

УДК 004. 415

А.Э. Бекбоева, aitursunbekboeva@gmail.com

А.Л. Лайлиева, lailievaikanysh@gmail.com

Кыргызско-Германский институт прикладной информатики

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КАЧЕСТВА ВОЗДУХА В ГОРОДЕ БИШКЕК

Загрязнение атмосферного воздуха мелкодисперсными частицами PM_{2.5} остаётся одной из наиболее острых экологических проблем городов Центральной Азии. Бишкек, расположенный в котловине и подверженный температурным инверсиям, ежегодно фиксирует экстремальные уровни загрязнения в зимний период. В статье представлен комплексный анализ динамики концентрации PM_{2.5} и индекса качества воздуха (AQI) за 2019–2024 годы на основе данных мониторинга посольства США в Бишкеке. Для прогнозирования индекса качества воздуха и классификации уровней загрязнения применены методы машинного обучения - Random Forest, Gradient Boosting, Logistic Regression, а также ансамблевая стекинг-модель, интегрирующая их преимущества. Проведено сравнение моделей по ключевым метрикам (R^2 , RMSE, MAE, Accuracy, Precision, Recall, F1-score). Результаты продемонстрировали, что стекинг-подход обеспечивает более устойчивое и точное прогнозирование по сравнению с базовыми моделями. Исследование подтверждает эффективность интеграции алгоритмов машинного обучения в систему экологического мониторинга и прогнозирования качества воздуха.

Ключевые слова: PM_{2.5}, качество воздуха, машинное обучение, Random Forest, Gradient Boosting, Logistic Regression, стекинг-модели, Бишкек, прогнозирование AQI.

Введение

Проблема загрязнения воздуха является одной из ключевых экологических угроз для городов Центральной Азии. Бишкек, столица Кыргызстана, на протяжении последних лет входит в число наиболее загрязнённых городов мира. В холодный период концентрация взвешенных частиц PM_{2.5} часто превышает безопасные пределы, установленные Всемирной организацией здравоохранения (25 $\mu\text{g}/\text{m}^3$), более чем в 20–30 раз. Эти частицы, проникая в дыхательные пути, вызывают хронические заболевания дыхательной и сердечно-сосудистой систем, сокращая продолжительность жизни населения [1].

Основные причины загрязнения в Бишкеке – использование угля для отопления частных домов, высокая плотность автотранспорта и неблагоприятные метеорологические условия. Географическое расположение города в котловине способствует накоплению загрязняющих веществ, а температурные инверсии препятствуют их рассеиванию [12].

Несмотря на очевидную актуальность проблемы, системный анализ многолетних данных о загрязнении воздуха с применением современных инструментов анализа и прогнозирования в Кыргызстане практически отсутствует. Развитие технологий больших данных и машинного обучения открывает возможности для создания предиктивных моделей, способных прогнозировать динамику качества воздуха и поддерживать принятие решений в сфере экологической политики.

Целью данного исследования является разработка и сравнительный анализ моделей машинного обучения, предназначенных для прогнозирования индекса качества воздуха (AQI) и классификации уровней загрязнения в Бишкеке. Для достижения этой цели проведена предварительная обработка и исследовательский анализ данных мониторинга концентрации частиц PM_{2.5}, что позволило выявить их динамику и закономерности. Особое внимание уделено изучению временных и сезонных особенностей загрязнения воздуха, отражающих влияние климатических и антропогенных факторов. В дальнейшем были применены и сопоставлены базовые модели машинного обучения, а на основе их анализа разработана ансамблевая стекинг-модель, направленная на повышение точности и устойчивости прогнозов [6].

Материалы и методы

Данные исследования

Исследование опирается на данные официального мониторинга атмосферного воздуха, предоставленные станцией при посольстве США в Бишкеке. Период наблюдений охватывает январь 2019 – декабрь 2024 годов, что обеспечивает достаточную репрезентативность и надёжность статистической выборки. Итоговый датасет включает 42 855 почасовых записей, содержащих сведения о текущей концентрации взвешенных частиц PM_{2.5} (NowCast Conc.), исходной концентрации (Raw Conc.), рассчитанном индексе качества воздуха (AQI) и соответствующей категории загрязнения согласно классификации Агентства по охране окружающей среды США (EPA) [2].

