

Команда 13
Проектный практикум

Эдуард Исаханов
Артем Смирнов
Яков Чибинин
Владимир Ульзутуев

Предсказание индекса качества воздуха по данным Мосэкомониторинга с помощью ARIMA

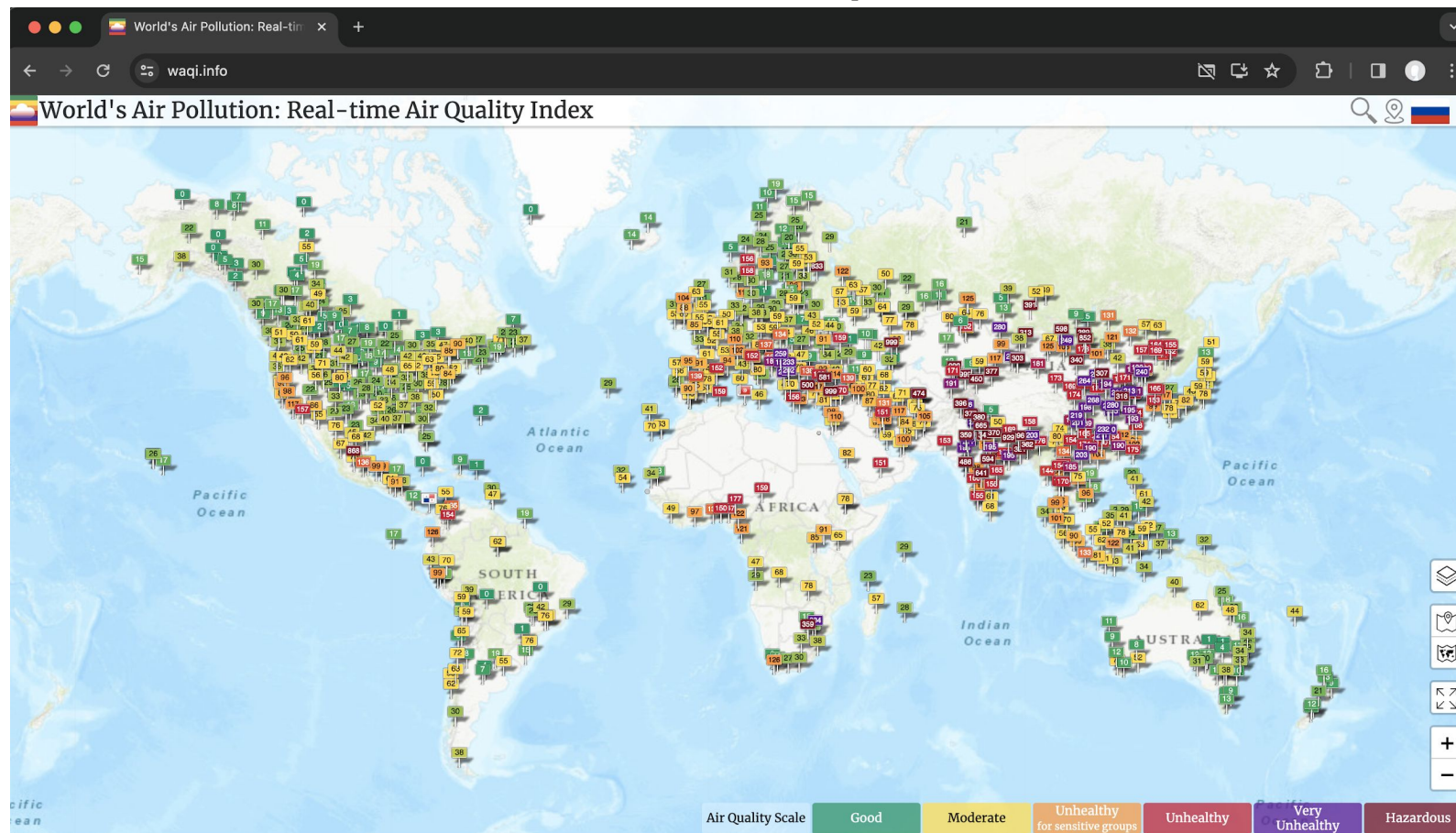
28 декабря 2023

Что такое индекс качества воздуха?

- Индекс качества воздуха (ИКВ, AQI) используется для того, чтобы сообщить людям о качестве воздуха в настоящий момент или о том, какое качество воздуха можно ожидать в будущем
- высокий ИКВ связано с рисками для общественного здоровья, и чем выше ИКВ, тем больше эти риски
- в разных странах существуют собственные ИКВ, которые соответствуют национальным стандартам качества воздуха

Какой ИКВ используем мы?

