

https://doi.org/10.51885/3134-8025_IICS_2026_3_11

XFTAP 87.15.09

ҚАЛАЛАРДАҒЫ АТМОСФЕРАЛЫҚ АУА САПАСЫН МОНИТОРИНГТЕУ МЕН БОЛЖАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕГІ БҰЛТТЫҚ SAAS ПЛАТФОРМАСЫН ӨЗІРЛЕУ

РАЗРАБОТКА ОБЛАЧНОЙ SAAS-ПЛАТФОРМЫ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ МОНИТОРИНГА И ПРОГНОЗИРОВАНИЯ КАЧЕСТВА АТМОСФЕРНОГО ВОЗДУХА В ГОРОДАХ

DEVELOPMENT OF A CLOUD SAAS PLATFORM BASED ON MACHINE LEARNING FOR MONITORING AND FORECASTING ATMOSPHERIC AIR QUALITY IN CITIES

А.Е. Тукушова ^{1*}, С.Ж. Рахметуллина ¹, М.Ж. Базарова ²,

А.В. Пененко ³, З.Т. Хасенова ¹

¹Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан

²Сәрсен Аманжолов атындағы Шығыс Қазақстан университеті, Өскемен қ., Қазақстан

³РГА СБ Есептеу математикасы және математикалық геофизика институты, Новосибирск қ., Ресей

*Жауапты автор: Тукушова Арайлым Ержанқызы, e-mail: araylym.tukushova@mail.ru

Түйінді сөздер:

бұлтты есептеу қызметі;
болжау; машиналық
оқыту; ауа сапасы; сәулет.

ТҮЙІНДЕМЕ

Бұл жұмыста жетілдірілген бөлшек-дифференциалдық стохастикалық модель негізінде сандық модельдеуді жүзеге асыруға арналған «SaaS» типіндегі есептеу сервисінің архитектурасы ұсынылады. Зерттеудің негізгі мақсаты – пайдаланушыларға жеке есептеу ресурстарын қажет етпей-ақ жоғары дәлдіктегі есептеулерді орындауға мүмкіндік беретін модульдік жүйені әзірлеу.

Ұсынылған шешім өзара байланысқан бірнеше ішкі жүйеден тұрады: деректерді қабылдау, тапсырмаларды кезекке қою және алдын ала өңдеу, математикалық есептеу модулі, нәтижелерді талдау мен визуализациялау, жүйені мониторингтеу, журнал жүргізу және қауіпсіздікті қамтамасыз ету. Әрбір ішкі жүйе қазіргі заманғы бағдарламалық құралдарды (Julia, Python, Django, Prometheus және т.б.) қолдану арқылы жүзеге асырылған, API арқылы өзара әрекеттеседі және PostgreSQL негізіндегі орталықтандырылған деректер қоймасына сүйенеді.

Нәтижесінде сұраныстарды автоматты түрде қабылдауды, есептеулер жүргізуді және алынған нәтижелерді түсіндіруді қамтамасыз ететін масштабталатын әрі істен шығуға төзімді архитектура қалыптастырылды. Қорытындыда мұндай үлестірілген тәсілдің тиімді екендігі және оны ғылыми модельдеу мен аналитиканың басқа да міндеттеріне бейімдеуге болатындығы атап көрсетіледі.



Ключевые слова:

облачный
вычислительный
сервис;
прогнозирование;
машинное обучение;
качество воздуха;
архитектура.

АННОТАЦИЯ

В данной работе предлагается архитектура вычислительного сервиса типа «SaaS» для реализации численного моделирования на основе усовершенствованной дробно-дифференциальной стохастической модели. Основная цель исследования-разработать модульную систему, которая позволяет пользователям выполнять высокоточные вычисления без необходимости использования собственных вычислительных ресурсов. Предлагаемое решение состоит из нескольких взаимосвязанных подсистем: прием данных, постановка задач в очередь и предварительная обработка, модуль математических вычислений, анализ и визуализация результатов, мониторинг системы, ведение журнала и обеспечение безопасности. Каждая подсистема, реализованная с использованием современных программных инструментов (Julia, Python, Django, Prometheus и т.д.), взаимодействует через API и полагается на централизованное хранилище данных на основе PostgreSQL. В результате была сформирована масштабируемая и отказоустойчивая архитектура, обеспечивающая автоматический прием запросов, проведение расчетов и интерпретацию полученных результатов. В заключении подчеркивается, что такой распределенный подход эффективен и может быть адаптирован к другим задачам научного моделирования и аналитики.

keywords:

cloud computing service;
forecasting; machine
learning; air quality;
architecture.

ABSTRACT

In this paper, the architecture of a computing service of the "SaaS" type is proposed for the implementation of numerical modeling based on an improved fractional-differential stochastic model. The main purpose of the research is to develop a modular system that allows users to perform high-precision calculations without the need to use their own computing resources. The proposed solution consists of several interconnected subsystems: data reception, queuing and preprocessing tasks, a mathematical calculation module, analysis and visualization of results, system monitoring, logging and security. Each subsystem implemented using modern software tools (Julia, Python, Django, Prometheus, etc.) interacts via APIs and relies on a centralized PostgreSQL-based data warehouse. As a result, a scalable and fault-tolerant architecture was created that provides automatic acceptance of requests, calculations, and interpretation of the results. In conclusion, it is emphasized that such a distributed approach is effective and can be adapted to other tasks of scientific modeling and analytics.

КІРІСПЕ

Қарқынды урбандалу және экологиялық мәселелердің күшеюі жағдайында атмосфералық ауаның сапасын бақылау аса өзекті міндетке айналып отыр. Дәстүрлі стационарлық бақылау жүйелері жоғары құнымен, кеңістіктік қамтылуының шектеулілігімен және деректердің кешігіп түсуімен сипатталады. Осыған байланысты зерттеудің мақсаты – машиналық оқыту алгоритмдерінің көмегімен ауадағы ластаушы заттардың деңгейін болжайтын бұлттық сервистің архитектурасын әзірлеу.

Жұмыстың ғылыми жаңалығы ауа сапасын бақылау жүйесіне машиналық оқыту әдістері мен бұлттық есептеулерді кіріктіруімен байланысты. Мұндай тәсіл ағымдағы көрсеткіштерді жедел талдауға және ластану деңгейін бірнеше сағат пен күн бұрын болжауға мүмкіндік береді. Бұл қоршаған орта жағдайының нашарлауына уақтылы әрекет ету үшін маңызды.

Экологиялық мониторингке арналған SaaS форматындағы бұлттық шешім инновациялық бағыт болып табылады. Қазіргі зерттеулер көрсеткендей, мұндай платформалар күрделі ғылыми-инженерлік есептеулер мен сандық модельдеудің тиімді инфрақұрылымы бола алады. (Zhang, Q., et al. (2010)) еңбекте бұлттық жүйелердің архитектурасы істен шығуға төзімді, ауқымдалатын және API арқылы модульдердің сенімді өзара әрекеттесуін қамтамасыз етуі тиіс екені көрсетілген. Ал (Buuya, R., et al. (2009)) есептеу ресурстарын автоматты түрде бөлудің маңызын атап өтеді. Осы қағидалар ұсынылып отырған сервистің негізіне алынды.

Мониторинг пен журналдаудың орталықтандырылған тәсілі жүйе сенімділігін арттырудың маңызды тетігі болып саналады (Ghobaei-Arani, M., et al. (2020)); сондықтан сервисте Prometheus және ELK стегі пайдаланылады. Үлкен көлемдегі деректерді сақтау, қауіпсіздік және бұлттағы аналитикалық өңдеу мәселелері (Hashem, I. A. T., et al. (2015)) еңбегінде қарастырылған, бұл тұжырымдар PostgreSQL мен BI-құралдарын қолдануда ескерілді. Микро-сервистік архитектураның артықшылықтары (Pahl, C., & Jamshidi, P. (2016)) жұмысында негізделген және олар әзірленіп жатқан сервистің модульдік құрылымында көрініс табады.

SaaS тәсілінің әртүрлі салалардағы, соның ішінде білім беру инфрақұрылымындағы тиімділігі (Smith, J., et al. (2020)) еңбегінде көрсетілген. Бұл тәжірибе экологиялық мониторинг жүйелерін құруға да бейімделе алады. Сонымен қатар жоғары жүктеме жағдайында ресурстарды басқару және өнімділікті оңтайландыру мәселелері (Smith, J., et al. (2020), Zhang, Y., et al. (2021)) еңбектерінде қарастырылған. Бұлттық есептеулер саласындағы өзге зерттеулер де ауқымдау, трафикті теңгеру және қауіпсіздік мәселелерінің маңызын атап өтеді (Mell, P., & Grance, T. (2011), Erl, T., Puttini, R., & Mahmood, Z. (2013), Shaban, K., Kadri, A., & Rezk, E. (2016), Zhang, Y., Wang, J., & Ma, L. (2017)).