Перед применением алгоритмов машинного обучения была выполнена комплексная предобработка данных. Исходные CSV-файлы, представляющие отдельные периоды наблюдений, были объединены в единый DataFrame. На этапе проверки структуры набора данных выявлены технические маркеры отсутствующих значений -999, используемые системой мониторинга для обозначения некорректных измерений. Эти значения были заменены на NaN, что позволило корректно обработать пропуски. Для восстановления числовых параметров применялись адаптивные стратегии: значения концентрации PM_{2.5}, относящиеся к текущим измерениям (NowCast Conc.), заполнялись средними значениями по соответствующему временному интервалу, а исходные концентрации (Raw Conc.) – медианными, что обеспечило устойчивость к выбросам. Категориальные пропуски в столбце AQI Category заменялись модальным значением – наиболее часто встречающейся категорией.

После обработки пропусков и исключения отрицательных значений, не имеющих физического смысла, данные были приведены к единому временному формату. Столбец Date (LT) был преобразован к типу datetime64[ns], после чего из него извлекались компоненты Year, Month, Day и Hour. Эти признаки отражают временные зависимости, включая суточные и сезонные циклы, и использовались в качестве входных переменных моделей машинного обучения.

На заключительном этапе подготовки данные были нормализованы с применением стандартизации (StandardScaler), при которой каждое значение приводилось к распределению с нулевым средним и единичной дисперсией. Процедура масштабирования выполнялась отдельно для обучающей и тестовой выборок во избежание утечки данных.

Исследовательский анализ данных

После этапа очистки и форматирования данных проведён исследовательский анализ, направленный на выявление закономерностей в распределении концентрации PM_{2.5}, как указано на рисунке 1 и индекса качества воздуха (AQI). Анализ выполнялся с использованием библиотек pandas, matplotlib и seaborn. Он охватывал оценку описательной статистики, исследование распределений, временных и сезонных тенденций, а также визуализацию выбросов [8-10].

Статистическая характеристика данных показала значительную изменчивость концентрации PM_{2.5} в пределах исследуемого периода. Среднее значение составило 31.83 $\mu\text{g}/\text{m}^3$ при стандартном отклонении 51.73 $\mu\text{g}/\text{m}^3$, медиана – 18 $\mu\text{g}/\text{m}^3$, а максимальные значения достигали 769.1 $\mu\text{g}/\text{m}^3$. Такое соотношение медианы и среднего указывает на правостороннюю асимметрию распределения, обусловленную частыми эпизодами экстремального загрязнения. Для уменьшения влияния выбросов и приведения распределения к нормальному виду использовалось логарифмическое преобразование данных, представленное на рисунке 1.

