

Одеський національний політехнічний університет

Інститут комп'ютерних систем

Комп'ютерні системи

Челишев Віталій Сергійович

**КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА**

ДОСЛІДЖЕННЯ СИСТЕМИ ОЦІНКИ ЗАБРУДНЕНOSTІ НАВКОЛИШНЬОГО

СЕРЕДОВИЩА НА SMART ДАТЧИКАХ

Спеціальність 123 – Комп'ютерна інженерія

Спеціалізація — Спеціалізовані комп'ютерні системи

Керівник: Яковлев Дмитро Павлович,

кандидат технічних наук, професор

Одеса – 2020

Міністерство освіти і науки України  
 Одеський національний політехнічний університет  
 Інститут комп'ютерних систем  
 Кафедра комп'ютерних систем

Рівень вищої освіти другий (магістерський)  
 Спеціальність 123 Комп'ютерна інженерія  
 Спеціалізація/ освітня програма Спеціалізовані комп'ютерні системи

ЗАТВЕРДЖУЮ

Завідувач кафедри

“ \_\_\_ ” \_\_\_\_\_ 20\_\_ року

**З А В Д А Н Н Я  
 НА КВАЛІФІКАЦІЙНУ РОБОТУ**

Челишев Віталій Сергійович

**(прізвище, ім'я, по батькові)**

1. Тема роботи Дослідження системи оцінки забрудненості навколишнього середовища на SMART датчиках

Керівник роботи Яковлев Дмитро Павлович, к.т.н., професор,

( прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом ректора ОНПУ від “ \_\_\_ ” \_\_\_\_\_ 20\_\_ року № \_\_\_

2. Зміст роботи

Розділ 1. Аналіз проблеми оцінки забрудненості навколишнього середовища

Розділ 2. Математичне моделювання процесу оцінки забрудненості повітря на основі даних SMART-датчиків

Розділ 3. Особливості реалізації моделі, що описує процес оцінки забрудненості навколишнього середовища

3. Перелік ілюстративного матеріалу

---

---

---

---

---

---

---

---

---

---



## ЗМІСТ

|   |    |
|---|----|
| ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ   | 6  |
| ВСТУП   | 7  |
| 1 АНАЛІЗ ПРОБЛЕМИ ОЦІНКИ ЗАБРУДНЕНOSTІ НАВКОЛИШНЬОГО<br>СЕРЕДОВИЩА  | 10 |
| 1.1 Обґрунтування вибору виду забруднення, що підлягає розгляду, та<br>його особливості                             | 10 |
| 1.2 Аналіз існуючих способів моніторингу та оцінки забруднення<br>обраного виду                                     | 14 |
| 1.3 Аналіз можливостей зведення систем оцінки забрудненості з<br>використанням SMART-датчиків різних типів          | 17 |
| 1.4 Уточнена постановка задачі дослідження  | 22 |
| 1.5 Висновки по розділу   | 23 |
| 2 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ОЦІНКИ<br>ЗАБРУДНЕНOSTІ ПОВІТРЯ НА ОСНОВІ ДАНИХ SMART-ДАТЧИКІВ                    | 25 |
| 2.1 Обґрунтування вибору математичного методу оцінки забрудненості  | 25 |
| 2.1.1 Методи теорії нейронних мереж   | 25 |
| 2.1.2 Методи теорії нечітких множин   | 28 |
| 2.1.3 Методи математичної статистики  | 33 |
| 2.2 Вибір розмірності та степеня регресійної моделі оцінки<br>забрудненості повітря на основі метеорологічних даних | 36 |
| 2.3 Проектування алгоритмічної складової процесу оцінки забрудненості   | 43 |
| 2.4 Висновки по розділу   | 50 |
| 3 ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ МОДЕЛІ, ЩО ОПИСУЄ ПРОЦЕС ОЦІНКИ<br>ЗАБРУДНЕНOSTІ НАВКОЛИШНЬОГО СЕРЕДОВИЩА                  | 51 |
| 3.1 Особливості програмної реалізації раніше розробленої<br>алгоритмічної складової процесу                         | 51 |

|   |    |
|---|----|
|   | 5  |
| 3.2 Опис інтерфейсу користувача системи   | 55 |
| 3.3 Опис та аналіз результатів програмної реалізації. Тестування продукту та аналіз ефективності розроблених рішень | 56 |
| 3.4 Висновки по розділу   | 59 |
| ВИСНОВОК  | 60 |
| ПЕРЕЛІК ПОСИЛАНЬ  | 61 |
| ДОДАТОК А. Вихідні тексти розробленого програмного забезпечення   | 64 |
| ДОДАТОК Б. Інструкція користувача для розробленого програмного продукту   | 71 |

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ГДК – Гранично допустима концентрація  
ІАД – Інтелектуальний аналіз даних  
ІТ – Інформаційні технології  
ПЗ – Програмне забезпечення  
ПК – Персональний комп’ютер  
ШНМ – Штучна нейронна мережа  
AJAX – Asynchronous Javascript And XML  
CSS – Cascading Style Sheet  
HTML – HyperText Markup Language  
UML – Unified Modeling Language  
XML – eXtensible Markup Language

## ВСТУП

Незважаючи на бурхливий розвиток технологій, людина як була кілька тисяч років тому назад, так і залишається досить ніжним створінням, яке для свого нормального існування потребує цілого комплексу безпечних, а краще, комфортних умов життя. В першу чергу, виставляються вимоги до (по мірі зменшенні нагальності, оперативності забезпечення):

- чистоти повітря, яке вона вдихає;
- чистоти води, яка використовується для пиття;
- біологічної чистоти рослин та тварин, які використовуються у їжу.

Менш очевидним, але безперечно необхідним є контроль наступних параметрів:

- хімічного складу ґрунтів, які використовуються для сільськогосподарських цілей;
- чистоти підземних, ґрунтових вод;
- радіаційної чистоти навколишнього середовища;
- глобального теплового забруднення місцевості, тощо.

Незважаючи на те, що усі вказані умови є важливими, все ж на першому місці знаходиться чистота повітря, для оцінки якої доцільно як розширювати існуючі мережі відповідних контролюючих пристроїв, так і розробляти нові. Сучасні датчики не тільки мають вимірюючий елемент, а й дозволяють використовувати автономний режим роботи, для якого необхідно вбудовувати у них комунікаційні підсистеми, розраховані на відповідні параметри передачі даних. Наприклад, сучасні датчики споживання електроенергії, встановлені у багатоквартирних будинках, можуть мати модуль Wi-Fi, за допомогою якого дані у повністю автоматичному режимі передаються на центральний блок, що встановлений зазвичай на даху будинку, який в свою чергу передає дані до дата-центру по GPRS з'єднанням. Описана система має усі ознаки «інтелектуальності»: автономний, повністю автоматичний режим роботи, наявність постійного комунікаційного зв'язку між

елементами системи з можливістю передачі змістовних даних, а також певних управлінських команд. Системи, що працюють за описаними принципами, називають інтелектуальними, розумними, або англійським терміном SMART. Відповідно, і побудова систем оцінки чистоти повітря саме на базі використання SMART-датчиків є **актуальною задачею** сучасної інженерної науки.

Відповідно до описаної актуальності можна сформулювати **мету дослідження** – проаналізувати системи оцінки забрудненості навколишнього середовища на SMART датчиках та запропонувати свою систему, що дозволить у випадку її реального впровадження більш адекватно реагувати (готуватися) до більших, чи менших рівнів забруднення.

Для досягнення такої мети необхідно вирішити певні **завдання дослідження**:

- проаналізувати існуючі види забруднень навколишнього середовища, обґрунтувати вибір одного з них та розглянути існуючі методи моніторингу та прогнозування його рівня;

- проаналізувати існуючі підходи до прогнозування фізичних величин, зокрема рівнів забруднення, та обґрунтувати вибір одного з методів для створення математичної моделі оцінки рівня забрудненості;

- запропонувати відповідну математичну модель та здійснити її реалізацію, провести тестування та дослідження отриманого продукту, оцінити його ефективність.

**Об'єкт дослідження** – процес оцінки рівня забрудненості навколишнього середовища.

**Предмет дослідження** – математичні моделі оцінки рівня забрудненості та їх реалізації.

**Наукова новизна** дослідження полягає в удосконаленні існуючої математичної моделі для оцінки загального рівня забрудненості повітря даної місцевості на базі окремих показників забрудненості місцевості, яка б враховувала різні ваги таких показників, та базувалася на використанні спеціалізованих SMART-датчиків.



**Практичне значення** роботи полягає у можливості створення повноцінного програмного продукту для оцінки інтегрального рівня забрудненості повітря на основі даних окремих датчиків, встановлених у даній місцевості.

**Методи, застосовані в роботі:** математичне моделювання, математична статистика (регресійний аналіз), технології програмування.

**Публікації:** за результатами дослідження підготовлена наукова стаття для публікації у фаховому виданні.

**В перспективі** математична модель, створена у роботі, може бути покращена шляхом залучення додаткових параметрів, окрім тих, що вже враховуються у роботі.

## 1 АНАЛІЗ ПРОБЛЕМИ ОЦІНКИ ЗАБРУДНЕНOSTІ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

1.1. Обґрунтування вибору виду забруднення, що підлягає розгляду, та його особливості

Безпека життя і здоров'я людини є однією з найвищих цінностей суспільства в сучасному світі. Нажаль, існує чимало факторів, що загрожують цим беззаперечним важливим складовим нашого життя. Цілий багатогранний комплекс утворюють різноманітні уражаючі фактори, пов'язані із забрудненням [1].

Забруднення є багатоплановим процесом і в галузі безпеки життєдіяльності та охорони праці виділяють наступні різні його типи:

а) хімічне – при неконтрольованому поширенні хімічної речовини у:

1) газоподібній фазі (забруднення повітря, яке може передаватися за рахунок інтенсивної у газах дифузії та конвективних потоків);

2) рідкій фазі (у 99,9% випадків несучою рідиною є вода; в цьому випадку також основну роль грають дифузія та конвекція, хоча не виключений варіант чисто механічного перенесення рідкого забруднювача з одного місця на інше за допомогою ще одного якогось носія);

3) твердому стані – за рахунок механічного перенесення (силою повітря, як при ураганах, води – як при повені, механічним перенесенням на якомусь носії).

б) фізичне – при розповсюдженні підвищених значень (у порівнянні із незбуреним фоном) різних фізичних величин. Для цього випадку можна виділити наступні підвиди:

1) радіаційне (виникає при поширенні у навколишньому середовищі радіоактивної речовини, що, відповідно, підлягає одному із трьох варіантів розпаду:  $\alpha$ ,  $\beta$  або  $\gamma$ ; надзвичайно небезпечне);

2) теплове (зважаючи, що більшість об'єктів оточуючого середовища мають суттєві теплоємності, цей вид забруднення досить довго не зникає навіть при повному зникненні джерел забруднення; крім того, нагріті маси повітря, води можуть переміщуватися на порівняно великі відстані);

3) шумове забруднення (його специфіка полягає у дуже малій інерційності: при зникненні джерела воно також миттєво пропадає, тому говорити у цьому випадку про стійке забруднення немає сенсу; крім того шумове забруднення не може поширюватися самовільно на невизначені відстані від джерела);

4) магнітне (також чітко прив'язане до джерела і пропадає при прибиранні самих магнітів, самовільно не поширюється);

5) електростатичне (при переміщенні статичних зарядів може змінювати свою конфігурацію, але зазвичай, у невеликих межах);

6) світлове (аналогічно неінерційне, самовільно не поширюється);

7) вібраційне (аналогічно).

в) біологічне забруднення – не тільки може поширюватися далеко від джерела, а й самовільно підвищувати власну інтенсивність, причому у досить широких межах (яка може характеризуватися, наприклад, масою живої речовини бактерій, або інших шкідливих організмів, на одиницю площі);

г) механічне забруднення великогабаритними відходами життєдіяльності людини (як-то при організації сміттєзвалищ). Характеризується тим, що самовільному поширенню не підлягає, може зростати лише контрольовано, «під наглядом» людини.

Аналіз особливостей усіх типів забруднень показує, що деякі з них підлягають самовільному поширенню, до того ж являються суттєвими небезпечними та шкідливими факторами. Так, особливу увагу слід приділяти питанням хімічного, радіаційного та біологічного забруднення – як небезпечним факторам для здоров'я (та тим більше, життя людини). Із них найбільш швидкоплинними, «мінливими» є процеси забруднення повітря. Дійсно, механічне забруднення взагалі не може виникнути і рухатися само по собі, забруднення води може поширюватися лише

повільно (як, наприклад, відбувається рух нафтових плям по поверхні води), так само як і забруднення ґрунту, а ось у випадку зараженого повітря ситуація є протилежною. В атмосфері майже завжди присутні досить суттєві процеси руху повітряних мас, в результаті чого рівень їх забрудненості у певній конкретній місцевості може різко змінюватися не тільки у різні дні, а й іноді – протягом однієї доби [2]. Важливу роль також грає середній за великі періоди часу рівень забрудненості повітря (наприклад, при виборі нового міста для переїзду, або при виборі розміщення майбутнього дитячого табору, і т.п.). Відповідно, бажаною є наявність інструменту, засобу, який би дозволяв проводити оцінку поточного та, особливо, майбутніх рівнів забрудненості на основі певної вхідної інформації.