Ауа сапасын бағалауға арналған SaaS-платформалар бойынша тікелей зерттеулер әлі де аз (Liu, X., Chen, H., & Li, J. (2018), Chen, J., Hu, Y., & Wu, L. (2019), Hosseini, S., Zhang, R., & Wang, P. (2020), Zhang, Q., Xiao, M., & Li, X. (2021)), алайда сабақтас салалардағы тәжірибе бұл бағыттың перспективалы екенін көрсетеді. Осыған байланысты архитектуралық негіз ретінде SaaS моделі таңдалды, мұнда пайдаланушы дайын нәтижелерді өзінде жеке есептеу жабдықтарынсыз алады.

Платформаны әзірлеу үшін икемділік пен жоғары өнімділікті қамтамасыз ететін технологиялар іріктелді: Docker, Kubernetes, Python, Julia, PostgreSQL және Java. REST API-ге негізделген микросервистік архитектура жүйенің модульдігін арттырып, оны қолдауды жеңілдетеді. Әрбір ішкі жүйеге деректерді жинау, сандық модельдеу, машиналық оқыту, визуализация және пайдаланушылық интерфейс үшін арнайы бағдарламалық құралдар таңдалды.

Машиналық оқыту әдістерін қолдану атмосфераның ластануын талдау және болжау мүмкіндіктерін кеңейтеді. Қазіргі зерттеулер дәстүрлі статистикалық әдістерден нейрожелілік және ансамбльдік модельдерге көшудің тиімді екенін көрсетеді (Kim, H., Park, J., & Seo, D. (2022), Wang, L., Yu, X., & Fang, Y. (2021)). Алғашқы жұмыстарда негізінен сызықтық регрессия қолданылса, кейін зерттеушілер шешім ағаштары (Xiao, M., Sun, Y., & Li, Z. (2023)), кездейсоқ ормандар (Wang, L., Yu, X., & Fang, Y. (2021)), градиенттік бустинг (Gao, P., Chen, X., & Liu, W. (2023)) және тірек векторлар әдісі (Jones, D. M. H., Skourides, P. R. J. S., & Hennessey, M. G. B. (2021)) секілді жетілдірілген алгоритмдерге назар аударды. Бұл әдістер күрделі бейсызық тәуелділіктерді жақсырақ ескеріп, болжам дәлдігін арттырды.

Соңғы жылдары кеңістіктік-уақыттық модельдер мен жасанды нейрондық желілерге, әсіресе сверткалық желілерге (Ma, H. T. M., Li, L., & Wang, Y. (2019)) және LSTM секілді рекурренттік архитектураларға қызығушылық артты. Мұндай модельдер ластану динамикасын дәл сипаттап, қысқа және орта мерзімді болжамдардың сапасын жақсартады (Adebayo, H. O., Olanrewaju, O. M., & Ajayi, J. I. (2021)). Сонымен қатар деректерді алдын ала өңдеу, спутниктік бақылауларды пайдалану және физикалық модельдермен біріктірілген гибридік тәсілдер де маңызды рөл атқарады (Ahmed, N., Shukla, S. K. R., & Gupta, A. K. (2022)).

Машиналық оқыту алгоритмдерін қолдану экологиялық ахуалдың өзгерісіне жедел әрекет етуге, ластануды азайтуға бағытталған проактивті шараларды әзірлеуге және халықтың экологиялық хабардарлығын арттыруға мүмкіндік береді. Практикалық енгізуде жоғары дәлдік пен есептеу ресурстарын тиімді пайдалану арасындағы тепе-теңдік маңызды. Бұл тұрғыда LightGBM, CatBoost, Ray Tune және Optuna секілді құралдар модельдерді жедел әрі тиімді баптауға мүмкіндік береді. Нәтижесінде нақты уақыт режимінде жұмыс істейтін, жылдам, бейімделгіш және ауқымдалатын бұлттық платформалар үшін қолайлы модельдер қалыптасады.

Зерттеудің ғылыми жаңалығы мен авторлық үлесі.

Осы зерттеудің ғылыми жаңалығы ауа сапасын мониторингтеу және болжау міндеттерін бұлттық есептеу ортасында кешенді түрде шешуге бағытталған. Жұмыста тек жеке машиналық оқыту моделі ғана емес, деректерді жинау, алдын ала өңдеу, есептеу, болжау, визуализация, мониторинг және қауіпсіздік модульдерін біріктіретін толық SaaS-платформа архитектурасы ұсынылады.

Зерттеудің негізгі үлесі келесі бағыттармен сипатталады:

Архитектуралық үлес: атмосфералық ауа сапасын мониторингтеу және болжау үшін модульдік бұлттық SaaS-платформа архитектурасы әзірленді және эксперименттік түрде тексерілді. Бұл архитектура әрбір ішкі жүйені жеке масштабтауға және оларды API арқылы байланыстыруға мүмкіндік береді.

Модельдік үлес: ауа ластануын болжау үшін XGBoost, LightGBM және HistGradientBoosting ансамбльдік модельдері салыстырылып, олардың тиімділігі MAE, RMSE және R^2 метрикалары бойынша бағаланды.

Интеграциялық үлес: ұсынылған платформа сыртқы дереккөздермен, орталықтандырылған деректер қоймасымен, мониторинг жүйелерімен және Smart City платформаларымен интеграциялануға бейімделген.

Осылайша, жұмыс қолданыстағы шешімдерден ауа сапасын болжау моделін бұлттық SaaS-инфрақұрылыммен біріктіруімен және нақты уақыт режимінде экологиялық мониторинг жүргізуге бағытталуымен ерекшеленеді.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӨДІСТЕРІ

Машиналық оқыту (МО) алгоритмдерін атмосфералық ауаның жағдайын талдауға қолдану ластану деңгейін алдын ала болжауға және қалалық экологиялық ахуалды жақсартуға бағытталған шараларды жедел әзірлеуге мүмкіндік береді. Сенімді болжамдық модельдер ансамбльдік әдістерді пайдалану және нәтижелерді түсіндіруді жеңілдететін визуализация құралдарын қолдану арқылы қалыптасады, бұл басқарушылық шешімдерді қабылдауды қолдайды.

Ұсынылған тәсілдің бастапқы кезеңінде МО алгоритмдері бақылау нүктелеріндегі ластаушы заттардың концентрациясын бағалайды. Модельдерді оқыту алдында деректерді міндетті түрде алдын ала өңдеу жүргізіледі. Бұл кезеңге келесі процедуралар кіреді:

- жетіспейтін мәндерді толықтыру;
- аномалиялық мәндерді анықтау және түзету;
- статистикалық корреляция негізінде ақпараттық белгілерді іріктеу.

Аталған рәсімдер, мысалы, (Kim, H., Park, J., & Seo, D. (2022), Xiao, M., Sun, Y., & Li, Z. (2023)) еңбектерінде егжей-тегжейлі сипатталған және нақты экологиялық деректер негізінде модель құру барысында стандартты тәжірибе болып саналады.

Модель параметрлерін оңтайландыру үшін Scikit-learn кітапханасының GridSearchCV құралы пайдаланылды. Бұл әдіс параметрлерді іздеуді кросс-валидациямен біріктіріп, болжам дәлдігін арттыруға мүмкіндік береді. Барлық модельдерді оқыту кезінде нәтижелердің қайта өндірілуін қамтамасыз ету үшін `random_state = 42` пайдаланылды.

Деректерді оқыту және тест жиындарына бөлу.

Модельдерді оқыту алдында деректер алдын ала өңдеуден өткізілді: жетіспейтін мәндер толықтырылды, аномалиялық бақылаулар түзетілді және ақпараттық белгілер қалыптастырылды. Осыдан кейін деректер жиыны оқыту және тест жиындарына бөлінді. Оқыту жиыны модель параметрлерін анықтау үшін, ал тест жиыны модельдің бұрын қолданылмаған деректердегі болжау қабілетін бағалау үшін пайдаланылды.

Деректерді бөлу кезінде уақыттық қатардың реттілігі сақталды, себебі атмосфералық ауа сапасының көрсеткіштері уақыт бойынша тәуелді болып келеді. Сондықтан болашақ кезеңдегі мәндердің оқыту жиынына өтіп кетуіне жол берілмеді. Жалпы деректердің 80%-ы оқыту жиынына, 20 %-ы тест жиынына бөлінді.

Гиперпараметрлерді таңдау кезінде 5-қатпарлы кросс-валидация қолданылды. Бұл тәсіл модель нәтижелерінің бір ғана кездейсоқ бөлініске тәуелді болмауын қамтамасыз етеді және қайта үйрену қаупін азайтады. Әрбір модель үшін сапа көрсеткіштері бірнеше қатпар бойынша есептеліп, кейін орташа мәндер негізінде ең тиімді параметрлер таңдалды.

