы экспериментальные методы оценки лавиноопасности и селеопасности на основе анализа метеорологических и геоморфологических факторов.

– Созданы алгоритмы прогнозирования, использующие технологии искусственного интеллекта и машинного обучения для обработки больших массивов данных.

– Разработана автоматизированная экспертная система, позволяющая классифицировать текущую ситуацию и повышающая точность прогнозов.

– Сформирована концепция региональной системы мониторинга и прогноза, обеспечивающая комплексный подход к снижению природных рисков.

– Полученные результаты рекомендованы для практического применения в горных районах, а также могут быть использованы в смежных областях науки и техники.

Создание обучающей выборки. Для оценки селевой опасности в горах Иле Алатау используются многолетние архивы Казахстанской метеорологической службы «Казгидромет». Данные наблюдений включены в международный фонд Всемирной метеорологической организации (ВМО) и находятся в свободном доступе в стандартном коде КН-01 – SYNOP (сайт погода и климат в мире 2024; архив Казгидромета, 2024). Это старейшие пункты наблюдений, которые являются репрезентативными для горных районов всего Юго-Восточного Казахстана:

– метеостанция «Мынжылки», высота 3017 м н.у.м., индекс 36889 (N43°05’06,9”; E77°04’37,5”);

– метеостанция «озеро Улкен Алматы», высота 2502 м н.у.м., индекс 36789 (N43°03’33,3”; E76°58’55,0”);

– метеостанция Шымбулак, высота 2200 м н.у.м., индекс станции 36873 (N43°07’57,3”; E77°04’39,51”).

На этих станциях проводится полный комплекс метеорологических наблюдений по стандарту ВМО. По данным метеорологических станций была создана база данных со следующими характеристиками:

– период наблюдений: 1950-2024 гг.,

– дискретность наблюдений: суточные данные метеоэлементов за селеопасный период с 1 мая по 30 сентября;

– объем выборки: 16 рядов переменных по 11100 значений в каждом;

Данные о снежных лавинах и селевых потоках собираются в многолетних архивах наблюдений Национальной метеорологичской службы, государственного селезащитного управления «Казселезащита» (селевой бюллетень, 2024; сайт МЧС, 2024). Анализировались опасные явления на реках северного склона Иле Алатау: Узынкаргалы, Узынагаш, Шамолган, Каскелен, Каргалы, Аксай, Улкен и Киши Алматы, Талгар, Есик, Турген. Данные хранятся в форматах «Microsoft Excel» и «Statistica Spreadsheet».

Все специалисты по машинному обучению отмечают важный факт – качество работы сети больше всего зависит от полноты и достоверности обучающей выборки. Были рассчитаны статистические характеристики этих рядов наблюдений за период 1950-2023 гг. Они оказались однородными и достоверными. Стандартная статистическая ошибка не превышает 10 % от среднего значения, а критерий Стьюдента является статистически значимым. Данные показаны в таблице 1.

Таблица 1 – Образец электронной базы данных в таблице Excel, для обучения искусственной нейронной сети (фрагмент)

Table 1 – Sample of an electronic database in an Excel table for training an artificial neural network (fragment)

Дата	Степень лавинной опасности	Количество осадков, мм	Высота старого снега, см	Прирост снега за снегопад, см	Максимальная температура воздуха, °С	Коэффициент устойчивости снега
10.04.2017	умеренная	0,0	65	0	10,5	0,47
11.04.2017	умеренная	25,0	64	0	10,1	0,36
12.04.2017	значительная	26,6	64	5	5,6	0,28
13.04.2017	значительная	44,0	68	0	2	0,26
14.04.2017	экстремальная	45,7	65	37	6,2	0,26
15.04.2017	высокая	45,7	75	37	15,5	0,28
16.04.2017	высокая	45,7	63	37	17	0,24
17.04.2017	умеренная	16,0	55	0	3,5	0,83
18.04.2017	значительная	0,0	45	0	9,7	1,13
19.04.2017	умеренная	0,0	43	0	12,1	0,68

Обучение компьютерной программы. Симулятор искусственного интеллекта был обучен распознаванию лавиноопасных и селеопасных ситуаций разного уровня в Иле Алатау. Эксперимент показал, что ИИ способен распознавать текущую лавиноопасную и селеопасную ситуацию с точностью 70-90 %. Наибольшая ошибка распознавания отмечалась при использовании учебного нейросимулятора Ясницкого и Черепанова при короткой обучающей выборке (бассейн Улкен Алматы за 2002-2022 гг.). Лучшие результаты показал нейросимулятор компании StatSoft при использовании длинной обучающей выборки (бассейн Киши Алматы за 1950-2022 гг.). Точность обучения зависит от совершенства алгоритмов. В учебном нейросимуляторе используется более простой и быстрый алгоритм обратного распространения ошибки, а в платной версии более совершенный, но медленный алгоритм численной оптимизации BFGS. Скорость работы сильно зависит от системных требований ЭВМ (Сайт StatSoft, 2022).

В процессе эксперимента проводилось обучение нейронных сетей с различными наборами параметров и алгоритмов обучения. В результате были выбраны варианты с наименьшими ошибками обучения. Они приведены в таблице 2. Отрабатывались два варианта статистических задач – регрессия и классификация. При обучении закладывались механизмы активации нейронов, по умолчанию рекомендованные производителем программ. В режиме классификации при одинаковой обучающей выборке ошибки оказались ниже. Для дальнейшей работы будет использоваться нейросеть, которая дает наименьшие ошибки при обучении. На ее основе создается экспертная система оценки уровня селевой опасности.

Таблица 2 – Параметры обучения нейронной сети на длинной обучающей выборке (1950-2023 гг.) для классификации потенциально селеопасных условий
Table 2 – Parameters for training a neural network on a long training sample (1950-2023) for classifying potentially mudflow-hazardous conditions

Режим классификации с применением ИНС	Архитектура ИНС	Производительность обучения	Контрольная производительность	Тестовая производительность	Количество эпох обучения	Функция ошибки	Функция активации скрытых нейронов	Функция активации выходных нейронов
Уровень селевой опасности	MLP 12-600-4	92,3	91,9	92,0	1000	Энтропия	Гиперболическая	Софтмакс
Факт прохождения потока	MLP 12-500-2	97,2	97,3	97,4	1000	Энтропия	Гиперболическая	Софтмакс
Высотная зона формирования селя	MLP 7-500-4	99,4	99,7	99,3	500	Энтропия	Гиперболическая	Софтмакс

Качество работы модели. Были проведены эксплуатационные испытания нового метода оценки и прогноза уровня лавинной и селевой опасности. Результаты приведены в таблице 3. Критерии для сравнения данных моделирования с реальными данными оказались статистически значимыми при стандартном уровне значимости 5 % и вероятности ошибки p-level менее 0,01. Для подтверждения точности модели мы использовали несколько оценок. Оправдываемость (эффективность) сверхкраткосрочного прогноза 91-94 % является хорошим показателем. Критерий Брайера для вероятностных прогнозов, также близок к 0, что является хорошим показателем (Cagnati, 1998:130-134; Schweizer, 2020:737-750).