Очевидно, що усі згадані процеси забруднення (як і майже будь-які процеси реального світу) можуть бути промодельовані на комп'ютері [3] – звичайно, за умови наявності відповідного програмного забезпечення. Результатами такого моделювання, наприклад, можуть бути масштаби забруднення при аваріях, катастрофах, чи інших нештатних ситуаціях (причому в динаміці, прив'язуючись до дат та, навіть, часу доби) – в першу чергу, питання територіального поширення забруднення. Такі відомості можуть допомогти спрогнозувати доцільну зону евакуації, зберігаючи людські життя та майно. Економічна обґрунтованість розробки відповідного програмного забезпечення є очевидною (як і його великий соціальний ефект). З точки зору математичної реалізації такої задачі, теоретично можливим є опис усіх трьох типів забруднення (хімічного, радіаційного та біологічного), однією системою рівнянь (можливо, із зануленням деяких коефіцієнтів у різних випадках із трьох згаданих).

Однак і задача прогнозування «неаварійних» рівнів забрудненості також не вкрай актуальною [4], оскільки дозволить вносити корективи у повсякденне функціонування суспільства з метою мінімізації шкоди для здоров'я людей. Наприклад, при високих рівнях забрудненості повітря діти (організми яких, як відомо, є найбільш вразливими для поганих умов навколишнього середовища) можуть проводити більше часу у кондиціонованих приміщеннях дитячих садків та шкіл, а у моменти чистого повітря на вулиці, очевидно більший час краще

проводити на відкритому просторі. При високих рівнях забрудненості повітря (які на сьогодні досить часто спостерігаються, наприклад, на території промислових районів Китаю – рисунок 1.1) доцільно переносити (звичайно, за наявності такої можливості) масові заходи, роботи на відкритому повітрі, і т.п., [5].



Рисунок 1.1 – Фото смогу (не туману!) в Пекіні у грудні 2015 р., який часто спостерігається у Китаї в різних провінціях по кілька тижнів без перерви

Зважаючи на вищесказане, можна зробити висновок, що задача моделювання процесу поширення забруднень в цілому є надзвичайно актуальною для сучасної галузі ІТ, а задача моделювання саме забруднення повітря (як, наприклад, описано у [6]) взагалі має стратегічний характер і пов'язана зі здоров'ям усієї людської цивілізації в майбутньому, саме тому вона і буде вирішуватися у даному дослідженні.

## 1.2. Аналіз існуючих способів моніторингу та оцінки забруднення обраного виду

При створенні нових наукових та інженерних розробок довільного призначення, по-перше, слід проаналізувати наявні у відкритому доступі аналогічні рішення та підходи.

Дуже близьким до даної роботи по своїй суті видається ресурс [7], стартова сторінка якого наведена на рисунку 1.2.

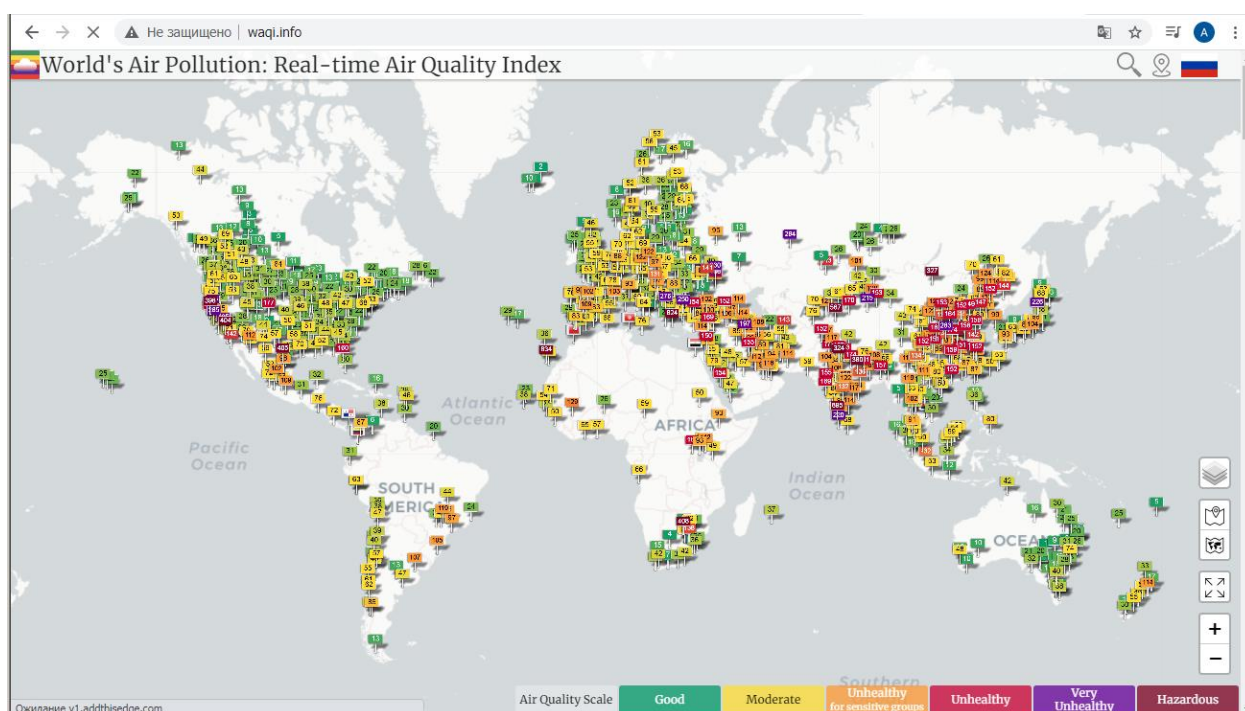


Рисунок 1.2 – Стартова сторінка ресурсу World's Air Pollution: Real-time Air Quality Index, присвяченого визначенню чистоти повітря

Ця система у реальному часі надає інформацію про параметри чистоти повітря по всьому світові (тобто це глобальна система), а саме у точках де встановлені спеціальні контролюючі станції, що популяризує даний сайт. На даному ресурсі є можливість передивлятися забруднення повітря в цілому, за значенням загального єдиного показника AQI (Air Quality Index), а також окремо по кільком важливим частинним показникам, серед яких є:

-  $PM_{2,5}$  – характеризує ступінь забруднення повітря малими частинками пилу (бруд), а саме діаметром до 2,5 мкм;

-  $PM_{10}$  – характеризує ступінь забруднення повітря великими частинками, а саме діаметром до 10 мкм;

-  $O_3$  – концентрація озону від промислових джерел;

-  $NO_2$  – концентрація оксиду азоту;

-  $SO_2$  – концентрація оксиду сірки;

-  $CO$  – концентрація чадного газу (моно оксиду вуглецю).

Згаданий вище загальний показник AQI обчислюється як зважена сума перелічених частинних показників:

$$AQI = \frac{\sum_{i=1}^N (w_i c_i)}{\sum_{i=1}^N w_i}, \quad (1.1)$$

де  $c_i$  – концентрація  $i$ -тої забруднюючої речовини;

$w_i$  – вага даної забруднюючої речовини, що обчислюються за формулою [8]:

$$w = \begin{cases} w^*, w^* > W_{\min}, w^* = \frac{c_{\min}}{c_{\max}}, \\ W_{\min}, w^* < W_{\min} \end{cases} \quad (1.2)$$

де  $w^*$  - перше можливе значення ваги, яке рівне відношенню мінімальної концентрації  $c_{\min}$  забруднюючої речовини до максимального значення  $c_{\max}$ , і це значення береться у якості ваги, якщо його значення більше порогу  $W_{\min}$ ;

$W_{\min}$  – друге можливе значення ваги, яке є постійним, і це значення береться у якості ваги, якщо  $w^*$  менше, за  $W_{\min}$  (тобто  $W_{\min}$  – порогове значення, нижче за яке вага  $w$  не може бути рівною).

Формула (1.2) побудована так, що мінімальне значення ваги рівне  $W_{\min}$ , а саме  $W_{\min}$  приймається рівним 0,5. Звичайно це є недоліком моделі, оскільки деякі частинні показники забрудненості повітря не є дуже суттєвими, але з іншого боку легко вимірюються, тому бажано було би включити їх у кінцевий результат. При підході (1.2), прийнятому при обчисленні AQI по (1.1), врахувати у малій мірі якийсь показник неможливо, тому доцільним є перегляд процедури вибору вагових коефіцієнтів (1.2). Взагалі ж ваговий коефіцієнт в (1.2) залежить від розкиду значень мінімальної  $c_{\min}$  та максимальної  $c_{\max}$  концентрацій забруднюючої речовини.

Ще однією загальною характеристикою забрудненості є «Інтегральний показник забруднення атмосфери», або по-іншому, «Індекс забруднення атмосфери» - ІЗА. Він розраховується по формулі, схожій на (1.1), яка має вид:

$$In = \sum_{i=1}^N \left( \frac{c_i}{ГДК_i} \cdot w_i \right), \quad (1.3)$$

де ваги  $w_i$  показують ступінь небезпеки і-тої речовини по відношенню до діоксиду сірки ( $SO_2$ );

ГДК – традиційна для екології аббревіатура, яка означає гранично допустиму концентрацію якоїсь речовини.

Прив'язка до порівняння з діоксидом сірки звичайно є недоліком моделі, оскільки кожна небезпечна (шкідлива) речовина має свою специфіку, дії по-різному, уражає різні органи і системи, тому доцільність порівняння з одним конкретним забруднювачем є як мінімум сумнівною.

Вказані приклади свідчать про те, що доцільним є розробка власної математичної моделі для виведення одного підсумкового показника забруднення на основі багатьох окремих частинних показників.



### 1.3. Аналіз можливостей зведення систем оцінки забрудненості з використанням SMART-датчиків різних типів

Для побудови системи оцінки забрудненості повітря, як уже зазначалося вище, доцільно використати не застарілі датчики, що потребують ручного (візуального) зчитування показів, а сучасні сенсори із розвинутими комунікаційними можливостями, що дозволяють автоматичну передачу власних показів до відповідних дата-центрів. Для цього можуть бути використані наступні рішення.

В першу чергу, поширеним та досить небезпечним забруднювачем є звичайний вуглекислий газ  $\text{CO}_2$ . Для визначення його рівня можна застосувати одне з численних рішень, присутніх на ринку, наприклад, від компанії Smart-MAC – рисунок 1.3. Він дозволяє вимірювати концентрацію вуглекислого газу від 0 до 5000 ppm (кількість часток  $\text{CO}_2$  на мільйон часток повітря), або у інших одиницях виміру – від 0 до 9129 мг/м<sup>3</sup>. Похибка вимірювання концентрації складає  $\pm 50$  ppm (або 91 мг/м<sup>3</sup>), чого цілком достатньо для побутових потреб щодо вимірювання рівня забруднення повітря. Даний датчик працює від джерела живлення 5 В, що дозволяє використати компактні автономні акумуляторні джерела живлення, а, отже, встановлювати датчик у будь-яких місцях, незалежно від наявності мережі 220 В поблизу. Цьому також сприяє досить широкий температурний режим роботи (0...+50 °С) та відносної вологості повітря (0...90 %). Комунікаційні можливості даного датчика полягають у його локальному підключенні до аналогового входу розумного лічильника імпульсів smart-MAC D105 – рисунок 1.4. Цей пристрій, в свою чергу, має можливості бездротової комунікації по мережі Wi-Fi, тобто утворює разом із самим датчиком дійсно розумний SMART-пристрій.



Рисунок 1.3 – Зовнішній вигляд датчику рівня вуглекислого газу компанії Smart-MAC



Рисунок 1.4 – Зовнішній вигляд розумного лічильника імпульсів smart-MAC D105

Більш небезпечною, хімічно схожою із вуглекислим газом речовиною, є чадний газ CO, для аналізу наявності якого також можна використати один із численних пристроїв, доступних на ринку, наприклад сигналізатор Honeywell XC100D – рисунок 1.5. Даний пристрій забезпечує вимірювання від 0 до 300 ppm, що у перекладі на масову концентрацію складає 349 мг/м<sup>3</sup>. Точність вимірювання,

що заявляється виробником, складає 1 ppm, що є достатнім для побутових пристроїв (слід сказати, що чадний газ є надзвичайно отруйною речовиною, тому навіть малі його концентрації до 10 ppm можуть негативно впливати на стан здоров'я людини при тривалій дії. Перевагою даного пристрою є використання вбудованої батареї, яка забезпечує термін безперервної роботи пристрою 10 років. Власне кажучи, він розрахований на однократне встановлення у верхніх частинах (адже чадний газ трохи легше повітря, тому концентрується у верхніх частинах приміщень) контрольованих приміщень та автономну роботу протягом вказаного періоду. Температурний діапазон ( $-10\dots+45$  °C) та відносна допустима вологість повітря (25...95 %) також сприяють широкому використанню даного датчику.



Рисунок 1.5 – Зовнішній вигляд сигналізатора чадного газу CO Honeywell XC100D

Для визначення частинок пилу (показники PM<sub>2,5</sub> та PM<sub>10</sub>) можна використати датчики на зразок PMS5003 – рисунок 1.6, що має провідний інтерфейс підключення UART (RS485), але по ньому може бути під'єднаний або до мікроконтролеру з можливістю подальшої бездротової передачі, або безпосередньо до модуля бездротової передачі даних Wi-Fi. Діапазон вимірювання приладу складає 0-1000 мкг/м<sup>3</sup> (у ppm не переводиться, оскільки мова йде не про газ, а про тверду речовину). Точність його вимірювання складає 1 мкг/м<sup>3</sup>. У якості елемента живлення використовується джерело на 4,5...5,5 В, що дозволяє використовувати комплект із 3 звичайних батарейок (чи акумуляторів) стандарту AA на 1,5 В.

Діапазон робочих температур є навіть більшим, ніж у попередніх пристроїв, і складає  $-10...+60$  °С, що дозволяє використовувати його не тільки у побутових приміщеннях, а й у більшості виробничих. Аналогічну властивість забезпечує і потрібна відносна вологість повітря, що для роботи даного датчика повинна знаходитися у межах  $0...99$  %.