Жұптық Student t-тесті қолданылды. Статистикалық маңыздылық деңгейі $p < 0,05$ деп қабылданды.

Зерттеудің мақсаты бір ғана «ең жақсы» модельді анықтау емес, заманауи ансамбльдік алгоритмдердің ауа ластануын болжаудағы тиімділігін бағалау болды. Алгоритмдерді таңдау ғылыми әдебиеттерді талдау негізінде жүзеге асырылды, онда бұл әдістер регрессиялық есептер мен үлкен көлемдегі деректерді өңдеу барысында жоғары нәтижелер көрсеткен.

Зерттеу аясында үш ансамбльдік алгоритм қарастырылды:

XGBoost – шешім ағаштарына негізделген стохастикалық градиенттік бустинг алгоритмі. Әрбір келесі ағаш алдыңғы модельдің қателерін түзету мақсатында құрылады және модель сапасы жақсармайынша оқыту жалғасады.

LightGBM – жоғары өнімді градиенттік бустинг фреймворкі. Ол оқыту жылдамдығының жоғары болуымен, жадыны тиімді пайдалануымен, дәлдігімен, параллель және үлестірілген есептеулерді қолдауымен ерекшеленеді. Соңғы зерттеулерге сәйкес, LightGBM салыстырмалы дәлдік жағдайында XGBoost алгоритмінен жылдамырақ жұмыс істейді.

HistGradientBoosting – гистограммалық тәсілге негізделген градиенттік бустинг алгоритмі. Бұл әдіс LightGBM қағидаларына сүйене отырып әзірленген, жетіспейтін мәндермен автоматты түрде жұмыс істей алады және белгілерді гистограммалық түрде көрсету арқылы есептеу шығындарын азайтады.

Аталған ансамбльдік әдістер (XGBoost, LightGBM және HistGradientBoosting) күрделі бейсызық тәуелділіктерді тиімді анықтайды және үлкен көлемдегі деректермен жұмыс істеуге жақсы бейімделген.

Модель гиперпараметрлерін баптау.

Ансамбльдік модельдердің сапасын арттыру үшін гиперпараметрлерді таңдау GridSearchCV (5-fold cross-validation) әдісі арқылы жүргізілді. Бұл әдіс әртүрлі параметрлер комбинацияларын тексеруге және кросс-валидация нәтижелері бойынша ең тиімді конфигурацияны таңдауға мүмкіндік береді.

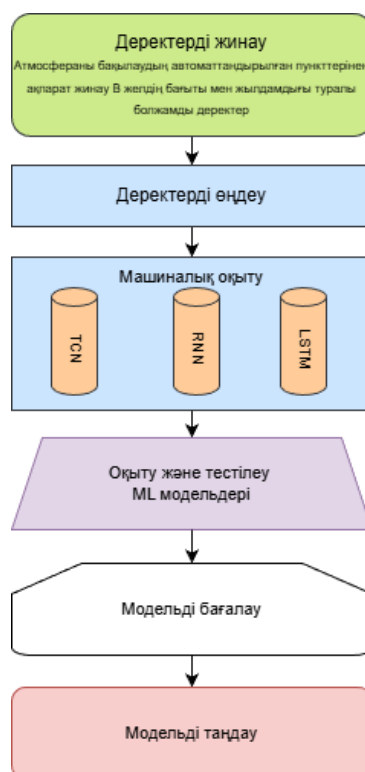
XGBoost моделі үшін келесі гиперпараметрлер қарастырылды: ағаштар саны ($n_estimators = 100, 200, 300$), ағаштың максималды тереңдігі ($max_depth = 4, 6, 8$), оқыту жылдамдығы ($learning_rate = 0.01, 0.05, 0.1$), бақылаулардың үлесі ($subsample = 0.8, 1.0$) және белгілердің үлесі ($colsample_bytree = 0.8, 1.0$).

LightGBM моделі үшін жапырақтар саны ($num_leaves = 31, 63$), ағаш тереңдігі ($max_depth = -1, 6, 10$), оқыту жылдамдығы ($learning_rate = 0.01, 0.05, 0.1$), ағаштар саны ($n_estimators = 100, 200, 300$) және белгілер үлесі ($feature_fraction = 0.8, 1.0$) тексерілді.

Ең тиімді параметрлер MAE және RMSE көрсеткіштері бойынша ең төмен қате берген конфигурациялар негізінде таңдалды.

Модель құру кезеңдері (1-сурет):

1. Деректерді жинау (Data collection)
 - ауа сапасын автоматты бақылау станцияларының өлшемдері;
 - метеорологиялық параметрлер (жел бағыты мен жылдамдығы және т.б.);
 - ластаушы заттардың концентрациялары (PM2.5, NO₂ және т.б.).
2. Деректерді алдын ала өңдеу (Data processing)
 - аномалияларды жою және жетіспейтін мәндерді толықтыру;
 - деректерді нормализациялау немесе стандарттау;
 - жаңа белгілерді генерациялау (жел векторы, температура, уақыт белгілері);
 - категориялық айнымалыларды кодтау.
3. Алгоритмді таңдау (Machine learning)
 - ықтимал модельдер: XGBoost, LightGBM, HistGradientBoosting.
4. Модельді оқыту және тексеру (Model training & testing)
 - модельді оқыту деректердің оқыту жиынтығында жүргізіледі;
 - нәтижелер тәуелсіз тест деректерінде тексеріледі.
5. Модель сапасын бағалау (Model evaluation)
 - бағалау метрикалары: MAE, RMSE және R².
6. Қорытынды модельді таңдау (Model selection)
 - болжам дәлдігі жоғары және қайта үйрену (overfitting) байқалмайтын модель анықталып, кейін ауа ластануын нақты уақыт режимінде болжау үшін қолданылады.



1-сурет. Машиналық оқыту моделін таңдау алгоритмі

Ескерту – авторлар құрастырған

XGBoost, LightGBM және HistGradientBoosting алгоритмдерінің үшеуі де ауа ластануын талдауға байланысты регрессиялық және классификациялық есептерде жоғары дәлдік көрсете алады.

XGBoost алгоритмі тұрақтылығы мен тиімді регуляризация механизмдерінің арқасында эталондық әдістердің бірі болып саналады. Алайда үлкен көлемдегі деректермен жұмыс істеген кезде оның оқыту жылдамдығы LightGBM алгоритмімен салыстырғанда баяуырақ болуы мүмкін.

LightGBM алгоритмі ірі және жоғары өлшемді деректер жиынтықтарында, сондай-ақ категориялық белгілер болған жағдайда ерекше тиімділік көрсетеді.

Ал HistGradientBoosting алгоритмі scikit-learn кітапханасына енгізілген және Python жобаларында тез прототип жасауға әрі оңай интеграциялауға ыңғайлы. Бұл әдіс әсіресе қарапайымдылық пен нәтижелердің қайта өндірілу мүмкіндігі маңызды болған жағдайда тиімді қолданылады.

Модельдердің нәтижелеріне сенімді болу және қайта үйрену (overfitting) қаупін азайту үшін олардың сапасын дұрыс және жан-жақты бағалау өте маңызды. Осы мақсатта модельдердің өнімділігі бірқатар статистикалық көрсеткіштер бойынша салыстырмалы түрде талданады (1-кесте).

1-кесте. MAE, RMSE және R^2 метрикалары бойынша модельдерді салыстыру

Әдіс	Модель түрі	Бейсызық тәуелділікті ескеру	Уақыттық тәуелділікті ескеру	Интерпретациялануы	Есептеу күрделілігі	Үлкен деректерге қолданылуы
XGBoost	Градиенттік бустинг	Иә	Жоқ	Орташа	Жоғары	Иә
LightGBM	Градиенттік бустинг	Иә	Жоқ	Орташа	Төмен (оңтайландырылған)	Иә
HistGradient Boosting	Градиенттік бустинг	Иә	Жоқ	Орташа	Орташа	Иә
Ескерту – авторлар Zhang, Y., et al. (2021) деректері негізінде жасаған						

Модельдердің сапасын сандық тұрғыдан бағалау үшін келесі метрикалар пайдаланылады (2-кесте):

2-кесте. Модель сапасын бағалау көрсеткіштері

Аббревиатура	Толық атауы	Сипаттамасы
MAE	Mean Absolute Error (Орташа абсолюттік қате)	Болжамдалған және нақты мәндер арасындағы абсолюттік қателердің орташа шамасын көрсетеді.
RMSE	Root Mean Squared Error (Орташа квадраттық қате)	Қателердің орташа квадраттық мәнінің түбірі. Үлкен қателіктерге MAE-ге қарағанда сезімталырақ.
R^2	Coefficient of Determination (Детерминация коэффициенті)	Тәуелді айнымалы дисперсиясының қандай үлесі модель арқылы түсіндірілетінін көрсетеді. 1-ге жақын мән модель сапасының жоғары екенін білдіреді.
Ескерту – автор Беспалов О.В. (2020) деректері негізінде жасаған		

1. Детерминация коэффициенті (R^2) – тәуелді айнымалы дисперсиясының модель арқылы түсіндірілетін үлесін көрсетеді. Бұл көрсеткіш 0 мен 1 аралығында өзгереді: 1 мәні болжамдардың нақты деректермен толық сәйкес келетінін білдіреді, ал 0 модельдің болжамдық қабілетінің жоқ екенін көрсетеді.