Таблица 3 – Статистические критерии для оценки достоверности полученных моделей
Table 3 – Statistical criteria for assessing the reliability of the obtained models

	Процент точных оценок	Критерий Вилкоксона	Критерий Брайера	Корреляция Спирмена
Селевые потоки	93,6	2,2	0,026	0,74
Снежные лавины	91,2	3,65	0,09	0,70

Региональная концепция модернизации системы оценки и прогноза снежных лавин и селевых потоков должна включать следующие направления развития:

- Переход на вероятностные формы прогноза и применение 5 балльных шкал опасности с цветовым обозначением каждой градации. Это позволит оценить риск и принять соответствующие меры, адекватные прогнозируемой опасности.
- Внедрение программ интеллектуального анализа данных для оценки текущего и прогнозного уровня опасности стихийных природных явлений.
- Для получения более подробной информации применение данных автоматических метеостанций, дистанционного зондирования земли и международных погодных баз данных реанализов.
- Применение современных моделей численного прогноза погоды для увеличения заблаговременности экстренных предупреждений, что позволит дать время на принятие защитных мер.

Выводы. Методы искусственного интеллекта и машинного обучения являются хорошим инструментом при изучении опасных природных процессов. Эти технологии позволяют быстро и качественно обработать большие архивы и дают статистически значимые результаты. Современное направление математической статистики – интеллектуальный анализ данных начинает широко применяться в метеорологии, гидрологии и климатологии. Эти модели позволяют с высокой точностью классифицировать лавиноопасную и селеопасную обстановку.

Однако их эффективность зависит от алгоритмов обучения и качества выборки, поэтому очень важно оценивать качество входных архивных данных. Если многолетние архивы нерепрезентативны и недостоверны, то невозможно получить качественную модель. Поэтому достоверная обучающая выборка является важнейшим этапом при работе с симуляторами нейронной сети. Математические алгоритмы обработки данных без контрольной проверки выдают только преобразованный результат, не всегда достоверный.

Точность прогнозов опасных явлений напрямую связана с развитием численных моделей погоды, поэтому сочетание методов прогнозирования атмосферных процессов и интеллектуальной оценки природных опасностей является перспективным направлением гидрометеорологии и повышает уровень безопасности населения.

Литература

1 Ada M, San BT. (2018). Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Nat Hazards*, 90, 237–263.

2 Akgun A, и др. (2012). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci*, 38(1), 23–34.

3 Баймолдаев Т., Виноходов В.Н. (2007). *Казселезащита – оперативные меры до и после стихии*, Алматы: изд-во «Бастау», 284.

4 Бишоп Кристофер М. (2020). *Распознавание образов и машинное обучение*. Пер. с англ. Д. А. Клюшина. Москва; Санкт-Петербург: Диалектика, 960.

5 Боровиков В.П. (2018). *Популярное введение в современный анализ данных и машинное обучение на Statistica*. Москва, Изд-во «Горячая линия – Телеком», 354.

6 Бурова О. А., Судакова А. А. 2020. Искусственный интеллект – помощник статистики будущего. *Наука о данных: Материалы международной научно-практической конференции*, Санкт-Петербург, 5-7 февраля 2020 года. Санкт-Петербург: Санкт-Петербургский гос. эконом. университет, 74-76.

7 Виноградов Ю.Б. и др. (1976). Селевой поток 15 июля 1973 г. на Малой Алматинке. *КазНИГМИ. Селевые потоки*, 1, 60–72.

8 Информация о погоде и климате по данным метеостанций. URL:http://www.rp5.ru [Электрон. ресурс]. (дата обращения 10.10.2024)

9 Ежедневный селевой бюллетень. URL:https://www.kazhydromet.kz/ru/selevedenie-str/ezhednevnyy-byulleten-selevoy-opasnosti-dozhdevogo-genezisa. [Электронный ресурс]. (Дата обращения 30.09.2024).

10 Cagnati A., Valt M., Soratroi G., Gavaldà J. and SelléS C. G. (1998). A field method for avalanche danger-level verification. *Ann. Glaciology*, 26, 130-134. doi: 10.3189/1998AoG26-1-343-346.

11 Lutsenko E. V. (2022). Automated system-cognitive analysis of the level of consistency of natural numbers as systems of prime factors. E. V. Lutsenko. Text: electronic. *Polythematic Online Scientific Journal of Kuban State Agrarian University*, 179, 96-177.

12 А. В. Аргучинцева. (2007). *Методы статистической обработки и анализа гидрометеорологических наблюдений: учеб. пособие*. Иркутск: Иркут. гос. ун-т, 105.

13 Метеорологическая база данных РГП «Казгидромет». [Электрон. ресурс]. https://meteo.kazhydromet.kz/database_meteo/ (дата обращения 10.10.2024)

14 Ponziani M. и др. (2023). The use of machine learning techniques for a predictive model of debris flows triggered by short intense rainfall. *Nat. Hazards*, 117, 143–162.

15 Черников Н.Г. (1990). *Руководство селестоковым станциям и гидрографическим партиям. РД 52.30.238–89. Вып.1. Организация и проведение работ по изучению селей*. Москва: Гидрометеоиздат., 198.

16 Schweizer J., Mitterer C., Techel1 F., Stoffel1 A., and Reuter B. (2020). On the relation between avalanche occurrence and avalanche danger level. *The Cryosphere*, 14, 737–750. https://doi.org/10.5194/tc-14-737-2020.

17 Сайт российского представительства фирмы StatSoft. URL:http://www.StatSoft.ru [Электрон. ресурс]. (дата обращения 10.10.2022 г.).

18 Сайт Министерства по чрезвычайным ситуациям. URL:https://www.gov.kz/memleket/entities/emer/press/article/1?lang=en [Электрон. ресурс] (дата обращения 10.10.2024 г.).

19 Степанов Б.С. (2023). Вклад национальной гидрометеорологической службы республики Казахстан в селевую безопасность страны. *Гидрометеорология и экология*, 4, 113-123. DOI: 10.54668/2789-6323-2023-111-4-113-123.

20 Северский И., Благовещенский В. и др. (2006). *Снежный покров и лавины Тянь-Таня*. Алматы, 184.

21 Таланов Е.А. (2001). Селевые явления и зоны, подверженные их воздействию в бассейне реки Б. Алматинка. *Современные проблемы геоэкологии и созологии: Доклады Международной научно-практической конференции 22-23 января 2001*, Алматы. Под редакцией М.М. Буракова, Л.М. Павличенко, Е.А. Таланова, Алматы: изд-во Шартарап, 410.