Також небезпечною речовиною, що активно виділяється промисловими технічними установками є озон. Виконуючи надзвичайно корисну функцію захисту нашої планети від космічного випромінювання та жорсткого ультрафіолету у верхніх шарах атмосфери, унизу він є досить небезпечним. Навіть невеликі концентрації озону можуть бути шкідливі для здоров'я. Ранні ознаки впливу підвищених концентрацій цього газу включають роздратування очей, сухість у роті, кашель, утруднення дихання, хрипи, задишку, біль в грудях і інші ознаки. Діти, а також люди, які страждають на астму, бронхіт, серцеві захворювання, особливо уразливі відносно дії цієї речовини.



Рисунок 1.6 – Зовнішній вигляд датчика пилу PMS5003

Деякі виробники обладнання вказують, що озон знищує віруси, бактерії і цвіль. Це так, але проблема в тому, що його концентрація, при якій газ починає деактивувати шкідливі біологічні об'єкти, в 5-10 разів більше рівня ГДК. Це означає,

що піддавати озонуванню можна виключно порожні (без людей) приміщення, після чого концентрація озону має бути зменшена в рази для перебування у них людей. Особливо небезпечним є тривалий вплив озону на організм людини, який може призводити до серйозних захворювань, включаючи хвороби легенів.

Це означає, що озон також слід контролювати у рамках створення комплексного показника чистоти повітря, для чого можна використати, наприклад, «розумний» датчик EnergoM-3001-O3 - рисунок 1.7. Діапазон вимірювань його складає 0-100 ppm, тобто 0-116 мг/м<sup>3</sup>. Точність вимірювання складає 1 ppm. Живлення цього пристрою реалізовано на не дуже зручному рівні 12-24 В, що вимагає використання спеціального блока живлення (перетворювача 220 В змінного струму у 12-24 В постійного струму), який поставляється разом із датчиком. Дані вимірювання датчик передає по інтерфейсу RS485, що дозволяє за необхідності підключати до одного кабелю декілька пристроїв (до 16 штук). Робоча температура знаходиться у межах -30...+50 °С, а необхідна відносна вологість повітря складає 15...95 %.



Рисунок 1.7 – Датчик промислового озону EnergoM-3001-O3

Сірчистий газ, безперечно, є одним із найнебезпечніших домішок, що може бути присутнім у повітрі біля великих підприємств, особливо пов'язаних із хімічною промисловістю. Для його контролю може бути використаний наприклад, датчик "Дозор-С" – рисунок 1.8. Він визначає небезпечну речовину у діапазоні 0-100

мг/м<sup>3</sup>, що еквівалентно 0-38 ppm. Точність вимірювання складає 1 мг/м<sup>3</sup>. Діапазон робочих температур складає -40...+50 °С, а необхідна відносна вологість повітря складає 10...95 %. Пристрій має розвинуті комунікаційні властивості і дозволяє підключатися до комп'ютерної техніки за інтерфейсом RS232 або RS485. Відповідно до документації виробника, особливості електричної схеми пристрою забезпечують можливість утворення лінії зв'язку довжиною до 1200 м. Живлення даного датчика може відбуватися від джерела постійного струму 24 В, а також від джерела змінного струму 220 В.



Рисунок 1.8 – Аналізатор газів «Дозор-С»: варіант для визначення сірчистого газу

Таким чином, можна констатувати, що на ринку присутні численні пристрої (причому із можливістю вибору: або придбати одразу більш дороге рішення із бездротовою передачею через Інтернет, або із більш дешевим дротовим підключенням до інших місцевих пристроїв, що вже мають можливості комунікації), серед яких є можливості вибору необхідних рішень, які у повній мірі відповідають вимогам кожного конкретного випадку проектування.

#### 1.4. Уточнена постановка задачі дослідження

Таким чином, беручи до уваги вищенаведену інформацію, можна сформулювати наступну уточнену задачу дослідження.

Зважаючи на те, що забруднення повітря, яким дихає людина, є одним із найбільш небезпечних видів забруднення (через його швидке поширення, негайну дію на організм та неможливість виведення багатьох типів забруднювачів із легенів), у даній роботі необхідно провести зведення системи оцінки забруднення саме повітря. Існуючі математичні моделі (такі, як AQI, ІЗА) мають певні недоліки, які обумовлюють необхідність розробки власного способу виведення інтегрального ступеня забрудненості повітря на основі частинних показників. Важливим елементом роботи є спирання на покази інтелектуальних (SMART) сучасних датчиків окремих забруднюючих речовин, які розглядаються у даному дослідженні.

Створену мат модель необхідно реалізувати програмно однією із мов програмування загального призначення (вибір засобів розробки має бути обґрунтований) та провести тестування отриманого продукту, оцінити його ефективність.

## 1.5 Висновки по розділу

Таким чином, у даному розділі розглянуто основні аспекти проблеми оцінки забруднення навколишнього середовища з використанням SMART-датчиків. В першу чергу описані різні типи забруднень та встановлено, що саме забруднення повітря є найнебезпечнішим, оскільки людина постійно і безперервно потребує повітря (отже, якщо воно заражене, то дія на організм починається негайно), і, крім того, забруднення у ньому можуть поширюватися із дуже великою швидкістю (на відміну, наприклад, від водних або ґрунтових). Наступним кроком став розгляд існуючих показників загального рівня забруднення повітря, до яких необхідно переходити, оскільки різних забруднюючих речовин існує багато, а пересічну людину цікавить ступінь небезпеки повітря в цілому, взагалі. Вияснено, що існуючі показники (AQI, ІЗА) мають певні недоліки, пов'язані із складністю вибору вагових коефіцієнтів моделей для окремих забруднювачів. Серед таких забруднювачів обрано основні та розглянуто прилади для вимірювання їх рівнів, що одночасно мають розвинуті інформаційно-комунікаційні можливості, необхідні для передачі їх

показів через мережні інтерфейси (тобто розглянуто існуючі SMART-датчики для цих речовин). На основі матеріалу даного розділу можна переходити до наступних досліджень.



## 2 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСУ ОЦІНКИ ЗАБРУДНЕНOSTІ ПОВІТРЯ НА ОСНОВІ ДАНИХ SMART-ДАТЧИКІВ

### 2.1. Обґрунтування вибору математичного методу оцінки забрудненості

Задача оцінки рівня забрудненості повітря може бути віднесена до класу задач прогнозування, адже мова йде про визначення рівня забрудненості (на основі значень забруднень по окремим показникам). Такі задачі сьогодні часто вирішують методами математичної статистики, а також штучного інтелекту, тому розглянемо їх докладніше.

#### 2.1.1. Методи теорії нейронних мереж

Окремим класом таких систем, які, в принципі, можуть застосовуватися для задач прогнозування, є інтелектуальні рішення на базі нейронних мереж, яким присвячена чи не найбільша відносна частка усіх наукових джерел останніх років з питань прогнозування. Така ситуація у великій мірі обумовлена певним «бумом», що спостерігається в галузі штучного інтелекту та суміжних напрямках, навколо поняття нейронної мережі. Цей об'єкт широко застосовується [9] для задач розпізнавання, класифікації, апроксимації, а, зважаючи на це – і для прогнозування.

Штучною нейронною мережею (далі – ШНМ), як відомо, називають систему, яка, приймаючи на вхід вектор вхідних величин, за складними (нелінійними) внутрішніми законами виробляє вектор вихідних величин, що є більш-менш задовільним розв'язком поставленої задачі.

Елементарною складовою ШНМ є один нейрон – рисунок 2.1. Це об'єкт, що має певну кількість входів (на рисунку –  $n$  входів), які називають дендритами та один вихід, що називається аксоном [9].

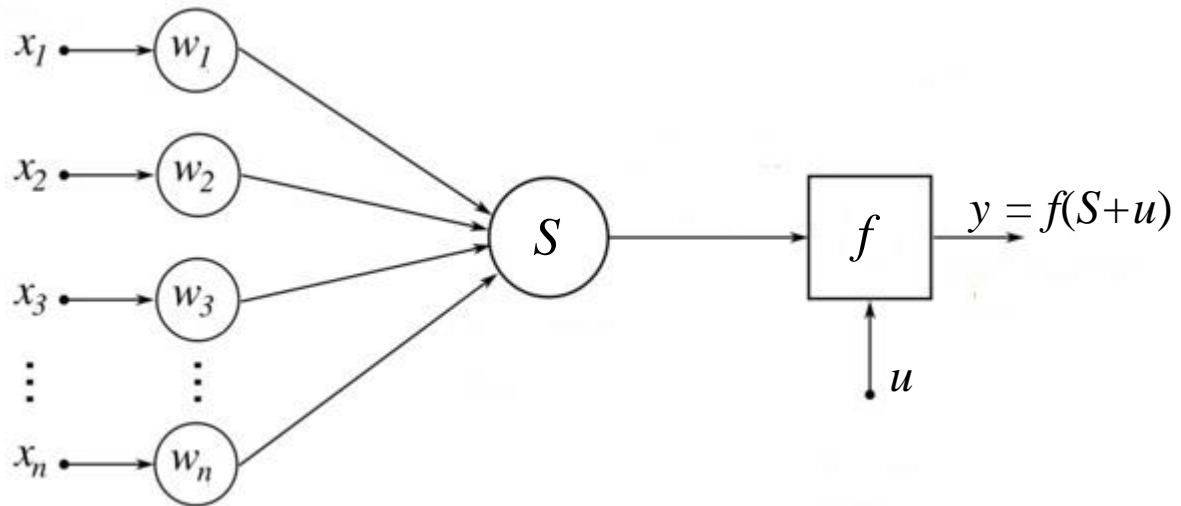


Рисунок 2.1 – Структура штучного нейрону

На входи нейрону подаються вхідні сигнали  $x_1, x_2, \dots, x_n$ , що є виходами попередніх нейронів, розташованих раніше по ходу проходження сигналу. Кожен вхідний сигнал  $x_i$  множиться на певне число  $w_i$ , що називають вагою (або синаптичною вагою) відповідного входу. Усі таким чином зважені вхідні сигнали подаються на суматор, який виробляє зважену суму  $S$  виду:

$$S = \sum_{i=1}^n w_i x_i . \quad (2.1)$$

До цієї зваженої суми вхідних сигналів може додаватися зсув або зміщення  $u$  (у поширеному частинному випадку зміщення відсутнє, тобто  $u = 0$ ) і від цієї суми виробляється певна функція  $f$ , що називається активаційною; так і формується вихідний сигнал нейрону  $y$ , що передається далі по нейронній мережі:

$$y = f(S + u). \quad (2.2)$$

Відповідно комбінуючи (2.1) та (2.2) маємо функцію, що здійснює штучний нейрон [10]:

$$y = f\left(\sum_{i=1}^n w_i x_i + u\right). \quad (2.3)$$

При аналізі такого способу завдання функції, що здійснює нейрон, як залежність (2.3), бачимо, що важливу роль для кожного нейрона відіграють три речі (особливо перші дві):

- набір конкретних значень ваг  $w_i$ , що описують кожен вхід нейрона;
- вид активаційної функції  $f$  даного нейрона;
- наявність зсуву або порогу активації  $u$  (який може бути як додатним, так і від'ємним).

Зважаючи на те, що звичайна ШНМ складається як мінімум із кількох нейронів (а частіше – з десятків, чи навіть сотень нейронів), а в кожного з них є багато входів, то задача вибору вагових коефіцієнтів  $w_i$  стає дуже серйозною за своєю складністю [11]. Власне, тривалий процес завдання конкретних «правильних» значень цих вагових коефіцієнтів називають процесом навчання нейронної мережі. Існують різні алгоритми навчання нейронних мереж, які укрупнено можна поділити на дві групи: «з учителем» та «без учителя». В цілому, навчання нейронних мереж є настільки складним процесом, що різні його варіанти мають досліджуватися в окремому великому блоці інформації і не можуть бути розглянуті в рамках даної роботи.

Вид активаційної функції  $f$ , очевидно, також є дуже важливим моментом для функціонування нейронів, хоча й у значно меншій мірі, ніж вибір синаптичних вагових коефіцієнтів [12]. На рисунок 2.2 наведено найбільш поширені види функцій активації, що найчастіше використовуються на практиці при роботі з нейронними мережами. З точки зору обчислювальної складності більш кращим є використання порогових функцій, що фактично являють собою константну залежність (або хоча б лінійних, але не сигмоїдних).

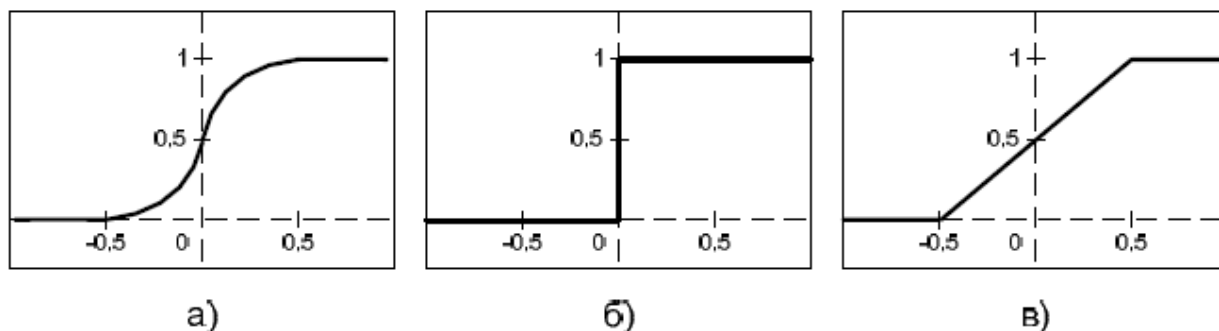


Рисунок 2.2 – Найбільш поширені типи функцій активації: а – сигмоїдна, б – порогова, в – лінійна

При побудові цілої нейронної мережі, крім характеристик окремих нейронів, важливим стає також спосіб їх з'єднання (так само, як і їх загальна кількість, організація нейронів у групи, шари, і т.д.).

