ПРИВАТНЕ АКЦІОНЕРНЕ ТОВАРИСТВО «ВИШИЙ НАВЧАЛЬНИЙ ЗАКЛАД

«МІЖРЕГІОНАЛЬНА АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ»»

Інститут комп’ютерно-інформаційних технологій

Кафедра комп’ютерних інформаційних систем і технологій

**КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА**

**Тема «Розробка системи оцінки ефективності нейромереж у задачах обробки та аналізу метеоданих»**

Виконав студент групи T16-9-17-Б1ППі(4.0д) Носов М.Б.

Керівник: к.т.н., доцент Рябий М.О.

Рецензент: к.т.н., доцент Рудніченко М.Д.

Допущено до захисту

Завідувач кафедри \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ д.е.н., професор Кавун С.В.

(підпис)

Одеса, 2023

РЕФЕРАТ / ABSTRACT

Пояснювальна записка до атестаційної роботи: 71 с., 24 рис., 3 додатки, 30 джерел.

МЕТЕОРОЛОГІЯ, ІНФОРМАЦІЙНА СИСТЕМА, АВТОМАТИЗАЦІЯ, SQLITE, PYTHON, DJANGO, WEB–САЙТ.

Об’єктом дослідження є методи інтелектуального аналізу данних, які використовуються в метеорології, та дозволяють автоматизувати робочі процеси, виконувати впровадження цифрових рішень до робочих процесів, адаптувати інформаційну систему для впровадження в різні напрямки метеорологічної діяльності.

Метою роботи є дослідження особливостей вирішення задачі аналізу метеоданих за допомогою використання алгоритмів машинного навчання. Реалізація такої системи дозволить спростити роботу синоптиків та метеорологів, а також зможе поширити інформацію про метеостан територій та прискорити аналіз доступної метеоінформації.

Для розробки використовуеться мова розробки Python включно з додатковим фреймворком для web-розробки Django 3, а також система керування базою даних SQLite 3.

Результатом роботи є розробка інформаційної системи для збору, обробки та агалізу метеоданих.

METEOROLOGY, INFORMATION SYSTEM, AUTOMATION, DOCUMENTATION, SQLITE, PYTHON, DJANGO, WEB SITE.

The object of research is the methods of intellectual analysis of data used in meteorology, which allow automating work processes, implementing digital solutions to work processes, and adapting the information system for implementation in various areas of meteorological activity.

The purpose of the work is to study the peculiarities of solving the problem of meteorological data analysis using machine learning algorithms. The implementation of such a system will simplify the work of forecasters and meteorologists, and will also be able to spread information about the weather conditions of the territories and speed up the analysis of available weather information.

For development, the Python development language is used, including the additional framework for web development Django, as well as the PostgreSQL database management system.

The result of the work is the development of an information system for the collection, processing and analysis of meteorological data.

ЗМІСТ

[ВСТУП 4](#__RefHeading___Toc9493_489360030)

[1 АНАЛІЗ ПРЕДМЕТНОЇ ТЕМАТИКИ ДОСЛІДЖЕННЯ 7](#__RefHeading___Toc9495_489360030)

[1.1 Аналіз особливостей метеорологічних задач 7](#__RefHeading___Toc9497_489360030)

[1.2 Аналіз специфіки машинного навчання 15](#__RefHeading___Toc9499_489360030)

[1.3 Аналіз можливостей сучасних засобів розрабки програмного забезпечення з використанням алгоритмів машинного навчання 21](#__RefHeading___Toc9501_489360030)

[1.4 Мета та завдання роботи 30](#__RefHeading___Toc9503_489360030)

[1.5 Висновки до першого розділу 31](#__RefHeading___Toc9507_489360030)

[2 ОБГРУНТУВАННЯ ВИБОРУ НАПРЯМКУ ДОСЛІДЖЕННЯ 33](#__RefHeading___Toc9509_489360030)

[2.1 Постановка завдання регресії 33](#__RefHeading___Toc9511_489360030)

[2.2 Порівняльний аналіз та вибір алгоритму регресії 37](#__RefHeading___Toc9513_489360030)

[2.3 Концепція попередньої обробки та аналізу даних 45](#__RefHeading___Toc9515_489360030)

[2.4 Висновок до другого розділу 49](#__RefHeading___Toc6381_2013452180)

[3 РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ЗБОРУ, ОБРОБКИ ТА АНАЛІЗУ МЕТЕОДАНИХ 51](#__RefHeading___Toc9517_489360030)

[3.1 Проектування розробки інформаційної системи 51](#__RefHeading___Toc9519_489360030)

[3.2 Опис розробленої бази даних інформаційної системи 60](#__RefHeading___Toc9521_489360030)

[3.3 Опис розробленого інтерфейсу до інформаційної системи 66](#__RefHeading___Toc9523_489360030)

[3.4 Практичне використання інформаційної системи 72](#__RefHeading___Toc9525_489360030)

[3.5 Висновки до третього розділу 74](#__RefHeading___Toc9527_489360030)

[ВИСНОВКИ 75](#__RefHeading___Toc9529_489360030)

[ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 76](#__RefHeading___Toc9531_489360030)

[ДОДАТОК А 79](#__RefHeading___Toc9533_489360030)

[ДОДАТОК Б 100](#__RefHeading___Toc9535_489360030)

[ДОДАТОК В 106](#__RefHeading___Toc9537_489360030)

# ВСТУП

Інформаційні технології активно використовуються в метеорології і роботі синоптиків, так як дозволяють не тільки автоматизувати їх роботу, а також підвищити зручність її виконання і значно підвищити точність результату.

Метеорологія - це наука, яка вивчає процеси, явища, що відбуваються в атмосфері.

Сучасна метеорологія не змогла б розвинутися без впровадження в неї інформаційних технологи і особливо суперкомп'ютерів, пов'язано це з тим що робота синоптиків, зокрема побудова прогнозу погоди, вимагає великих як тимчасових так і людських ресурсів. Починаючи від простого побудови прогнозу погоди закінчуючи попередженням про насування погодних катастроф, які могли б привести до великих людських жертв і масштабних руйнувань.

Використання обчислювальної техніки і відповідних програм дозволяє вирішувати завдання метеорології з великою точністю, для значно більших територій, при цьому за більш короткий термін і з залученням менших людських ресурсів. У зв'язку з цим дана галузь вкрай сильно потребує розробки зручних і високопродуктивних програмних засобах.

Тема більш ніж актуальна так як розвинена метеорологія дозволить поліпшити важливі аспекти життя людей, наприклад: збільшити ефективність сільськогосподарської галузі, дозволить визначати безпечні для життя людей місця, прокладати більш оптимальні маршрути кораблям і літакам, мінімізувати наслідки природних катастроф.

Зв'язок роботи з програмами наукових досліджень кафедри ПІ. Для поліпшення загальних процесів автоматизації та створення єдиного інформаційного простору для сфери метеорології повинні використовуватися сучасні інформаційні технології, які дозволяють розробити сучасний програмний засіб, здатний виконати поставлене завдання за менший проміжок часу і забезпечити більш високу точність, ніж людина.

Щодо алгоритмів застосовуваних у цій сфері можна виділити проблеми складності алгоритму, довгого часу його виконання, а також гнучкості алгоритму. Сожность алгоритму виявляється у тому що сформована математична модель грамозка і має оптимальну складність алгоритму виконання. Проблема гнучкості виникає внаслідок того, що складена математична модель пов'язана на певних константах прив'язаних до конкретної території або регіону.

Використання нейронних мереж може вирішити обидві ці проблеми, нейронна мережа простіше у використанні, має універсальний інтерфейс отримання вхідних даних і легко адаптується до змін завдяки своїй структурі. У результаті можна отримати універсальний алгоритм, який зможе обробляти дані з будь-яких регіонів.

Розширення погляду залежність набору вхідних даних. Поточні алгоритми, що використовуються в метеорології, використовують для своєї роботи набір вхідних параметрів, що складається з температури повітря, його вологості, напряму вітру та його швидкості. Ці параметри просто збирати, але аналіз погоди на їх підставі може мати серйозну похибку через "прогалини" в моделі. В даному випадку під “пробілами” маються на увазі метеорологічні змінні, які не враховуються в даній моделі, але можуть значно впливати на результат. Нейронні мережі дозволяють легко оцінити значущість своїх вхідних параметрів на результат і могли б допомогти у вирішенні цього завдання у сфері метеорології.

Мета і задачі дослідження. Метою роботи є дослідження ефективності роботи різних варіантів нейронних мереж на задачах обробки та аналізу метеоданих. Для виконання поставленої задачі, необхідно виконати:

* виконати аналіз поточних методів вирішення метеорологічних завдань;
* виконати аналіз впливу впровадження нейронних мереж у предметну область;
* виконати порівняльний аналіз схожих програмних систем;
* визначити програмні засоби реалізації інформаційної системи;
* виконати проектування інформаційної системи;
* виконати розробку інформаційної системи та моделей нейромереж;
* дослідження ефективності роботи різних моделей нейронних мереж;
* дослідження ступеня впливу різних вхідних параметрів на результат виконання завдання;

Об’єкт дослідження. Об'єктом дослідження є різні варіації нейромереж та набори вхідних параметрів для них.

Методи дослідження. В якості наукової основи дослідження використовують: алгоритми машинного навчання, методи оцінки якості моделей прогнозування та класифікації даних.

Наукова новизна одержаних результатів. Наукова новизна отриманих в результаті виконання даної атестаційної роботи полягає в агрегації та адаптації існуючих моделей прогнозування та класифікації, створення єдиного програмного застосування для їх використання з метою аналізу метеоданих. Розроблене програмне забезпечення дозволяє проводити аналіз метеоданих, переглядати результати оцінки якості роботи нейромереж та забезпечувати інтерактивну взаємодію між користувачем та інтерфейсом.

Практичне значення одержаних результатів. Розроблена інформаційна система може використовуватися як система автоматизації робочих процесів метеорологічних центрів, а також може виступати в якості початкової системи прогнозування погоди. Головною особистістю такої інформаційної системи є її простота використання та гнучкість до алгоритмів обробки та аналізу даних.

Результатом розробки є інформаційна система для метеорологічних центрів, яка автоматизує процеси збору метеоданих, та дозволяє зручно і швидко проводити порівняльний аналіз моделей нейромереж та наборів їх параметрів.

Публікації. Результати розробки та дослідження створеного програмного застосування опубліковано у двох науково-практичних конференціях [38, 39], за результатами яких сформовані збірники тез (додаток В).

# АНАЛІЗ ПРЕДМЕТНОЇ ТЕМАТИКИ ДОСЛІДЖЕННЯ

* 1. Аналіз особливостей метеорологічних задач

Метеорологія є комплексом наук все атмосферні явища, такі як: погода, клімат, хмари, склад атмосфери, її будова, тепловий режим і влагообмін, оптичні та акустичні, поведінка фронтів, вітру і електричного поля. Всі ці дослідження допомагають знайти і зрозуміти фізичні залежності які відбуваються в атмосфері, що в кінцевому рахунку наблизить людство до розуміння життєвих циклів планети.

Сучасних метеорологів можна розділити по виду діяльності на наступні категорії або наукові напрямки:

* Фізична метеорологія – займаються розробкою радіолокаційних і космічних методів дослідження атмосферних явищ.
* Динамічна метеорологія – вивчення фізичних механізмів атмосферних процесів.
* Синоптична метеорологія – наука про закономірності зміни погоди.
* Кліматологія – наука вивчаюча клімат як сукупність погодних характеристик за багаторічний період і їх зміна властивих певної території.
* Аерологія – наука вивчає верхні шари атмосфери для декількох десятків кілометрів від поверхні землі.

Прикладні напрямки сучасної метеорології:

* Авіаційна метеорологія – наука вивчає вплив погоди на діяльність авіації.
* Агрометеорологія – наука яка вивчає вплив погоди на сільське господарство.
* Біометеорологія – наука вивчає вплив атмосферних процесів на людину та інші живі організми.
* Ядерна метеорологія – наука вивчає природну і штучну радіоактивність, поширення в атмосфері радіоактивних домішок, вплив ядерних вибухів.
* Радіометеорологія – наука вивчає поширення радіохвиль в атмосфері.

Якщо узагальнити те кожен розділ займається пошуком деяких погодних залежностей і вивчення впливу погодних явище на деяку галузь. Для виконання цієї роботи необхідно взаємодіяти злагоджено і задіяти безліч фахівців по всьому світу, які кожні три години збирають інформацію про погоду зі всіх доступних джерел, далі проводять їх обробку і будують різні карти описують графічним методом поточний погодний стан атмосфери в різних регіонах. Далі за отриманими картками проводиться аналіз виявляє необхідну інформацію, наприклад це може бути аналіз погодних аномалій для розуміння того чи можуть вони перерости в природні катастрофи і якщо це так варто попередити населення цих регіонів.

Метеорологічні данні які найчастіше збирають для досліджень [3]:

* Місце та час коли були зняті метеодані, базова інформація яка потрібна для прив’язки метеоданих на карті.
* Висота хмар від рівня моря.
* Дальність видимості. Кількість метрів далі котрих не можна побачити наземні об’єкти.
* Вид хмар, якій преобладає в даному секторі.
* Напрям вітру.
* Швидкість вітру.
* Температура повітря в час збору даних.
* Точка роси, це температура повітря при якій починає випадати роса.
* Атмосферний тиск на рівні метеостанції.
* Атмосферний тиск на рівні моря.
* Максимальна та мінімальна температура за добу.
* Погодне явище.
* Кількість опадів.

На базі цих метеоданих будуються прогнози погоди та більшості з усіх метеорологічних розрахунків.

Основою для прогнозу погоди є облік періодичних і неперіодичних змін метеорологічних величин і явищ погоди [4]. Періодичні зміни тієї чи іншої метеорологічної величини обумовлюються добовим і річним ходом цієї величини, неперіодичні – еволюцією і переміщенням синоптичних об'єктів: циклонів і антициклонів, повітряних мас і атмосферних фронтів. Ось чому прогнозом погоди завжди передує прогноз синоптичного положення. Найбільшу трудність і практичний інтерес представляє облік імені не періодичних змін.

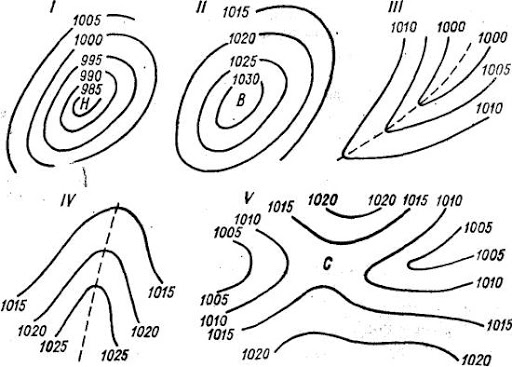


Рисунок 1.1 – Приклад відображення баричних систем на картах.