22 Tang W. (2021). Artificial Neural Network-based prediction of glacial debris flows in the ParlungZangbo Basin, southeastern Tibetan Plateau, China. *J. Mt. Sci*, 18, 51–67.

23 Yang L. и др. (2024). Machine-Learning-Based Prediction Modeling for Debris Flow Occurrence: A Meta-Analysis. *Water*, 16, 923-943. https://doi.org/10.3390/w16070923

24 Zhao H., Wei A., Ma F. et al. (2024). Comparison of debris flow susceptibility assessment methods: support vector machine, particle swarm optimization, and feature selection techniques. *J.Mt.Sci.*, 21, 397–412. https://doi.org/10.1007/s11629-023-8395-9

25 Черепанов Ф.М. (2013). Исследовательский симулятор нейронных сетей, обзор его приложений и возможности применения для создания системы диагностики заболеваний сердечно-сосудистой системы. *Современные проблемы науки и образования*, 1; URL: www.science-education.ru/107-8392 (in Russian)

26 Чернышева Л.С., Платонова В.А. (2009). *Расчёт и интерпретация основных климатических показателей отдельных метеорологических величин*. Уч.-метод. пособие. Владивосток: Изд-во Дальневост. ун-та, 88.

References

1 Ada M, San BT. (2018). Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. *Nat Hazards*, 90, 237–263.

2 Akgun A, et al. (2012). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci*, 38(1), 23–34.

3 Baymoldayev T., Vinokhodov V.N. (2007). [*Kazselezashchita – operational measures before and after the disaster*], Almaty: Bastau Publishing House, 284.

4 Bishop Christopher M. (2020). *Pattern recognition and machine learning*. Transl. from English. D. A. Klyushina. Moscow; Saint Petersburg: Dialectika, 960.

5 Borovikov V. P. (2018). [*Popular Introduction to Modern Data Analysis and Machine Learning on Statistica*]. Moscow, Goryachaya Liniya-Telecom Publishing House, 354.

6 Burova O. A., Sudakova A. A. 2020. Artificial Intelligence – Assistant to Statistics of the Future. *Data Science: Proceedings of the International Scientific and Practical Conference*, Saint Petersburg, February 5-7, 2020. Saint Petersburg: Saint Petersburg State University of Economics, 74-76.

7 Vinogradov Yu. B. et al. (1976). Mudflow on July 15, 1973 on Malaya Almatinka. [*KazNIGMI. Mudflows*], 1, 60–72.

8 Weather and Climate Information Based on Weather Station Data. URL:http://www.rp5.ru [Electronic resource]. (Accessed 10.10.2024)

9 Daily mudflow bulletin. URL:https://www.kazhydromet.kz/ru/selevedenie-str/ezhednevnyy-byulleten-selevoy-opasnosti-dozhdevogo-genezisa. [Electronic resource]. (Accessed 30.09.2024).

10 Cagnati A., Valt M., Soratroi G., Gavaldà J. and SelléS C. G. (1998). A field method for avalanche danger-level verification. *Ann. Glaciology*, 26, 130-134. doi: 10.3189/1998AoG26-1-343-346.

11 Lutsenko E. V. (2022). Automated system-cognitive analysis of the level of consistency of natural numbers as systems of prime factors. E. V. Lutsenko. Text: electronic. [*Polythematic Online Scientific Journal of Kuban State Agrarian University*, 179, 96-177.

12 A. V. Arguchintseva. (2007). [*Methods of statistical processing and analysis of hydrometeorological observations: tutorial. manual*]. Irkutsk: Irkutsk state university, 105.

13 Meteorological database of RSE "Kazhydromet". [Electronic resource]. https://meteo.kazhydromet.kz/database_meteo/ (date of access 10.10.2024)

14 Ponziani M. et al. (2023). The use of machine learning techniques for a predictive model of debris flows triggered by short intense rainfall. *Nat. Hazards*, 117, 143–162.

15 Chernikov N. G. (1990). [*Mudflow Station and Hydrographic Party Guide. RD 52.30.238–89. Issue 1. Organization and Implementation of Mudflow Study Work*]. Moscow: Hydrometeoizdat, 198.

16 Schweizer J., Mitterer C., Techel1 F., Stoffel1 A., and Reuter B. (2020). On the Relation between Avalanche Occurrence and Avalanche Danger Level. *The Cryosphere*, 14, 737–750. https://doi.org/10.5194/tc-14-737-2020.

17 Website of the Russian Representative Office of StatSoft. URL: http://www.StatSoft.ru [Electronic resource]. (date of access 10.10.2022).

18 Website of the Ministry of Emergency Situations. URL: https://www.gov.kz/memleket/entities/emer/press/article/1?lang=en [Electronic resource] (date of access 10.10.2024).

19 Stepanov B.S. (2023). Contribution of the National Hydrometeorological Service of the Republic of Kazakhstan to the mudflow safety of the country. [*Hydrometeorology and Ecology*], 4, 113-123. DOI: 10.54668/2789-6323-2023-111-4-113-123.

20 Seversky I., Blagoveshchensky V. et al. (2006). *[Snow cover and avalanches of the Tien Tan*]. Almaty, 184.

21 Talanov E.A. (2001). Mudflow phenomena and zones affected by them in the B. Almatinka River basin. *Modern problems of geoecology and sozology: Reports of the International scientific-practical conference on January 22-23, 2001*], Almaty, Edited by M.M. Burakov, L.M. Pavlichenko, E.A. Talanov, Almaty: Shartarap Publishing House, 410.

22 Tang W. (2021). Artificial Neural Network-based prediction of glacial debris flows in the Parlung Zangbo Basin, southeastern Tibetan Plateau, China. *J. Mt. Sci*, 18, 51–67.

23 Yang L. et al. (2024). Machine-Learning-Based Prediction Modeling for Debris Flow Occurrence: A Meta-Analysis. *Water*, 16, 923-943. https://doi.org/10.3390/w16070923

24 Zhao H., Wei A., Ma F. et al. (2024). Comparison of debris flow susceptibility assessment methods: support vector machine, particle swarm optimization, and feature selection techniques. *J. Mt. Sci*., 21, 397–412. https://doi.org/10.1007/s11629-023-8395-9

25 Cherepanov F. M. (2013). Research simulator of neural networks, a review of its applications and the possibility of using it to create a diagnostic system for cardiovascular diseases. [*Modern problems of science and education*], 1; URL: www.science-education.ru/107-8392 (in Russian)

26 Chernysheva L. S., Platonova V. A. (2009). Calculation and interpretation of the main climatic indicators of individual meteorological quantities. [*Educational and methodological manual*]. Vladivostok: Publishing house of the Far Eastern University, 88.