Найпростішим і в той же час найпоширенішим способом з'єднання нейронів у конкретну ШНМ є перцептрон (перцептрон), що обумовлено широким спектром задач, які можна вирішувати за допомогою ШНМ з такою внутрішньою структурою [12].

Слід відмітити, що нейронні мережі мають недолік, який полягає у поганій передбачуваності результатів їх висновків. Існують повідомлення [13] про різку зміну висновку нейронної мережі при дуже незначних змінах вхідних величин. Зважаючи на це, використовувати нейронну мережу у даній роботі не будемо.

### 2.1.2. Методи теорії нечітких множин

Одним із таких популярних методів [14] отримання кінцевого рішення в задачах моделювання складного характеру, зокрема, таких як прогнозування в екологічній сфері, є застосування апарату нечіткої логіки (fuzzy logic), що є досить перспективним механізмом, який також широко застосовується в галузі інтелектуального аналізу даних (ІАД). Перевагою таких систем є можливість роботи із нечисловими, розмитими, або взагалі якісними, вираженими словами, лінгвістичними даними. Процедура фазифікації вхідної інформації, в якій

використовуються лінгвістичні змінні дозволяє поставити у відповідність елементам деякої множини їх текстові описи, а потім використати простий та зрозумілий набір правил нечітких продукцій для вироблення кінцевого рішення у нечіткій, розмитій формі. Після застосування процедури дефазифікації отримуємо конкретне значення прогнозованої величини. Головною перевагою при такому підході є відсутність необхідності зведення складної математичної моделі (які часто реалізують на базі методів багатовимірної регресійного аналізу, інших статистичних методів, що потребує нетривіального математичного підґрунтя), а використання простої бази правил, яку може розробити на основі аналізу логіки роботи системи (або ходу протікання процесу) будь-яка тверезо мисляча людина, що має хоча б невеликий досвід спостереження або роботи із системою. Правила мають структуру на зразок «якщо дебіторська заборгованість є високою, то швидкість відвантаження готової продукції має бути малою», «якщо борг по зарплатні середній, то імовірність придбання нового обладнання є низькою», і т.п.[15].

Зважаючи на широку поширеність систем нечіткого виводу, розглянемо докладніше весь процес їх використання (зокрема, для задач прогнозування, як у даній роботі) [16]. Він полягає у циклічному виконанні процедури нечіткого виводу (див. нижче) за допомогою системи нечіткого виводу. Єдиним кроком, що не виконується під час робочого часу системи, а реалізується лише один раз – при її розробці – є створення бази нечітких правил продукцій (бази правил нечітких продукцій).

Під системою нечітких правил продукцій мається на увазі узгоджена множина окремих правил нечітких продукцій, побудованих на основі виразів виду:

$$\text{ПРАВИЛО } <1>: \text{ЯКЩО } \langle \beta_1 \in \alpha_{11} \rangle \text{ ТО } \langle \beta_2 \in \alpha_{21} \rangle, \quad (2.4)$$

де « $\in$ » означає «являється», або англійською «IS».

В (2.4) нечіткі вирази антецеденту  $A$  та консеквенту  $B$  розписані у формі « $\beta_i \in \alpha_{ij}$ » через лінгвістичні змінні  $\beta_i$ , деякі з яких є вхідними (це змінні, що беруть участь у антецедентах), деякі – вихідними (змінні, які входять до консеквентів).  $\alpha_{ij}$  – який-небудь конкретний терм із терм-множини лінгвістичної змінної  $\beta_i$ .

Таким чином, першим кроком процедури нечіткого виводу, який до того ж виконується лише один раз, є створення бази правил нечітких продукцій. Це задача, яку має виконувати експерт (експерти) у даній предметній галузі[17], оскільки від якості цієї системи (її повноти та адекватності кожного правила) залежить успішна робота системи прогнозування в майбутньому. Помилки при проектуванні бази правил можуть приводити до катастрофічних наслідків, причому найбільша частка складних проблем має за причину саме недоліки системи нечітких правил продукцій.

Важливою властивістю бази правил має бути її узгодженість, тобто правила не повинні суперечити одне одному (так завжди і є, коли розробку виконує один експерт, що має закінчений, усталений і ефективний погляд на проблему заданого класу). При розробці бази правил командою експертів, перевірка на узгодженість має відбуватися окремим її підетапом.

Після зведення бази правил нечітких продукцій та її реалізації у певному електронному апаратному забезпеченні розпочинається процес нормального функціонування системи прогнозування, що є ітеративним, циклічним і кожна ітерація складається зі складових, показаних на рисунок 2.3:

- фазифікація вхідних величин, що входять до умов правил нечітких продукцій;
- агрегування підумов для правил, що мають складені антецеденти;
- активізація підвисновків;
- акумулювання висновків (для тих вихідних змінних, різні терми яких входять до різних активних правил нечітких продукцій, а зазвичай для кожної вихідної змінної так і є);
- дефазифікація вихідних змінних.



Рисунок 2.3 – UML-діаграма діяльності процесу нечіткого виводу

Розглянемо докладніше кожен із п'яти повторюваних етапів процедури нечіткого виведення.

Першим із них є фазифікація, суть якої полягає у обчисленні значень функцій належності усіх активних термів усіх вхідних лінгвістичних змінних[18]. Даний процес виконується на основі звичайних (не нечітких) даних і має на меті їх

переведення у терміни теорії нечітких множин, саме тому фазифікацію ще називають введенням нечіткості.

Ціллю фазифікації є встановлення відповідності між конкретним (зазвичай, числовим) значенням окремої вхідної змінної системи нечіткого виводу та значенням функції належності відповідного їй терма вхідної лінгвістичної змінної. Після завершення цього етапу для всіх вхідних змінних мають бути визначені конкретні значення функцій належності по кожному із лінгвістичних термів, що використовуються у підумовах бази правил системи нечіткого виводу.

Наступним етапом процедури нечіткого виводу є агрегування. Воно застосовується до тих правил, які мають складені умови, що складаються з декількох підумов. Для кожної підумови шляхом виконання процесу фазифікації встановлюється ступінь істинності[19]. По цим даним для кожного правила встановлюється ступінь його істинності в цілому, що виконується на основі об'єднання ступенів істинності окремих підумов, що входять до даного правила (якщо якесь правило побудовано на основі елементарної умови, то ступінь істинності всього правила приймається рівним ступеню істинності цієї єдиної умови). Об'єднання підумов одного правила здійснюється за допомогою операції нечіткої кон'юнкції ТА чи нечіткої диз'юнкції АБО.

Результатом етапу агрегування є встановлення ступеня істинності (у вигляді числа в діапазоні від 0 до 1 включно) для кожного із правил, що утворюють базу нечітких продукцій.

Активізація є третім із повторюваних етапів процедури нечіткого виводу і полягає у встановленні ступеня істинності кожного із підвисновків усіх правил, що були активними на цій ітерації процесу (тобто таких правил, які на попередньому кроці агрегування мали ненульові результуючі ступені істинності). Ціллю активізації є вироблення ступенів істинності для кожного підвисновку, в результаті чого на наступному кроці можна буде перейти до визначення цілих лінгвістичних змінних, що є вихідними.

Акумуляція є передостаннім етапом процедури нечіткого виводу і необхідна через той факт, що різні терми однієї і тієї ж лінгвістичної змінної можуть входити



до підвисновків різних правил нечітких продукцій, і тому мають бути якимсь чином об'єднані для формування значення цієї конкретної лінгвістичної змінної.

Результатом акумуляції є сформований набір функцій належності для кожної лінгвістичної змінної, які враховують і відповідають усім термам, задіяним у активованих правилах бази правил нечітких продукцій.

Останній етап процедури нечіткого виводу називається дефазифікацією, а полягає він у тому, щоби на основі розмитої інформації, якою є отримані на попередньому етапі види функцій належності усіх вихідних змінних, обчислити конкретні значення цих змінних, яких їм треба надати[20].

Для виконання дефазифікації існує багато різноманітних методів, але найбільш популярними є метод центру тяжіння (який також називають просто словом «центроїд») та центру площі, відповідно до яких треба досліджувати геометричну фігуру, обмежену прямими  $x = 0$ ,  $x = x_{\max}$ ,  $\mu = 0$ ,  $\mu = 1$ , а також – лінією, що зображує функцію належності даної вихідної лінгвістичної змінної. У першому випадку слід знаходити координату  $x$  центра тяжіння цієї двовимірної фігури, а у другому – координату  $x$  вертикальної прямої, що ділить площу цієї фігури на 2 рівних по площі частини.

Після виконання етапу дефазифікації для кожної вихідної змінної буде наявним конкретне числове значення, якому вона повинна дорівнювати для того, щоби забезпечити найкращий результат моделювання.

Нечіткі системи мають недолік складності контролю кінцевого результату моделювання традиційними математичними методами, а сам цей результат сильно залежить від вибору функцій належності (який, взагалі кажучи, є вільним). Через ці причини нечіткі множини у даній роботі застосовувати не будемо.

### 2.1.3. Методи математичної статистики

Окрім розглянутих методів із галузі штучного інтелекту, які з'явилися порівняно нещодавно (у другій половині ХХ ст.), також для описаних вище задач можна застосовувати методи математичної статистики, розроблені дещо раніше, а саме більше 100 років тому (початок ХХ ст.). Серед усього блоку відомостей про

математичну статистику для задач прогнозування особливо виділяється розділ «Аналіз часових рядів».

Аналіз часових рядів - сукупність математико-статистичних методів аналізу, призначених для виявлення структури часових рядів і для їх прогнозування. Сюди відносяться, зокрема, методи регресійного аналізу.

Виявлення структури часового ряду необхідно для того, щоб побудувати математичну модель того явища, яке є джерелом аналізованого часового ряду. Прогноз майбутніх значень часового ряду використовується для ефективного прийняття рішень. Прогнозування часових рядів має широке застосування: від прогнозу температури до передбачення значення біржових котирувань і валютних пар.

Часові ряди складаються з двох елементів:

- періоду часу, за який або за станом на який наводяться числові значення;
- числових значень того чи іншого показника, званих рівнями ряду.

Часові ряди класифікуються за такими ознаками:

а) за формою подання рівнів:

- ряди абсолютних показників;
- відносних показників;
- середніх величин.

б) за кількістю показників, для яких визначаються рівні в кожен момент часу: одномірні і багатовимірні тимчасові ряди;

в) за характером тимчасового параметра: моментні і інтервальні часові ряди. У моментних часових рядах рівні характеризують значення показника станом на певні моменти часу. В інтервальних рядах рівні характеризують значення показника за певні періоди часу. Важлива особливість інтервальних часових рядів абсолютних величин полягає в можливості підсумовування їх рівнів. Окремі ж рівні моментного ряду абсолютних величин містять елементи повторного рахунку. Це робить безглуздим підсумовування рівнів моментних рядів;

г) по відстані між датами і інтервалами часу виділяють рівновіддалені - коли дати реєстрації або закінчення періодів йдуть один за одним з рівними інтервалами і неповні (нерівновіддалені) – коли принцип рівних інтервалів не витримується;

д) за наявністю пропущених значень: повні і неповні часові ряди;

е) часові ряди бувають детермінованими і випадковими: перші отримують на основі значень деякої не випадковою функції (ряд послідовних даних про кількість днів у місяцях); другі є результатом реалізації деякої випадкової величини.

є) залежно від наявності основної тенденції виділяють стаціонарні ряди, в яких середнє значення і дисперсія постійні, і нестационарні, які містять основну тенденцію розвитку.

Часові ряди, як правило, виникають в результаті вимірювання деякого показника. Це можуть бути як показники (характеристики) технічних систем, так і показники природних, соціальних, економічних та інших систем (наприклад, погодні дані). Типовим прикладом тимчасового ряду можна назвати біржовий курс, при аналізі якого намагаються визначити основний напрям розвитку (тенденцію або тренд).

Найбільш поширеним методом аналізу часових рядів є регресійний аналіз. При його застосуванні метою є встановлення виду та параметрів функції, яка відповідає закономірності зміни однієї вихідної контрольованої величини від одного (одновимірний регресійний аналіз та регресія) або декількох (багатовимірний регресійний аналіз та багатовимірна регресія) вхідних параметрів. При цьому у більшості випадків залежність, що встановлюється, обирається лінійною (отже, визначенню підлягають коефіцієнти такої лінійної залежності), хоча в окремих випадках (коли відомо про суттєву не лінійність моделі) застосовують і вищі степені апроксимуючого поліному (квадратична регресія, і т.д.). У дуже рідких випадках використовують якісь інші функціональні залежності, окрім поліномів. Конкретний вид регресійної залежності буде обґрунтований у наступному підрозділі.

## 2.2. Вибір розмірності та степеня регресійної моделі оцінки забрудненості повітря на основі метеорологічних даних

Задача відтворення функціональної залежності між однією величиною (аргументом) та іншою (функцією), так само як і між кількома аргументами та функцією, є типовою для розділу математичної статистики, що називається регресійним аналізом. Сам вид цієї функціональної залежності називається регресією (або рівнянням регресії). Іноді назву уточнюють словом «одновимірна», якщо треба підкреслити, що вихідна величина залежить тільки від однієї вхідної, або «багатовимірна», якщо вхідних величин декілька.