Синоптичний метод в даний час є основним при розробці довгострокового прогнозу погоди.

Суть методу в тому, що на підставі аналізу карт погоди за кілька послідовних термінів складають прогноз синоптичного положення, який заклю чає в прогнозі виникнення, переміщення і еволюції повітряних мас, атмосферних фронтів, баричних систем, приклад відображення баричних систем на картах рис 1.1.

Карта, на яку наносять передбачуване положення синоптичних об'єктів, називається прогностичною, рис 1.2.

Прогностичну карту складають на певний момент часу наступної доби, а іноді на двоє або більше доби вперед.

На підставі прогностичної карти складають прогноз погоди в тому чи іншому районі. Причому прогноз погоди є логічним продовженням прогнозу синоптичного положення і виходить з основного принципу, що полягає в припущенні, що з переміщенням і еволюцією синоптичних об'єктів переносяться з певними змінами і властиві їм умови погоди.

Тому за прогностичне значення метеорологічної величини в першому наближенні приймаються їх значення в районі, звідки очікується переміщення синоптичного об'єкту, в район, для якого складається прогноз погоди. При такому великому обсязі роботи та даних необхідних до зберігання дана галузь не може обійтися без використання інформаційних технологій. Особливо новітніх програмних засобів, які би спростили роботу з розрахунками та потужних комп’ютерів, які би ці алгоритми оброблювали.

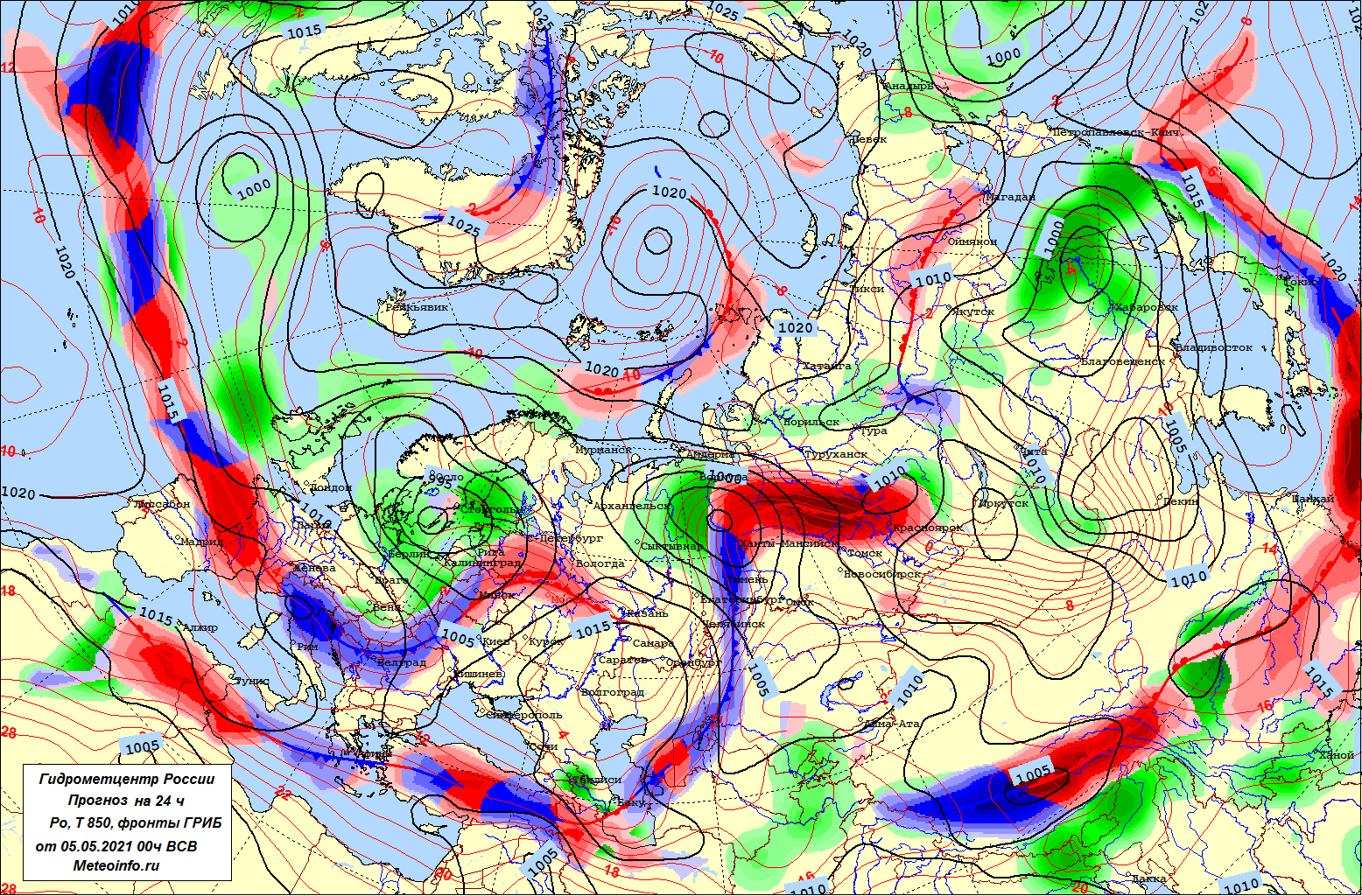


Рисунок 1.2 – Приклад прогностичної метеорологічної карти

Інформаційні технології щільно увійшли і намертво закріпилися в житті сучасних метеорологів.

У своїй щоденній роботі їм необхідно використовувати масивні програмні комплекси які виконують завдання збору, зберігання і перегляду інформації з джерел метеоданих, програми для побудови погодних карт, для визначення і відстеження баричних систем і атмосферних фронтів, побудови прогностичних карт, виявлення погодних аномалій, аналізу погодних даних, побудови графіків залежностей, отримання статистики і графічного відображення метеоданих, а також прогнозування катастроф і інші.

Вплив інформаційних технологій на галузь досить велике, так як без них виконання більшості завдань сучасної метеорології було б неможливим. Якби метеорологи використовували старі або "класичні" методи, то всьому людству довелося б освоїти цю професію і спільно виконувати її завдання. Сам процес збору даних залучає більше кількість людей, а її аналіз зажадав би все доступне в добі час.

Для вирішення та прискорення цих процесів людство прийшло до використання суперкомп'ютерів [5].

Суперкомп'ютером називається комплекс електронних обчислювальних машин працюють як одне ціле, що дає всій системі велику обчислювальну потужність.

Однак навіть суперкомп'ютерів необхідно достатньо часу для розрахунків еволюції баричних систем і прогнозу погоди.

Наприклад, для розрахунку короткострокового прогнозу погоди для всієї території України, з точністю в п'ять квадратних кілометрів, середньостатистичному суперкомп'ютера знадобиться близько години.

На даний момент процес побудови прогнозу запускається не частіше ніж раз на три години, цього цілком достатньо для більшості територій.

Винятками ж є території з високою ймовірністю появи атмосферних катаклізмів. Як приклад таких областей можна привести

Японію, з їх штормами і цунамі, південні штати Америки, з місцевими повенями і торнадо і т.д.

З кожним днем ​​необхідних метеорологічних розрахунків стає все більше, через те що все більше галузей виявилися залежні від інформації про стан атмосфери або про її вплив на свою галузь і отже отримали необхідність в аналізі метеоданих.

Завдяки цій тенденції сильно збільшилося фінансування метеорології що в кінцевому рахунку дозволило метеорологам почати безліч досліджень, які в майбутньому допоможуть краще розуміти процеси атмосфери і їх вплив на наше життя.

Основна частина фінансової підтримки необхідна для придбання обладнання та розробки програмного забезпечення для всіх галузей даної сфери.

При активному розвитку інформаційних технологій в сфері метеорології для людства можуть відкритися можливості:

* + Ідеальна організація і оптимальний розподіл вирощуваних культур. При точному знанні того які метео-параметри були раніше і які будуть в кожній точці світу люди зможуть розподілити висаджуються культури ідеально по необхідним умовам. Дане нововведення дозволить збільшити обсяги вирощуваних культур, але також пристосуватися до змін клімату і вчасно реагувати на випадки коли земля стає менш придатною для вирощування конкретної культури.
  + Оптимізація і полегшення маршрутів суден і літаків. Дане нововведення дозволяє запобігти небезпеці в дорозі і потенційно врятували б життя людей, мінімізували б витрати на ремонтні роботи транспорту, а також не допустити забруднення природи від можливих аварій.
  + Можливість побудови різних архітектурних споруд. Для побудови різних споруд важливо знати які погодні умови переважають в районі будівлі щоб зрозуміти як сильно треба зміцнити конструкцію будівлі для її стійкості.
  + Ремонтні роботи. Частково відноситься в попередньому пункту, для деяких ремонтних робіт важливі певні погодні умови з чим і можуть допомогти нові системні комплекси метеорологів, точно вибравши проміжок часу з ідеальною погодою для ремонтних робіт.

Не менш значущий вплив інформаційні технології привнесли в системи для вимірювання метеоданих.

Раніше використовувані аналогові прилади, рис 1.3, хоч і відрізнялися дешевизною і примітивністю в роботі з ними, проте були недостатньо точними і досить громіздкими.



Рисунок 1.3 – Класична метеостанція обладнана аналоговою системою вимірювання

Сучасні прилади компактні, мають можливість передавати результати вимірювань по лініях зв'язку і при цьому мають велику точність. Це дозволяє прискорити процес збору даних, а їх автономність дає можливість вільно масштабувати мережу метеостанцій, рис 1.4.

Подібні системи мають основний стовп на якому закріплені різні датчики, наприклад датчики напряму вітру, кількості опадів, атмосферного тиску та різні датчики для аналізу повітря.

Система автономного живлення представляє собою джерело енергії, систему стабілізації енергії та систему накопичування енергії.

Як джерело енергії може бути сонячні панелі чи вітряки. Система накопичування енергії це акумуляторна батарея яка буди живити станцію в час коли енергії системи живлення не буде достатньо енергії [5].



Рисунок 1.4. – Приклад сучасної метеостанції та функцій які вона може виконувати

* 1. Аналіз специфіки машинного навчання

Проаналізувавши завдання метеорології можна дійти висновку, що в ній можна застосувати майже всі види та варіації нейромереж та алгоритмів машинного навчання.. Щодо алгоритмів застосовуваних у цій сфері можна виділити проблеми складності алгоритму, довгого часу його виконання, а також гнучкості алгоритму. Сожность алгоритму виявляється у тому що сформована математична модель грамозка і має оптимальну складність алгоритму виконання. Проблема гнучкості виникає внаслідок того, що складена математична модель пов'язана на певних константах прив'язаних до конкретної території або регіону.

Використання нейронних мереж може вирішити обидві ці проблеми, нейронна мережа простіше у використанні, має універсальний інтерфейс отримання вхідних даних і легко адаптується до змін завдяки своїй структурі. У результаті можна отримати універсальний алгоритм, який зможе обробляти дані з будь-яких регіонів.

Розширення погляду на залежність набору вхідних даних. Поточні алгоритми, що використовуються в метеорології, використовують для своєї роботи набір вхідних параметрів, що складається з температури повітря, його вологості, напряму вітру та його швидкості. Ці параметри просто збирати, але аналіз погоди на їх підставі може мати серйозну похибку через "прогалини" в моделі. В даному випадку під “пробілами” маються на увазі метеорологічні змінні, які не враховуються в даній моделі, але можуть значно впливати на результат. Нейронні мережі дозволяють легко оцінити значущість своїх вхідних параметрів на результат і могли б допомогти у вирішенні цього завдання у сфері метеорології.

Починаючи від найскладнішого і найкомплекснішого штучного інтелекту, для обробки загальної картини поведінки метеорологічних об'єктів, до простих математичних моделей для аналізу даних конкретного датчика на вузькому участі землі. Для завдань прогнозування використовуються відповідні нейронні мережі, навчені на сотнях тисяч записів метеоданих з різних точок світу, завдання визначення погодних явищ вирішуються алгоритмами класифікації. Також значний вплив на результат надає процес підступу та аналізу аномалій. Аномаліями називають будь-яке значне відхилення в даних, їх поява може будувати висновки про різного роду сторонньому вплив на метеоситуацію в місці їх збору. У цьому контексті, в першу чергу варто уточнити значущість і види сил, які можуть мати вплив.

Безліч різних факторів може впливати на метеорологічну ситуацію, наприклад, це можуть бути різні природні катастрофи, наприклад, потопи і пожежі або техногенні катастрофи у вигляді тих же пожеж або витоків різних речовин. Подібні явища безпосередньо впливають на метеоситуацію в певній області, безпосередньо впливає на температуру, швидкість і напрям вітру, тиск, вологість. Залежно від масштабів це може вплинути і на загальну метеорологічну картину у світі.



Рисунок 1.5 – Приклад метеостанції домашнього типу

Якщо природні катастрофи відрізнити досить просто, то техногенні можна виділити тільки завдяки датчикам аналізу склад повітря і різні домішки в ньому. Подібні станції сильно відрізняються від професійних аналогів але можуть збирати дані про температуру, вологість повітря і атмосферний тиск, рис 1.5. Деякі моделі дозволяють аналізувати якість повітря, конкретно його склад і кількість домішок. Питання відстеження та аналізу стану повітря не менш важливий для метеорології. У повітрі можуть міститися різні домішки які можуть той чи інший ефект на погоду в регіоні. Таким чином вивчаючи склад повітря можна в майбутньому передбачити згубні явища в регіоні по зміні складу його повітря і навпаки, припустити склад повітря через погодні явища в регіоні, рис 1.6.