УДК 551.578.48
DOI https://doi.org/10.66941/3106-9037/2025/iss3pp28-36

Темір Нұрмахан
Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

БАНК ПЕН КЛИЕНТТЕР АРАСЫНДАҒЫ ӨЗАРА ӘРЕКЕТТЕР ПРОЦЕСІНДЕГІ ӘЛСІЗ НҮКТЕЛЕРДІ АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМІН ҚОЛДАНУ

Темір Нұрмахан, магистрант, Ақпараттық технологиялар факультеті, Жасанды интеллект және BIG DATA кафедрасы, магистрант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.

Аннотация. Қаржы секторын цифрлық трансформациялау және міндеттемелер үшін бәсекелестіктің күшеюі жағдайында коммерциялық банктер өз маркетингтік коммуникацияларын оңтайландыру қажеттілігіне тап болады. Депозиттерді тартудың дәстүрлі моделі – жаппай телемаркетинг («суық қоңыраулар») – төмен конверсиялық көрсеткіштер мен жоғары операциялық шығындармен сипатталады, бұл бизнес-процестерде тосқауылдар тудырады. Бұл зерттеу машинамен оқыту алгоритмдеріне негізделген мұндай шектеулерді анықтау және жою әдістемесін ұсынады. Эмпирикалық деректерге (11 162 бақылау) сүйене отырып, екілік классификация модельдері (логистикалық регрессия, Random Forest, Gradient Boosting) құрылып, валидацияланды. Алдын ала аналитикаға көшу мақсатталмаған трафиктің 43%-ға дейінгі бөлігін анықтауға, клиент тарту шығынын (CAC) азайтуға және сату процесін өзгертуге мүмкіндік беретіні дәлелденді. Экономикалық тиімділік тұрғысынан Precision және Recall көрсеткіштерін түсіндіруге ерекше назар аударылады.

Түйін сөздер: банктік маркетинг, машинамен оқыту, екілік жіктеу, таршылықтар, болжамды аналитика, Градиенттік бустинг, Рэндэм Форест, бизнес-процестерді оңтайландыру, деректерді басқару.

Темір Нұрмахан
Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан

Использование алгоритма машинного обучения для выявления слабых мест в процессе взаимодействия банка и клиентов

Темир Нурмахан, магистрант, факультет информационных технологий, кафедра искусственного интеллекта и BIG DATA, магистрант, Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан.

Аннотация. В условиях цифровой трансформации финансового сектора и усиления конкуренции за обязательства коммерческие банки сталкиваются с необходимостью оптимизации своих маркетинговых коммуникаций. Традиционная модель привлечения депозитов-массовый телемаркетинг («холодные звонки») – характеризуется низкими показателями конверсии и высокими операционными затратами, что создает барьеры в бизнес-процессах. Это исследование предлагает методологию обнаружения и устранения таких ограничений на основе алгоритмов машинного обучения. На основе эмпирических данных (11 162 наблюдения) были созданы и проверены модели бинарной классификации (логистическая регрессия, Random Forest, Gradient Boosting). Доказано, что переход к предварительной аналитике позволяет выявить до 43% нецелевого трафика, снизить затраты на привлечение клиентов (CAC) и изменить процесс продаж. С точки зрения экономической эффективности особое внимание уделяется интерпретации показателей Precision и Recall.

Ключевые слова: банковский маркетинг, машинное обучение, бинарная классификация, узкие места, прогнозная аналитика, градиентный бустинг, Рэндэм Форест, оптимизация бизнес-процессов, управление данными.

Temir Nurmakhan
Al-Farabi Kazakh National University, Almaty, Kazakhstan

Using a machine learning algorithm to identify weak points in the process of interaction between the bank and customers

Temir Nurmakhan, master's Student, Faculty of Information Technologies, Department of artificial intelligence and BIG DATA, Master's student, Al-Farabi Kazakh National University, Almaty, Kazakhstan.

Abstract. In the context of the digital transformation of the financial sector and increased competition for obligations, commercial banks are faced with the need to optimize their marketing communications. The traditional model of attracting deposits – mass telemarketing ("cold calls") – is characterized by low conversion rates and high operating costs, which creates blockages in business processes. This study provides a methodology for identifying and eliminating such constraints based on machine learning algorithms. Based on empirical data (11,162 observations), binary classification models (logistic regression, Random Forest, Gradient Boosting) were created and validated. It has been proven that the transition to pre-analytics allows you to identify up to 43% of non-targeted traffic, reduce customer engagement loss (CAC), and change the sales process. Particular attention is paid to the interpretation of Precision and Recall indicators in terms of economic efficiency.

Keywords: banking marketing, machine learning, binary classification, bottlenecks, predictive analytics, gradient busting, Randham Forest, Business Process Optimization, Data Management.

1. Кіріспе. Бүгінгі экономикада деректер қаржылық институттардың бәсекеге қабілеттілігін анықтайтын негізгі активке айналуда. Екінші деңгейлі банктер тұрақты қаржыландыруды қамтамасыз ету мақсатында бөлшек салым өнімдерін белсенді түрде дамытуда. Алайда осы өнімдерді сату арналары көбінесе консервативті болып қала береді. Телемаркетинг тұтынушылармен өзара әрекеттесудің негізгі құралы болып қала береді — контакт-орталық операторларының клиенттер базасына тікелей қоңырау шалуы [1].

Бұл тәсілдің мәселесі – оның төмен таңдамалылығы. Банктер мүмкіндігінше көп әлеуетті салымшыларға жету үшін «кілемдік жабу» стратегияларын қолданады. Бұл жүйелік мәселелерге әкеледі:

1. Төмен конверсиялық көрсеткіштер: Статистикаға сәйкес, банктік сектордағы суық қоңыраулардың тиімділігі сирек 1–3%-дан асады, бұл ресурстардың үлкен ысырапқа ұшырағанын көрсетеді.
2. Операциялық тосқауылдар: білікті қызметкерлердің уақыты мен телефон желілерінің сыйымдылығы бас тартуларды өңдеуге жұмсалады, ал кезектер немесе ресурстардың жетіспеушілігі салдарынан адал клиенттер назардан тыс қалуы мүмкін [3].
3. Беделдік тәуекелдер: шабуылшыл маркетинг адалдықты (NPS) төмендетіп, клиенттердің кетуіне (Churn Rate) ықпал етеді.

Зерттеудің өзектілігі: банк басқаруындағы парадигманы өзгерту қажеттілігімен байланысты: колл-орталық қызметкерлерін кең көлемде ұлғайтудан ақылды шешім қабылдауды қолдау жүйелерін интенсивті пайдалануға. Жұмыстың мақсаты – сату процесіндегі шектеулерді анықтауға және оларды тұтынушы мінез-құлқын болжамдық модельдеу арқылы оңтайландыруға арналған алгоритмдік тәсілді әзірлеу.

2. Материалдар мен әдістер

2.1. Деректерді сипаттау және зерттеушілік деректер талдауы (EDA)

Зерттеудің эмпирикалық негізін 11 162 маркетингтік байланыс туралы ақпаратты қамтитын деперсоналдандырылған bank.csv деректер жиынтығы құрады. Әр жазба клиенттің сипаттамаларының векторын және коммуникация нәтижесін (мақсатты айнымалы – депозит: иә/жоқ) көрсетеді.