2. Орташа абсолюттік қате (MAE) – болжамдалған және нақты мәндер арасындағы абсолюттік ауытқулардың орташа шамасын сипаттайды.

3. Орташа квадраттық қате (RMSE) – ауытқулардың квадраттарын ескере отырып есептеледі. Сондықтан бұл көрсеткіш үлкен қателіктерге анағұрлым сезімтал болып келеді және болжам мен бақылау арасындағы орташа қашықтықты сипаттайды.

НӘТИЖЕЛЕР ЖӘНЕ ОЛАРДЫ ТАЛҚЫЛАУ

Өзірленген бұлттық платформа икемділігімен, ауқымдалу мүмкіндігімен және тұрақтылығымен ерекшеленеді. Жүйе жаңа дереккөздерді оңай қосуға мүмкіндік береді, атмосфералық жағдай туралы деректерді жинау, өңдеу және талдау үдерістерін толық автоматтандырады және пайдаланушылар саны мен телеметриялық деректер көлемі артқан жағдайда да жоғары өнімділікті сақтайды. Нәтижесінде пайдаланушылар нақты уақыт режимінде сенімді әрі жедел ақпарат алады, ал болжам жасау және визуализация модульдері сервисті мамандар үшін де, кең аудитория үшін де қолайлы құралға айналдырады.

Атмосфералық ауаның ластануын болжау және мониторингтеу функцияларын іске асыру үшін бұлттық сервис құрамында келесі негізгі ішкі жүйелер анықталды:

1. Деректерді жинау ішкі жүйесі

Бұл ішкі жүйе атмосфера жағдайын бақылауға арналған автоматтандырылған станциялардан алынатын деректерді, сондай-ақ желдің бағыты мен жылдамдығына қатысты болжамдық ақпаратты тұрақты түрде жинауға жауап береді. Деректер белгіленген уақыт кестесіне сәйкес арнайы сыртқы сервистер арқылы алынады.

2. Математикалық модельді шешу ішкі жүйесі

Ұсынылған бөлшек-дифференциалдық стохастикалық модель негізінде атмосферадағы ластанушы заттардың таралуын сандық әдістермен есептейтін алгоритмдерді жүзеге асырады. Бұл модуль уақыт пен кеңістіктегі қоспалардың таралу динамикасын модельдейді.

3. Кезектерді басқару серверінің ішкі жүйесі

Параллель есептеулердің орындалуын үйлестіреді. Ол жоғары өнімді есептеу тораптарында әртүрлі бастапқы деректермен математикалық модель шешушісін бірнеше рет іске қосады. Есептеулер аяқталғаннан кейін алынған нәтижелерді өңдеу және статистикалық агрегаттау процесін бастайды.

4. Турбуленттілік коэффициентін анықтау ішкі жүйесі

Атмосферадағы турбулентті тасымал коэффициенттерін есептеу үшін машиналық оқыту алгоритмдері мен соңғы элементтер әдісіне негізделген сандық модельдеуді біріктіретін гибридік тәсілді қолданады.

5. Пайдаланушылармен өзара әрекеттесу ішкі жүйесі

Сервис нәтижелерін көрсету және басқару үшін арналған веб-интерфейстен, сондай-ақ сыртқы әзірлеушілер мен жүйелерге деректерге қол жеткізу мүмкіндігін беретін бағдарламалық API интерфейсінен тұрады.

6. Талдау және болжау ішкі жүйесі

Математикалық модель шешушісінен алынған нәтижелерді өңдейді: статистикалық талдау жүргізеді, деректерді біріктіреді және визуализациялайды. Сонымен қатар метеорологиялық факторларды ескере отырып ластанушы заттардың таралуын болжайды және пайдаланушылар үшін есептер мен графикалық материалдар дайындайды.

7. Мониторинг және журнал жүргізу ішкі жүйесі

Жүйенің жұмыс көрсеткіштерін бақылап, сервистің барлық әрекеттерін тіркейтін журналдарды жинайды. Сонымен қатар ықтимал қателер немесе ақаулар туралы әкімшілерге хабарлама жіберуді қамтамасыз етеді.

8. Қауіпсіздік және қолжетімділікті басқару ішкі жүйесі

Деректердің қорғалуын қамтамасыз етеді, пайдаланушыларды аутентификациялау және авторизациялау механизмдерін іске асырады. Қолжетімділік құқықтары белгіленген рөлдер мен қауіпсіздік саясаттарына сәйкес басқарылады.

Жүйе компоненттерінің өзара әрекеттесу сұлбасы (2-сурет)

1. Деректерді жинау

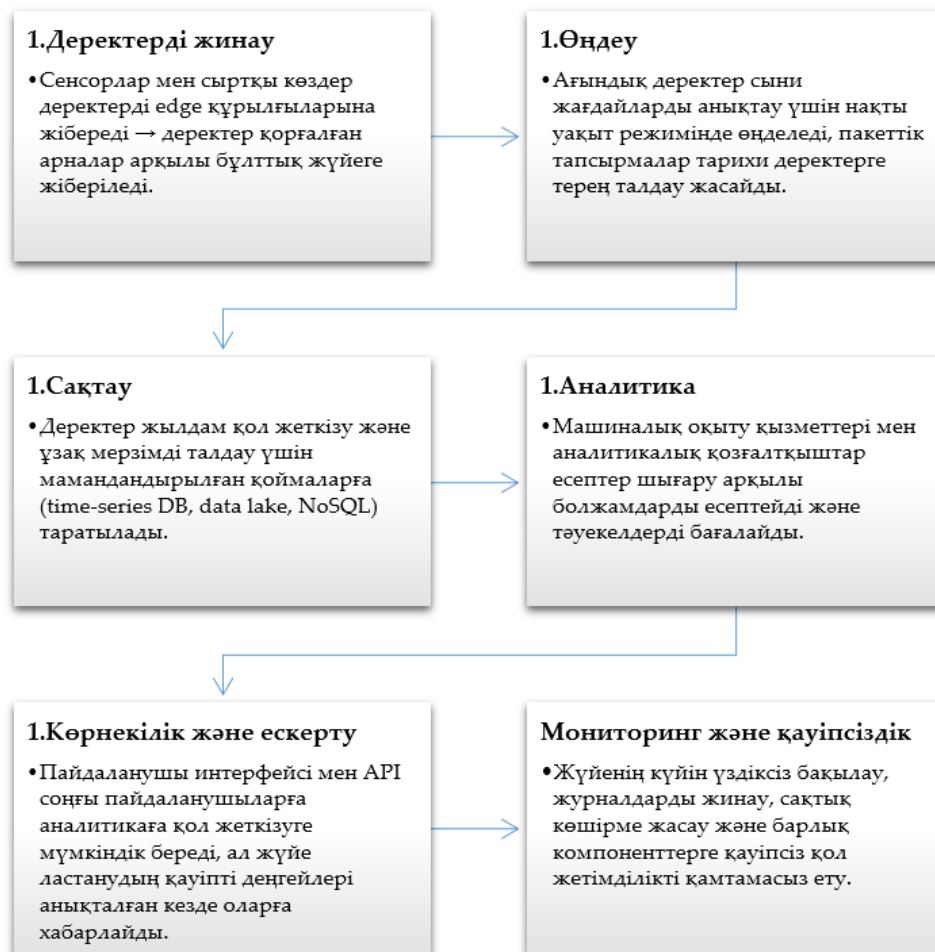
Датчиктер мен сыртқы ақпарат көздері өлшеу нәтижелерін edge-тораптарға жібереді, содан кейін деректер қорғалған байланыс арналары арқылы бұлттық ортаға жеткізіледі.

2. Деректерді өңдеу

Ағындық деректер жедел түрде талданып, қауіпті жағдайларды дереу анықтауға мүмкіндік береді, ал пакет режиміндегі есептер тарихи деректер жиынтығын терең талдауға бағытталған.