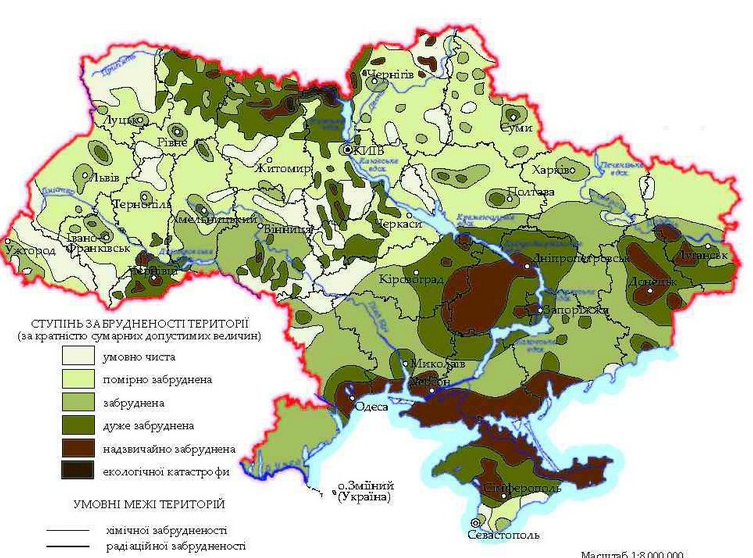


Рисунок 1.6 – Карта забруднення повітря і грунту України

Факторами зміни можуть бути, як досить очевидні, наприклад заводи або аварії, так і менш очевидні і внаслідок чого більш небезпечні, наприклад розкриття покладів метану внаслідок землетрусів. Аналіз складу повітря більш актуальний для другого випадку, конкретно для систем запобігання катастроф. Система вивчає аномалії в повітрі і може виявити виниклі проблеми, після чого попередити про це відповідні служби. Існують приклади місць які вимагають постійного моніторингу ситуації, наприклад озера в парку Йеллоустоун, які є потенційно небезпечною зоною і можуть завдати серйозної шкоди регіону. Також подібні системи фахівці можуть відслідковувати так звані «мертві зони» в морях і океанах, це зони в яких сталася якась аномалія в слідстві якої живі організми в ній вимерли або покинули її, рис. 1.7. Найчастіше подібний ефект викликаний забрудненнями вод і подібні місця відслідковуються за складом повітря.

Все це створює велику кількість параметрів та їх комбінацій, які треба враховувати при аналізі метеоданих, а про багатьох важливих метеорологічних змінних можуть навіть не підозрювати. У вирішенні цих завдань використання нейронних мереж є вкрай ефективним. Подібні алгоритми можуть визначати різні види аномалій залежно від величених відхилень у даних. У деяких випадках навіть стабільні дані можуть бути ознакою аномалії. На планеті відбувається багато внутрішніх цикорів, які мають свій вплив на метеоситуацію, наприклад такі як зміна пір року та цикли кліматичних періодів.

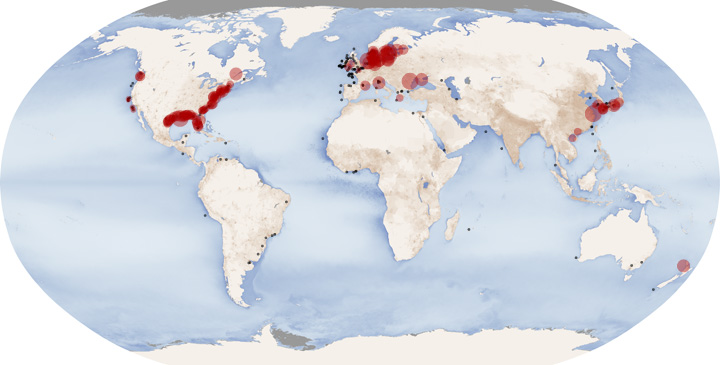


Рисунок 1.7 – Карта світового океану із зазначеними на ній мертвими зонами

Ефективність розвитку сфери сильно залежить від стандартизації її елементів, наприклад розробки визначених стандартів для метеостанцій, що дозволить максимально упростити процес розробки ПЗ для них, вибрати найоптимальніші алгоритми роботи [9]. Створення специфічних програмних бібліотек і фреймворків дозволить полегшити написання програм, оскільки базові алгоритми будуть вже готові, це зменшить шанс помилки і збільшить швидкість розробки шляхом перевикористання коду.

Кастально циклів кліматичних періодів, рис. 1.8, варто описати докладніше. Так як кругообіг циклів кліматичних періодів, як наприклад льодовиковий період, тривають велику кількість років, то варто також враховувати факти еволюції метеорологічних змінних. Під еволюцією у разі мається на увазі природне зміни очікуваних значень метеорологічних змінних протягом тривалого часу. Враховуючи це, можна зрозуміти, що доведеться регулярно адаптувати алгоритми обробки та аналізу під відповідні норми метеоданих, щоб уникнути випадкового визначення їх як аномальних.

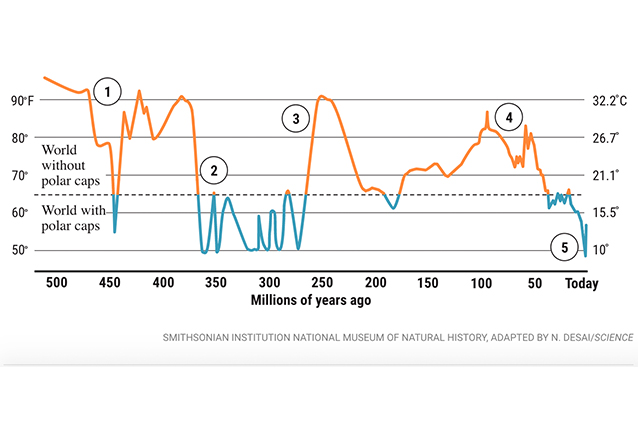


Рисунок 1.8 – Цикли леднікового періоду

Існують методики для визначення значущості вхідних параметрів на результат роботи мережі, для цього можна запитати коефіцієнти, що склалися у нейронів при навчанні і на їх підставі можна зробити висновок. Необхідно це для того, щоб дати оцінку тому, наскільки ефективно складено набір вхідних даних. За допомогою цього можна або виключити параметри, які мінімально впливають на результат, що вкрай важливо, так як це дає можливість прискорити алгоритм і зменшити обсяг даних, що зберігаються. Також завдяки цій можливості можна збільшити точність роботи свого алгоритму, для цього необхідно виділити всі теоретично можливі метеорологічні змінні, які можуть брати участь у процесі, який ви аналізуєте, після чого пробувати підставляти їх комбінації у вигляді вхідних параметрів.

Дана методика лізу даних за допомогою нейронних мереж, з подальшим аналізом ваг нейронів у них дозволяє виявити навіть найбільш неочевидні залежності. Наприклад, у 2018 році нейромережа виявила залежність між парою людських генів і схильністю людини до раку. Це відкриває можливість до нових досліджень і може в майбутньому зробити значний внесок у лікування раку. Подібні взаємозв'язки можуть бути виявлені і у сфері метеорології, наприклад взаємозв'язок виду поверхні на те, як поводиться тепература повітря.

Іноді дана методика може застосовуватися для аналізу впливу будь-яких сучасних розробок на загальну метеорологічну ситуацію. У серпні 2022 року був розроблений матеріал, що відображає 97.9% світла, що дозволяє знизити нагрівання покритих нею об'єктів. У кінцевому підсумку це світло розсіюється, отже впливає навколишні об'єкти, отже нагріває їх, в такий спосіб загальна картина змінюється і може призвести до несподіваних наслідків.

* 1. Аналіз можливостей сучасних засобів розрабки програмного забезпечення з використанням алгоритмів машинного навчання

Сучасні засоби розробки систем з використанням алгоритмів машинного навчання рясніють різними функціями на вирішення будь-яких заадч. До таких систем належать TensofFlow, PyTourch, sklearn.

TensorFlow – це бібліотека для машинного навчання, групи технологій, яка дозволяє навчати штучний інтелект вирішенню різних завдань. Бібліотека спочатку розроблена для Python і найчастіше використовується з ним.

Існують продаж TensorFlow для інших мов: C#, C++, Go, Java, Swift і так далі. Вони використовуються рідше за основну — головним чином для написання коду під специфічні платформи. Сама бібліотека написана мовою Python із використанням швидкого та продуктивного C++ для вирішення математичних завдань. Тому вона ефективно працює зі складними обчисленнями. Бібліотека розроблена Google як продовження внутрішньої бібліотеки компанії. TensorFlow безкоштовна, вона має відкритий вихідний код, який можна переглянути на GitHub, її активно підтримує спільнота ентузіастів. Назва читається як «тензор флоу» та утворена від двох понять: тензор та потік даних.

Для чого використовується TensorFlow Сама бібліотека включає безліч інструментів для різних напрямків ML, але найчастіше використовується для роботи з нейронними мережами. Це структури, натхненні пристроєм мереж нейронів у людській нервовій системі. Нейронні мережі складаються із програмних елементів-«нейронів» та зв'язків між ними, і такий пристрій дозволяє їм навчатися. TensorFlow працює зі звичайними та глибокими нейронними мережами різних типів: рекурентними, згортковими тощо. Також вона використовується для машинного та глибокого навчання.

Приклади використання технологій — розпізнавання природної мови, зображень та рукописних текстів, різноманітні завдання класифікації чи кластеризації, обробка великих даних, як-от метеоданих.

Можна виділити такі особливості TensorFlow:

* У TensorFlow моделі представлені за допомогою графів математичних абстракцій, які складаються з вершин і шляхів між ними. Граф можна порівняти із схемою доріг між різними точками. У програмуванні це зазвичай потрібно при вирішенні «маршрутних» завдань та при створенні нейронних мереж.
* TensorFlow працює з тензорами - багатовимірними структурами даних у векторному, тобто спрямованому просторі. Вони використовуються в лінійній алгебрі та фізиці. Звідси походить назва бібліотеки. За допомогою тензорів описуються шляхи графа, а вершини це математичні операції.
* Обчислення TensorFlow виражаються як потоки даних через граф. Це означає, що інформація «рухається» по графу, передається шляхами від вершини до вершини.
* Бібліотека може працювати на потужностях звичайного центрального процесора (CPU) або використовувати потужності графічного процесора (GPU). Режим перемикається у коді. Існує спеціальний тензорний процесор TPU, створений розробниками бібліотеки, - ним можна скористатися через хмарні сервіси Google.

Перевагами TensorFlow можна виділити високий рівень абстракції. Бібліотека написана так, що не потрібно думати про технічну реалізацію абстрактних понять. Можна зосередитись на описі логіки програми та на математиці, а спосіб реалізації обчислень – завдання TensorFlow, а не програміста. Це полегшує розробку та дозволяє сконцентруватися на важливих завданнях.

Інтерактивна розробка TensorFlow дозволяє працювати з компонентами моделі окремо і створювати її на ходу, при цьому окремо перевіряти кожен елемент. Це зручніше, ніж описувати граф як єдину монолітну структуру. Підхід робить розробку більш інтерактивною — структуру можна гнучко налаштовувати та змінювати.

Гнучкість TensorFlow можна використовувати для створення нейронних мереж, для глибокого навчання та інших напрямів ML. Його функції різноманітні на вирішення широкого спектра завдань. Завдяки графам та тензорам у TensorFlow можна легко зобразити складну математичну структуру. Гнучкість стосується як функцій, а й технічної боку питання: бібліотека працює і з центральним, і з графічним процесором, її легко використовувати з іншими інструментами для машинного навчання. Наприклад, TensorFlow застосовують із API Keras.

Кросплатформенність TensorFlow, як і сам Python, працює в популярних операційних системах, локально або в хмарі. Вона має розширення для мобільних пристроїв, IoT та браузерних додатків. Для мобільних пристроїв та інтернету речей можна скористатися середовищем TensorFlow Lite. А якщо модель машинного навчання має працювати у браузері, підійде TensorFlow.js - версія для використання з JavaScript та Node.js.

Велика спільнота TensorFlow — популярна бібліотека, тому на багато питань легко можна знайти відповідь у спільноті. Воно розвиває технологію, створює нові продукти та доповнення, пов'язані з TensorFlow, пише документацію та туторіали. Все це полегшує старт і дозволяє отримати з бібліотеки максимум користі.

Власні стандарти TensorFlow – продукт Google. Компанія відома своїми стандартами для технологій. Перша версія бібліотеки була призначена для внутрішнього використання. Досі в TensorFlow зустрічається неочевидна поведінка, через яку код може бути складніше налагоджувати. Складнощі можна компенсувати, якщо уважно вивчати документацію та користуватися додатковими інструментами для налагодження.

Високе споживання пам'яті, якщо TensorFlow використовується з графічним процесором, вона забирає всю його пам'ять. Це знижує продуктивність. Наприклад, якщо моделей кілька і вони написані на різних фреймворках, TensorFlow забере відеопам'ять у інших – виникне помилка. Щоб цього не було, споживання пам'яті бібліотекою треба обмежувати вручну. Складність у вивченні машинного навчання загалом. З TensorFlow складності можуть виникнути через її специфічні стандарти — це не найприязніша до новачків бібліотека.

TensorFlow встановлюється за допомогою pip, пакетного менеджера Python. Версії для роботи з CPU та GPU завантажуються окремо. Також для коректної роботи потрібна Anaconda – спеціальний дистрибутив Python для машинного навчання.

Більш детально про внутрішній пристрій TensorFlow можна сказати що TensorFlow — це наскрізна платформа для машинного навчання. Він підтримує наступне:

• Числове обчислення на основі багатовимірного масиву (подібно до NumPy.)

• GPU і розподілена обробка

• Автоматична диференціація

• Побудова моделі, навчання та експорт

Тензори, TensorFlow працює з багатовимірними масивами або тензорами, представленими як об’єкти tf.Tensor. Найважливішими атрибутами tf.Tensor є його форма і dtype:

• Tensor.shape: повідомляє розмір тензора вздовж кожної з його осей.

• Tensor.dtype: повідомляє вам тип усіх елементів у тензорі.

TensorFlow реалізує стандартні математичні операції над тензорами, а також багато спеціалізованих операцій для машинного навчання.

Змінні, звичайні об’єкти tf.Tensor є незмінними. Щоб зберегти вагові коефіцієнти моделі (або інший змінний стан) у TensorFlow, використовуйте tf.Variable.

Автоматична диференціація

Градієнтний спуск і пов’язані з ним алгоритми є наріжним каменем сучасного машинного навчання. Щоб увімкнути це, TensorFlow реалізує автоматичне диференціювання (autodiff), яке використовує обчислення для обчислення градієнтів. Зазвичай ви використовуєте це для обчислення градієнта помилки або втрати моделі відносно її ваг.

Хоча можно використовувати TensorFlow в інтерактивному режимі, як і будь-яку іншу бібліотеку Python, TensorFlow також надає інструменти для:

* Оптимізація продуктивності: для прискорення навчання та висновків.
* Експорт: щоб ви могли зберегти свою модель після завершення навчання.

Для цього потрібно використовувати tf.function, щоб відокремити чистий код TensorFlow від Python. Під час першого запуску функці tf.function, хоча вона виконується на Python, вона фіксує повний оптимізований графік, що представляє обчислення TensorFlow, виконані в межах функції. Під час наступних викликів TensorFlow виконує лише оптимізований графік, пропускаючи будь-які кроки, не пов’язані з TensorFlow. Графік може бути непридатним для повторного використання для вхідних даних із іншою підписом (форма і dtype), тому натомість створюється новий графік. Ці отримані графіки дають дві переваги:

• У багатьох випадках вони забезпечують значне прискорення виконання (хоча це не тривіальний приклад).