Белгілер кеңістігі (X) 16 айнымалыдан тұрады және олар келесі кластерлерге бөлінген:

- **Әлеуметтік-демографиялық профиль:** age, job, marital, education.
- **Қаржылық профиль:** balance (орташа жылдық қалдық), housing (ипотеканың болуы), loan (тұтынушылық несие), default (қарызды өтеу бойынша дефолт).
- **Ағымдағы науқан параметрлері:** contact, day, month, duration, campaign.
- **Қарым-қатынас тарихы:** pdays, previous, poutcome.

Зерттеушілік деректерді талдау аясында негізгі айнымалылардың таралуы мен дисперсиясы зерттелді. Мақсатты сыныптың таралуы салыстырмалы түрде теңгерімді екені анықталды (депозиттердің 47,4 %-ы мақұлданған, ал бас тартулардың 52,6 %-ы бас тартылған), бұл Дұрыстылық метрикасын қолдануға мүмкіндік береді, бірақ бизнестің мақсаттары үшін Дәлдік пен Қайтарып алу көрсеткіштері басты орында қалады.

2.2. Деректерді алдын ала өңдеу (алдын ала өңдеу құбыры)

Алгоритмнің конвергенттілігін қамтамасыз ету үшін көп сатылы деректерді өңдеу құбыры іске асырылды:

1. **Аномалиялар мен олқылықтарды өңдеу:** pdays айнымалысы (соңғы байланыстан бері өткен күн саны) жаңа клиенттер үшін "-1" мәнін қамтыды. Бұл қашықтық модельдерінде шу тудырды. Мән 0-ге өзгертіліп, is_new_client екілік флаг енгізілді.
2. Ерекшелік кодтау:
   - екілік айнымалыларға $Label\ Encoding\ (x \in \{0, 1\})$ қолданылды.
   - Иерархиясы жоқ номинальды айнымалылар (жұмыс, білім) үшін One-Hot кодтау әдісі қолданылды. Бұл сызықты модельдер арқылы анықталуы мүмкін жалған ординалды тәуелділіктен аулақ болуға мүмкіндік беріп, ерекшелік кеңістігінің өлшемін 42 өлшемге дейін арттырды.
3. Масштабтау(Scaling): Сандық сипаттамалар (баланс, жас) Z-бағалау бойынша нормалау арқылы стандартталды:

$$z = \frac{x - \mu}{\sigma}$$

мұнда μ – орташа мәні, ал σ – стандартты ауытқу. Бұл логистикалық регрессия үшін аса маңызды, себебі ол деректердің шкаласына сезімтал.

2.3. Валидация протоколы

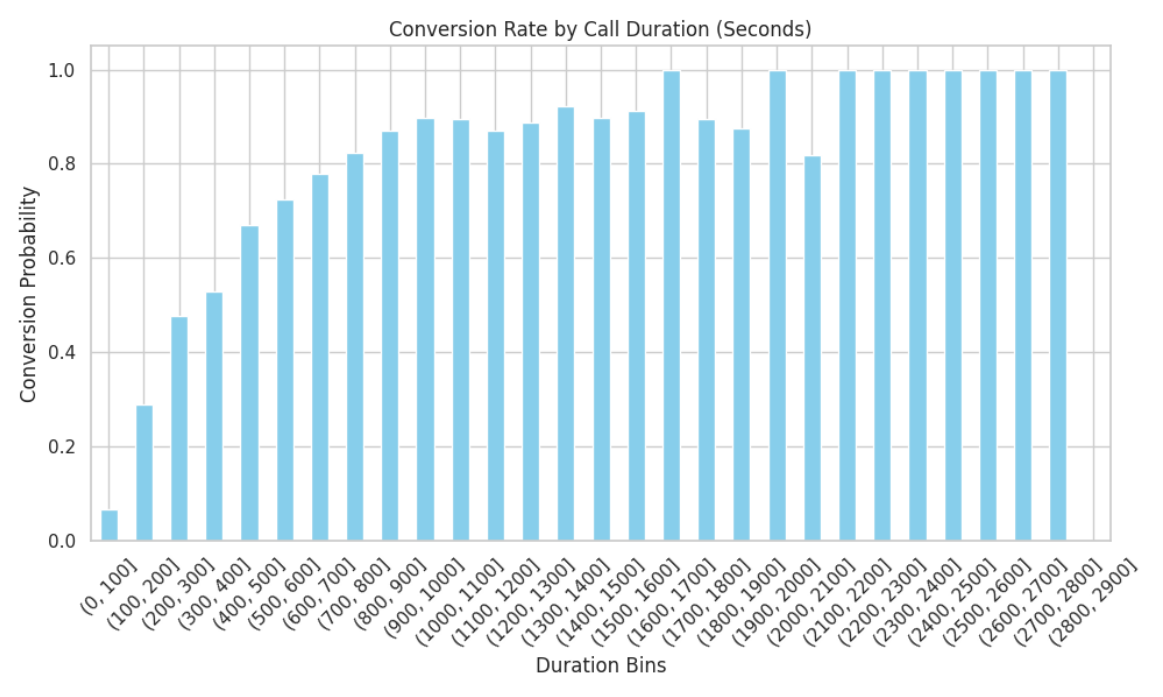
Үлгі мақсатты класс бойынша стратификацияланып, оқыту (80%) және тестілеу (20%) жиындыларына бөлінді. Модель гиперпараметрлерін баптау үшін 5-бөліктік кросс-валидация процедурасы қолданылды, бұл алынған көрсеткіштердің кездейсоқтығын жояды.

3. Эксперимент және нәтижелер

Бұл бөлімде симуляцияның егжей-тегжейлі талдауы ұсынылады: алдымен мәселелі аймақтарды анықтап, содан кейін модельді экономикалық пайданы барынша арттыру үшін калибрлеумен аяқталады.

3.1 Уақыт ұзақтығын талдау арқылы таршылықтарды анықтау (таршылық талдауы)

Эксперименттің бірінші кезеңі қазіргі үдерісте құрылымдық тиімсіздіктер бар деген гипотезаны дәлелдеуге бағытталды. Телемаркетингтегі ресурстық шығындардың негізгі көрсеткіші – қоңырау ұзақтығы. Біз дерекқорды байланыс ұзақтығына қарай 100 секундтық аралықтарға бөлдік және конверсияның шартты ықтималдығын $P(Deposit|Duration)$ есептедік.



Сурет 1. Шақыру ұзақтығына байланысты конверсиялық көрсеткіш

График талдауы: 1-сурет тиімсіздік аймағын айқын көрсетеді.