3. Деректерді сақтау

Ақпарат әртүрлі мақсаттарға арналған арнайы сақтау жүйелеріне бөлінеді: кейбірі жылдам қол жеткізу үшін, ал басқалары ұзақ мерзімді архив ретінде пайдаланылады.



2-сурет. Өзара әрекеттесу схемасы

Ескерту – авторлар құрастырған

4. Аналитика

Машиналық оқыту модельдері мен аналитикалық жүйелер болжау жүргізеді, ықтимал тәуекелдерді бағалайды және есептер қалыптастырады.

5. Визуализация және хабарландырулар

Веб-интерфейс пен API пайдаланушыларға талдау нәтижелеріне қол жеткізуге мүмкіндік береді. Егер көрсеткіштер белгіленген шектен асса, жүйе автоматты түрде хабарламалар жібереді.

6. Мониторинг және қауіпсіздік

Барлық сервистердің жұмысы тұрақты түрде бақыланып, журналдар жиналады, резервтік көшіру жүзеге асырылады және жүйе компоненттеріне қауіпсіз қолжетімділік аутентификация мен авторизация құралдары арқылы қамтамасыз етіледі.

Деректер базасының сызбасы қалыптастырылды. Есептеу сервисін жобалау барысында оның барлық ішкі жүйелері ТОФИ бағдарламалық платформасының негізінде жұмыс істейтіні қарастырылды. Бұл платформа интерфейстерді құруды, визуализация параметрлерін баптауды, пайдаланушылардың қолжетімділік құқықтарын басқаруды және жүйені әкімшілендіруді жеңілдетеді. Сонымен қатар компоненттер арасындағы деректер алмасу мен байланыстарды сипаттайтын архитектура әзірленді. Ішкі жүйелер нақты анықталған функциялық сигнатуралары бар API арқылы өзара әрекеттеседі, ал орталықтандырылған сақтау орны ретінде ортақ деректер базасы пайдаланылады. Сервис SaaS үлгісіне сәйкес құрылған, сондықтан пайдаланушы жетілдірілген бөлшек-дифференциалдық стохастикалық модель бойынша дайын есептеу нәтижелерін өз аппараттық ресурстарын пайдаланбай-ақ алады.

Деректерді қабылдау модулі пайдаланушыдан алынған ақпаратты жинап, оны бастапқы өңдеуден және форматтаудан өткізеді, содан кейін деректер ағынын кезектер серверіне жібереді. Кезектер сервері тапсырмаларды бос есептеу тораптарына үлестіріп, сандық есептеу шешушісін іске қосады. Бұл кезеңде деректер соңғы элементтер әдісі арқылы өңделеді. Есептеу барысында шешуші арнайы модуль арқылы анықталған турбуленттілік коэффициенттерін пайдаланады. Алынған нәтижелер веб-интерфейс арқылы визуализацияланып, пайдаланушыларға түсінікті түрде ұсынылады және сервисті басқаруды жеңілдетеді.

Деректер базасын басқару жүйесі ретінде PostgreSQL қолданылды. Бұлттық есептеу сервисінің негізгі компоненттері, олардың атқаратын қызметтері мен өзара әрекеттесу байланыстары 3-кестеде көрсетілген.

3-кесте. Бұлттық есептеу сервисінің компоненттері

Компонент	Мақсаты	Өзара әрекеттесу
1. Пайдаланушының сервис функцияларымен өзара әрекеттесу ішкі жүйесі	Деректерді енгізу және нәтижелерді көрсету	API шлюзімен, кезектер серверімен және қауіпсіздік жүйесімен байланысады
2. Қауіпсіздік ішкі жүйесі (OAuth 2.0, JWT, TLS)	Аутентификация, авторизация және деректерді қорғау	Кезектерге, деректер базасына және API-ға қолжетімділікті басқарады
3. Деректерді алу ішкі жүйесі (Requests, AIOHTTP)	Пайдаланушы деректерін қабылдау және валидациялау	Тапсырмаларды кезектер серверіне жібереді
4. Кезектер серверінің ішкі жүйесі (RabbitMQ, Celery)	Тапсырмаларды есептеу тораптарына асинхронды түрде бөлу	Математикалық шешушіні іске қосады
5. Шешуші ішкі жүйе (Julia, Ferrite, CUDA, Pardiso)	Сандық модельдеу және соңғы элементтер әдісі арқылы есептеу	Кезектер серверінен және турбуленттілік модулінен деректер алады
6. Турбуленттілік коэффициентін анықтау ішкі жүйесі (Python, XGBoost, Scikit-learn)	Атмосфера параметрлерін болжау	Нәтижелерді шешушіге жібереді
7. Талдау және болжау ішкі жүйесі (Python, R, Pandas, Matplotlib, Tableau, Power BI)	Нәтижелерді өңдеу, визуализациялау және есептер дайындау	Нәтижелерді UI, деректер базасы және журнал жүйесіне жібереді

3-кестенің жалғасы

Компонент	Мақсаты	Өзара әрекеттесу
8. Мониторинг және журнал жүргізу ішкі жүйесі (Prometheus, Grafana, ELK)	Метрикаларды, қателерді және журналдарды бақылау	Барлық ішкі жүйелерден деректер қабылдайды
9. Қауіпсіздік және қолжетімділікті басқару ішкі жүйесі	Қауіпсіздік саясатын және қолжетімділікті басқару	Барлық компоненттерде қолданылады
10. Деректер базасы (PostgreSQL)	Барлық деректерді орталықтандырылған сақтау	Барлық ішкі жүйелер пайдаланады
Ескерту – авторлар Zhang, Y., et al. (2021). деректері негізінде жасаған		

Негізгі өзара әрекеттесу арналары

1. Пайдаланушы интерфейсі (UI) → API → Деректерді алу модулі → Кезек → Сандық шешуші → UI.

Пайдаланушы веб-интерфейс арқылы жүйеге сұраныс жібереді. Бұл сұраныс API арқылы деректерді қабылдау модуліне жеткізіледі, кейін тапсырма кезекке орналастырылады. Одан кейін шешуші модуль есептеуді орындайды және нәтижелер қайтадан пайдаланушы интерфейсіне жіберіледі.

2. Турбуленттілік модулі → Сандық шешуші → Аналитика → UI / ДБ / Мониторинг.

Турбуленттілік коэффициенттері есептеліп, шешушіге беріледі. Шешушінің нәтижелері талдау жүйесінде өңделіп, деректер базасына сақталады, пайдаланушы интерфейсінде көрсетіледі және мониторинг жүйесі арқылы бақыланады.

3. Барлық компоненттер → Мониторинг және журнал жүйесі.

Әрбір сервис өзінің жұмыс көрсеткіштері мен журналдарын орталық мониторинг жүйесіне жіберіп отырады, бұл жүйенің жұмысын бақылауға және аудит жүргізуге мүмкіндік береді.

4. Ортақ PostgreSQL деректер базасы.

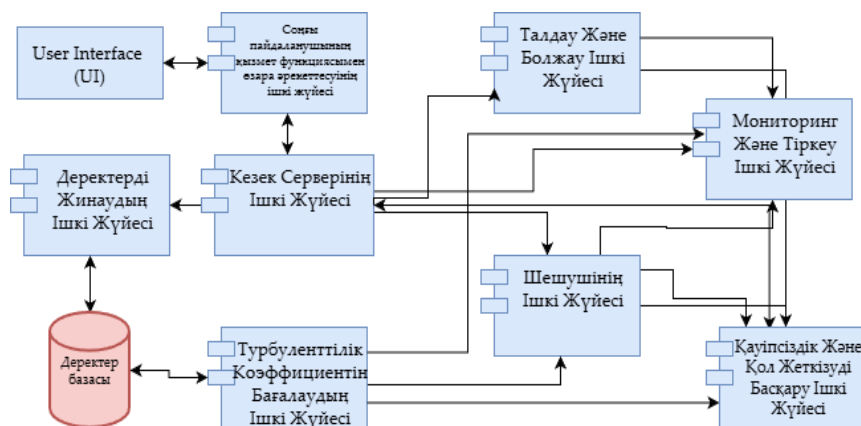
Барлық ішкі жүйелер бірыңғай деректер қоймасын пайдаланып, ақпарат алмасуды қамтамасыз етеді.

Аталған байланыстарды көрнекі түрде көрсету үшін 4-суретте бұлттық есептеу сервисінің компоненттік диаграммасы ұсынылған.

Әрбір ішкі жүйе үшін арнайы бағдарламалық құралдар жиынтығы таңдалды:

1. Деректерді алу ішкі жүйесі

Requests кітапханасы (синхронды сұраныстар) және AIOHTTP (асинхронды шақырулар) негізінде жүзеге асырылған.