• Можна експортувати ці графіки за допомогою tf.saved\_model для запуску в інших системах, як-от сервер або мобільний пристрій, встановлення Python не потрібне.

Модулі, шари та моделі. tf.Module – це клас для керування вашими об’єктами tf.Variable та об’єктами tf.function, які з ними працюють. Клас tf.Module необхідний для підтримки двох важливих функцій:

Можна зберігати та відновлювати значення своїх змінних за допомогою tf.train.Checkpoint. Це корисно під час навчання, оскільки можна швидко зберегти та відновити стан моделі.

Можна імпортувати та експортувати значення tf.Variable і графіки tf.function за допомогою tf.saved\_model. Це дозволяє запускати вашу модель незалежно від програми Python, яка її створила.

Отримана SavedModel не залежить від коду, який її створив. Ви можете завантажити SavedModel з Python, інших мовних прив’язок або TensorFlow Serving. Ви також можете конвертувати його для запуску з TensorFlow Lit  або TensorFlow JS. Класи tf.keras.layers.Layer та tf.keras.Model побудовані на tf.Module забезпечуючи додаткову функціональність і зручні методи для створення, навчання та збереження моделей. Деякі з них демонструються в наступному розділі.

Однак альтернативою TensorFlow можна представити PyTourch. PyTorch визначається як бібліотека машинного навчання з відкритим кодом для Python. Він використовується для програм, таких як обробка природної мови. Спочатку він був розроблений дослідницькою групою зі штучного інтелекту Facebook та програмним забезпеченням Uber Pyro для ймовірнісного програмування, яке ґрунтується на ньому.

Спочатку PyTorch був розроблений Х'ю Перкінсом як оболонка Python для LusJIT, заснована на платформі Torch. Є два варіанти PyTorch.

PyTorch перепроектує та впроваджує Torch у Python, спільно використовуючи ті ж основні бібліотеки C для внутрішнього коду. Розробники PyTorch налаштували цей внутрішній код для ефективної роботи з Python. Вони також зберегли апаратне прискорення на основі графічного процесора, а також функції розширення, які зробили Torch на базі Lua.

Простий інтерфейс – PyTorch пропонує простий використання API; отже, він вважається дуже простим у роботі та працює на Python. Виконання коду у цьому середовищі досить просто.

Використання Python – це бібліотека вважається Pythonic, яка плавно інтегрується зі стеком даних Python. Таким чином, він може використовувати всі сервіси та функціональні можливості, які пропонують середовищем Python.

Обчислювальні графи PyTorch надає відмінну платформу, яка пропонує динамічні обчислювальні графи. Таким чином користувач може змінити їх під час виконання. Це дуже корисно, коли розробник не знає скільки пам'яті потрібно для створення моделі нейронної мережі.

PyTorch відомий тим, що має три рівні абстракції, як зазначено нижче –

* Тензор - імперативний n-вимірний масив, що працює на GPU.
* Змінна – вузол у обчислювальному графі. Це зберігає дані та градієнт.
* Модуль – Рівень нейронної мережі, в якому зберігатимуться дані про стан або ваги, що вивчаються.

Переваги PyTorch –

* Код легко налагоджувати та розуміти.
* Він включає багато шарів, як Torch.
* Включає безліч функцій втрати.
* Це можна як розширення NumPy для графічних процесорів.
* Це дозволяє будувати мережі, структура яких залежить самих обчислень.

Нижче наведено порівняння між TensorFlow та PyTorch

|  |  |
| --- | --- |
| PyTorch | TensorFlow |
| PyTorch тісно пов'язаний із заснованим на lua фреймворком Torch, який активно використовується у Facebook. | TensorFlow розроблено Google Brain і активно використовується в Google. |
| PyTorch є відносно новим у порівнянні з іншими конкурентними технологіями. | TensorFlow не є новим і розглядається багатьма дослідниками та професіоналами як інструмент, необхідний для роботи. |
| PyTorch включає все в імперативній і динамічній манері. | TensorFlow включає статичні та динамічні графіки у вигляді комбінації. |
| Граф обчислень у PyTorch визначається під час виконання. | TensorFlow не включає опцію часу виконання. |
| PyTorch включає можливість розгортання для мобільних і вбудованих платформ. | TensorFlow краще працює для вбудованих фреймворків. |

Чим швидше виходить обчислювати свою функцію і чим гнучкіші можливості для її визначення, тим краще. Тепер, коли кожен фреймворк вміє використовувати всю потужність відеокарт, перший критерій перестав відігравати значну роль.

Тензорні обчислення - основа PyTorch, каркас, навколо якого збільшується вся інша функціональність. На жаль, не можна сказати, що міць і виразність бібліотеки в даному аспекті збігаються з такою у NumPy. Щодо роботи з тензорами, PyTorch керується принципом максимальної простоти та прозорості, надаючи тонку обгортку над викликами BLAS. Цей фреймворк є низькорівневим у сенсі реалізації інтерфейсу його використання. Однак його більш ніж достатньо для вирішення майже будь-яких задач машинного навчання.

Ще однією цікавою бібліотекою для розв'язання задач машинного навчання є sklearn. Це бібліотека яка допомагає створювати прості моделі для машинного навчання, вони застосовуються для обчислення моделей, що легко збираються, для алгоритмів де потрібна швидкості і мала витрата ресурсів пам'яті.

Які завдання вирішує бібліотека, наприклад, препроцессинг. термін "препроцессинг" (preprocessing) означає попередню обробку даних. Вони приводяться до виду, необхідного для подачі на вхід будь-якого алгоритму. Наприклад, зображення підганяються під один розмір та колірну схему. З даних вилучаються ключові ознаки, якими буде вчитися модель, і також призводять до потрібного формату.

Зменшення розмірності. Часто у даних міститься надмірна інформація. Наприклад, деякі ознаки можна отримати з інших. Щоб зробити подальший аналіз ефективнішим, розмірність вибірки зменшується — так, щоб зберегти максимум корисних даних. І тому використовуються спеціальні методи, наприклад метод головних компонент.

Виявлення аномалій. Алгоритми «відсікають» з набору даних помилкові чи дивні записи, які лише додають зайві похибки. Це потрібно, щоб аналіз та навчання працювали точніше.

Вибір датасету. Якщо потрібно познайомитися з бібліотекою, можна скористатися одним із готових навчальних наборів. Вони теж є у бібліотеці.

Вибір моделі. Функції та алгоритми допомагають оцінити ефективність різних моделей для вирішення задачі. Їх можна порівнювати один з одним, проводити валідацію результатів, вибирати точніші. Це потрібно для покращення якості навчання.

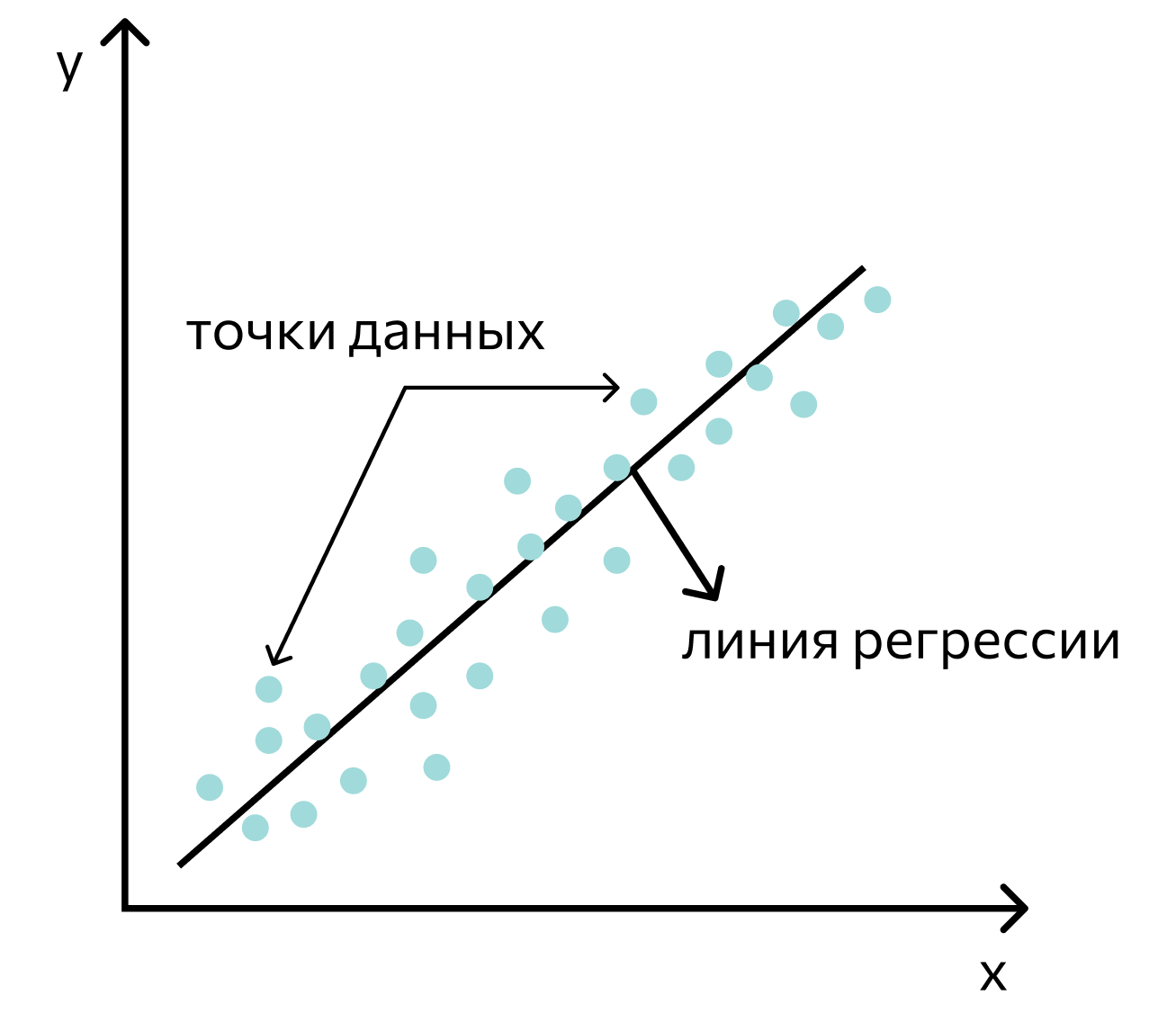
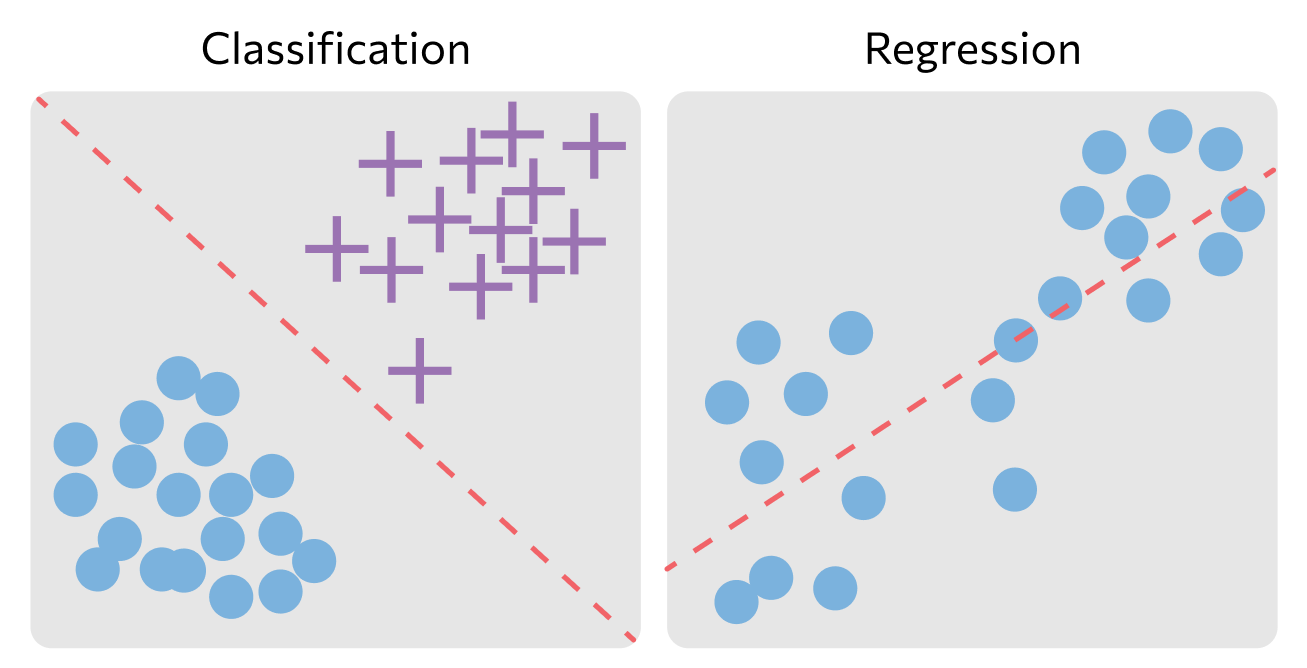


Рисунок 1.8 — Приклад регресії

Регресія. Це прогнозування показників за наявними даними, які можуть приймати нескінченну кількість різних значень. Ці показники би мало бути пов'язані з будь-яким об'єктом, тобто. бути його атрибутами. Наприклад, прогноз кількості користувачів на сайті у різні дні – це завдання регресії.

Класифікація - прогнозування показника з кінцевою кількістю значень. Простіше визначення — прогнозування категорії, до якої належить об'єкт. Розпізнавання жанру тексту чи об'єктів на зображенні – це завдання класифікації.

Рисунок 1.9 — Приклад регресії та класифікації

Кластеризація. Це розподіл даних з датасету за великими групами — кластерами, тобто їх угруповання. Схожі об'єкти об'єднуються у класи. Критерії, якими визначається схожість, залежить від моделі та умов завдання.

* 1. Мета та завдання роботи

Мета і задачі дослідження. Метою роботи є дослідження ефективності роботи різних варіантів нейронних мереж на задачах обробки та аналізу метеоданих. Для виконання поставленої задачі, необхідно виконати:

* виконати аналіз поточних методів вирішення метеорологічних завдань;
* виконати аналіз впливу впровадження нейронних мереж у предметну область;
* виконати порівняльний аналіз схожих програмних систем;
* визначити програмні засоби реалізації інформаційної системи;
* виконати проектування інформаційної системи;
* виконати розробку інформаційної системи та моделей нейромереж;
* дослідження ефективності роботи різних моделей нейронних мереж;
* дослідження ступеня впливу різних вхідних параметрів на результат виконання завдання.