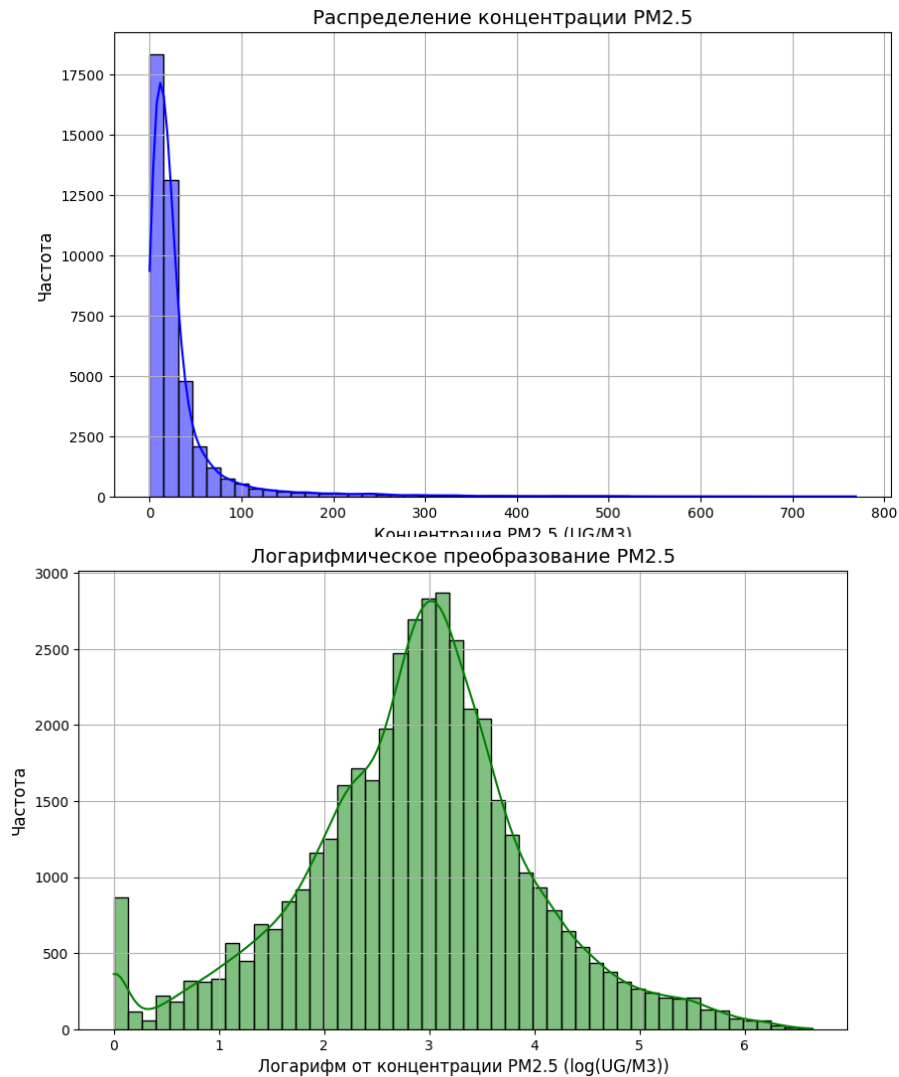


Рисунок 1 – Распределение концентрации PM2.5 с наложенной кривой плотности (до и после логарифмического преобразования)

Анализ выбросов проводился с использованием метода межквартильного размаха (IQR) и Z-оценки. Пороговое значение по IQR определялось как $Q3 + 1.5 \times IQR$, что позволило выявить около 10% наблюдений, выходящих за пределы нормальных диапазонов. Эти данные сохранялись в наборе как отражающие реальную экологическую динамику, а не ошибки измерений, поскольку пики загрязнения являются характерными для зимних месяцев.

Для визуализации выбросов применялась коробчатая диаграмма, приведенная на рисунке 2, демонстрирующая превышения над безопасным уровнем, установленным ВОЗ ($25 \mu\text{g}/\text{m}^3$) [3].

Исследование временных зависимостей, визуализированное на рисунке 3, позволило определить суточные, месячные и годовые паттерны загрязнения. Анализ показал чётко выраженные пики концентрации в вечерние часы (20:00–22:00), что совпадает с увеличением транспортной активности и началом отопления жилых зданий. Минимальные значения фиксировались утром (8:00–10:00), когда происходит активное перемешивание воздушных масс [12].