Дійсно, в загальному випадку предметом пошуку у регресійному аналізі може бути не тільки функція однієї змінної. З математичного аналізу добре відомі функції багатьох змінних, які також широко використовуються у регресійному аналізі, якщо необхідно встановити залежність однієї вихідної величини від декількох вхідних аргументів. Така задача називається множинною регресією (або встановленням рівняння множинної регресії). Розглянемо ці задачі докладніше.

Як зазначено вище, у практичних дослідженнях часто виникає необхідність встановлення залежності між однією вихідною величиною та набором вхідних параметрів, які впливають на значення цієї величини. Результатом є функція багатьох змінних:

$$y = f(x_1, x_2, \dots, x_n).$$

У найпростішому і найбільш поширеному випадку розглядається залежність вихідної змінної від одного аргументу:

$$y = f(x). \quad (2.5)$$

Функція (2.5) може бути більш чи менш складною, залежно від явища, що розглядається.

Наприклад, експериментально досліджується залежність середньої кількості опадів ( $y$  мм), що випадають у певній місцевості, від кількості днів, що пройшли від початку сезону дощів. На основі реальних опадів побудовано графік, на якому кожна точка відповідає одному дощу: по осі абсцис відкладається номер дня від початку сезону дощів, на який спостерігався дощ, а по осі ординат – кількість мм води, що випала під час цього дощу – рисунок 2.4. Шляхом регресійного аналізу можна приблизно побудувати функціональну залежність, за допомогою якої для інших днів можна спрогнозувати кількість опадів, які в середньому випадатимуть у даній місцевості.

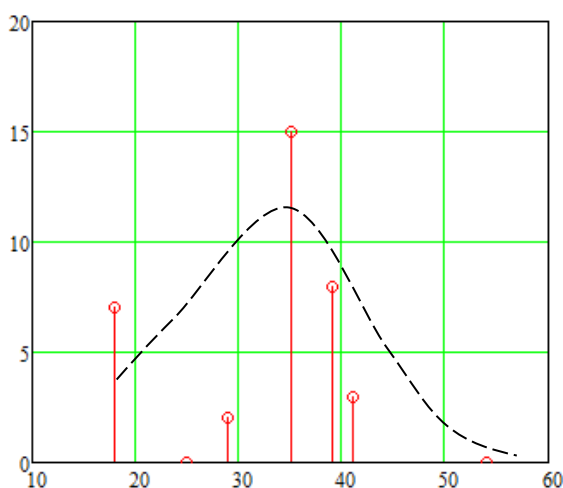


Рисунок 2.4 – Приклад побудови регресії для залежності середньої кількості опадів ( $y$  мм), залежно від часу ( $x$  днів), що пройшов від початку сезону дощів (нелінійна регресія)

Очевидно, залежність рисунок 2.4 є досить складною з математичної точки зору. У найпростішому випадку функція (2.5) представляється лінійною залежністю:

$$y = ax + b, \quad (2.6)$$

де коефіцієнти регресії  $a$  і  $b$  мають бути визначені на основі експериментальних даних про процес, що моделюється.

Наприклад, виконано декілька експериментальних замірів величини  $y_i$  при заданій величині  $x_i$  – рисунок 2.5, що показані точками.

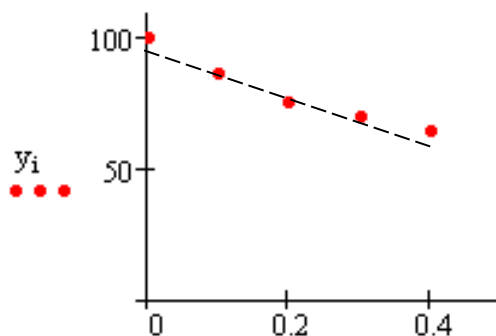


Рисунок 2.5 – Функціональна залежність, близька до лінійної

Після виконання експериментів постає задача відтворення функціональної залежності величини  $y$  від  $x$ . Як уже сказано вище, можливо виконати апроксимацію лінійною залежністю (2.6) (ї яка показана пунктиром на рисунок 2.5). Можливо, більш адекватно відображувала би явище квадратична залежність (або які-небудь інші варіанти функціональної залежності), але з математичної точки зору, як найпростіша, узята саме лінійна залежність (2.6).

Обрати конкретні коефіцієнти можна за методом найменших квадратів: слід розглянути відхилення  $\varepsilon_i = (y_i - (ax_i + b))$  та знайти мінімум суми квадратів таких нев'язок:

$$\sum \varepsilon_i^2 \rightarrow \min.$$

Із цієї вимоги слідує система двох рівнянь, з якої можна визначити коефіцієнти лінійної регресії:

$$\begin{cases} a \sum x_i^2 + b \sum x_i = \sum x_i y_i \\ a \sum x_i + b n = \sum y_i \end{cases} \rightarrow a = \frac{\sum x_i y_i - x_{cp} y_{cp}}{\sum x_i^2 - (x_{cp})^2}; \quad b = y_{cp} - a x_{cp}. \quad (2.7)$$

де  $x_i, y_i$  – пара значень відповідно незалежної змінної та функції, що відповідають  $i$ -тому вимірюванню;

$x_{cp}, y_{cp}$  – середні арифметичні усіх значень  $x_i$  та  $y_i$  відповідно;

$n$  – кількість вимірювань, тобто пар  $x_i, y_i$ .

Аналогічну до (2.7) систему можна отримати і для коефіцієнтів квадратичної регресії:

$$\begin{cases} a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i \\ a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i + cn = \sum y_i. \end{cases}$$

Невідомі коефіцієнти  $a, b, c$  можна знайти, розв'язуючи отриману систему лінійних алгебраїчних рівнянь, наприклад, методом Крамера (звичайно, можливим є застосування і інших підходів):

$$a = \frac{\Delta_a}{\Delta} = \frac{\begin{vmatrix} \sum x_i^2 y_i & \sum x_i^3 & \sum x_i^2 \\ \sum x_i y_i & \sum x_i^2 & \sum x_i \\ \sum y_i & \sum x_i & n \end{vmatrix}}{\begin{vmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{vmatrix}},$$

$$b = \frac{\Delta_b}{\Delta} = \frac{\begin{vmatrix} \sum x_i^4 & \sum x_i^2 y_i & \sum x_i^2 \\ \sum x_i^3 & \sum x_i y_i & \sum x_i \\ \sum x_i^2 & \sum y_i & n \end{vmatrix}}{\begin{vmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{vmatrix}},$$

$$c = \frac{\Delta_c}{\Delta} = \frac{\begin{vmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 y_i \\ \sum x_i^3 & \sum x_i^2 & \sum x_i y_i \\ \sum x_i^2 & \sum x_i & \sum y_i \end{vmatrix}}{\begin{vmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{vmatrix}}.$$

Зважаючи на складність (непрактичність) формул квадратичної регресії, в роботі будемо застосовувати лінійну регресію, розмірність якої встановимо нижче.

Переходячи до питання розмірності, можна згадати, що на практиці одна контрольована величина може бути функцією багатьох вхідних параметрів, як в принципі і є у даному дослідженні:

$$y = f(x_1, x_2, \dots, x_m).$$

Форма вказаної залежності може бути досить складною (залежно від конкретного процесу, що досліджується), але з математичної точки зору найпростішим є відтворення рівняння лінійної множинної регресії:

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m. \quad (2.8)$$

Рівняння (2.8) називають гіперплощиною, зважаючи на те, що таке рівняння із двома незалежними аргументами у правій частині визначає реальну площину у тривимірному фізичному просторі.

Задача багатовимірного регресійного аналізу полягає у визначенні коефіцієнтів  $a_i$  на основі доступних експериментальних даних. Під експериментальними даними мається на увазі таблиця відповідності виміряних значень вхідних параметрів  $x_{ij}$  та значення контрольованої величини  $y_i$ , загальний вид якої показано як таблиця 2.1.



Таблиця 2.1 - Загальний вид таблиці експериментальних висхідних даних для зведення рівняння множинної регресії

| № експ | $x_1$    | $x_2$    | ... | $x_m$    | $y$   |
|--------|----------|----------|-----|----------|-------|
| 1      | $x_{11}$ | $x_{12}$ | ... | $x_{1m}$ | $y_1$ |
| 2      | $x_{21}$ | $x_{22}$ | ... | $x_{2m}$ | $y_2$ |
| ...    | ...      | ...      | ... | ...      | ...   |
| $n$    | $x_{n1}$ | $x_{n2}$ | ... | $x_{nm}$ | $y_n$ |

Для виконання конкретних математичних операцій працюють з трьома матрицями, що описуються нижче.

Першою записується регресійна матриця:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}.$$

Далі береться матриця-стовпець значень контрольованої величини:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Останньою формується матриця-стовпець шуканих коефіцієнтів:

$$A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}.$$

Використовуючи метод найменших квадратів, операції диференціювання за компонентами вектора, та виконуючи необхідні перетворення, отримують наступну формулу для знаходження невідомих коефіцієнтів рівняння (2.8) множинної лінійної регресії:

$$A = (X^T X)^{-1} X^T Y, \quad (2.9)$$

де літерою  $T$  позначена операція транспонування (заміна рядків на стовпці, або навпаки);

степенем мінус 1 позначена обернена матриця.

Як бачимо, залежність (2.9) є порівняно складною, оскільки вимагає перемноження матриць, а головне, знаходження оберненої матриці. Для цих операцій доцільним є використання спеціалізованого математичного програмного забезпечення, але можна і використати мови програмування загального призначення. Іншим варіантом є вирішення системи рівнянь, які отримуються методом найменших квадратів за допомогою метода Крамера (а не матричного):

$$\left\{ \begin{array}{l} na_0 + a_1 \sum x_{1i} + a_2 \sum x_{2i} + \dots + a_m \sum x_{mi} = \sum y_i \\ a_0 \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i} x_{2i} + \dots + a_m \sum x_{1i} x_{mi} = \sum x_{1i} y_i \\ a_0 \sum x_{2i} + a_1 \sum x_{2i} x_{1i} + a_2 \sum x_{2i}^2 + \dots + a_m \sum x_{2i} x_{mi} = \sum x_{2i} y_i \\ \dots \\ a_0 \sum x_{mi} + a_1 \sum x_{mi} x_{1i} + a_2 \sum x_{mi} x_{2i} + \dots + a_m \sum x_{mi}^2 = \sum x_{mi} y_i \end{array} \right. \quad (2.10)$$

При цьому головний визначник системи матиме вид:

$$\Delta = \begin{vmatrix} n & \sum x_{1i} & \sum x_{2i} & \dots & \sum x_{mi} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \dots & \sum x_{1i}x_{mi} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 & \dots & \sum x_{2i}x_{mi} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{mi} & \sum x_{mi}x_{1i} & \sum x_{mi}x_{2i} & \dots & \sum x_{mi}^2 \end{vmatrix}. \quad (2.11)$$

Міняючи за методом Крамера відповідний стовпець на стовпець вільних членів системи (праві частини кожного рівняння із (2.10)), отримуватимемо частинні визначники  $\Delta_i$ , на основі яких за формулами Крамера визначатимемо шукані коефіцієнти рівняння множинної лінійної регресії (2.8).

Для розрахунків корисним буде спочатку визначити усі суми, присутні у системі (2.10). Саме ці залежності є основою для побудови конкретної математичної моделі процесу оцінки забруднення повітря на основі відомостей про температуру, тиск та рух повітряних мас, що буде зводитися у наступному викладенні.

### 2.3 Проектування алгоритмічної складової процесу оцінки забруднення

Виділимо у складі даної роботи два типи алгоритмів: алгоритм роботи з програмою та алгоритм математичних розрахунків значення забруднення.

Програма призначена для отримання одного підсумкового значення ступеня забрудненості повітря, яке вираховується виключно на основі вхідних даних. Такий режим роботи може бути названий пакетним і алгоритм роботи з програмою тоді буде наступним – рисунок 2.6.