4-сурет. Бұлтты есептеу қызметінің компоненттік диаграммасы

Ескерту – авторлар құрастырған

2. Математикалық модельді шешу ішкі жүйесі

Толығымен Julia тілінде жүзеге асырылған және Ferrite, CUDA, Pardiso сияқты кітапханаларды пайдаланады.

3. Турбуленттілік коэффициентін анықтау ішкі жүйесі

Python тілінде жазылған және XGBoost, scikit-learn және қосымша кітапханалар қолданылған.

4. Пайдаланушымен өзара әрекеттесу ішкі жүйесі

Python тіліндегі Django REST Framework платформасы негізінде жүзеге асырылған.

5. Талдау және болжау ішкі жүйесі

Python және R тілдері, сондай-ақ Jupyter, Pandas, Matplotlib, SciPy құралдары пайдаланылады. Деректерді визуализациялау үшін Tableau және Power BI қолданылған, ал болжау кіріктірілген машиналық оқыту модельдері арқылы жүзеге асырылады.

Бұл ішкі жүйе деректерді математикалық шешуші модулінен және турбуленттілік коэффициентін есептеу модулінен қабылдайды. Алынған нәтижелер келесі жүйелерге жіберіледі:

1. Пайдаланушымен өзара әрекеттесу ішкі жүйесіне (нәтижелерді көрсету үшін);

2. Мониторинг және журнал жүргізу ішкі жүйесіне (есептеу барысын және нәтижелерін тіркеу үшін);

3. Деректер қоймасына (есептеу нәтижелерін ұзақ мерзімді сақтау қажет болған жағдайда).

Мониторинг және журнал жүргізу ішкі жүйесі

Жүйе жұмысының негізгі метрикалары (жүктеме, кідірістер, қателер) Prometheus, Grafana немесе соған ұқсас бұлттық сервистер арқылы жиналады.

1. Журнал жазбалары ELK стеги (Elasticsearch + Logstash + Kibana) негізінде орталықтандырылып сақталады, бұл жүйедегі оқиғаларды жедел талдауға мүмкіндік береді.

2. Электрондық пошта, SMS және мессенджерлер арқылы автоматты хабарламалар жіберу механизмдері орнатылған, бұл сыни жағдайларға жылдам әрекет етуге мүмкіндік береді.

3. Бұл ішкі жүйе шешуші модульден, турбуленттілік модулінен, аналитика жүйесінен және басқа да компоненттерден деректер қабылдайды.

Қауіпсіздік және қолжетімділікті басқару ішкі жүйесі

1. Пайдаланушыларды аутентификациялау және авторизациялау OAuth 2.0 / JWT протоколдары арқылы жүзеге асырылады, барлық API сұраныстары қорғалған.

2. Деректерді беру кезінде TLS шифрлау әдісі қолданылады, ал сақтау барысында деректер базасы мен резервтік көшірмелер шифрланады.

3. Қауіпсіздік оқиғаларын бақылау және аномалияларды анықтау үшін SIEM жүйелері пайдаланылады.

Интеграция мүмкіндіктері

1.Пайдаланушы интерфейсі (UI) және бағдарламалық интерфейс (API);

2.Орталықтандырылған деректер қоймасы (деректерге қолжетімділікті басқару);

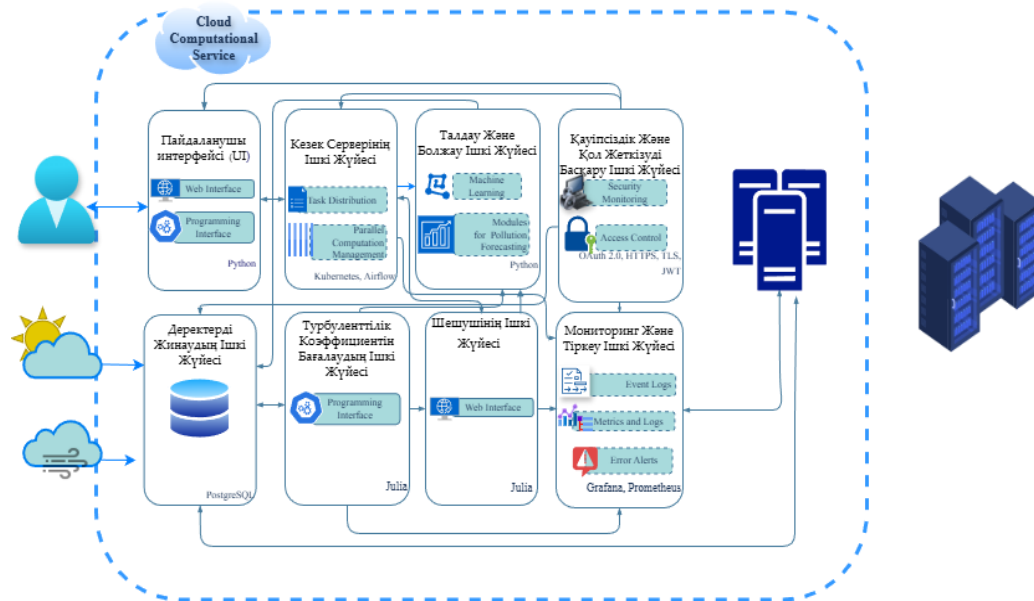
3.Кезектер сервері (пайдаланушы тапсырмаларын тексеру және басқару).

Осы компоненттердің барлығы бірігіп 5-суретте көрсетілген бұлттық есептеу сервисінің толық архитектурасын қалыптастырады.

Бұлттық ауа сапасын мониторингтеу платформасының мүмкіндіктері

Ауа ластануын бақылауға арналған бұлттық платформаны құру келесі мүмкіндіктерді қамтамасыз етеді:

1. Атмосфераның жай-күйі туралы деректерді әртүрлі көздерден - қалалық датчиктерден, метеостанциялардан және спутниктерден - жинақтап, оларды бір орталықта талдауға мүмкіндік береді.



5-сурет. Бұлтты есептеу қызметінің архитектурасы

Ескерту – авторлар құрастырған.

2. Ластану деңгейіне қатысты аналитикалық мәліметтер мен болжамдарды мемлекеттік органдарға, экологтарға, ғалымдарға және тұрғындарға жедел ұсынады.

3. Зиянды заттардың концентрациясын дәл болжау үшін есептеу ресурстарын көп қажет ететін машиналық оқыту алгоритмдерін қолдануға мүмкіндік береді.

4. Нақты уақыт режимінде мониторинг жүргізіп, қауіпті деңгейдегі шекті мәндердің асуын бірден тіркеп, жедел әрекет ету механизмдерін іске қосады.

5. Жергілікті серверлерге кететін шығындарды азайтып, бұлттық инфрақұрылым арқылы жүйені оңай масштабтауға мүмкіндік береді.

6. «Ақылды қала» платформалары, денсаулық сақтау жүйелері және басқа қалалық сервистермен интеграциялануға мүмкіндік береді.

7. Веб-порталдар мен мобильді қосымшалар арқылы көрнекі визуализациялар мен жедел хабарламалар ұсынып, азаматтардың экологиялық сауаттылығын арттырады.

Ұсынылған архитектура икемді, ауқымды және сенімді жұмыс істеуге бағытталған. Жүйе әртүрлі дереккөздерді біріктіріп, атмосфералық деректерді жинау, өңдеу және талдау үдерістерін автоматтандырады.

Эксперименттік деректер және модельдерді тексеру

Аналитика және болжау ішкі жүйесі болашақтағы ауа ластану деңгейін есептеу үшін машиналық оқыту алгоритмдерін пайдаланады. Ұсынылған әдістеме Өскемен қаласындағы ауа сапасының ұзақ мерзімді бақылау деректерінде тексерілді.

Бастапқы деректер өндірістік қала аумағының әртүрлі бөліктерінде орналасқан бес автоматтандырылған мониторинг станциясынан жиналды. Осылайша өндірістік аймақтар да, тұрғын аудандар да қамтылды. Өскемен қаласы үшін солтүстік бөліктегі өнеркәсіптік кәсіпорындар, қарқынды автокөлік қозғалысы және жеке сектордағы жылыту жүйелері жоғары экологиялық жүктеме туғызады. Датчиктердің орналасуы қала аумағындағы зиянды қоспалардың таралуын салыстырмалы түрде толық бағалауға мүмкіндік береді.

Зерттеуде Шығыс Қазақстан өңірінің атмосфералық ауа сапасына қатысты тарихи бақылау деректері пайдаланылды. Деректер автоматтандырылған мониторинг станцияларынан алынды және алдын ала тазалау, қалыпқа келтіру және аномалияларды өңдеу кезеңдерінен өтті.

Зерттеу деректер жиыны 2005–2025 жылдарды қамтиды және мыналарды қамтиды:
– жеті негізгі атмосфералық ластаушы заттың концентрациялары;
– метеорологиялық параметрлер (температура, қысым, жел бағыты мен жылдамдығы, салыстырмалы ылғалдылық);

– ластаушы заттардың таралуына әсер ететін ауа райы құбылыстарының кодтары.