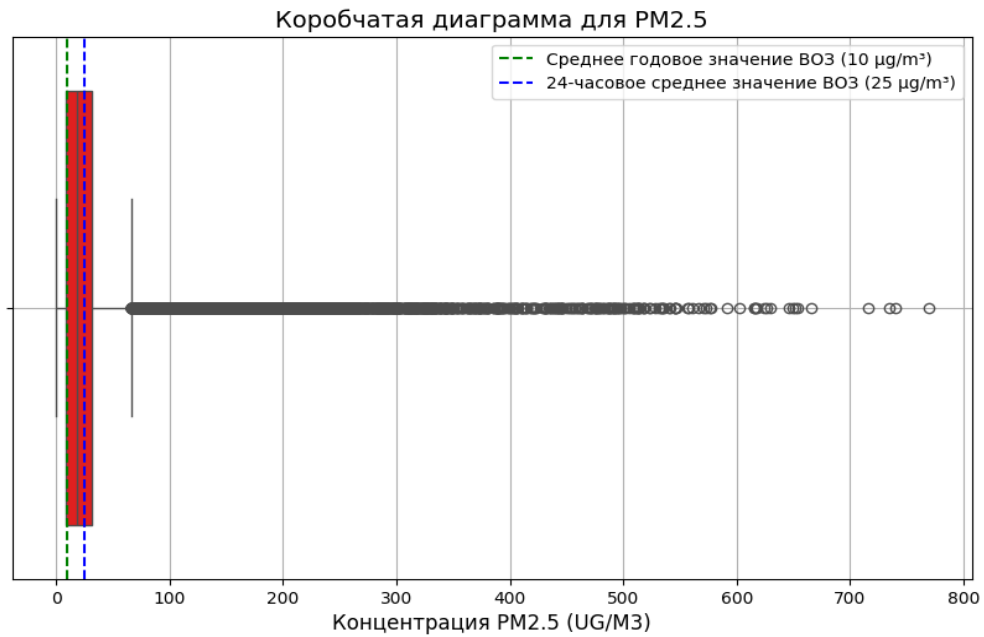


Рисунок 2 – Коробчатая диаграмма концентрации PM2.5 с выделением уровней ВОЗ

Сезонная динамика выявила сильную зависимость уровня загрязнения от времени года. Наибольшие значения наблюдались зимой – с декабря по февраль, когда концентрации PM2.5 превышали среднегодовой уровень в несколько раз. Весной и летом показатели снижались до минимальных значений. Эта закономерность подтверждает, что основной вклад в загрязнение воздуха вносят отопительные процессы и метеорологические факторы, характерные для холодного сезона.

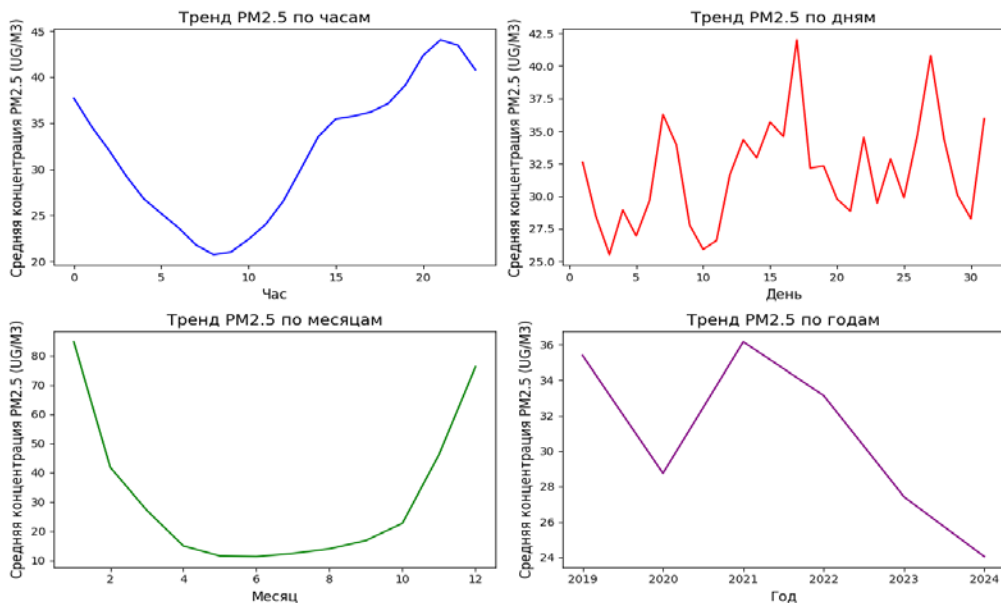


Рисунок 3 – Временные тренды концентрации PM2.5 по часам, дням, месяцам и годам

Для наглядного представления суточных и сезонных закономерностей использовались тепловые карты, приведенные ниже на рисунке 4, показывающие распределение концентраций PM2.5 в зависимости от времени суток и месяца. На диаграмме заметны чёткие вечерние пики в зимний период, а также устойчивая тенденция снижения концентраций летом [2].