Розроблюване програмне забезпечення повинно підтримувати можливості запуску в будь-якій операційній системі, яка має веб-барузер.

Мінімальні апаратні вимоги до сторони користувача:

* Процесор Celeron 2.9 GHz.
* Обсяг оперативної пам'яті 2 Gb..
* Доступ до інтернету

Програмно системі потрібен тільни будь-якій веб-брауезер.

Мінімальні апаратні вимоги до сторони серверу:

* Процесор і5, 4 ядра по 3.2 GHz.
* Обсяг оперативної пам'яті 16 Gb..
* Відеокарта Nvidia з 4 гб памяті, моделі не старіше 1050
* Доступ до інтернету

Програмно системі потрібен Python 3.8, Django3, SQLite3, TensorFlow2.

* 1. Висновки до першого розділу

Результатом проведеного аналізу предметної області, було визначено що сучасна метеорологія є добрим напрямком для впровадження сучасних інформаційних технологій [5]. Це пов’язано з актуальним використанням засобів комунікації та інформаційного обміну між метеостанціями та метеоцентрами [7].

Проведений аналіз інформаційних матеріалів показав, що сфера потребуе нові алгоритми аналізу метеоданих, які будут більш гнучкими та будуть аналізувати більше різноманітниї метеорологічниї змінних. Результати аналізу існуючих алгоритмів прогнозу метеоданих з ціллю найти серед них більшт оптимальний та точний. Також потрібно провести достідження з ціллю знайти найкращу комбінацію метеорологічних смінних для розрахунку прогнозу погоди.

Розглянувши найпопулярніші засоби розробки, що використовуються в галузі машинного навчання, можна дійти висновку, що найбільш зручною для вирішення задач роботи є мова Python та фреймворк для створення нейромереж TensorFlow 2, aдже саме в ній є всі необхідні функції.

Отримані в рамках розділу результати будуть використані під час обґрунтування вибірку напрямку дослідження та формалізації завдання з розробки програмного забезпечення у подальших розділах роботи.

# ОБГРУНТУВАННЯ ВИБОРУ НАПРЯМКУ ДОСЛІДЖЕННЯ

2.1 Постановка завдання регресії

З метою визначення усіх функціональних можливостей інформаційних систем для метеорологів необхідно проаналізувати існуючі рішення та сформувати головні вимоги до таких систем. Для об’єктивного порівняння, необхідно використовувати тільки web-рішення, які автоматизують робочі задачі метеорологів, або спрощують роботу метеорологічного центру.

Відсутність системи збору метеоданих це критичний недолік для майбутнього метеорології, який не дасть їй можливості розвиватися.

Основним завданням системи є аналіз метеорологічних даних з різних джерел і побудова короткострокового прогнозу погоди з прогнозуванням можливих надзвичайних погодних явищ. Гроза і сильна злива, град і шквальний вітер, сніговий буран або екстремальна температура часто формуються протягом короткого часу і часто мають відносно невеликі розміри, що не дозволяє звичайним інструментам створення прогнозу погоди.

Система короткострокового прогнозу погоди і попередження про надзвичайні погодні умови аналізує метеорологічні дані в режимі реального часу в рамках заданої місцевості обмеженою відносно невеликою територією і будує прогноз виникнення надзвичайних погодних умов в рамках декількох годин, що дозволяє надати більш точний прогноз.

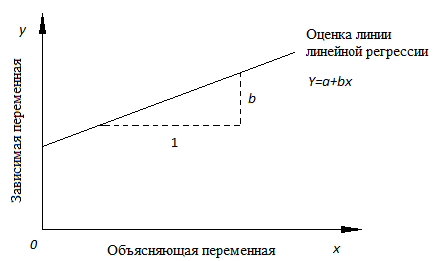
Починається весь процес роботи системи зі збору даних з якими далі доведеться працювати, після чого вони обробляються та зберігаються у системі. Після цього в дію вступає модуль аналізу цих даних, він складається з алгоритму розрахунку регресії та алгоритму класифікації. На даний момент розглянемо алгоритм розрахунку регресії, адже саме це завдання вирішує система під час побудови короткострокового прогнозу погоди.

Розрахунок алгоритму регресії починається з побудови математичного рівняння, яке оцінює лінію простої (парної) лінійної регресії:

Y=a+bx.

x називаєтся незалежною змінною чи предиктором. Y – залежна змінна або змінна відгук. Це значення, що ми очікуємо для y (у середньому), якщо знаємо величину x, тобто. це «передбачене значення y»

* a – вільний член (перетин) лінії оцінки; це значення Y коли x=0, рис 2.1.
* b – кутовий коефіцієнт чи градієнт оціненої лінії; вона є величину, яку Y збільшується загалом, якщо ми збільшуємо x однією одиницю.
* a та b називають коефіцієнтами регресії оціненої лінії, хоча цей термін часто використовують тільки для b.

Рисунок 2.1 - Лінія лінійної регресії, що показує перетин a та кутовий коефіцієнт b (величину зростання Y при збільшенні x на одну одиницю)

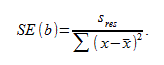
Парну лінійну регресію можна розширити, включивши до неї більше однієї незалежної змінної; у цьому випадку вона відома як множинна регресія.

Аномальні значення (викиди) та точки впливу. "Впливове" спостереження, якщо воно опущене, змінює одну або більше оцінок параметрів моделі (тобто кутовий коефіцієнт або вільний член). Викид (спостереження, що суперечить більшості значень у наборі даних) може бути "впливовим" спостереженням і може добре виявлятися візуально, під час огляду двовимірної діаграми розсіювання або графіка залишків. І для викидів, і для "впливових" спостережень (крапок) використовують моделі як з їх включенням, так і без них звертають увагу на зміну оцінки (коефіцієнтів регресії). При проведенні аналізу не варто відкидати викиди або точки впливу автоматично, оскільки звичайне ігнорування може вплинути на отримані результати. Завжди вивчайте причини появи цих викидів та аналізуйте їх.

Гіпотеза лінійної регресії при побудові лінійної регресії перевіряється нульова гіпотеза про те, що генеральний кутовий коефіцієнт лінії регресії β дорівнює нулю. Якщо кутовий коефіцієнт лінії дорівнює нулю, між  і  немає лінійного співвідношення: зміна  не впливає на . Для тестування нульової гіпотези про те, що дійсний кутовий коефіцієнт  дорівнює нулю можна скористатися наступним алгоритмом:

Обчислити статистику критерію, що дорівнює відношенню , яка підпорядковується  розподілу з  ступенями свободи, де  стандартна помилка коефіцієнта 



,

 - оцінка дисперсії залишків.

Зазвичай, якщо досягнутий рівень значущості  нульова гіпотеза відхиляється. Можна розрахувати 95% довірчий інтервал для генерального кутового коефіцієнта :



де  відсоткова точка  розподілу зі ступенями свободи  що дає можливість двостороннього критерію . Це той інтервал, який містить генеральний кутовий коефіцієнт з ймовірністю 95%.

Для великих вибірок, скажімо,  ми можемо апроксимувати  значенням 1,96 (тобто статистика критерію прагнутиме до нормального розподілу). Оцінка якості лінійної регресії: коефіцієнт детермінації R2. Через лінійне співвідношення  і  очікуємо, що  змінюється, у міру того як змінюється , і називаємо це варіацією, яка обумовлена чи пояснюється регресією. Залишкова варіація має бути якнайменше. Якщо це так, то більшість варіації  пояснюватиметься регресією, а точки лежатимуть близько до лінії регресії, тобто. лінія добре відповідає даним.

 Частка загальної дисперсії , яка пояснюється регресією називають коефіцієнтом детермінації, зазвичай виражають через відсоткове співвідношення та позначають R2 (у парній лінійній регресії це величина r2, квадрат коефіцієнта кореляції), що дозволяє суб'єктивно оцінити якість рівняння регресії.

Різниця  є відсотком дисперсії який не можна пояснити регресією. Немає формального тесту для оцінки  ми змушені покластися на суб'єктивне судження, щоб визначити якість припасування лінії регресії.

2.2 [Порівняльний аналіз та вибір алгоритму](#_Toc62808901) регресії

У разі під алгоритмом регресії мається на увазі збірний образ всіх алгоритмів прогнозування будь-яких даних. Далі розглянемо різні варіанти архітектур нейронних мереж, які застосовуються в алгоритмах прогнозування.

Першою розглянутою архітектурою буде багатошаровий перцептрон, рис 2.2.

Рисунок 2.2 - багатошаровий перцептрон

Багатошаровий перцептрон складається з трьох або більше шарів. Кожен вузол у шарі з'єднаний з кожним вузлом у наступному шарі, що робить мережу повністю пов'язаною. Приховані шари містять вузли мережі (одиниці), що не піддаються спостереженню. Кожна прихована одиниця є функцією виваженої суми вхідних даних. Ця функція є функцією активації, при цьому значення ваги визначаються алгоритмом оцінки. Мережа містить другий прихований рівень, кожна прихована одиниця у якому є функцією від виваженої суми одиниць у першому прихованому рівні. В обох рівнях використовують однакову функцію активації. Кількість прихованих шарів. Багатошаровий перцептрон може мати один або два приховані рівні.

функція активації. Функція активації "зв'язує" шар зважених сум об'єктів із наступним шаром значень даних об'єктів.

* Гіперболічний тангенс. Це така функція: γ(c) = tanh(c) = (e c−e −c)/(e c+e −c). Вона визначена для дійсних змінних та переводить їх у діапазон (–1, 1). При використанні автоматичного вибору архітектури це функція для всіх нейронів у прихованих шарах.
* Сігмоїд. Це така функція: γ(c) = 1/(1+e−c). Вона визначена для дійсних змінних та переводить їх у діапазон (0, 1).

Кількість нейронів. Кількість одиниць у кожному прихованому шарі може бути задано явно або автоматично визначено алгоритмом оцінки. Вихідний шар містить цільові (залежні) змінні. функція активації. Функція активації "зв'язує" шар зважених сум об'єктів із наступним шаром значень даних об'єктів.

* Тотожність. Це функція γ(c) = c. Вона визначена всім аргументів і повертає їх незміненими. При використанні автоматичного вибору архітектури це функція активації для нейронів у прихованих шарах, якщо не існує кількісних залежних змінних.
* Софтмакс. Це така функція: γ(c k) = exp(c k)/Σjexp(c j). Вона приймає вектор дійсних аргументів і перетворює їх на вектор, компоненти якого укладені в діапазоні (0, 1), а їх сума дорівнює 1. Функція софтмакс доступна тільки в тому випадку, якщо всі залежні категоріальні змінні. При використанні автоматичного вибору архітектури це функція активації для нейронів у вихідному шарі, якщо всі залежні категоріальні змінні.
* Гіперболічний тангенс. Це така функція: γ(c) = tanh(c) = (e c−e −c)/(e c+e −c). Вона визначена для дійсних змінних та переводить їх у діапазон (–1, 1).
* Сігмоїд. Це така функція: γ(c) = 1/(1+e−c). Вона визначена для дійсних змінних та переводить їх у діапазон (0, 1).

Така архітектура частіше знаходить застосування у завданнях розпізнавання мови та машинному перекладі.

Рекурсивні нейронні мережі - вид нейронних мереж, що працюють з даними змінної довжини. Моделі рекурсивних мереж використовують ієрархічні структури зразків під час навчання. Наприклад, зображення, що складаються зі сцен, що поєднують підсцени, що включають багато об'єктів. Виявлення структури сцени та її деконструкція-нетривіальне завдання. При цьому необхідно ідентифікувати окремі об'єкти, так і всю структуру сцени.

У рекурсивних мережах нейрони з однаковими вагами активуються рекурсивно відповідно до структури мережі. У процесі роботи рекурсивної мережі виробляється модель для передбачення як структур змінної розмірності, так і скалярних структур через активацію структури відповідно до топології. Мережі RvNNs успішно застосовуються при навчанні послідовних структур та дерев у завданнях обробки природної мови, при цьому фрази та речення моделюються через векторне уявлення слів. RvNNs спочатку з'явилися для розподіленого уявлення структур, використовуючи предикати математичної логіки. Розробки рекурсивних мереж та перші моделі почалися в середині 1990-х.

У найпростішій архітектурі вузли мережі сходяться до батьків через матрицю ваги прихованого шару, що використовується багаторазово через всю мережу, і нелінійну функцію активації типу гіперболічного тангенсу.



Рисунок 2.3 - Архітектура простої рекурсивної мережі, тут W - навчена матриця ваг

Ця архітектура з деяким удосконаленням використовується для послідовного дешифрування натуральних сцен зображення або для структурування речень природної мови, рис 2.3.

Рекурсивна каскадна кореляція RecCC - це підхід до конструювання рекурсивних мереж, що оперують із трьома доменами, перші додатки такого роду з'явилися в хімії, а розширення утворює спрямований ациклічний граф. У 2004 році було запропоновано систему навчання рекурсивної мережі без вчителя. Тензорні рекурсивні мережі використовують одну функцію тензора для всіх вузлів дерева.

Рекурентна нейронна мережа, на відміну прямої нейронної мережі, є варіантом рекурсивної ІНС, у якій зв'язки між нейронами — спрямовані цикли. Останнє означає, що вихідна інформація залежить не тільки від поточного входу, але також станів нейрона на попередньому кроці. Така пам'ять дозволяє користувачам вирішувати завдання NLP: розпізнавання рукописного тексту чи мови. У статті Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks автори показують модель рекурентної мережі, яка генерує нові пропозиції та короткий зміст текстового документа. Siwei Lai, Liheng Xu, Kang Liu, та Jun Zhao у своїй роботі Recurrent Convolutional Neural Networks for Text Classification створили рекурентну згорткову нейромережу для класифікації тексту без рукотворних ознак. Модель порівнюється з існуючими методами класифікації тексту - Bag of Words, Bigrams + LR, SVM, LDA, Tree Kernels, рекурсивними та згортковими мережами. Описана модель перевершує за якістю традиційні методи для всіх використовуваних датасетів.

Існує багато різновидів, рішень та конструктивних елементів рекурентних нейронних мереж. Проблема рекурентної мережі полягає в тому, що якщо враховувати кожен крок часу, то стає необхідним для кожного кроку часу створювати свій шар нейронів, що викликає серйозні обчислювальні складності. Крім того, багатошарові реалізації виявляються обчислювально нестійкими, тому що в них зазвичай зникають або зашкалюють ваги. Якщо обмежити розрахунок фіксованим тимчасовим вікном, то отримані моделі не відображатимуть довгострокових трендів. Різні підходи намагаються вдосконалити модель історичної пам'яті та механізм запам'ятовування та забування.