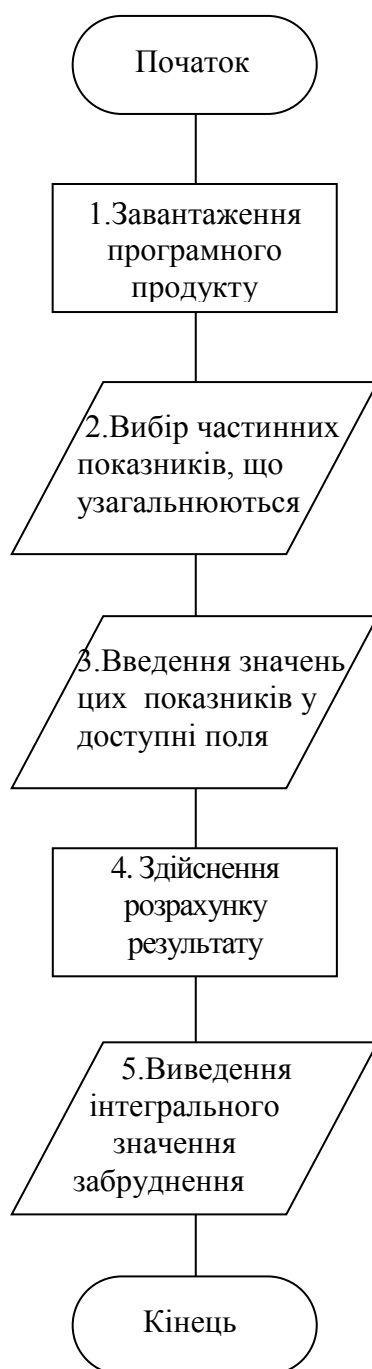


Рисунок 2.6 – Алгоритм роботи з програмою оцінки забруднення повітря

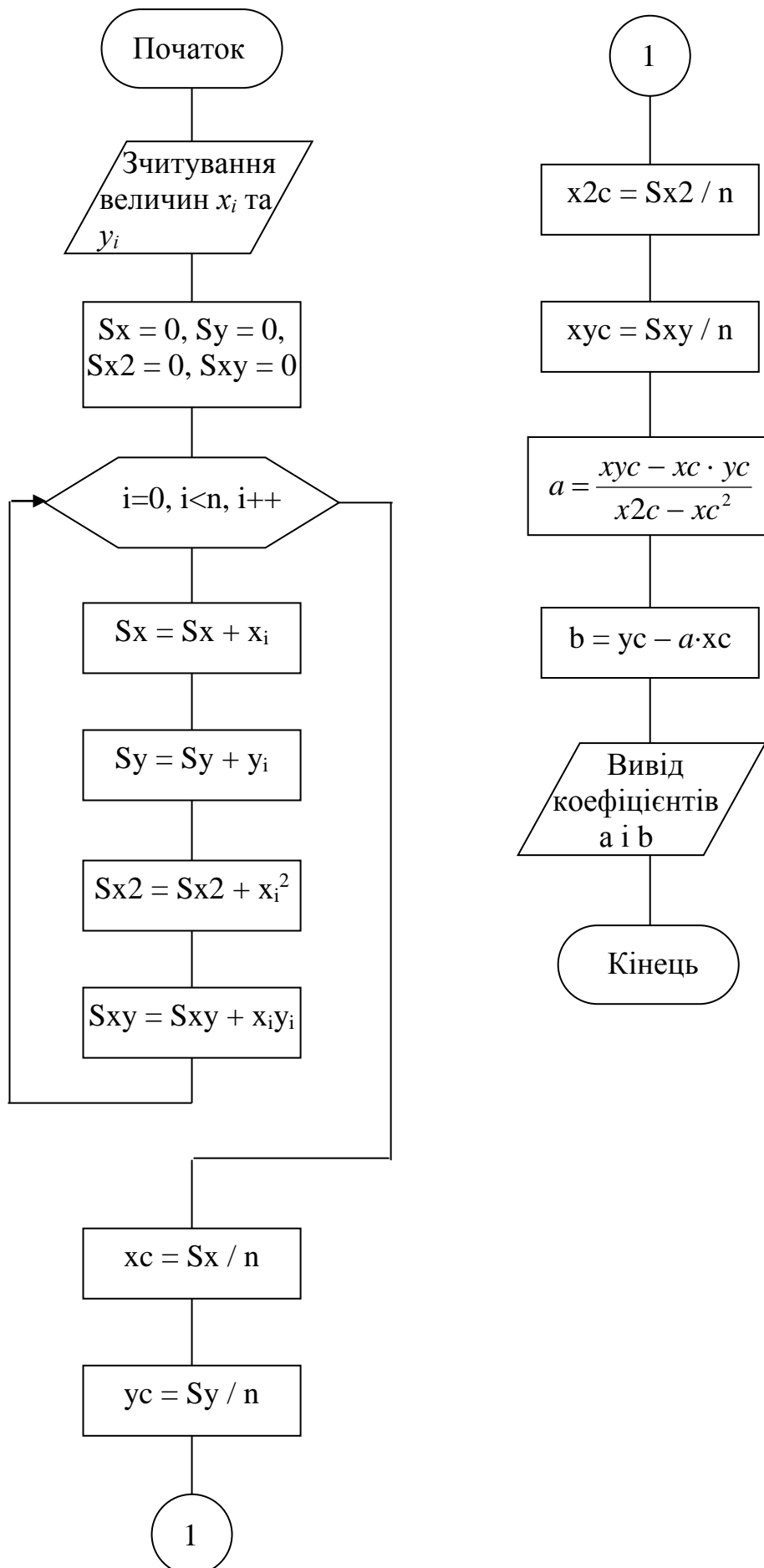


Рисунок 2.7 – Алгоритм розрахунку лінійної регресії

На рисунку 2.6 пункт 4 – «здійснення розрахунку» втілює усі математичні формули розрахунку коефіцієнтів множинної регресії, записані у попередньому підрозділі. Для лінійної регресії алгоритм має вид рисунку 2.7.

Для задачі множинної регресії алгоритм є схожим за тим виключенням, що в його основі будуть лежати формули (2.8)-(2.11).

Керуючись наведеними алгоритмами, можна передбачити наступну структуру математичної моделі системи, яка показана на рисунку 2.8.

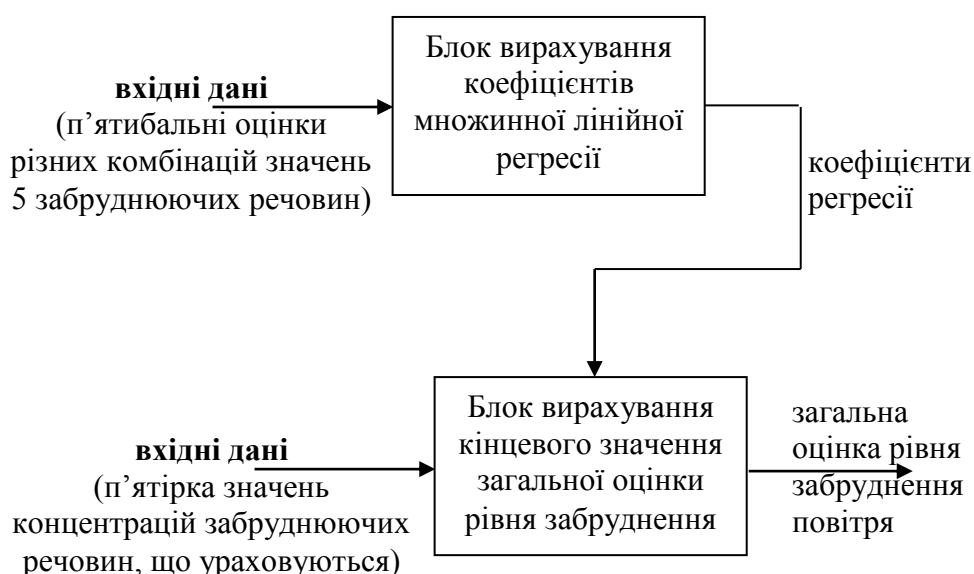


Рисунок 2.8 – Структура математичної моделі системи оцінки забрудненості навколишнього середовища

Ще одним важливим питанням, яке має бути вирішене під час розробки алгоритмічної складової, є отримання масиву вхідних даних для роботи алгоритму лінійної регресії. Його пропонується вирішувати за допомогою експертного опитування спеціалістів, які залучені до вирішення проблем забруднення навколишнього середовища, зокрема повітря. Процедуру опитування опишемо докладніше.

Кожному експертові пропонувалося декілька десятків наборів із п'яти чисел кожний, де:

- перше число відповідає концентрації чадного газу CO;

- друге число відповідає концентрації вуглекислого газу  $\text{CO}_2$ ;
- третє число є концентрацією озону  $\text{O}_3$ ;
- четверте є концентрацією сірчистого газу  $\text{SO}_2$ ;
- п'яте – вміст частинок пилу у повітрі, а саме, показник  $\text{PM}_{2,5}$ .

У шостому порожньому полі експерт повинен був проставити число від 1 до 5, де:

- оцінка «5» на його думку відповідає відмінній якості повітря із даними показниками, яким можна дихати без будь-яких обмежень по тривалості, під час спортивних занять, інтенсивної фізичної роботи, особам із захворюваннями дихальної системи, і т.п.;

- оцінка «4» відповідає хорошому повітрю, яким може без обмежень по тривалості дихати здорова людина, але яке може викликати негативні реакції у осіб із проблемами дихальної системи, а також у деяких здорових осіб під час інтенсивної дихальної активності;

- оцінка «3» відповідає задовільному повітрю, яким без негайних наслідків для здоров'я можна дихати протягом середніх за тривалістю інтервалів часу порядку тижнів (після яких мають наступати періоди покращення), і яке викликає негативні реакції у осіб із вадами дихальної системи, а також може викликати хвороби дихальної системи у первинно здорових осіб при тривалому диханні (протягом місяців та, тим більше, років);

- оцінка «2» відповідає поганому, шкідливому повітрю, вдихання якого протягом середніх по тривалості інтервалів часу порядку тижнів може спричинити хвороби дихальної системи у початково здорової людини, а також негайно викликає серйозні порушення дихання у осіб, які вже мають проблеми з дихальною системою.

- оцінка «1» відповідає надзвичайно небезпечному повітрю, вдихання якого без захисних засобів (типу респіратора) практично зразу викликає негативні реакції і може спричинити серйозні захворювання навіть у здорових людей, що невеликі інтервали часу порядку годин вдихають таке повітря, а також загострення (аж до загибелі) у початкового вже хворих осіб.

Результатом опитування стала таблиця виду 2.1 із численним набором вхідних даних, наведена у Додатку А, яку далі було використано для формування рівняння лінійної множинної регресії.

Таблиця 2.1 – Приклад заповнення таблиці вхідних даних для пошуку рівняння множинної лінійної регресії

| №  | CO, ppm | CO <sub>2</sub> , ppm | O <sub>3</sub> , ppm | SO <sub>2</sub> , ppm | PM <sub>2,5</sub> ,<br>мг/м <sup>3</sup> | Оцінка<br>експерта<br>(1-5) |
|----|---------|-----------------------|----------------------|-----------------------|--|-----------------------------|
| 22 | ...     | ...                   | ...                  | ...                   | ...                                      | ...                         |
| 23 | 18      | 620                   | 0,05                 | 1,8                   | 0,02                                     | <b>3</b>                    |
| 24 | ...     | ...                   | ...                  | ...                   | ...                                      | ...                         |

Зважаючи на те, що вхідних параметрів у моделі буде всього 5, головний визначник системи (2.11) прийматиме вид:

$$\Delta = \begin{vmatrix} n & \sum x_{1i} & \sum x_{2i} & \sum x_{3i} & \sum x_{4i} & \sum x_{5i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \sum x_{1i}x_{3i} & \sum x_{1i}x_{4i} & \sum x_{1i}x_{5i} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 & \sum x_{2i}x_{3i} & \sum x_{2i}x_{4i} & \sum x_{2i}x_{5i} \\ \sum x_{3i} & \sum x_{3i}x_{1i} & \sum x_{3i}x_{2i} & \sum x_{3i}^2 & \sum x_{3i}x_{4i} & \sum x_{3i}x_{5i} \\ \sum x_{4i} & \sum x_{4i}x_{1i} & \sum x_{4i}x_{2i} & \sum x_{4i}x_{3i} & \sum x_{4i}^2 & \sum x_{4i}x_{5i} \\ \sum x_{5i} & \sum x_{5i}x_{1i} & \sum x_{5i}x_{2i} & \sum x_{5i}x_{3i} & \sum x_{5i}x_{4i} & \sum x_{5i}^2 \end{vmatrix}. \quad (2.12)$$

Розрахунки цього та шести частинних визначників дозволили вирішити задачу пошуку невідомих коефіцієнтів у рівнянні множинної лінійної регресії (2.8), де  $m = 5$ . В результаті шукане рівняння отримано у вигляді:

$$y = 4,717 - 0,089x_1 - 0,00004x_2 - 0,502x_3 - 0,085x_4 - 0,0031x_5. \quad (2.13)$$



З інформаційними цілями нагадаємо діапазони концентрацій, що їх вимірюють датчики відповідних забруднюючих величин:

- датчик CO<sub>2</sub>: від 0 до 5000 ppm;
- датчик CO: від 0 до 300 ppm;
- датчик O<sub>3</sub>: від 0 до 100 ppm;
- датчик SO<sub>2</sub>: від 0 до 38 ppm;
- датчик частинок пилу PM<sub>2,5</sub>: 0-1000 мкг/м<sup>3</sup>.

На основі математичної моделі (2.13) можна проводити програмну реалізацію, що і буде виконуватися у наступному розділі. За необхідності можна виконувати коригування моделі, вводячи до неї нові дані від експертів.

Також наведемо тут структурну схему комп'ютерної системи, що включатиме описану математичну модель, вказані раніше датчики, інтерфейсну частину і т.п. – рисунок 2.9.