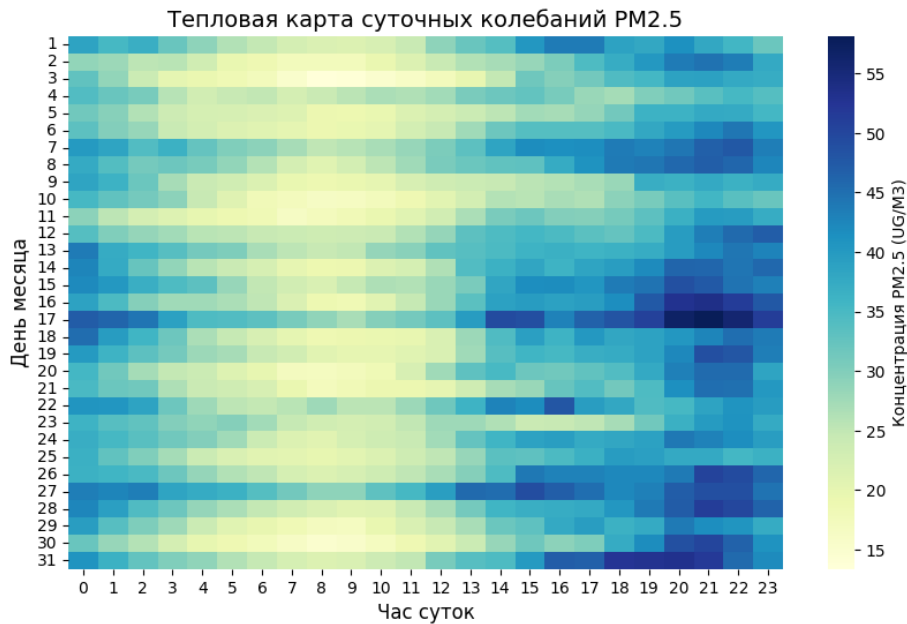


Рисунок 4 – Тепловая карта суточных и сезонных колебаний PM2.5

Анализ категорий индекса качества воздуха (AQI Category) показал, что значительная часть наблюдений относится к классам Moderate и Unhealthy for Sensitive Groups, что свидетельствует о частом превышении безопасных норм загрязнения. Это означает, что жители Бишкека на протяжении большей части года подвергаются воздействию воздуха, качество которого не соответствует рекомендациям Всемирной организации здравоохранения. Более высокие уровни загрязнения — Unhealthy, Very Unhealthy и Hazardous — встречаются реже, но фиксируются в периоды экстремального ухудшения экологической ситуации, главным образом в зимние месяцы, когда активизируется отопление углём и наблюдается температурная инверсия [12].

На заключительном этапе исследовательского анализа были выявлены ключевые статистические взаимосвязи между параметрами. Коэффициент корреляции между концентрацией PM2.5 и индексом AQI составил 0.97, что указывает на практически линейную зависимость между этими показателями. Вместе с тем зависимость AQI от времени суток и сезона носит нелинейный характер, отражающий влияние метеорологических и антропогенных факторов. Эти результаты подтвердили необходимость применения методов машинного обучения, способных учитывать сложные взаимосвязи между признаками и обеспечивать точное прогнозирование динамики загрязнения.

Таким образом, проведённый исследовательский анализ позволил сформировать комплексное представление о динамике и сезонности загрязнения воздуха в Бишкеке. Полученные результаты стали основой для выбора признаков и последующего построения моделей машинного обучения, направленных на прогнозирование индекса качества воздуха и классификацию уровней загрязнения.

Алгоритмы машинного обучения

После завершения исследовательского анализа данных была проведена разработка и обучение моделей машинного обучения, направленных на прогнозирование индекса качества воздуха и классификацию уровней загрязнения в Бишкеке. Для решения этих задач применялись методы машинного обучения, способные учитывать сложные взаимосвязи между концентрацией загрязняющих веществ и временными характеристиками. Установленные на этапе предварительного анализа нелинейные и сезонные зависимости потребовали использования алгоритмов, способных моделировать как линейные, так и нелинейные структуры данных.

Исходный набор данных был разделён на обучающую и тестовую выборки в соотношении 80:20 при фиксированном параметре `random_state=42`, что обеспечивало воспроизводимость экспериментов. В качестве входных признаков использовались числовые параметры – концентрация PM2.5 (NowCast Conc.) и временные характеристики (год, месяц, день, час), отражающие сезонные и суточные изменения загрязнения. Целевыми переменными выступали индекс AQI (для регрессионной задачи) и категория AQI Category (для классификации).

Выбор алгоритмов определялся их различными свойствами – устойчивостью, интерпретируемостью и способностью выявлять нелинейные закономерности. Модель Random Forest использовалась как ансамбль деревьев решений, формирующих итоговое предсказание путём усреднения результатов множества случайно построенных деревьев. Такой подход обеспечивает низкую чувствительность к выбросам и устойчивую работу при наличии коррелированных признаков [4]. Gradient Boosting представляет собой метод последовательного построения деревьев, где каждое новое дерево исправляет ошибки предыдущих, что делает алгоритм эффективным для сложных зависимостей, но требующим тщательной настройки параметров для предотвращения переобучения [5].