WAQI.Info: World Air Quality Index (на основе US EPA Air Now)



Базовые показатели ИКВ

Дневной цвет ИКВ	Уровень опасности	Значения ИКВ	Описание качества воздуха
Зеленый	Низкий	0-50	Качество воздуха удовлетворительное, а загрязнение воздуха представляет незначительный риск или не представляет его вовсе
Желтый	Умеренный	51-100	Качество воздуха приемлемое. Может представлять опасность для тех, кто особенно чувствителен к загрязнению воздуха
Оранжевый	Нездоровый для уязвимых групп	101-150	Представители уязвимых групп могут испытывать последствия для здоровья. Влияние на обычных людей менее вероятно
Красный	Нездоровый	151-200	Обычные люди могут испытывать последствия для здоровья; представители уязвимых групп могут испытывать более серьезные последствия для здоровья
Фиолетовый	Крайне нездоровый	201-300	Здоровье под угрозой: риск последствий для здоровья возрастает для всех
Бордовый	Опасный	301 и выше	Чрезвычайная ситуация: высока вероятность негативных последствий для здоровья всех

Пять основных загрязнителей

- Твердые частицы (PM2.5 и PM10)
- Приземный озон (O_3)
- Монооксид углерода (CO)
- Диоксид серы (SO_2)
- Диоксид азота (NO_2)

Наиболее полные данные – по твердым частицам, получаемые с датчиков частных лиц, негосударственных организаций и официальных служб.

Композитный ИКВ

World Air Quality Index использует композитный ИКВ по пяти загрязнителям, рассчитываемый по формуле:

$$\text{ИКВ} = \max(\text{ИКВ}_{\text{PM2.5}}, \text{ИКВ}_{\text{PM10}}, \text{ИКВ}_{\text{O}_3}, \dots)$$

Используемые данные

Исторические данные для 26 станций [“Мосэкомониторинга”](#),
бесплатно доступные на [WAQI.Info](#).

Ежедневный композитный ИКВ бесплатно доступен через API
[WAQI.Info](#).

Воздух ⇒ Данные измерительных станций

МОСЭКОМОНИТОРИНГ

Главная **Актуальное** Воздух Зелёные насаждения Метеоданные

← Воздух

ИЗМЕРИТЕЛЬНЫЕ СТАНЦИИ

Поиск по станциям

По округам **Списком** Показать на карте

Название	Адрес	Округ
АКАДЕМИКА АНОХИНА	город Москва, улица Академика Анохина, дом 38, корпус 1	ЗАО
БАЗОВСКАЯ	город Москва, Базовская улица, дом 15, корпус 5	САО
БИРЮЛЁВО	город Москва, Востряковский проезд, дом 11, корпус 1	ЮАО

JSON API

The JSON API can be used for advanced programmatic integration:

- Access to more than 11000 station-level and 1000 city-level data
- Geo-location query (based on latitude/longitude or IP address)
- Individual AQI for all pollutants (PM2.5, PM10, NO2, CO, SO2, Ozone)
- Station name and coordinates
- Originating EPA name and link
- Current weather conditions
- Stations within a map lat/lng bounds
- Search stations by name

```
$ curl -i "http://api.waqi.info/feed/shanghai/?to"
```

```
{
  status: "ok",
  data: {
    aqi: 70,
    time: {
      s: "2023-12-28 11:00:00"
    },
    city: {
      name: "Shanghai",
      url: "http://aqicn.org/city/shanghai/",
      geo: [
        "31.2047372",
        "121.4489017"
      ]
    }
  }
}
```

Сравнение методов предсказания

	Инерционный прогноз	Регрессия	Нейронные сети
Усилия по разработке	Низкие	Умеренные	Умеренные/высокие
Усилия по поддержанию	Низкие	Умеренные	Умеренные
Точность	Низкая	Умеренная/высокая	Умеренная/высокая
Метод описания	Сегодняшняя (вчерашняя) концентрация загрязнителя – это прогноз концентрации загрязнителя на завтра.	Уравнение регрессии предсказывает концентрацию загрязнителей	Нелинейные уравнения и веса предсказывают концентрацию загрязнителей
Сильные стороны	Хорошо работает там, где на протяжении нескольких дней подряд были высокие или низкие концентрации загрязнителей	Часто используется и проста в использовании	Позволяют находить нелинейные зависимости
Потенциальные ограничения	Не предсказывает начало или конец эпизода; низкая точность	Не способна точно предсказывать крайние концентрации	Не способны точно предсказывать крайние концентрации. Требуют в 1,5 раза больше усилий для разработки (в сравнении с регрессией) при незначительном улучшении точности предсказания

Модель ARIMA

Модель ARIMA (авторегрессионная интегрированная скользящая средняя) - это статистическая модель, которая используется для анализа временных рядов.

Она представляет собой комбинацию авторегрессии (AR), интегрирования (I) и скользящей средней (MA). Модель ARIMA позволяет учитывать тренды, сезонность и другие особенности временных рядов, что делает ее полезной для прогнозирования будущих значений временных рядов.