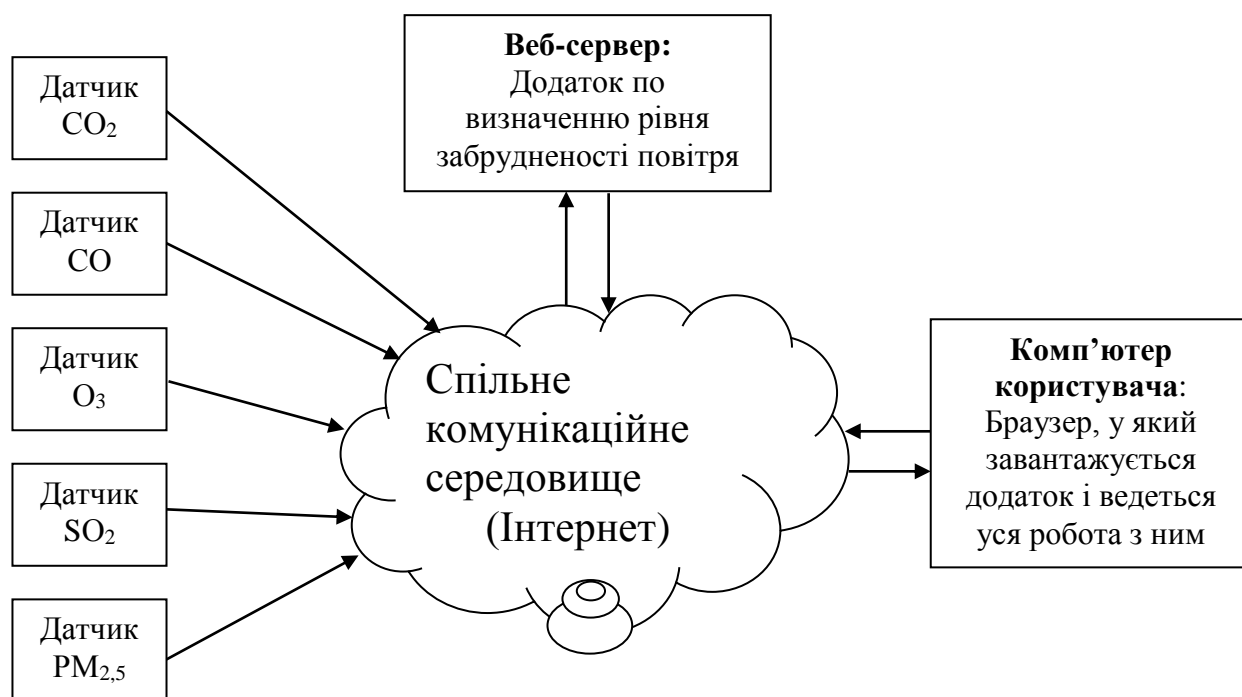


Рисунок 2.9 – Структурна схема комп'ютерної системи, в якій можна втілити розроблену математичну модель та програмну реалізацію системи

Відмітимо, що на рисунку 2.9 зображений найбільш загальний варіант схеми, а у частинних реалізаціях схема може спрощуватися, наприклад, веб-сервер може бути безпосередньо завантажений на комп'ютері користувача, а окремі датчики можуть бути напряму під'єднані до цього комп'ютера, без використання мережі Інтернет.

## 2.4 Висновки по розділу

Таким чином, у даному розділі розроблено математичну модель, що далі буде покладена в основу системи оцінки забрудненості навколишнього середовища, яка побудована на базі використання показів віддалених SMART-датчиків.

Докладно проаналізовані можливості двох магістральних напрямків галузі штучного інтелекту – теорії нечітких множин та нейронних мереж, а також математичної статистики. Встановлено, що для даної задачі більш доцільним є використання добре передбачуваних методів статистики, а саме, регресійного аналізу. Конкретно, модель будувалася у вигляді рівняння множинної регресії із п'ятьма вхідними показниками, якими є рівні забруднення по обраним частинним забруднюючим речовинам. Зважаючи на загальну математичну складність роботи із множинною регресією, ступінь відповідного рівняння обрано рівним 1 (лінійна множинна регресія). Також у розділі розроблено алгоритми, на основі яких далі будуватиметься програмна реалізація системи, що досліджується.

### 3 ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ МОДЕЛІ, ЩО ОПИСУЄ ПРОЦЕС ОЦІНКИ ЗАБРУДНЕНОСТІ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

3.1. Особливості програмної реалізації раніше розробленої алгоритмічної складової процесу

Першочерговим питанням, яке постає перед розробниками практично будь-якого програмного забезпечення, є вибір моделі або технології його розробки. Такими, що реально використовуються на сьогоднішній день у виробничій практиці, є технології структурного (процедурного) та об'єктно-орієнтованого програмування [21]. Кожна з них має свої особливості, переваги і недоліки.

Структурне, або як його ще називають практикуючі програмісти, процедурне програмування засноване на використанні окремих структурних блоків - в першу чергу, підпрограм (процедур і функцій). Структурне програмування передбачає побудову програми відповідно до трьох основних принципів: слідування, розгалуження, повторення. Об'єктно-орієнтоване програмування застосовується для обширних за змістом задач, коли складність сприйняття проекту, побудованого за структурним підходом стає завеликою.

Ґрунтуючись на особливостях двох існуючих методів програмування, вибираємо структурний підхід як більш простий, що відповідає невеликим (не промисловим) масштабам проектного ПЗ, а також вимогам до складності (програма не повинна бути складною через цільову аудиторію користувачів, якими є не спеціалісти з галузі ІТ, а екологи чи «домашні» користувачі, які піклуються про чистоту повітря, що вдихають).

Після вибору технології програмування наступним кроком є вибір мови програмування (або комплексу мов – за умови, наприклад, написання програмного продукту у вигляді розподіленого клієнт-серверного додатку, а також при деяких інших умовах).

Найбільш укрупнено за типом платформи кінцевого програмного продукту усі мови програмування можна розподілити на засоби для розробки настільних додатків, засоби веб-розробки та мобільної розробки. Мобільні технології в даній роботі взагалі не можуть стати в нагоді, оскільки розроблюваний програмний продукт не потребує мобільності, а обчислювальні ресурси мобільних систем в рази менше, ніж у настільних, чи тим більше, клієнт-серверних. Що ж стосується альтернативи «веб-додаток» чи «настільне програмне забезпечення», перший варіант видається зручнішим через наступні причини:

- є повністю кросплатформенним (якісні веб-додатки однаково функціонально працюють у будь-яких сучасних веб-браузерах, незалежно від того, на якій операційній системі вони запуснені);

- дозволяє використовувати обчислювальні ресурси серверу, які завжди значно потужніші, ніж у настільних ПК;

- можуть мати досить швидкий та зручний інтерфейс, за умови використання у них сучасних веб-технологій (типу AJAX, кешування, локальної доробки даних, предзавантаження сторінок, і т.д.).

Таким чином, обираємо веб-розробку та проаналізуємо засоби, що можуть при такому випадку бути використані. Розглянемо найпоширеніші засоби розробки, що є популярними на даний момент у відповідній галузі.

В цілому увесь процес веб-розробки традиційно ділять на дві компоненти: front-end та back-end. Зважаючи на те, що у даній роботі не буде використовуватися база даних, то використання бек-енд компонентів, взагалі кажучи, не є потрібним, а розглядати слід лише фронт-енд рішення, можливостей яких буде цілком достатньо для досягнення цілей даної роботи.

Усі програмні коди та висхідні тексти, що виконуються у браузері, тобто на стороні клієнта, відносяться до front-end частини. З іншого боку, все, що виконується на сервері, відноситься до back-end складової.

Якщо глибоко не вдаватися у детальний аналіз, то коротко можна сказати, що набір інструментів для front-end розробки значно менший, ніж для беку. Серед них можна виділити:

– HTML – п'ятої версії (HTML5), найсучасніший стандарт мови гіпертекстової розмітки (HyperText Markup Language), де традиційно під гіпертекстом мають на увазі текст, оснащений гіперпосиланнями на інші частини документів та супроводжуваний картинками, звуками, відео, і т.п.;

– CSS – третьої версії (CSS3) – правила запису стилів різноманітних елементів веб-сторінки, або, якщо говорити точно – каскадні таблиці стилів;

– JavaScript – це спеціалізована мова програмування, яка традиційно використовується для управління об'єктною моделлю документа у браузері під час перегляду сторінок мережі Інтернет (існують продукти, що перетворюють JavaScript на мову загального призначення, команди якої виконуються на сервері, наприклад, Node.JS; але такий підхід сильно протирічить початковій концепції використання цієї мови програмування, і, за умови наявності значної кількості традиційних засобів back-end-розробки перетворення на серверну мову ще й JavaScript значна кількість програмістів вважає абсолютно недоцільною, тому про цей варіант використання говорити більше не будемо).

Особливістю JavaScript є те, що його висхідні програмні коди інтерпретуються, що є досить гнучким, але повільним рішенням. Однак, у значній частині усіх прикладних задач (як і у випадку зведення програмної складової системи оцінки забрудненості повітря), необхідність у швидкій роботі програми не є критичною. Взагалі, розглянувши особливості мови JavaScript (разом із основою у вигляді HTML та CSS) можна із впевненістю сказати, що її доцільно узяти в якості основного засобу для зведення системи оцінки забрудненості навколишнього середовища на SMART датчиках.

Обравши мову програмування, розглянемо структуру програмного продукту, що розробляється. Він має чимало досить складних особливостей, серед яких, найголовніше, виділяється власне розрахункова процедура, яка виконується за формулою (2.13). Вона, як і багато інших функціональних речей, реалізована функцією мови JavaScript (відповідно до обраного вище структурного підходу до програмування). Наведемо повний перелік функцій, що використовуються у програмному продукті разом із їхніми короткими описами.

а) `filldiv()` – заповнює основний шар (`maindiv`), вибудовуючи в ньому таблицю для введення даних потрібної форми та розміру. Говорячи точніше, основний шар містить дочірній шар `paramlist`, який при завантаженні сторінки спочатку є порожнім (у статичному HTML), а потім заповнюється за допомогою перезапису його властивості `innerHTML`. Як відомо, перезапис цього поля для будь-якого блокового тегу одразу призводить до зміни його відображення браузером, тому це є досить зручною можливістю для оновлення даних таблиці з параметрами розрахунку рівня забрудненості. Ця функція `filldiv()` видає у веб-сторінку назви окремих частинних показників та їх ваги (коефіцієнти) у рівнянні лінійної регресії. Її основною частиною є цикл по змінній `j`, який заповнює рядки таблиці з параметрами. Всередині цього циклу є цикл по змінній `i`, який створює 10 варіантів тегу `<option>` для випадваючого списку ваг даного параметру, що можна міняти. Після завершення великого циклу по змінній `j`, наприкінці функції `filldiv()`, впроваджено ще один цикл по цій же змінній в якому усім полям для вибору значення даного частинного показника встановлюється обробник, за допомогою якого перевіряється введене значення, чи є воно числом та відповідає необхідним вимогам (наприклад, є додатним);

б) `startcalc()` – реалізує формулу (2.13) і побудована на базі алгоритмів, розглянутих у попередньому розділі. Алгоритмічно функція не є складною, оскільки фактично призначена для реалізації формули зваженої суми лінійної регресії (яка рахується досить просто, якщо коефіцієнти регресії уже визначені попередньо, як власне і є у даному випадку);

в) `checknum()` – сервісна функція, яка перевіряє, щоби користувач вводив у поля тільки числові дані, а у протилежному випадку очищує введене (нечислове) значення і видає відповідне попередження. Функція виконана універсальною, що реалізується за допомогою об'єкта `this`: оцінюється та корегується поле `this.value`, що дозволяє використовувати одну і ту ж саму функцію для контролю практично будь-якого текстового поля, що має властивість `value` (тобто усіх тегів типу `INPUT`);

г) `bindhandlers()` – стандартна для JavaScript функція, яка викликається уже коли всі елементи управління створені, і, отже, для них можна задати обробники

методом `addEventListener`. Використання такої функції замість прописування усіх обробників безпосередньо в самих тегах (за допомогою атрибутів `onload`, `onclick`, `onkeydown`, і т.д.) дозволяє використовувати повністю автономні скрипти, що розміщуються в окремих js-файлах.

### 3.2. Опис інтерфейсу користувача системи

При створенні програмного продукту наступним кроком після вибору технології програмування, мови програмування та основного засобу розробки є проектування інтерфейсу користувача системи, тобто її зовнішнього вигляду. Він формується набором та розміщенням необхідних елементів управління, які призначені для збору вхідної інформації від користувача та для відображення йому вихідної інформації по роботі програмного продукту.

Беручи до уваги специфіку проектного програмного продукту (який має надати оцінку загального рівня забрудненості на основі набору частинних показників забрудненості, які є вхідними даними), можна зробити висновок, що у нього має бути 2 основні робочі зони: для збору інформації та для відображення результатів оцінки. Таким чином, якщо класифікувати інтерфейс користувача, то, незважаючи на складність самої задачі, що вирішується, інтерфейс користувача хоча і є графічним, але досить простим і може бути віднесений до пакетних.

В інтерфейсі передбачені наступні елементи:

- текстові мітки для пояснень, про який саме параметр йде мова у даному рядку;
- прапорці, знімаючи та встановлюючи які, можна включати або виключати із розрахунку той, або інший частинний показник. Дана можливість є необхідною у випадку, коли покази одного із датчиків відсутні, але існує необхідність оцінки рівня забрудненості хоча би на основі даних наявних працюючих датчиків;
- поле для редагування ваги того, чи іншого частинного показника. Поле необхідне у зв'язку з тим, що в майбутньому можливим є корегування моделі

лінійної регресії та деяка зміна її коефіцієнтів, які таким чином можна змінити безпосередньо у вікні програмного продукту, а не у його вихідному тексті;

- текстове поле для введення значення того, чи іншого частинного показника;
- кнопка активізації розрахунку;
- поле для виведення результуючого значення показника забруднення.

Кінцевий вигляд інтерфейсу представлений на рисунку 3.1.