Осылайша деректер жиыны өнеркәсіптік қаладағы ауа сапасын кеңістіктік-уақыттық болжау үшін репрезентативті негіз қалыптастырады.

Деректер жиынының көлемі.

Алдын ала өңдеу, жетіспейтін мәндерді жою және барлық ластаушы заттар бойынша бақылауларды біріктіру нәтижесінде қорытынды деректер жиыны 201854 бақылау жазбасынан тұрды. Оның ішінде 161483 жазба (80%) модельдерді оқыту үшін, ал 40371 жазба (20%) тәуелсіз тестілеу үшін пайдаланылды. Зерттеуде 6 кіріс белгісі қолданылды, ал мақсатты айнымалы ретінде күкірт диоксидінің (SO_2) концентрациясы таңдалды. Кіріс белгілеріне бақылау постының идентификаторы, CO , NO_2 , PM_{10} , $\text{PM}_{2.5}$ концентрациялары және уақыт белгісі негізінде қалыптастырылған уақыттық сипаттамалар енгізілді. Мұндай бөлу модельдердің тек оқыту деректерінде ғана емес, жаңа деректерде де қаншалықты тұрақты жұмыс істейтінін бағалауға мүмкіндік берді.

Модельдерді салыстыру нәтижелері

Болжамдардың дәлдігі мен тұрақтылығын бағалау үшін градиенттік бустингке негізделген үш алгоритм қолданылды: XGBoost; LightGBM; HistGradientBoosting.

Модельдердің сапасы MAE, RMSE және R^2 көрсеткіштері бойынша бағаланды.

4-кесте. MAE, RMSE және R^2 метрикалары бойынша модельдерді салыстыру

Модель	MAE	RMSE	R^2
XGBoost	2.13×10^2	2.73×10^2	0.887
LightGBM	2.15×10^2	2.83×10^2	0.896
HistGradientBoosting	2.32×10^2	2.91×10^2	0.880

Ескерту – авторлар Zhang, Y., et al. (2021). деректері негізінде жасаған

XGBoost көрсеткіші MAE және RMSE бойынша ең жақсы нәтиже көрсетті, ал LightGBM жоғарырақ R^2 мәніне ие болды. Сондықтан модельді таңдау нақты қолдану мақсатына байланысты жүзеге асырылады.

MAE. Ең төмен орташа абсолюттік қате XGBoost моделінде байқалды (2.13×10^2), бұл оның болжамдарының орташа қателігі ең аз екенін көрсетеді.

RMSE. Орташа квадраттық қате бойынша да XGBoost ең жақсы нәтиже көрсетті (2.73×10^2), яғни ол ірі қателіктердің әсерін тиімді төмендетеді.

R^2 . Ең жоғары детерминация коэффициенті LightGBM моделінде тіркелді (0.896), бұл модельдің деректер өзгергіштігін жақсырақ түсіндіретінін көрсетеді.

Нәтижелердің статистикалық тұрақтылығын бағалау.

Модельдердің нәтижелерін неғұрлым сенімді бағалау үшін 5-қатпарлы кросс-валидация нәтижелері қолданылды. Әр қатпарда модельдер қайта оқытылып, MAE, RMSE және R^2 көрсеткіштері есептелді. Кейін әрбір модель үшін осы көрсеткіштердің орташа мәндері анықталды.

Сонымен қатар модельдер арасындағы айырмашылықтың кездейсоқ емес екенін тексеру мақсатында жұптық статистикалық салыстыру жүргізілді. Нәтижелерді бағалау кезінде маңыздылық деңгейі $p < 0.05$ ретінде қабылданды. Бұл тәсіл алынған айырмашылықтардың статистикалық тұрғыдан сенімді екенін көрсетуге мүмкіндік береді.

Модельдер бойынша қорытынды:

1. XGBoost – MAE және RMSE көрсеткіштері бойынша ең дәл модель, сондықтан нүктелік болжамдардың дәлдігі маңызды болған жағдайда тиімді.

2. LightGBM – R^2 көрсеткіші жоғары болғандықтан, деректер құрылымын жақсы түсіндіретін модель.

3. HistGradientBoosting – қарастырылған метрикалар бойынша алдыңғы екі алгоритмнен төмен нәтиже көрсетті, бірақ басқа параметрлер немесе жағдайлар кезінде тиімді болуы мүмкін.

Модельді таңдау зерттеу мақсатына байланысты:

– егер негізгі мақсат - қателіктерді барынша азайту болса, XGBoost таңдалады;

– ал деректердің жалпы заңдылықтарын жақсы түсіндіру қажет болса, LightGBM тиімдірек.

ҚОРЫТЫНДЫ

Бұл жұмыста индустриялық аймақтардағы атмосфералық ауа сапасын мониторингтеу және болжауға арналған бұлттық есептеу сервисінің кешенді архитектурасы ұсынылып, тәжірибелік түрде тексерілді. Платформа SaaS үлгісі негізінде жасалған және келесі артықшылықтарды қамтамасыз етеді:

1. Икемділік пен ауқымдылық – микросервистік құрылым, контейнеризация және оркестрация есептеу ресурстарын оңай кеңейтуге және жаңа дереккөздерді қосуға мүмкіндік береді.

2. Әртүрлі деректер ағындарын біріктіру – жерүсті бақылау станцияларының деректері, метеорологиялық ақпарат және спутниктік бақылаулар PostgreSQL деректер базасында біріктіріледі.

3. Кеңейтілген аналитика – Julia тілінде жазылған шешуші атмосферадағы ластаушы заттардың таралуын сипаттайтын бөлшек-дифференциалдық стохастикалық модельді жүзеге асырады.

4. Жоғары болжам дәлдігі – XGBoost, LightGBM және HistGradientBoosting алгоритмдері 2005–2025 жылдардағы Өскемен қаласының деректерінде тексерілді. XGBoost ең төмен MAE = 2.13×10^2 және RMSE = 2.73×10^2 мәндерін көрсетті, ал LightGBM ең жоғары $R^2 = 0.896$ нәтижесіне жетті.

5. Жеделдік және қолжетімділік – сервис веб-интерфейс және API арқылы нақты уақыт режимінде ақпарат пен интерактивті визуализацияларды ұсынады.

6. Сенімділік және қауіпсіздік – Prometheus, Grafana және ELK негізіндегі мониторинг, резервтік көшіру және SIEM бақылауы жүйенің тұрақтылығы мен деректер қауіпсіздігін қамтамасыз етеді.

Ұсынылған шешімнің практикалық маңызы – басқару органдарына, зерттеушілерге және қалалық қызметтерге күрделі есептеу модельдерін жеке суперкомпьютерлік инфрақұрылымсыз қолдануға мүмкіндік беруінде. Бұл экологиялық шешімдерді қабылдауды жеделдетіп, халықтың ақпараттану деңгейін арттырады.

Ұсынылған тәсілдің негізгі артықшылығы – оның модульдік құрылымы, масштабталу мүмкіндігі және сыртқы ақпараттық жүйелермен интеграцияға бейімділігі. XGBoost, LightGBM және HistGradientBoosting модельдерін салыстыру нәтижелері ансамбльдік әдістердің ауа сапасын болжау міндеттерінде тиімді екенін көрсетті. Сонымен қатар кросс-валидация және статистикалық тексеру модель нәтижелерінің тұрақтылығын бағалауға мүмкіндік береді.

Осы нәтижелер ұсынылған архитектураның экологиялық мониторинг, Smart City және қалалық басқару жүйелері үшін практикалық маңызы бар екенін көрсетеді.

Болашақ зерттеулер бағыттарына мыналар жатады:

– мониторинг желісін жаңа датчиктермен кеңейту;

– платформаны басқа қалалар мен климаттық аймақтарға бейімдеу;

– модельдерді үздіксіз жаңарту үшін active learning әдістерін енгізу;

– платформаны көлік, денсаулық сақтау және басқа «ақылды қала» жүйелерімен интеграциялау.

МҮДДЕЛЕР ҚАЙШЫЛЫҒЫ: Авторлар мүдделер қайшылығы жоқ екенін мәлімдейді.

ЗЕРТТЕУГЕ ДЕЙІНГІ ЕРІКТІ КЕЛІСІМ: Зерттеуге адам қатысушылары тартылған жоқ, сондықтан ерікті ақпараттандырылған келісім алу талап етілмеді. Бұл бөлім қолданылмайды.

ДЕРЕКТЕРДІҢ ҚОЛЖЕТІМДІЛІГІ ТУРАЛЫ МӘЛІМДЕМЕ: Осы зерттеудің нәтижелерін растайтын деректер автордың негізделген сұрауы бойынша ұсынылады. Материалдарды алу үшін мақала авторына хабарласу қажет.