Цілком рекурентна мережа, ця базова архітектура розроблена у 1980-х. Мережа будується з вузлів, кожен із яких з'єднаний з усіма іншими вузлами. У кожного нейрона поріг активації змінюється з часом і є речовим числом. Кожна сполука має змінну речовинну вагу.

Вузли поділяються на вхідні, вихідні та приховані для навчання з учителем з дискретним часом, кожен (дискретний) крок часу на вхідні вузли подаються дані, а інші вузли завершують свою активацію, і вихідні сигнали готуються для передачі нейроном наступного рівня. Якщо мережа відповідає за розпізнавання мови, в результаті на вихідні вузли надходять вже мітки (розпізнані слова).

У навчанні з підкріпленням (reinforcement learning) немає вчителя, що забезпечує цільові сигнали для мережі, натомість іноді використовується функція пристосованості (придатності) або функція оцінки (reward function), за якою проводиться оцінка якості роботи мережі, при цьому значення на виході впливає на поведінка мережі на вході Зокрема, якщо мережа реалізує гру, на виході вимірюється кількість пунктів виграшу чи оцінки позиції. Кожен ланцюжок обчислює помилку як сумарну девіацію за вихідними сигналами мережі. Якщо є набір зразків навчання, помилка обчислюється з урахуванням помилок кожного зразка.

Мережа довгої короткострокової пам'яті (Long Short-Term Memory, LSTM) - різновид архітектури рекурентної нейромережі, створена для більш точного моделювання часових послідовностей та їх довгострокових залежностей, ніж традиційна рекурентна мережа.

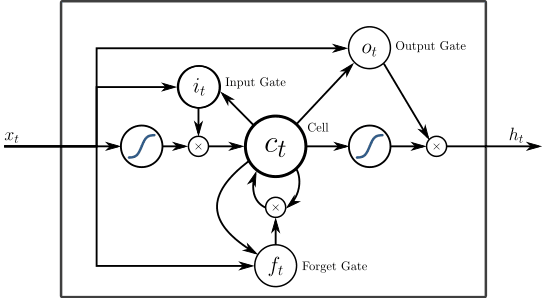


Рисунок 2.4 - структура LSTM сети

LSTM-мережа не використовує функцію активації в рекурентних компонентах, збережені значення не модифікуються, а градієнт не прагне зникнути під час тренування. Часто LSTM застосовується у блоках по кілька елементів. Ці блоки складаються з 3 або 4 затворів (наприклад, вхідного, вихідного та гейту забування), які контролюють побудову інформаційного потоку по логістичній функції.

У Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling автори показують архітектуру глибокої LSTM рекурентної мережі, яка досягає хороших результатів для великомасштабного акустичного моделювання.

У роботі Part-of-Speech Tagging with Bidirectional Long Short-Term Memory Recurrent Neural Network представлена ​​модель автоматичної морфологічної розмітки. Модель показує точність 97.4 % у розмітці. Apple, Amazon, Google, Microsoft та інші компанії впровадили у продукти LSTM-мережі як фундаментальний елемент.

Модель послідовності, якщо розглядати на високому рівні, модель seq2seq має кодер, декодер і проміжний крок як основні компоненти:

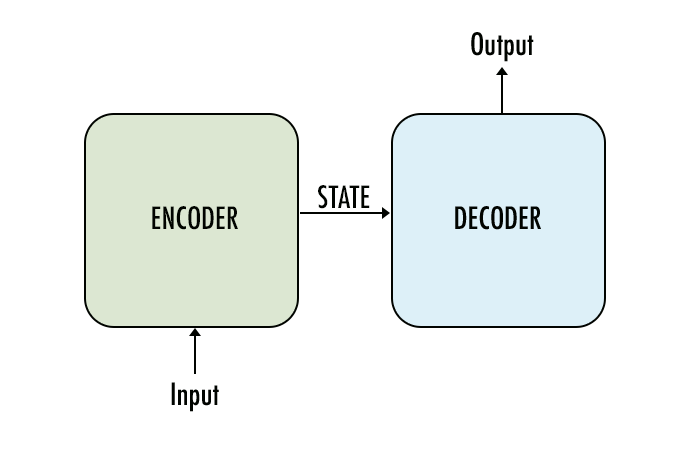


Рисунок 2.5 - схема архітектури автоенкодера

Використовуємо вбудовування, тому спочатку маємо скласти «словниковий» список, який містить усі слова, які ми хочемо, щоб наша модель могла використовувати або читати. Вхідні дані моделі повинні бути тензорами, що містять ідентифікатори слів у послідовності. Однак існує чотири символи, які ми повинні містити в своєму словнику.

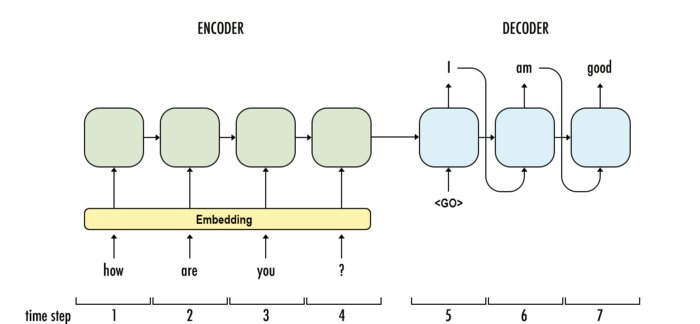


Рисунок 2.6 - логіка роботи автоенкодера

Словники Seq2seq зазвичай резервують перші чотири місця для цих елементів:

• <PAD>: під час навчання нам потрібно буде передавати наші приклади в мережу пакетами. Усі вхідні дані в цих пакетах мають бути однакової ширини, щоб мережа могла виконати обчислення. Однак наші приклади не мають однакової довжини. Ось чому нам потрібно буде доповнити коротші вхідні дані, щоб привести їх до однакової ширини пакета

• <EOS>: це ще одна необхідність пакетування, але більше з боку декодера. Це дозволяє нам повідомляти декодеру, де закінчується речення, а також дозволяє декодеру вказувати те саме у своїх виводах.

• <Nk>: Якщо ви тренуєте свою модель на реальних даних, ви виявите, що можете значно підвищити ефективність ресурсів своєї моделі, ігноруючи слова, які не часто виявляються у вашому словникові запаси, щоб гарантувати розгляд. Ми замінюємо їх на.

• <GO>: Це вхідні дані для першого тимчасового кроку декодера, щоб повідомити декодеру, коли почати генерувати вихідні дані.

Примітка, для представлення цих функцій можна використовувати інші теги. Наприклад, я бачив <s> і </s> замість <GO> і <EOS>. Тому переконайтеся, що все, що ви використовуєте, узгоджено через попередню обробку та навчання моделі/висновок.

Підготовка вхідних даних для навчального графіка трохи складніша з двох причин:

1. Ці моделі працюють набагато краще, якщо ми передаємо декодеру нашу цільову послідовність, незалежно від того, які її часові кроки фактично виводяться під час тренування. Отже, на відміну від графіка, ми не будемо передавати вихідні дані декодера самому собі на наступному часовому етапі.

2. Дозування, одна з оригінальних статей про послідовність дій, Sutskever et al. 2020, повідомляє про кращу продуктивність моделі, якщо вхідні дані змінені. Таким чином, ви також можете змінити порядок слів у послідовності введення.

Під час попередньої обробки ми робимо наступне: ми створюємо свій словниковий запас з унікальних слів (і підраховуємо випадки, коли ми це робимо), створити копію розмов із заміною слів на їхні ідентифікатори, ми можемо додати ідентифікатори слів <GO> і <EOS> до цільового набору даних зараз або зробити це під час навчання.

2.3 Концепція попередньої обробки та аналізу даних

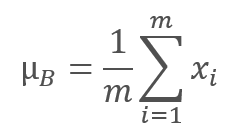
У практиці використання нейронних мереж використовують різні підходи до нормалізації даних. Але всі вони спрямовані на утримання даних навчальної вибірки та вихідних даних прихованих шарів нейронної мережі в заданому діапазоні та з певними статистичними характеристиками вибірки, такими як дисперсія та медіана. У нейронах мережі застосовуються лінійні перетворення, які у процесі навчання зміщують вибірку у бік антиградієнта.

Розглянемо повнозв'язковий перцептрон з двома прихованими шарами. При прямому проході кожен шар генерує деяку сукупність даних, які є навчальною вибіркою для наступного шару. Результат роботи вихідного шару порівнюється з еталонними даними і зворотному проході поширюється градієнт помилки від вихідного шару через приховані шари до вихідних даних. Отримавши кожному нейроні свій градієнт помилки оновлюємо вагові коефіцієнти, підлаштовуючи нейронну мережу під навчальні вибірки останнього прямого проходу. І тут виникає конфлікт: підлаштовуючи другий прихований шар (H2 на малюнку нижче) під вибірку даних на виході першого прихованого шару (на малюнку H1), в той час як змінивши параметри першого прихованого шару змінили масив даних. Тобто. підлаштовуючи другий прихований шар під вже неіснуючу вибірку даних. Аналогічна ситуація з вихідним шаром, який підлаштовується під вже змінений вихід другого прихованого шару. А якщо ще врахувати спотворення між першим і другим прихованими шарами, масштаби помилки збільшуються. І чим глибша нейронна мережа, тим сильніший прояв цього ефекту. Це було названо внутрішнім підступним зрушенням.

У класичних нейронних мережах зазначена проблема частково вирішувалася зменшенням коефіцієнта навчання. Невеликі зміни вагових коефіцієнтів не дуже змінюють розподіл вибірки на виході нейронного шару. Але такий підхід не вирішує масштабування проблеми зі зростанням кількості шарів нейронної мережі та знижує швидкість навчання. Ще одна проблема невеликого коефіцієнта навчання - застрявання в локальних мінімумах.

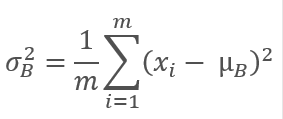
У лютому 2015 року Sergey Ioffe та Christian Szegedy запропонували метод пакетної нормалізації даних (Batch Normalization) для вирішення проблеми внутрішнього зсуву коваріації. Суть методу полягала у нормалізації кожного окремого нейрона на певному часовому інтервалі зі зміщенням медіани вибірки до нуля та приведенням дисперсії вибірки к1.

Алгоритм проведення нормалізації наступний. Спочатку за вибіркою даних вважається середнє значення.

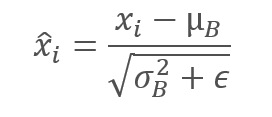


де m – розмір вибірки (batch).

Потім вважаємо дисперсію вихідної вибірки.



І нормалізуємо дані вибірки, привівши вибірку до нульового середнього та одиничної дисперсії.



Зверніть увагу, що в знаменнику до дисперсії вибірки додається константа ϵ, невелике позитивне число з метою виключити поділ на нуль.

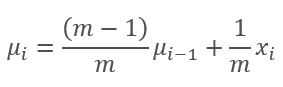
Така нормалізація може спотворити вплив вихідних даних. Тому автори методу додали ще один крок — масштабування та усунення. Було введено 2 змінні γ і β, які навчаються разом із нейронною мережею методом зворотного градієнтного спуску.



Застосування цього методу дозволяє кожному етапі навчання отримувати вибірку даних з однаковим розподілом, що практично робить навчання нейронної мережі більш стабільним і дозволяє збільшити коефіцієнт навчання. Загалом це дозволить підвищити якість навчання за менших витрат часу на навчання нейронної мережі.

Але в той же час зростають витрати на зберігання додаткових коефіцієнтів. А також для розрахунку ковзної середньої та дисперсії потрібно зберігання в пам'яті історичних даних кожного нейрона на весь розмір пакету. І тут можна подивитися у бік експоненційної середньої. На малюнку нижче наочно представлені графіки ковзної середньої та ковзної дисперсії на 100 елементів у порівнянні з експоненційної ковзної середньої та експоненційної ковзної дисперсії на ті ж 100 елементів. Графік був побудований для 1000 випадкових елементів діапазону від -1.0 до 1.0.

Як видно на графіку, ковзна середня та експоненційна ковзна середня зближуються після 120-130 ітерацій і далі відхилення мінімально, яким можна знехтувати. До того ж графік експоненційної ковзної середньої має більш згладжений вигляд. А ось для розрахунку EMA достатньо попереднього значення функції та поточного елемента послідовності. Нагадаю формулу експоненційної ковзної середньої.

 ,

де

* m - розмір вибірки (batch),
* i – ітерація.

Для зближення графіків ковзної дисперсії та експоненційної ковзної дисперсії потрібно трохи більше ітерацій (310-320), але в цілому картина схожа. Що стосується дисперсією застосування експоненціальної дає як економію пам'яті, а й значно знижує кількість обчислень, т.к. для ковзної дисперсії ми перераховували б відхилення від середньої для всього batch-а.

Експерименти, проведені авторами методу, показують, що застосування методу Batch Normalization виступає у ролі регуляризатора. Це дозволяє відмовитися від інших методів регуляризації, зокрема від розглянутого раніше Dropout. Більше того, є пізніші роботи, в яких показано, що спільне використання Dropout та Batch Normalization негативно позначається на результатах навчання нейронної мережі.

У сучасних архітектурах нейронних мереж запропонований алгоритм нормалізації можна зустріти у різних варіаціях. Автори пропонують використовувати Batch Normalization безпосередньо перед нелінійністю (формулою активації). Як одну із варіацій даного алгоритму можна розглядати метод Layer Normalization, представлений у липні 2016 року.

Головна думка цього розділу в тому, що процес нормалізації вкрай важливий, щоб навчання нейронної мережі проходило ефективно і різні похибки даних не впливали на її роботу.

# 2.4 Висновок до другого розділу

У другому розділі проведено постановку завдання прогнозування. Зроблено порівняльний аналіз та вибір алгоритму регресії. Розглянуто процес попередньої обробки і аналізу даних. Виконавши аналіз сучасних інформаційних систем, які використовуються в якості головних електронних інструментів в метеорології, було сформовано концепцію розробки програмного засобу.

Таке рішення дозволить розробити сучасну та функціональну систему, а також покращити показники ефективності алгоритмів обробки та аналізу.

Так як даних буде багато то програмний комплекс повинен бути оптимізований для роботи з великими даними, в даному випадку кращим рішенням буде використовувати систему управління базами даних SQLite.