Для объединения преимуществ отдельных моделей была реализована ансамблевая стекинг-архитектура (Stacking). В её структуре базовыми уровнями служили Random Forest и Gradient Boosting, а роль мета-модели выполняли Linear Regression (для регрессии) и Logistic Regression (для классификации). Такое сочетание позволило использовать как точность нелинейных ансамблей, так и интерпретируемость линейных алгоритмов. Обучение стекинг-моделей проводилось с применением пятифолдинговой кросс-валидации, обеспечивающей надёжную оценку качества и снижение зависимости от случайного разбиения данных.

С целью повышения устойчивости к шуму, характерному для экологических данных, в процессе обучения использовалось добавление слабого гауссовского шума ($\sigma=0.05$), что выполняло роль регуляризации и способствовало улучшению обобщающей способности моделей. Качество предсказаний оценивалось на тестовой выборке с использованием стандартных метрик: для регрессии – средней абсолютной ошибки (MAE), среднеквадратичной ошибки (RMSE) и коэффициента детерминации (R^2); для классификации – точности (Accuracy), полноты (Recall), точности предсказаний (Precision) и F1-меры [7].

Таким образом, описанный подход к построению моделей машинного обучения позволил сформировать универсальную методологию прогнозирования индекса качества воздуха, объединяющую интерпретируемость линейных моделей и точность ансамблевых алгоритмов.

Результаты

Эффективность регрессионных моделей

В задаче регрессии рассматривались три модели: Random Forest Regressor, Gradient Boosting Regressor и Stacking Regressor, объединяющий их предсказания с использованием линейной мета-модели. Все алгоритмы показали высокое соответствие между предсказанными и фактическими значениями AQI, однако стекинг-ансамбль обеспечил наилучшие показатели точности и устойчивости. В таблице 1 приведены результаты эффективности регрессионных моделей.

Таблица 1 – Метрики

Модель	Выборка	MAE	RMSE	R^2
Random Forest Regressor	Обучающая	8.392	11.186	0.969
Random Forest Regressor	Тестовая	8.307	11.114	0.967

Gradient Boosting Regressor	Обучающая	11.717	18.531	0.915
Gradient Boosting Regressor	Тестовая	11.294	17.595	0.917
Stacking Regressor	Обучающая	4.130	6.617	0.989
Stacking Regressor	Тестовая	4.006	6.577	0.988

Снижение среднеквадратичной ошибки (RMSE) почти в два раза по сравнению с базовыми моделями подтверждает, что стекинг-архитектура эффективно компенсирует индивидуальные недостатки отдельных алгоритмов. Коэффициент детерминации R^2 близок к единице, что отражает высокую степень совпадения между прогнозами и наблюдаемыми значениями AQI.

Эффективность классификационных моделей

Во второй задаче – классификации категорий AQI – использовались Random Forest Classifier, Logistic Regression и Stacking Classifier, объединяющий их сильные стороны. Результаты оценки моделей по ключевым метрикам представлены ниже в таблице 2.

Таблица 2 – Метрики эффективности классификационных моделей

Model	Dataset	Accuracy		Precision	Recall	F1-Score
Random Forest Classifier	Train	0.875		0.586	0.508	0.525
Random Forest Classifier	Test	0.880		0.592	0.513	0.531
Logistic Regression	Train	0.961		0.965	0.928	0.946
Logistic Regression	Test	0.963		0.967	0.930	0.947
Stacking Classifier	Train	0.982		0.981	0.975	0.978
Stacking Classifier	Test	0.984		0.983	0.982	0.982

Как видно, стекинг-модель достигла наилучших результатов по всем метрикам. Высокие значения Precision и Recall указывают на способность корректно различать все категории загрязнения, включая редкие и экстремальные состояния воздуха.

Интерпретация и практическая значимость

Высокие значения коэффициента детерминации и метрик классификации подтверждают эффективность ансамблевых подходов при анализе экологических данных. Применение стекинг-моделей обеспечивает устойчивое прогнозирование даже при наличии выбросов и сезонных колебаний концентрации загрязняющих веществ. Пример тестирования на новых данных приведен ниже на рисунке 5.

```
Predicted AQI (RandomForestRegressor): [128.01537472]
Predicted AQI (GradientBoostingRegressor): [128.01537472]
Predicted AQI Category (RandomForestClassifier): [4]
Predicted AQI Category (LogisticRegression): [4]
```

Рисунок 5 – Пример предсказания индекса качества воздуха (AQI) и категории загрязнения воздуха, выполненного финальной стекинг-моделью машинного обучения при вводе реальных данных ($PM_{2.5} = 46.5 \mu g/m^3$, декабрь 2024 года).