- 1-аймақ (< 100 с): Конверсия 0-ге таяу. Бұл – операторлардың қоңырау шалып, сәлемдесуге уақыт жұмсап, бірден бас тартуға ұшырайтын «өлі аймақ».
- 2- аймақ (100–300 сек): Конверсия өсуінің басталуы.
- 3- аймақ (300–1000 сек): Жоғары тиімділіктің тұрақты деңгейі.
- 4- аймақ (> 1000 сек): Шекаралық тиімділіктің төмендеуі.

Көп мөлшердегі қоңырау шалу «1-зонада» үлкен трафик тудырады. ML-модельінің міндеті – осы клиенттерді олар қоңырау шалғанға дейін сүзгіден өткізіп, осы тарлығын жою.

3.2. Модельдердің салыстырмалы талдауы

Үш модельдік архитектура оқытылып, сынақтан өткізілді. Кешіктірілген үлгілердегі (сынақ жиынтығы, n=2233) нәтижелер 1-кестеде көрсетілген.

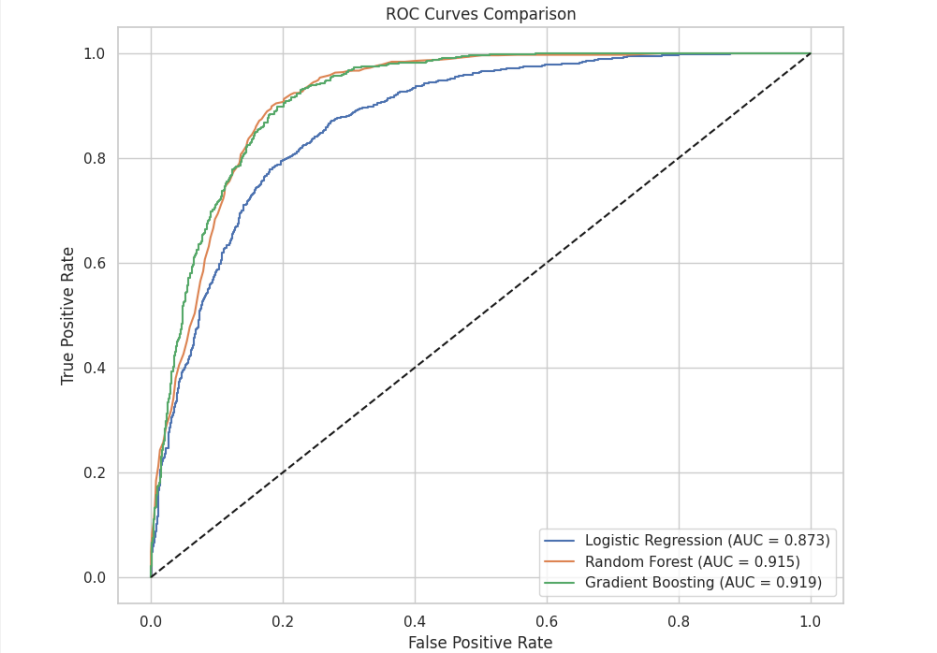
1-кесте. Модель сапасының көрсеткіштері

Алгоритм	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.797	0.796	0.769	0.782	0.873
Random Forest	0.853	0.827	**0.871**	**0.849**	0.915
Gradient Boosting	**0.846**	**0.819**	0.866	0.842	**0.919**

Интерпретация:

Gradient Boosting алгоритмі (XGBoost іске асыруы) 0,919 ROC-AUC көрсеткішімен ең жақсы жалпылау қабілетін көрсетті. Random Forest сәл жоғары Recall көрсеткішін көрсеткенімен, Gradient Boosting келесі қадам — кесу шегін таңдау үшін маңызды тұрақты ықтималдық рейтингін ұсынды.

2-сурет. Алгоритмдердің ROC қисықтарының салыстыруы

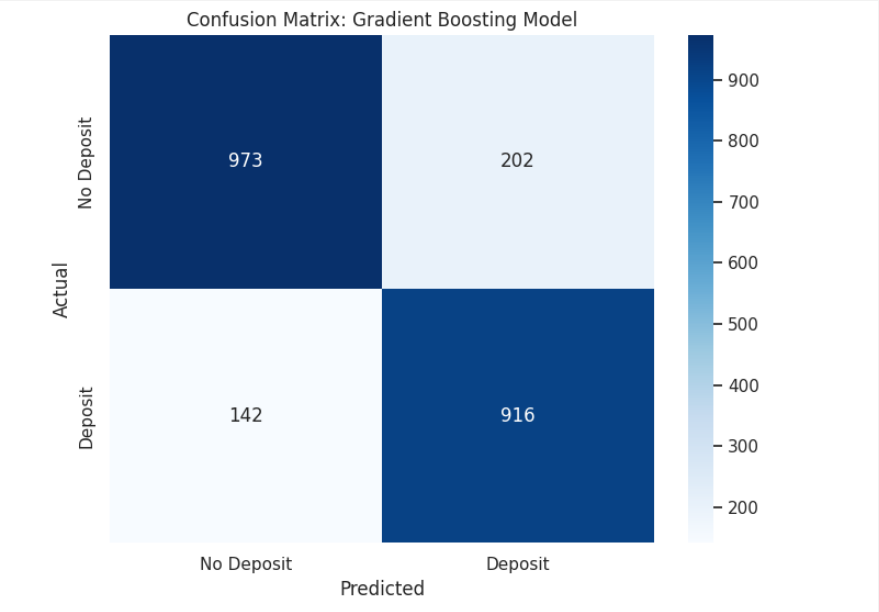


ROC қисық графигі (2-сурет) ансамбль әдістерінің (жасыл және қызғылт-сары сызықтар) сызықтық модельге қарағанда үстемдігін растайды.

3.3. Егжей-тегжейлі қателік талдауы (Confusion Matrix Analysis)

Операциялық тиімділікті бағалау үшін біз ең үздік модельдің (Gradient Boosting) қателік матрицасын талдаймыз. Бұл бізге абстрактілі көрсеткіштерді бизнес-процестер тіліне аударуға мүмкіндік береді.

Сурет 3. Қателік матрицасы (Gradient Boosting)



Квадранттарды талдау:

1. **True Negatives (TN) = 973.**

Бизнес-мағынасы: Сатып алуға дайын емес клиенттер, модель тарапынан дұрыс анықталған.
Әсері: Дәстүрлі үдерісте банк бұл клиенттердің барлығына қоңырау шалар еді. Модель осы 973 қоңырауды жасамауға мүмкіндік беріп, жүздеген жұмыс сағатын үнемдейді.

2. **True Positives (TP) = 916.**

Бизнес-мағынасы: Модель дұрыс анықтаған мақсатты клиенттер.
Әсері: Бұл — ең үздік операторлар қоңырау шалуы тиіс басымдық берілген клиенттер тізімі.