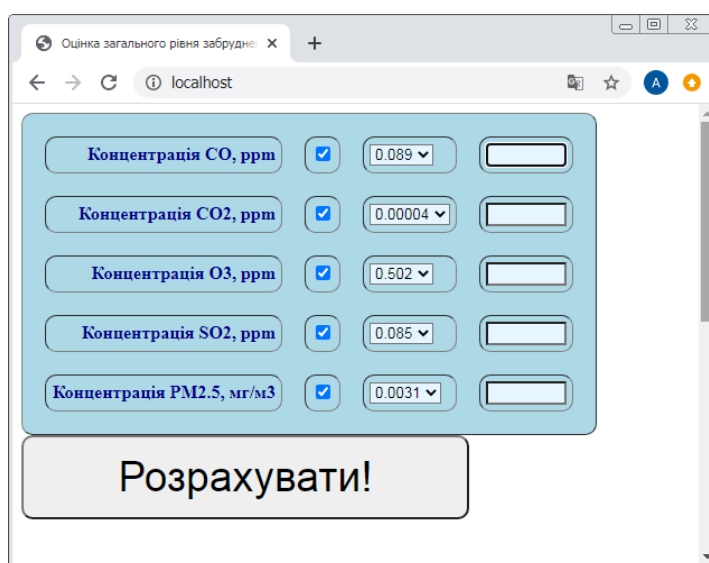


Рисунок 3.1 – Інтерфейс користувача програми для оцінки загального рівня забрудненості на основі показів 5 окремих SMART-датчиків

Інструкція користувача до розробленого програмного продукту наведена у Додатку Б.

### 3.3 Опис та аналіз результатів програмної реалізації. Тестування продукту та аналіз ефективності розроблених рішень

Для створеної програмної реалізації було проведено тестування, за результатами якого встановлено наступні факти:

- програмний продукт працює без виникнення системних помилок, аварійних завершень, «синіх» екранів, інших порушень доступу до пам'яті чи інших частин системи;



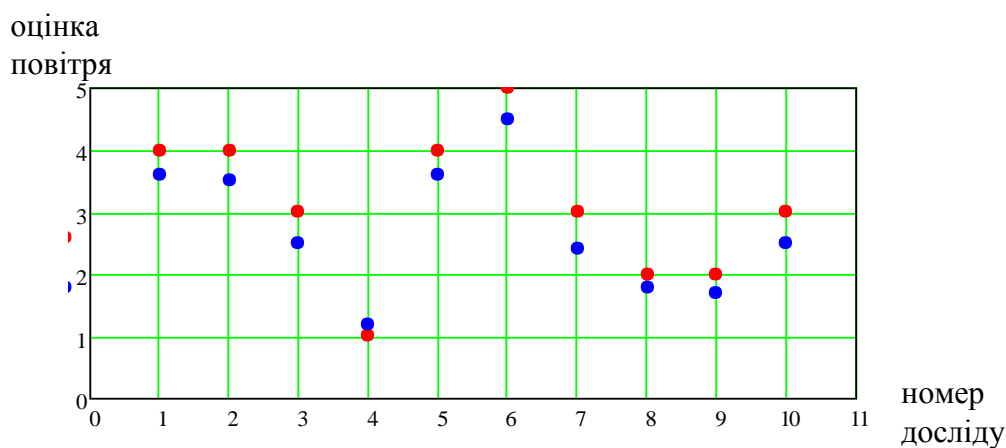
- логіка роботи програми відповідає початково заявленим цілям, що було встановлено в результаті порівняння результатів оцінки рівня забрудненості програмою та експертами.

Для перевірки останньої тези перед формуванням коефіцієнтів (2.13) із початкового набору вхідних даних були випадково вилучені 10 записів, які пізніше використовувалися для оцінки ефективності висновків системи. Порівняння оцінок, наданих системою по цим 10 записам із оцінками експертів показали цілком задовільне середнє відхилення біля 14 % (таблиця 3.1), що свідчить про досить високий рівень достовірності рішень програми і можливість за допомогою неї ефективно оцінювати загальний рівень забрудненості на основі частинних показників, що надаються окремими SMART-датчиками кожний.

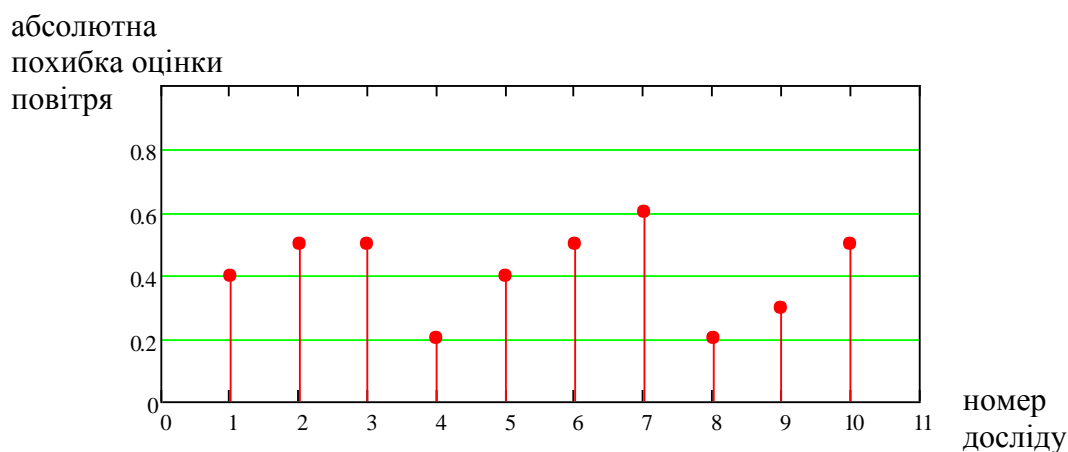
Таблиця 3.1 – Порівняння оцінок, виконаних спроектованою системою та еталонних значень.

| №      | Еталонна оцінка | Оцінка, виставлена системою | Похибка, % |
|--------|-----------------|-----------------------------|------------|
| 1      | 4               | 3,6                         | 10         |
| 2      | 4               | 3,5                         | 13         |
| 3      | 3               | 2,5                         | 17         |
| 4      | 1               | 1,2                         | 20         |
| 5      | 4               | 3,6                         | 10         |
| 6      | 5               | 4,5                         | 10         |
| 7      | 3               | 2,4                         | 20         |
| 8      | 2               | 1,8                         | 10         |
| 9      | 2               | 1,7                         | 15         |
| 10     | 3               | 2,5                         | 17         |
| Всього | -               | -                           | <b>14</b>  |

Інформація, наведена у таблиці може бути наочно представлена у графічній формі на рисунку 3.2. На рисунку 3.2, *а* у відповідному масштабі наведено обидва значення оцінки повітря: червоним – еталонна оцінка, а синім – передбачення створеної у даній роботі системи. На рисунку 3.2, *б* можна бачити абсолютне значення похибки, що виникає при порівнянні еталонного та отриманого системою значень. Бачимо, що похибка не виходить за абсолютні межі 0,6 (по шкалі від 1 до 5), що у відсотковому вираженні для середнього значення похибки складає розраховане у таблиці 3.1 значення 14 %.



*а)*



*б)*

Рисунок 3.2 – Порівняння еталонних оцінок та оцінок, *а* – виставлених системою; *б* – відповідні похибки у абсолютному вираженні

### 3.4 Висновки по розділу

Таким чином, у даному розділі розглянуто особливості програмної реалізації системи оцінки забрудненості навколишнього середовища. По-перше, обґрунтовано вибір програмних засобів для можливості виконання запропонованої реалізації і для більшої кросплатформенності вирішено використовувати мову JavaScript з основою на HTML. Через це, при виникненні потреби доробки цього продукту, розміщення його в мережі Інтернет буде простим процесом, що не потребує робочого часу через будь-яку переробку створеного ПЗ.

Розроблені вище алгоритми впроваджено у програмних кодах і на основі експертних даних проведено визначення коефіцієнтів рівняння регресії. Запропоноване програмне забезпечення із вбудованою математичною моделлю піддано тестуванню, яке показало його стабільність (відсутність системних помилок при його роботі), а також належне виконання покладеної на нього функціональності: середнє відхилення показів програми відхиляється від еталонних даних на величини до 14%.

## ВИСНОВОК

Таким чином, у даній роботі було запропоновано систему оцінки забрудненості навколишнього середовища на SMART датчиках, що дозволяє адекватно реагувати (готуватися) до більших, чи менших рівнів забруднення (переносити заходи та роботи, планувати дозвілля дітей, і т.п.). Для цього у роботі було виконано наступні дії:

- проаналізовано існуючі види забруднень навколишнього середовища, та обґрунтовано, що саме забруднення повітря є найбільш небезпечним і таким, якому слід приділяти максимальну увагу;

- проаналізовано існуючі підходи до прогнозування фізичних величин, зокрема рівнів забруднення, та обрано метод математичної статистики із розділу регресійного аналізу, а саме зведення лінійної множинної регресії;

- було запропоновано відповідну математичну модель, причому на основі масиву експертних оцінок загального рівня забрудненості повітря по відповідним комбінаціям обраних п'яти частинних показників (CO, CO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, PM<sub>2.5</sub>) визначено коефіцієнти моделі.

- проведено тестування та встановлено ефективність моделі (шляхом порівняння її оцінок із оцінками експертів, що відповідають невеликій частині початкового масиву вхідних даних, було встановлено середнє розходження порядку 14 %).

Таким чином, запропонована система може ефективно використовуватися для оцінки загального рівня забрудненості навколишнього середовища, а саме повітря, надаючи результат у зрозумілій та звичній 5-бальній шкалі.

Також, за результатами дослідження була підготовлена наукова стаття для публікації у фаховому виданні.

## ПЕРЕЛІК ПОСИЛАНЬ

- 1 Біляєв М. М. Моделювання і прогнозування стану довкілля: підручник для студентів вищих навчальних закладів / М. М. Біляєв, В. В. Біляєва, П. С.; Дніпропетровський національний університет залізничного транспорту імені академіка В. Лазаряна МОН України. - Кривий Ріг: Вид. Р. А. Козлов, 2016. - 207 с.
- 2 Бисчоков Р. М. Методы проведения анализа и прогнозирования изменения динамики агроклиматических ресурсов Северного Кавказа: монография. – Нальчик: КБГСХА, 2012. – 190 с.
- 3 Ковальчук П. І. Моделювання і прогнозування стану навколишнього середовища: Навч. Посібник. – К.: Либідь, 2003. – 208 с.
- 4 Попов О. О. Математичне моделювання розповсюдження техногенного забруднення від підприємств паливної енергетики / О. О. Попов. // Збірник наукових праць Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України. – К.: ІПМЕ ім. Г.Є. Пухова НАН України, 2009. – Вип. 51. – С. 73-84.
- 5 Олійник Р.В. Моделювання оперативного прогнозування забруднення атмосферного повітря автотранспортними потоками / Р.В. Олійник, С.О. Никонович, С.М. Тарабан // Моделювання оперативного прогнозування забруднення атмосферного повітря автотранспортними потоками С.95-98.
- 6 Бідюк П.І. Моделювання забруднення атмосфери та поверхневих вод в Україні / П.І. Бідюк, О.І. Савенков, О.М. Трофимчук // Екологічна безпека та природокористування, 2011 р. - С.16-34.
- 7 World's Air Pollution: Real-time Air Quality Index [Електронні дані]. - Режим доступу <https://waqi.info/>. – Заголовок з екрана.
- 8 A Beginner's Guide to Air Quality Instant-Cast and Now-Cast. [Електронні дані]. - Режим доступу <https://aqicn.org/faq/2015-03-15/air-quality-nowcast-a-beginners-guide/>:. – Заголовок з екрана.
- 9 Марчук Г.И. Математическое моделирование в проблеме окружающей среды. – М.: Наука, 1982. – 319 с.

10 Царегородцев В.Г. Нейросетевые анализ и моделирование современных связей климата и растительности / В.Г. Царегородцев, Д.И. Назимова // Современные подходы к интеграции информационных технологий. Труды V Всеросс. семинара "Информационные технологии в энергетике", Иркутск, 2000. Иркутск: ИСЭМ СО РАН, 2001. - С.157-165.

11 Басс Л.П. Сверточные нейронные сети с глубоким обучением в задачах обработки гиперспектральных спутниковых данных / Л.П. Басс, М.Г. Кузьмина, О.В. Николаева // Препринты ИПМ им. М.В. Келдыша. 2018. № 282. – 32 с.

12 Bengio Y., Lamblin P., Popovici D., and Larochelle H.. Greedy layer-wise training of deep networks, in Proc. Neural Inf. Process. Syst., Cambridge, MA, USA, pp. 153–160, 2007.

13 Gangwar R. K., Mathur A. K., Gohil B. S., and Sujit Basu. Neural network based retrieval of atmospheric temperature profile using AMSU-A observations, International Journal of Atmospheric Sciences, vol. 2014, Article ID 763060, 8 p.

14 Goodfellow, Ian & Shlens, Jonathon & Szegedy, Christian. (2014). Explaining and Harnessing Adversarial Examples. arXiv 1412.6572.

15 Матвеев М. Г. Методические аспекты применения аппарата нечеткой логики при обработке и анализе метеорологической информации / М. Г. Матвеев, В. В. Михайлов, М. Е. Семенов, Р. Е. Первезенцев // Вестник Военного авиационного инженерного университета. – 2011. – №3 (14). - С. 35–42.

16 Тэрано Т. Прикладные нечёткие системы / Тэрано Т., Асаи К., Сугэно М. — М.: Мир, 1993. — 368 с.

17 Ярушкина Н.Г. Основы теории нечетких и гибридных систем. Учебное пособие. Финансы и статистика, 2004.- 320с.

18 Круглов В.В. Нечеткая логика и искусственные нейронные сети / Круглов В.В., Дли М.И., Голунов Р.Ю. – Физматлит, 2001. - 224 с.

19 Кофман А. Введение в теорию нечетких множеств.- М.: Радио и связь, 1982.- 432 с.

20 Усков А.А. Интеллектуальные системы управления на основе методов нечеткой логики / Усков А.А., Круглов В.В. ; Смоленск: Смоленская городская типография, 2003. - 177 с. - ISBN 5-94223-038-2.

21 Павловская Т.В. Процедурное и объектно-ориентированное программирование. Учебник. - М.: Питер, 2015. - 496 с.