Практическая проверка показала, что при вводе реальных данных (например, $PM_{2.5} = 46.5 \mu g/m^3$, декабрь 2024 года) финальная стекинг-модель предсказала значение $AQI \approx 128$, что соответствует категории *Unhealthy for Sensitive Groups*. Это демонстрирует согласованность алгоритмов и их готовность к использованию в системах оперативного мониторинга качества воздуха.

Таким образом, ансамблевая стекинг-модель показала наилучшие результаты как по точности прогнозирования индекса AQI , так и по классификации категорий загрязнения, что делает её перспективным инструментом для интеллектуальных систем экологического анализа и раннего предупреждения.

Научная новизна исследования

Полученные результаты послужили основой для формулировки научной новизны исследования. Научная новизна проведенного исследования заключается в комплексном применении методов машинного обучения для анализа и прогнозирования качества воздуха в условиях города Бишкек, где ранее подобные подходы использовались ограниченно. Впервые была построена ансамблевая стекинг-модель, объединившая преимущества *Random Forest* и *Gradient Boosting* с мета-моделью линейной регрессии. Такой подход позволил добиться высокой точности предсказаний как для регрессионных, так и для классификационных задач, обеспечив более надежную оценку динамики загрязнения воздуха по сравнению с отдельными моделями.

Новизна работы также состоит в адаптации алгоритмов машинного обучения к специфике климатических и экологических условий Центральной Азии, где значительное влияние на загрязнение оказывают сезонные факторы и особенности рельефа. Разработанная методология может быть применена не только к Бишкеку, но и к другим регионам со схожими проблемами – высоким уровнем загрязнения в зимний период и выраженными температурными инверсиями.

Кроме того, в исследовании предложен единый аналитический *pipeline*, включающий автоматизированную предобработку данных, исследовательский анализ, обучение и оценку ансамблевых моделей. Это делает подход воспроизводимым и расширяемым для задач экологического мониторинга на уровне муниципалитетов и исследовательских центров.

Практическое применение результатов

Благодаря этой воспроизводимости и масштабируемости результаты исследования приобретают реальную практическую ценность. Разработанная стекинг-модель может использоваться как интеллектуальный инструмент прогнозирования качества воздуха, предназначенный для оперативного оповещения населения о превышении предельно допустимых концентраций загрязняющих веществ. На основе такого решения возможно создание прототипа веб-платформы или мобильного приложения, автоматически формирующего краткосрочные прогнозы AQI и информирующего граждан о потенциальных рисках для здоровья.

Предложенная методика может применяться в рамках городских и региональных систем экологического мониторинга, обеспечивая анализ данных о загрязнении в режиме реального времени. Использование модели позволяет оценивать динамику загрязнения в отопительный сезон, выявлять закономерности и тенденции, а также прогнозировать периоды с повышенным уровнем загрязнения.

В перспективе результаты исследования могут быть интегрированы в инфраструктуру умных городов (*Smart City*), где анализ данных с IoT-датчиков качества воздуха и погодных станций будет выполняться автоматически. Это создаёт возможности для формирования адаптивных прогнозов, поддержки управленческих решений и разработки эффективных мер по снижению уровня загрязнения воздуха в городских условиях.

Ограничения исследования

Несмотря на достигнутые результаты, исследование имеет ряд ограничений, связанных с качеством и объёмом исходных данных. Используемый набор данных охватывает только один источник мониторинга – станцию при посольстве США в Бишкеке, что ограничивает пространственное представление загрязнения. Расширение сети сенсоров и включение данных с других станций позволило бы повысить пространственную точность прогнозов.

Вторым ограничением является отсутствие метеорологических параметров (температуры, скорости ветра, влажности, атмосферного давления), которые оказывают значительное влияние на рассеивание загрязняющих веществ. Включение этих факторов в будущие модели позволит повысить качество предсказаний и уточнить сезонные зависимости.