Отримані в рамках виконання даного розділу результати будуть застосовані під час практичної реалізації програмного забезпечення.

# РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ЗБОРУ, ОБРОБКИ ТА АНАЛІЗУ МЕТЕОДАНИХ



## Проектування розробки інформаційної системи

Початковим етапом при розробці програмного забезпечення є проектування. Проектування дозволяє визначити головні функціональні можливості програмного забезпечення, його архітектуру, складові, а також компоненти з якими взаємодіє програмне забезпечення.

В більшості випадків проект складається з розробки головних діаграм, які в графічному вигляді дозволяють визначити усі складові.

В якості засобу проектування використовується мова UML, яка є спеціалізованою нотаціонною мовою, яку використовують розробники для графічного представлення програмного забезпечення. В якості середовищ проектування може бути використано різні редактори, які дозволяють взаємодіяти зі спеціалізованими об’єктами проектування.

Одним із засобів проектування, є створення діаграми варіантів використання, завдяки якій, визначаються усі вимоги до кінцевої системи.

Діаграма варіантів використання складається з об’єктів, які мають вигляд овалу, та запису функціональної ролі. Зовнішні об’єкти, якими можуть бути як користувачі так і сервера, зображаються у вигляді чоловічка.

Одним з вайжливіших елементів на діаграмі варіантів використання є використання стрілок, кожна з яких має своє спеціалізоване призначення. В більшості випадків, саме стрілки, дозволяють розробникам орієнтуватися в програмній системі, а також створювати систему зв’язків елементів між собою.

Діаграма варіантів використання для інформаціійнох системи для збору, обробки та аналізу метеоданих, зображено на рис. 3.1.

Програмне рішення має web-архітектуру, отже дозволяє користувачам отримувати усі функціональні можливості без інсталяції.

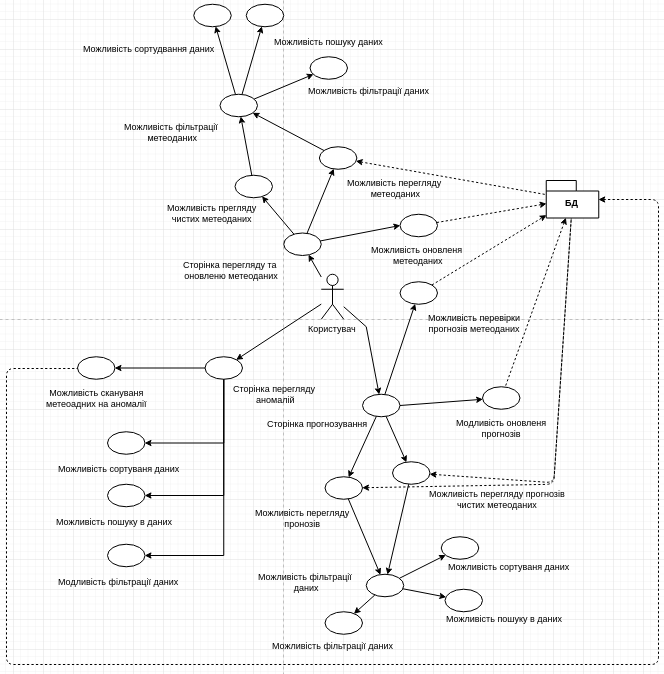


Рисунок 3.1 – Діаграма варіантів використання до інформаційної системи

Після виконання всіх етапів проектування, здійснюється перехід до стадії розробки програмного забезпечення.

Добре обґрунтований проект, дозволяє виявити усі недоліки системи, та виправити їх, ще до початка кодінгу.

Також, проектування є допоміжним засобом, для подальшого тестування, за рахунок того, що воно поєднує у собі усі особистості роботи системи, та допомагає проаналізувати схожість розробленої системи з проектом.

Усі елементи програмного забезпечення є взаємопов’язаними з роботою з метеорологічними даннми. Це дозволяє отримати повний огляд метеорологічної ситуації.

Кожен із варіантів використання має спеціалізоване призначення, і має бути описан індивідуально.

Перегляд знайдених аномалій. Для перегляду знайдених аномалій користувачу потрібно в шапці сайту виьрати пункт «перегляд аномалій». Після переходу на данну сторінку користувач бачить таблицью з усіми знайденими аномаліями в метеоданих, доступна функція фільтрації, пошуку та сортування.

Перегляд зібраних метеоданих. Для перегляду зібраних метеоданих користувачу потрібно в шапці сайту виьрати пункт «метеоданні». Після переходу на данну сторінку користувач бачить таблицью з усіми зібраними метеоданими, доступна функція фільтрації, пошуку та сортування.

Перегляд розрахованих прогнозів. Для перегляду розрахованих прогнозів користувачу потрібно в шапці сайту виьрати пункт «перегляд прогнозів». Після переходу на данну сторінку користувач бачить таблицью з усіма прогнозами метеоданих та погодних явищь, доступна функція фільтрації, пошуку та сортування.

Перегляд данних в таблицях. Користувач може переглядати дані виведені в таблиці на сторінках сервісу. Перший рядок таблиці це рядок з іменами стовпчикиів, при натисканні на ім’я виконуеться сортування таблиці в алфавітному порядку за цим ім’ям. При повторному натисканні буде відортивовано в обратному алфавітному порядку.

Пошук. Є розширенням інформаційної системи та дозволяє відшукувати данні в таблицях за будь-яким ключем. Ключем може бути будь-який текст, будь-то цифри, символи чи букви.

Оновлення метеоданих. На сторінці прегляду метеоданниз користувач може натиснути кнопку оновлення метеоданих, це передасть сигнал на сервер та він почне викачувати нові метеоданні з ресурсів. Після оновлення система запускае алгоритм пошуку аномалій в цих данних, далі на вилучених та «чистих» данних система робить прогнозування метеоданих на наступні 3 доби і вже по цим метеоданни система погнозуе погодні явища. Усі данні які оброблюються на данному етапі записуються в базу данних.

Оновлення прогнозу. На сторінці перегляду прогнозів користувач має можливість оновити зроблений системою прогноз, ця функція потрібна для тестування алгоритму прогнозування, та використовуется коли нових данних не має або сервер оновив нейромережі які працювали з даними.

Запуск алгоритма пошуку аномалій. На сторінці перегляду аномалій користувач має можливість оновити список знайдений аномалій, ця функція потрібна для тестування алгоритму пошуку аномалій, та використовуется коли нових данних не має або сервер оновив нейромережі які працювали з даними.

Можливість проведення тестування ефективності навчання нейромереж на сирих та «чистих» метеоданих. Процесс тестування починается з того що нейромережа для пошуку аномалій пробігає данні, формує два списки данних, на цих списках вчаться нейромережі прогнозування метеоданих та погодних явищь, пілся цього ці нейромережі проходять тестування на сирих та «читстих» данних де і порівнюется їх точність.

Після визначення усіх функціональних можливостей програмного забезпечення, є необхідність виконати побудову головних класів. Діаграма класів дозволяє визначити розробникам усі складові елементи системи, та їх особливості роботи. В більшості випадків діаграма класів будується до початка розробки, а потім модифікується, в залежності від зміни умов, або функціоналу програмного забезпечення.

Інформаційну систиему можно розділити на два головних компоненти для кожного з них була розроблена діаграма класів. Компонент інтерфейсу та контрелеру та компонент роботи з алгоритмами обробки та нейромережами. Діаграма класів компонента інтерфейсу та контролеру, зображено на рис. 3.2.

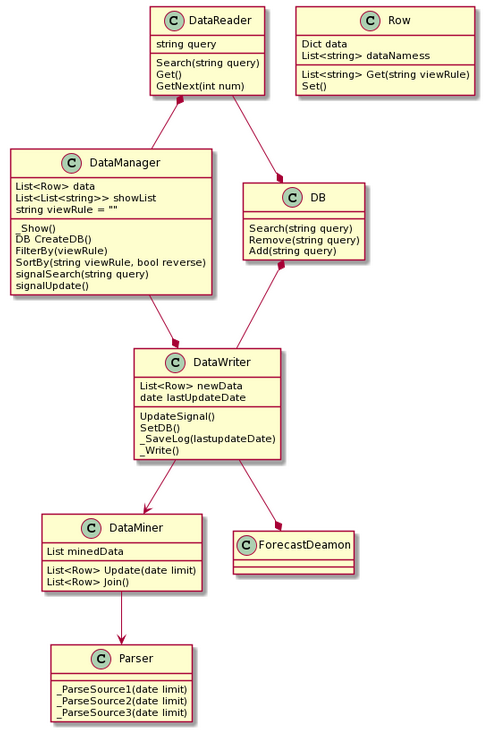


Рисунок 3.2 – Діаграма класів компонента інтерфейсу та контролеру інформаційної системи

На діаграмі зображено функціональну залежність компонентів. У зв’язку з тим, що це web-система то простір класів розроблюється з точки зору окремих компонентів.

В залежності від типу проекту діагарма класів може відрізнятися одна від одної, але в більшості випадків вона має єдину логічну струкруту.

Також на діаграмі визначається залежність компонентів один від одного та вказується система зв’зку між ними.

Реляційні зв’язки є досить суттєвими та необхідні для функціонального аналізу проекту системи.

На цій диаграмі можна розглянути усі задіяні класи. Клас DataReader, DB та DataWriter це підкомпонент для роботи з базою данних, Вони використовуються для запису та зчитування данних.

Кожен з цих класів є обгорткою над інтегрованих в Django класів та пракює за данимим через спеціальний контейнер, за ради більш зручної роботи з ними.

Клас DataReader тісно пов’язаний з інтерфейсом сторінок, він витягує потрібні данні з бази данних та передає їх до головного класу контролера DataManager.

Клас DataReader реалізуе функціонал фільтрації, пошуку, сортування та зміни сторінок.

Клас DataWriter реалізує функціонал запису данних в базу. Цей клас бере данні з двох джерел, з компонента обробки та аналізу метеоданих та компонента парсінгу даних з зовнішніх джерел. Клас DB реалізує базову логіку роботи з базою даних.

Підкомпонент для парсінгу данних з зовнішніх джерел складається з двох класів DataMiner і Parser, клас Parser реаоізує функціонал збору данних з зовнішніх джерел, він використувую https запроси та бібліотеку BS4 для парсінгу. Клас DataMiner використовується для обєднання усіх вилучених данних в єдиний список для подальшого завантащення в базу данних.

Діаграма класів компонента алоритмів розрахунку та нейромереж на рис 3.3.

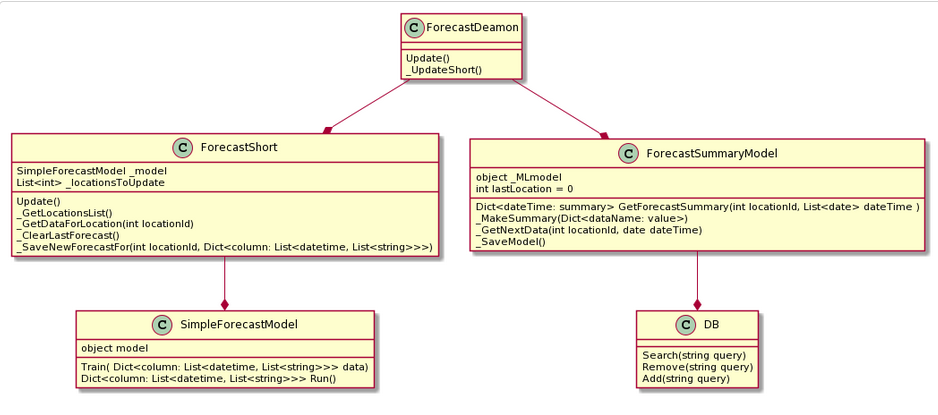


Рисунок 3.3 – Діаграма класів компонента алоритмів розрахунку та нейромереж

Компонент алоритмів розрахунку та нейромереж складається з класів ForecastDeamon, ForecastShort та ForecastSummaryModel. ForecastDeamon це клас контролер усіх алгоритмів обробки та аналізу, він отримує сигнал на запуск алгоритмів обробки потім передає його далі в класи ForecastShort та ForecastSummaryModel. Клас ForecastShort потрібен для прогнозування метеорологічних данних та роботи з нейромережею SimpleForecastModel, яка прогнозує метеодані. Клас ForecastSummaryModel реалізує функціонал прогнозування погодних явищ використовуючи нейромережу для аналізу метеоданих.

Для більш покращеного розуміння алгоритмічного складу розробленого програмного забезпечення, використовується побудова алгоритмічних блок-схем які дозволяють зрозуміти користувачам та розробникам, як саме працює програмне забезпечення.

В першу чергу слід описати схему роботи алгоритму збору та аналізу даних на аномалії, рис. 3.4. Даний алгоритм наводиться в дію при нажаніі на кнопку «оновити дані». Дана блок-схема чітко відображає наступний порядок дій, спершу активується система збору даних, вона опитує наявні у неї джерела даних і виробляє їх завантаження, далле за допомогою нейромережі для посіка аномалій виявляються проблемні дані які потім записують у різні списки, сирих і очищених дані, ці дані заносяться в базу для подальшого аналізу і прогнозування.

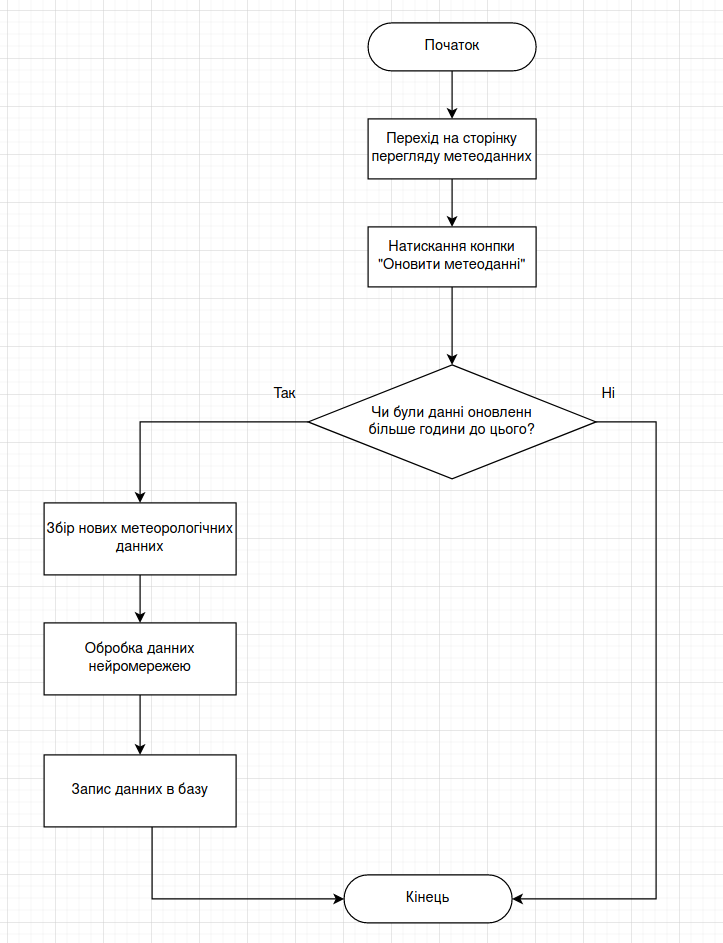


Рисунок 3.4 – Схема работи алгоритму збору и аналізу данних на аномалії

Друга блок-схема представлена ​​в даному розділі це схема обробки отриманих даних і побудова прогнозу на слушні три доби, рис. 3.5, кожна запість прогнозу містить спрогнозовані інформацію про погоду і прогнозоване за цими даними погодне явище, різниця між записами о третій годині часу.