Формула ARIMA

Модель ARIMA(p, d, q) имеет следующую формулу:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

где:

- y'_t - значение временного ряда в момент времени t
- c - константа
- $\phi_1, \phi_2, \dots, \phi_p$ - коэффициенты авторегрессии (AR)
- $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ - значения временного ряда в предыдущие моменты времени
- $\theta_1, \theta_2, \dots, \theta_q$ - коэффициенты скользящей средней (MA)
- $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ - остатки модели в предыдущие моменты времени
- ε_t - остаток модели в момент времени t
- p - порядок авторегрессии
- d - порядок интегрирования
- q - порядок скользящей средней

Эта формула позволяет учитывать предыдущие значения временного ряда, а также остатки модели для прогнозирования будущих значений.

Реализация приложения на сервере Streamlit

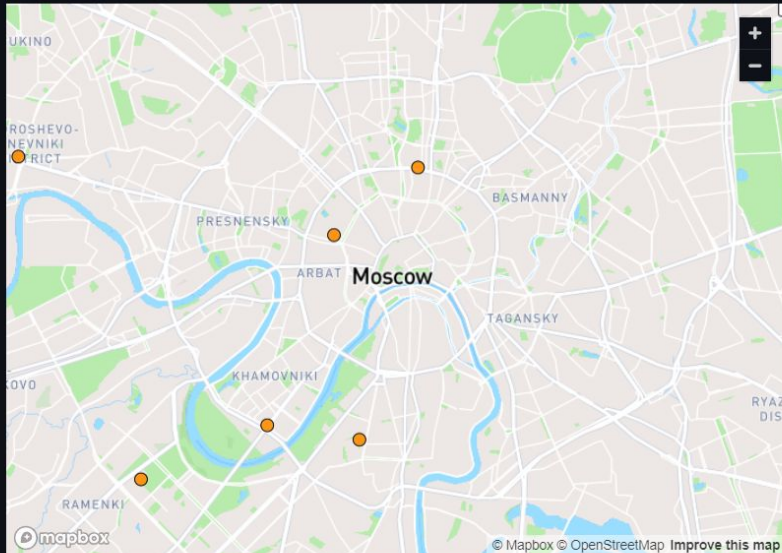
Для визуализации на сервере Streamlit был реализован функционал:

- Наложение данных по координатам на карту с цветовой дифференциацией по уровню загрязнения
- Реализация приложения на сервере Streamlit

Уровень загрязнения воздуха в городе Москва

Карта График

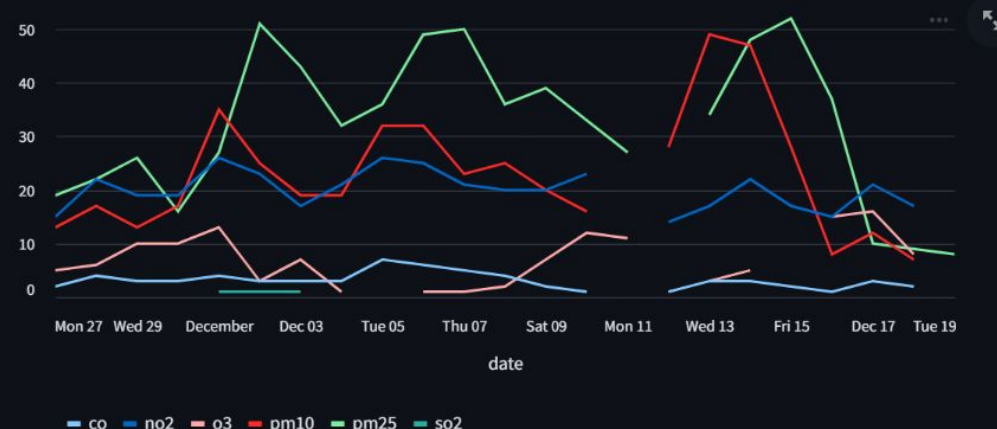
Карта Москвы



Уровень загрязнения воздуха в городе Москва

Карта График

График загрязнения воздуха на станции: Novokosino, Moscow, Russia (Новокосино, Москва)



Что дальше?

- Использование в модели новых данных от пользовательских датчиков, собираемых глобальной сетью sensor.community, для Москвы
- Создание гибридных моделей (ARIMA и нейронные сети), сочетающих лучшие качества обоих подходов, для предсказания качества воздуха*

* См., напр.: Duan, J., Gong, Y., Luo, J. *et al.* Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. *Sci Rep* **13**, 12127 (2023). <https://doi.org/10.1038/s41598-023-36620-4>; Luo, J., Gong, Y., Air pollutant prediction based on ARIMA-WOA-LSTM model, *Atmospheric Pollution Research* **14** (2023). <https://doi.org/10.1016/j.apr.2023.101761>.