Кроме того, текущие модели ориентированы на краткосрочное прогнозирование AQI и не учитывают долгосрочные тренды, связанные с изменением климата, ростом населения и инфраструктуры. Для решения этой задачи перспективным направлением является использование гибридных моделей, объединяющих машинное обучение и методы временных рядов (например, LSTM или Prophet) [11].

Наконец, модели могут испытывать трудности при экстремальных значениях концентрации PM_{2.5}, когда наблюдаются выбросы за пределами обучающего диапазона. Для повышения устойчивости к подобным ситуациям в будущем планируется применение более сложных ансамблевых структур, включающих нейронные сети и методы оптимизации гиперпараметров.

Заключение

Проведённое исследование направлено на анализ загрязнения атмосферного воздуха в Бишкеке и построение моделей машинного обучения для оценки и прогнозирования индекса качества воздуха. Результаты обработки и анализа данных подтвердили наличие устойчивых закономерностей, связанных с сезонными изменениями концентраций PM_{2.5} и влиянием метеорологических факторов, определяющих динамику загрязнения.

В ходе работы разработаны и протестированы модели машинного обучения, обеспечивающие как прогнозирование значений AQI, так и классификацию уровней загрязнения. Применение ансамблевых подходов, основанных на стекинговой архитектуре, позволило повысить точность и устойчивость результатов. Полученные показатели подтверждают эффективность предложенного метода и его применимость для решения задач экологического моделирования и мониторинга.

Разработанная методика может быть использована при создании интеллектуальных систем анализа и краткосрочного прогнозирования качества воздуха в городских условиях. Дальнейшее развитие исследования связано с интеграцией пространственных и метеорологических данных, что обеспечит повышение точности прогнозов и расширит возможности практического применения модели в области управления качеством атмосферного воздуха.

Литература

1. Анализ качества воздуха в Бишкеке: распределение источников PM_{2.5} и меры по сокращению выбросов : [Электронный ресурс]. — Climate and Clean Air Coalition, 2019. — URL: <https://www.ccacoalition.org/ru/resources/air-quality-analysis-bishkek-pm25-source-apportionment-and-emission-reduction-measures> (дата обращения: 15.09.2024).
2. AirNow. U.S. Embassies and Consulates: Kyrgyzstan — Bishkek (historical air quality data, 2019–2024) : [Электронный ресурс]. — URL: [https://www.airnow.gov/international/us-embassies-and-consulates/#Kyrgyzstan\\$Bishkek](https://www.airnow.gov/international/us-embassies-and-consulates/#Kyrgyzstan$Bishkek) (дата обращения: 10.09.2024).

3. World Health Organization. WHO global air quality guidelines: particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. — Geneva : World Health Organization, 2021. — URL: <https://www.who.int/> (дата обращения: 30.10.2024).
4. Breiman L. Random Forests // Machine Learning. — 2001. — Vol. 45, No. 1. — P. 5–32. — DOI: 10.1023/A:1010933404324.
5. Chen T., Guestrin C. XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). — San Francisco, CA, 2016. — P. 785–794. — DOI: 10.1145/2939672.2939785.
6. Stacking to improve model performance: a comprehensive guide on ensemble learning in Python : [Электронный ресурс]. — Medium. — URL: https://medium.com/@brijesh_soni/stacking-to-improve-model-performance-a-comprehensive-guide-on-ensemble-learning-in-python-9ed53c93ce28 (дата обращения: 15.10.2024).
7. Performance metrics in machine learning: complete guide : [Электронный ресурс]. — Neptune.ai. — URL: <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide> (дата обращения: 01.10.2025).
8. McKinney W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. — 2nd ed. — Sebastopol, CA : O'Reilly Media, 2017. — 544 p.
9. Hunter J. D. Matplotlib: A 2D graphics environment // Computing in Science & Engineering. — 2007. — Vol. 9, No. 3. — P. 90–95. — DOI: 10.1109/MCSE.2007.55.
10. Waskom M. L. seaborn: statistical data visualization // Journal of Open Source Software. — 2021. — Vol. 6, No. 60. — Art. 3021. — DOI: 10.21105/joss.03021.
11. Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. — 1997. — Vol. 9, No. 8. — P. 1735–1780.
12. Казанцев М. Р., Верзунов С. Н. Методологические подходы к оценке влияния автотранспорта на состояние атмосферного воздуха в городе Бишкеке // Проблемы автоматизации и управления. — 2025. — № 2(53). — С. 74–87.