АЛҒЫС БІЛДІРУ: Авторлар мақаланың сапасын жақсартуға ықпал еткен құнды ескертулер үшін анонимді рецензенттерге алғыс білдіреді

ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫН ПАЙДАЛАНУ ТУРАЛЫ ХАБАРЛАМА: Осы мақаланы дайындау кезінде авторлар ғылыми мәтінді тұжырымдауға және редакциялауға, материалды құрылымдауға, сондай-ақ академиялық стиль талаптарына сәйкес тұжырымдарды нақтылауға көмектесу үшін генеративті жасанды интеллект технологияларын пайдаланды. Жасанды интеллект технологияларын қолдану көмекші сипатта болды және авторлық ғылыми талдауды, нәтижелерді түсіндіруді және тұжырымдарды тұжырымдауды алмастырмады. Мақаланың мазмұны, ұсынылған деректердің дұрыстығы және Академиялық этика нормаларының сақталуы үшін авторлар толығымен жауап береді (OpenAI (2026)).

ӘДЕБИЕТТЕР ТІЗІМІ

- Adebayo, H. O., Olanrewaju, O. M., & Ajayi, J. I. (2021). Evaluation of air quality prediction using machine learning methods. *Environmental Engineering Science*, 38(2), 87–94. <https://doi.org/10.1089/ees.2020.0429>
- Ahmed, N., Shukla, S. K. R., & Gupta, A. K. (2022). Evaluation of air pollution models for long-term forecasting in urban cities. *Environmental Pollution*, 289, 117905. <https://doi.org/10.1016/j.envpol.2021.117905>
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the fifth utility. *Future Generation Computer Systems*, 25(6), 599–616. <https://doi.org/10.1016/j.future.2008.12.001>
- Chen, J., Hu, Y., & Wu, L. (2019). Gradient boosting for air pollution forecasting. *Journal of Environmental Sciences*, 75, 101–112. <https://doi.org/10.1016/j.jes.2018.03.022>
- Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: Concepts, technology & architecture*. Prentice Hall.
- Gao, P., Chen, X., & Liu, W. (2023). Hybrid models combining machine learning and atmospheric physics. *Atmospheric Pollution Research*, 14, 221–239. <https://doi.org/10.1016/j.apr.2023.02.009>
- Ghobaei-Arani, M., Souri, A., & Rahmanian, A. (2020). Resource management approaches in fog computing: A comprehensive review. *Journal of Supercomputing*, 76(7), 4814–4850. <https://doi.org/10.1007/s11227-018-2710-2>
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of big data on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. <https://doi.org/10.1016/j.is.2014.07.006>
- Hosseini, S., Zhang, R., & Wang, P. (2020). Support vector regression for air quality prediction. *Applied Intelligence*, 50, 2948–2960. <https://doi.org/10.1007/s10489-020-01732-0>
- Jones, D. M. H., Skourides, P. R. J. S., & Hennessey, M. G. B. (2021). Application of machine learning algorithms for air quality forecasting: A review. *Environmental Modelling & Software*, 141, 105034. <https://doi.org/10.1016/j.envsoft.2021.105034>

- Kim, H., Park, J., & Seo, D. (2022). LSTM-based models for air quality forecasting. *Journal of Machine Learning Research*, 23, 1–17.
- Liu, X., Chen, H., & Li, J. (2018). Decision tree models for air quality assessment. *Science of the Total Environment*, 612, 1242–1253. <https://doi.org/10.1016/j.scitotenv.2017.09.014>
- Ma, H. T. M., Li, L., & Wang, Y. (2019). Air pollution forecasting using Random Forest and XGBoost. *Atmospheric Environment*, 215, 116909. <https://doi.org/10.1016/j.atmosenv.2019.116909>
- Mell, P., & Grance, T. (2011). The NIST definition of cloud computing (NIST Special Publication 800-145). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-145>
- Pahl, C., & Jamshidi, P. (2016). Microservices: A systematic mapping study. *Proceedings of the 6th International Conference on Cloud Computing and Services Science (CLOSER)*, 137–146. <https://doi.org/10.5220/0005785501370146>
- Shaban, K., Kadri, A., & Rezk, E. (2016). Air quality prediction using linear regression models. *Environmental Modelling & Software*, 83, 127–141. <https://doi.org/10.1016/j.envsoft.2016.05.017>
- Smith, J., et al. (2020). Cloud computing architectures: A comprehensive review. *IEEE Transactions on Cloud Computing*, 8(3), 456–469. <https://doi.org/10.1109/TCC.2020.2968459>
- Wang, L., Yu, X., & Fang, Y. (2021). Data preprocessing techniques in air pollution studies. *Environmental Data Science*, 5, 99–114. <https://doi.org/10.1017/eds.2021.8>
- Xiao, M., Sun, Y., & Li, Z. (2023). Integration of satellite data for air pollution mapping. *Advances in Environmental Monitoring*, 12, 87–102.
- Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18. <https://doi.org/10.1007/s13174-010-0007-6>
- Zhang, Q., Xiao, M., & Li, X. (2021). CNN models for spatial analysis of air pollution. *Remote Sensing of Environment*, 255, 112123. <https://doi.org/10.1016/j.rse.2020.112123>
- Zhang, Y., Wang, J., & Ma, L. (2017). Multiple linear regression models for air pollution prediction. *Atmospheric Environment*, 150, 273–285. <https://doi.org/10.1016/j.atmosenv.2016.11.027>
- Zhang, Y., et al. (2021). Scalability and performance optimization in cloud services. *Journal of Cloud Computing*, 10(1), 1–15. <https://doi.org/10.1186/s13677-021-00255-0>
- OpenAI. (2026, March 16). Assistance with academic writing, methodological structuring, and editing of a scientific article [Large language model]. ChatGPT. <https://chat.openai.com/>

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Тукушова Арайлым Ержанкызы – информатика магистрі, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан

Тукушова Арайлым Ержанкызы – магистр информатики, Восточно-Казахстанский технический университет им. Д.Серикбаева, г. Усть-Каменогорск, Казахстан, E-mail:

Tukushova Arailym – master of computer sciences, D. Serikbayev East Kazakhstan technical university, Ust-Kamenogorsk, Kazakhstan, E-mail: araylym.tukushova@gmail.com, ORCID: <https://orcid.org/0009-0002-7562-3897>



Рахметуллина Сауле Жадыгеровна – техникалық ғылымдар кандидаты, қауымдастырылған профессор Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан
Рахметуллина Сауле Жадыгеровна – кандидат технических наук, ассоциированный профессор, Восточно-Казахстанский технический университет им. Д. Серикбаева, г. Усть-Каменогорск, Казахстан,

Rakhmetullina Saule – PhD in Technical sciences, associate professor, D. Serikbayev East Kazakhstan technical university, Ust-Kamenogorsk, Kazakhstan,

E-mail: SRakhmetullina@edu.ektu.kz,

ORCID: <https://orcid.org/0000-0002-3142-0249>



Базарова Мадина Жомартовна – PhD, қауымдастырылған профессор, С. Аманжолов атындағы Шығыс Қазақстан университеті, Өскемен қ., Қазақстан

Базарова Мадина Жомартовна – PhD, ассоциированный профессор, Восточно-Казахстанский университет им. С.Аманжолова, г. Усть-Каменогорск, Казахстан

Bazarova Madina Zhomartovna – PhD, associate professor, East Kazakhstan University named after S. Amanzholov, Ust-Kamenogorsk, Kazakhstan,

e-mail: madina_vkgtu@mail.ru,

ORCID: <https://orcid.org/0000-0003-2580-6580>,



Пененко Алексей Владимирович – физика-математика ғылымдарының докторы, РҒА СБ Есептеу математикасы және математикалық геофизика институты, Новосибирск қ., Ресей,

Пененко Алексей Владимирович – доктор физико-математических наук, Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск, Россия,

Penenko Aleksey – Doctor of Physical and Mathematical Sciences, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia,

E-mail: a.penenko@yandex.ru,



Хасенова Зарина Төлеубековна – PhD, Цифрлық технологиялар және жасанды интеллект мектебінің деканы, Д.Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен қ., Қазақстан,

Хасенова Зарина Төлеубековна – PhD, декан Школы цифровых технологий и искусственного интеллекта, Восточно-Казахстанский технический университет имени Д. Серикбаева, г. Усть-Каменогорск, Казахстан,

Khassenova Zarina – PhD, Dean of the School of Digital Technologies and Artificial Intelligence, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan,

E-mail: zkhasanova@edu.ektu.kz,

ORCID: <https://orcid.org/0009-0001-1417-7793>