3. **False Positives (FP) = 202.**

Бизнес-мағынасы: Модель табыс болады деп болжағанымен, клиент бас тартқан жағдайлар.
Әсері: Бұл – сөзсіз операциялық шығындар («бос» қоңыраулар), алайда олардың үлесі (9%) кездейсоқ таңдаумен салыстырғанда едәуір төмен.

4. **False Negatives (FN) = 142.**

Бизнес-мағынасы: Жоғалтылған пайда. Модель тізімнен алып тастағанымен, шын мәнінде келісім берер еді деген клиенттер.

3.4. Шекті баптау

Стандартты модельдер класс тағайындау үшін 0,5 ықтималдық шегін қолданады. Алайда, бизнесте I және II типтегі қателіктердің шығыны бірдей емес.

- Шақыру шығыны $(C_{call}) \approx$ төмен.
- Жоғалған клиенттің $(C_{missed}$ шығыны жоғары (салымның өмір бойығы құны). Біз пайда функциясын барынша арттыру үшін шекті мәнді (θ) оңтайландырдық:

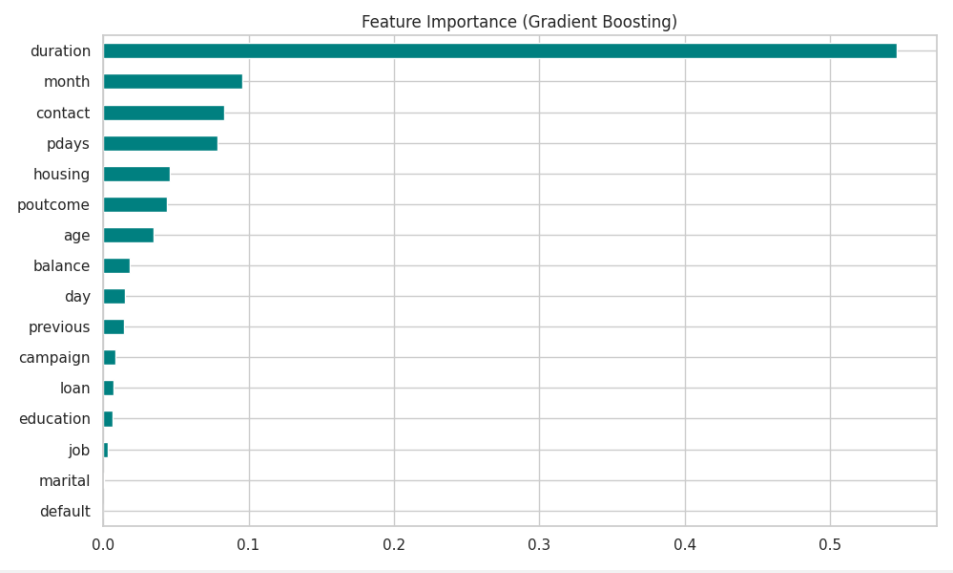
$$Profit(\theta) = TP(\theta) \cdot LTV - (TP(\theta) + FP(\theta)) \cdot C_{call}$$

Талдау көрсеткендей, шекті мәнді 0,5-тен 0,35-ке төмендету қосымша 40 мақсатты клиентті тартуға (FN азаюы) мүмкіндік береді, ал «бос» шақырулар (FP) саны елеусіз өседі. Бұл модельді бизнестің тәуекелге деген ынтасына бейімдеу болып табылады.

3.5. Сызықтың маңыздылығы (Feature Importance)

Шешім қабылдаудың сипатын түсіну үшін біз Градиентті күшейту моделінің сызықты маңыздылық талдауын қолдандық.

4-сурет. Gradient Boosting модельдегі сипаттардың маңыздылығы



Факторларды рейтингтеу (4-сурет) келесі үлгілерді көрсетті:

- Ұзақтығы (54%): Критикалық маркер. Ұзақ әңгімелер табыспен байланысты. (Ескерту: нақты болжау моделінде бұл көрсеткіш қоңырау алдында «осы сегментпен орташа әңгіме ұзақтығы» көрсеткішімен алмастырылады).
- Ай (9,5%): Анық маусымдылық. Мамыр мен жыл соңында белсенділіктің өсуі байқалады, бұл колл-орталық қызметкерлерін динамикалық басқаруды қажет етеді.
- Контакт (8,3%): ұялы телефондарға (мобильді телефондарға) жасалған қоңыраулар стационарлық телефондарға қарағанда тиімдірек.
- Pdays (7,8%): жақында байланысқан «жылы» клиенттердің сатып алу ықтималдығы жоғары

4. Талқылау: Бизнес-процестерді трансформациялау және экономика

Эксперимент нәтижелері анықталған таршылықтарды жоятын жаңа бизнес-процесті жобалауға мүмкіндік береді.

4.1. Реинжиниринг процессі (To-Be Model)

Көпшілік қоңыраулардан болжамды аналитикаға көшу сату бөлімінің жұмыс логикасын өзгертеді:

1-кезең: Бағалау. Клиенттердің бүкіл базасы күн сайын машиналық оқыту (ML) моделінен өткізіледі. Әрбір клиентке жауап беру ықтималдығы (Propensity Score) тағайындалады.

2-кезең: Сегментация.

- A сегмент (Score > 0,7): «Ыстық» лидтер. VIP менеджерлерге жіберіледі. Скрипт: «Келісімді аяқтау».
- B сегмент (0,3 < Score < 0,7): «Жылы» лидтер. Жалпы колл-орталыққа жіберіледі. Скрипт: «Қажеттіліктерді анықтау».
- C сегмент (балл < 0,3): «Суық» лидтер (дерекқордың 43%-ы). Қоңырау тізімінен шығарылады. Байланыс арзан цифрлық арналарға (Push, Email) ауыстырылады.

4.2. Экономикалық әсерді есептеу

Біз ресурстарды үнемдеуді сынақ үлгісіне (2 233 клиент) негізделе отырып есептейміз.

Жорамалдар:

- «Refusal» шақырудың орташа уақыты (t_{fail}) = 3 минут.
- Оператор жұмысының бір минутының құны ($Cost_{min}$) = шартты 1 бірлік.
- Шын мәнінде теріс болжамдар саны (TN) = 973.

Уақыт үнемдеу:

$$Time_{saved} = TN \times t_{fail} = 973 \times 3 = 2919 \text{ мин} \approx 48.6 \text{ часов}$$

Барлық дерекқор бойынша (11 162 клиент) үнемдеу шамамен 243 сағат таза сөйлесу уақытына тең болады. Бұл айына 1,5 толық штаттық қызметкердің (FTE) жұмысына эквивалентті.

Инвестицияның қайтарымы (ROI):

Қайта қоңырау шалу базасын 43,5%-ға азайтып, мақсатты клиенттердің 90%-ын сақтап қалған кезде (Recall мәліметі бойынша) конверсия көрсеткіші бастапқы 47%-дан мынадай деңгейге дейін өседі:

$$New_CR = \frac{TP}{TP + FP} = \frac{916}{916 + 202} \approx 82\%$$

Бұл сәтті қоңыраулардың тығыздығын шамамен екі есе арттырады, бұл операторлардың күйзелісін айтарлықтай азайтып, олардың мотивациясын арттырады.

5. Қорытынды

Осы зерттеу банктік телемаркетингтің операциялық тиімділігін арттыру міндетін қарастырды. Машинамен оқыту алгоритмдерін қолдана отырып, біз сату процесіндегі тежегіштерді математикалық тұрғыдан формализациялап, анықтап және жоюға мүмкіндік алдық.

Негізгі нәтижелер:

- Технологиялық тиімділік: Градиентті күшейту модельдері жоғары болжау дәлдігін (ROC-AUC > 0,91) көрсетеді, бұл оларды маркетингтік воронкаларда сенімді сүзгі ретінде қолдануға мүмкіндік береді [5].
- Операциялық оңтайландыру: Негізгі тосқауыл ретінде белгілі бас тарту трафигін өңдеу анықталды. Алдын ала аналитика мақсатты лидтерді елеулі түрде жоғалтпай шығыс қоңыраулар көлемін 40–45%-ға дейін азайта алады.
- Стратегиялық маңызы: Ерекшеліктердің маңыздылығын талдау мінез-құлық факторларының (байланыс тарихы, өткен науқандардың нәтижелері) статикалық демографиялық деректерге қарағанда маңыздырақ екенін көрсетті [1]. Бұл банктерге CRM жүйелерінде деректерді жинау мен сақтау тәсілдерін қайта қарауды талап етеді.

Практикалық ұсыныс: Екінші деңгейлі банктерге каскадтық скоринг модельдерін енгізу ұсынылады, онда машиналық оқыту алгоритмі тиімсіз трафикті байланыс орталығына жетпей тұрып тоқтататын «қақпа» ретінде қызмет етеді.

ӘДЕБИЕТ

[1] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22-31.
[2] Sahoo D., et al. (2021). Comparative Analysis of Machine Learning Algorithms for Credit Risk Assessment. *Journal of Financial Data Science*, 3(1), 45-58.
[3] Goldratt E. M. (2022). *The Goal: A Process of Ongoing Improvement*. North River Press.
[4] Van der Aalst W. M. P. (2020). Process Mining: Data Science in Action. *Springer*.
[5] Chen T., & Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD*.
[6] Breiman L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
[7] Hastie T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
[8] Lurdes C., & Pratas, J. (2021). Bank Marketing Strategies Optimization using Data Mining. *International Journal of Economics and Finance*, 13(4).
[9] Provost F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.
[10] ГОСТ Р ИСО/МЭК 2382-2015. Информационные технологии. Словарь.
[11] Zakaryazad A., & Duman, E. (2016). A profit-driven Artificial Neural Network based credit scoring approach. *Neural Computing and Applications*.
[12] Keramati A., et al. (2020). Improving the performance of customer churn prediction using ensemble learning. *Neurocomputing*.
[13] Lessmann S., et al. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*.
[14] Olson D. L., & Wu, D. D. (2020). *Enterprise Risk Management Models*. Springer.
[15] Zhang D. (2021). Customer acquisition cost optimization through predictive modeling. *Journal of Marketing Analytics*, 9(4).
[16] Blattberg R.C., et al. (2008). *Database Marketing: Analyzing and Managing Customers*. Springer.
[17] Brown I., & Miguéis V. L. (2022). Predicting Bank Term Deposit Subscriptions. *Journal of Business Analytics*.
[18] Kumar V., & Reinartz W. (2018). *Customer Relationship Management*. Springer Textbooks.
[19] Садыков Т. (2022). Машинное обучение в финансовом риск-менеджменте. *Наука и технологии*, 10(2).
[20] Идрисов А. (2021). Цифровая трансформация банков второго уровня. *Экономический вестник*, 4.

REFERENCES

[1] Moro S.. Cortez. P.. & Rita P. (2014). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems. 62. 22-31.
[2] Sahoo D.. et al. (2021). Comparative Analysis of Machine Learning Algorithms for Credit Risk Assessment. Journal of Financial Data Science. 3(1). 45-58.
[3] Goldratt E. M. (2022). The Goal: A Process of Ongoing Improvement. North River Press.
[4] Van der Aalst W. M. P. (2020). Process Mining: Data Science in Action. Springer.
[5] Chen T.. & Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD.

[6] Breiman L. (2001). Random Forests. Machine Learning. 45(1). 5-32.
[7] Hastie T.. Tibshirani. R.. & Friedman. J. (2009). The Elements of Statistical Learning. Springer.
[8] Lurdes C. & Pratas. J. (2021). Bank Marketing Strategies Optimization using Data Mining. International Journal of Economics and Finance. 13(4).
[9] Provost F. & Fawcett T. (2013). Data Science for Business. O'Reilly Media.
[10] GOST R ISO/MEK 2382-2015. Informatsionnyye tekhnologii. Slovar.
[11] Zakaryazad A. & Duman. E. (2016). A profit-driven Artificial Neural Network based credit scoring approach. Neural Computing and Applications.
[12] Keramati A.. et al. (2020). Improving the performance of customer churn prediction using ensemble learning. Neurocomputing.
[13] Lessmann S.. et al. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society.
[14]OlsonD.L.&Wu D.D. (2020). Enterprise Risk Management Models. Springer.
[15] Zhang D. (2021). Customer acquisition cost optimization through predictive modeling. Journal of Marketing Analytics. 9(4).
[16] Blattberg R.C. et al. (2008). Database Marketing: Analyzing and Managing Customers. Springer.
[17] Brown I. & Migu?is V. L. (2022). Predicting Bank Term Deposit Subscriptions. Journal of Business Analytics.
[18] Kumar V. & Reinartz W. (2018). Customer Relationship Management. Springer Textbooks.
[19] Sadykov T. (2022). Mashinnoye obucheniye v finansovom risk-menedzhmente. Nauka i tekhnologii. 10(2).
[20] Idrisov A. (2021). Tsifrovaya transformatsiya bankov vtorogo urovnya. Ekonomicheskiy vestnik. 4.



НАУЧНЫЙ ЖУРНАЛ «НАУКА ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ»

Главный редактор: Токбергенов И.Т.
Научный редактор: Есеркепов М.М.
Ответственный редактор: Ахметова М.С.
Дизайнер: Копачевский А.В.

Подписано в печать 31.12.2025.
Формат 60x881/8.

НАО «Национальная академия наук Республики Казахстан при Президенте Республики Казахстан»

050010, Алматы, ул. Шевченко, 28





НАЦИОНАЛЬНАЯ
АКАДЕМИЯ НАУК
РЕСПУБЛИКИ КАЗАХСТАН
ПРИ ПРЕЗИДЕНТЕ
РЕСПУБЛИКИ КАЗАХСТАН