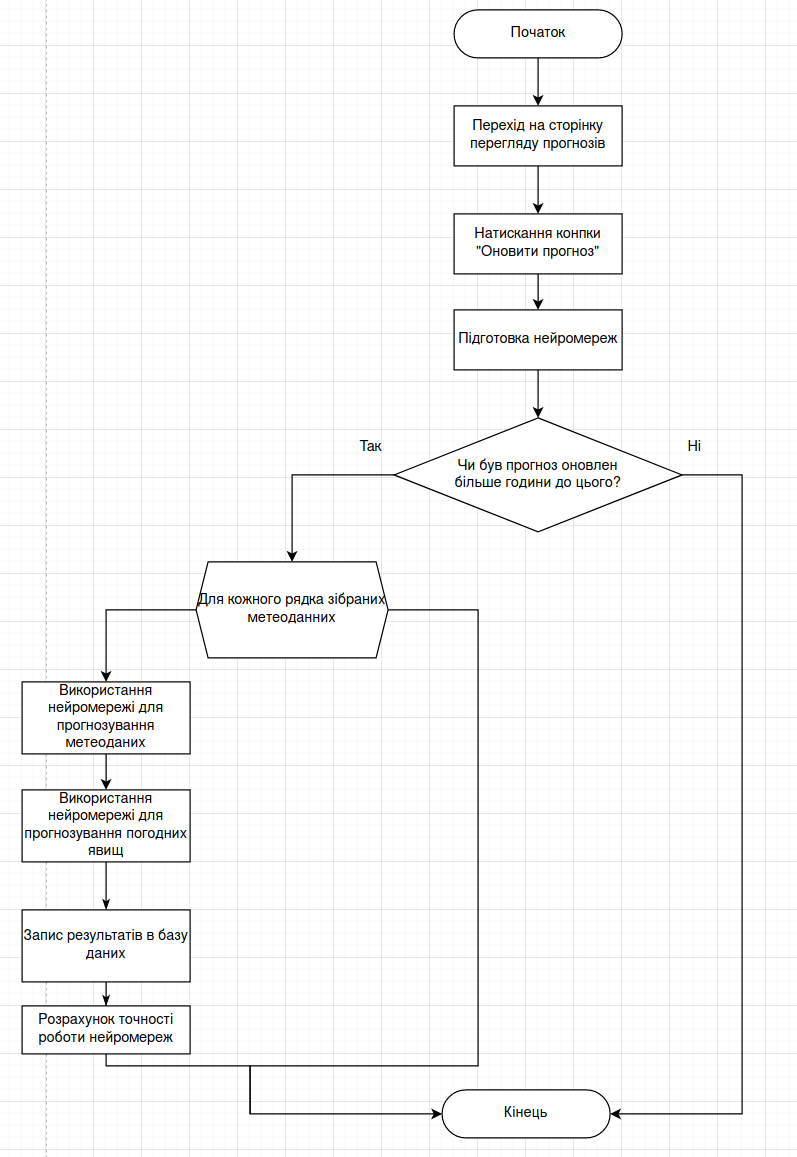


Рисунок 3.5 – Схема обробки отриманих даних і побудова прогнозу

Кроки по блок-схемі наступні, вичітваніе даних з обох списків, далі підготовка нейромереж під ці дані і складання прогнозу метеоданих, слідом інша нейросеть за прогнозом робить прогноз погодного явища, в процесі система аналізує точність роботи всіх нейромереж і порівнює дані їх точності.

Ці результати потрібні для основної мети даної роботи, аналіз впливу систем пошуку аномалій в метеорологію шляхом порівняння точності алгоритмів прогнозування навчених на даних з аномаліями і без.

Алгоритм представлений з додатковими оптимізаціями, внаслідок чого алгоритм простий для розуміня і при тому швидкий та зручний в використані.

Інші алгоритми використовані в системі більш прості та не потребують додаткового зобарження у вигляді блочної схеми.

Це алгоритми пошуку, фільтрації, сортування, вичитки даних, вичитки данних з зовнішніх джерел, роботи з базою даних та розміщення даних та збору сторінок виводу інформації.

## Опис розробленої бази даних інформаційної системи

Під час розробки інформаційної ситсеми було визначено перелік таблиць та їх структуру, які є необхідними для доброго функціонування усієї системи.

Ці талиці складають структуру організації даних в базі з якою і працює система.

Кожна таблиця складаеться з стовпців, кожен з яких має своє ім’я та тип даних, це частини даних які записуються в базу. Кожен рядок таблиці це окремих запис з даних.

Однією з перших таблиць, яка була розроблена є таблиця метеорологічних даних, рис. 3.6.

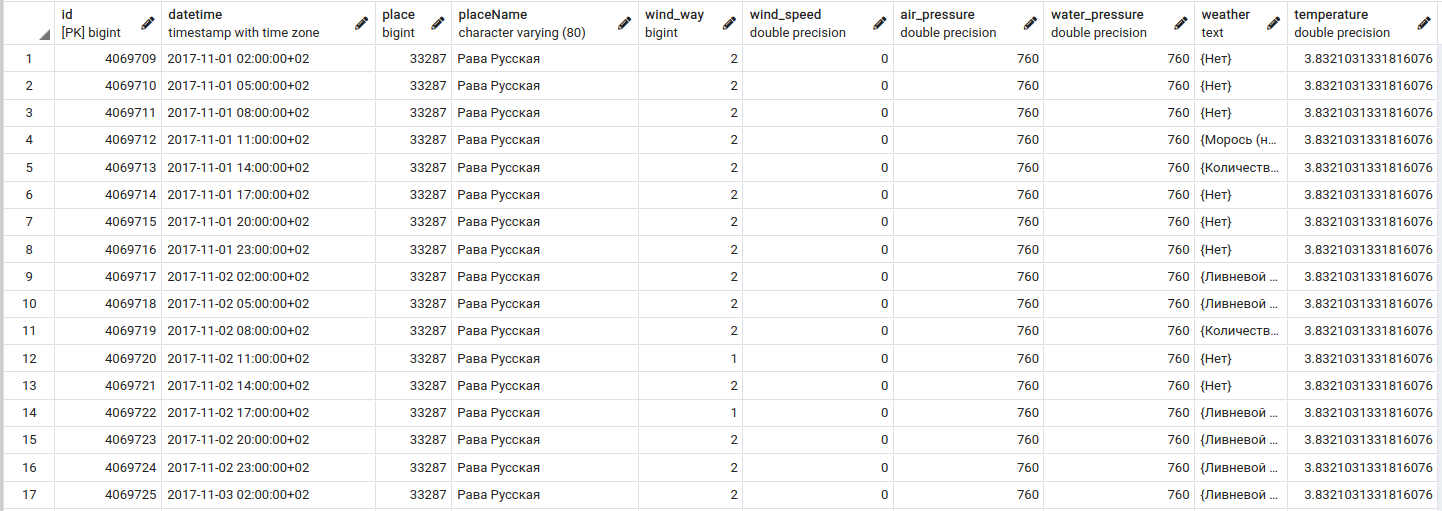


Рисунок 3.6 – Вікно таблиці метеоданих в базі даних

Дана таблиця створена засобами фреймворка Django і виконується наступним кодом.

class Meteodata(models.Model):

id = models.BigAutoField("id", primary\_key=True)

datetime = models.DateTimeField("Дата и время")

place = models.BigIntegerField("Код места")

placeName = models.CharField("Место", max\_length=80)

temperature = models.FloatField("Температура воздуха")

wind\_way = models.BigIntegerField("Направление ветра")

wind\_speed = models.FloatField("Скорость ветра")

air\_pressure = models.FloatField("Атмосферное давление")

water\_pressure = models.FloatField("Давление моря")

weather = models.TextField("Погодное явление")

class Meta:

verbose\_name = "Метеоданные"

verbose\_name\_plural = "Метеоданные"

Детально по кожному полю таблиці метеоданих:

* Id це поле целочисленном типу зберігає унікальний номер запису.
* Datetime це підлозі в якому зберігається дата і времм коли було зроблено запис.
* Place це поле в якому зберігається назва регіону в якому був зроблений запис.
* placeNumber це поле в якому зберігається номер регіону в якому був зроблений запис.
* temperature поле температури яка була зафіксована в момент запису.
* wind\_way поле зберігає напрямок вітру.
* wind\_speed поле зберігає значення швидкості вітру.
* air\_pressure поле зберігає значення атмосферного тиску.
* water\_pressure поле зберігає значення атмосферного тиску на рівні моря.
* weather поле зберігає погодне явище відбувалося в момент запису.

Відповідно до цієї таблиці створена таблиця для записів яка має однаковий набір полів, рис 3.7.

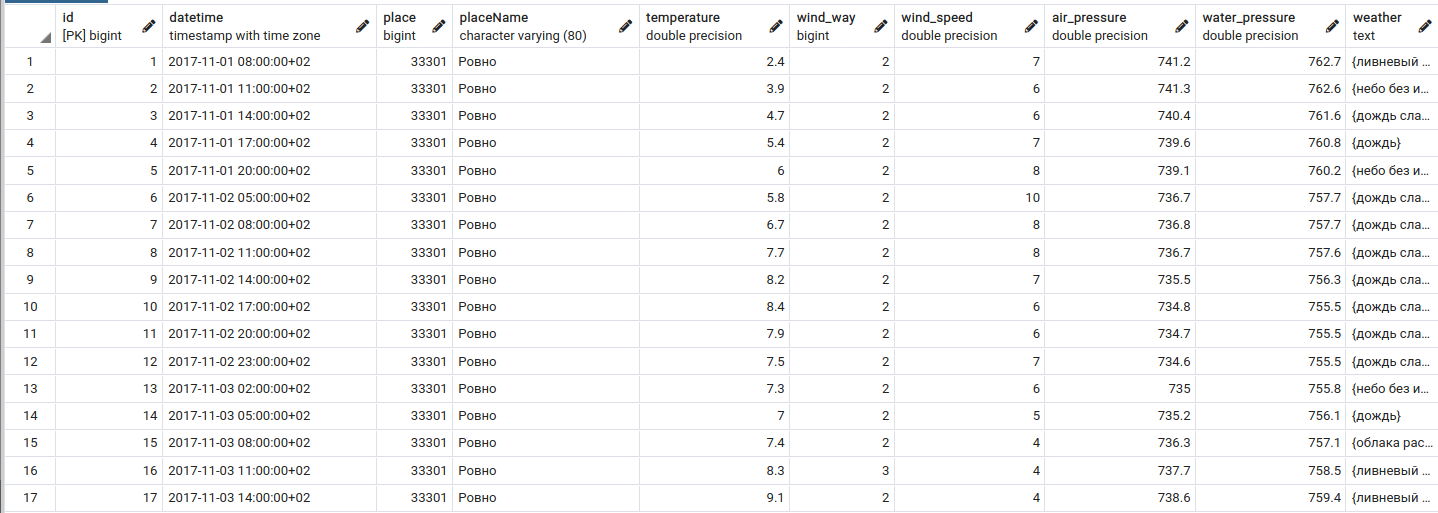


Рисунок 3.7 – Таблиця очищених метеоданих

Таблиця прогнозів про погоду містить збігаються поля з таблицею з інформацію про погоду.

Таблиця зберігає список знайдених аномалій містить наступні поля:

* Id поле зберігає унікальний номер запису.
* Meteodata\_Id містить унікальний номер запису з таблиці мететоданних.
* Fieldname містить назву поля в якому була обноружена аномалія.
* Value зберігає в собі значення поля в якому була виявлена аномалія.
* Anomaly містить кратне опис аномалії або її тип.

Дана табіца створюється наступним кодом:

class MeteodataAnomalies(models.Model):

id = models.BigAutoField("id", primary\_key=True)

meteodata\_id = models.BigIntegerField("Код записи")

fieldname = models.TextField("Название поля")

value = models.TextField("Значение")

anomaly = models.TextField("Аномалия")

class Meta:

verbose\_name = "Аномалии"

verbose\_name\_plural = "Аномалии"

Також варто звернути увагу що були отримані раніше таблиці є додатковими, крім них Django створює радий основних таблиць, це таблиці зберігають в собі дані про підключених до системи з'єднань, їх унікальні коди, паролі адміністратора, таблиці логування дій в системі, таблиця логування помилок в системі, таблиця підключених модулів і т.д.

База даних вкрай чутливий елемент системи і требудет до себе великої уваги. Так як система працює з великими даними повинна бути можливість легко взаємодіяти з нею, здійснювати запис в базу даних з високою швидкістю. Для цього раздаботанние таблиці даних повинні бути схожі, а дані в них однорідні. Дане завдання оптимізації процесів вирішується на рівні програмного інтерфейсу надається фреймворком Django 3, а також додатковими бібліотеками, які максимально спрощують створення системи.

Як і було позначена раніше Django генерує таблиці в базі даних, для цього використовується python код який описує структуру яку повинна мати таблиця. Для цього використовується наступні типи даних:

* BinaryField. Створює BLOB-поле для зберігання двійкових даних (наприклад, зображень, аудіо чи інших мультимедійних об'єктів)
* BooleanField. Створює логічне поле для зберігання значень True / False (або 0/1)
* DateField. Створює поле дати для зберігання дат.
* FloatField. Створює стовпець для зберігання чисел з плаваючою комою.
* TextField. Створює текстове поле для зберігання тексту.
* BigIntegerField. Створює велике ціле число, щоб вмістити числа від –9223372036854775808 до 9223372036854775807. Цей діапазон може змінюватися залежно від марки БД
* EmailField. Примушує текст – це дійсний електронний лист із внутрішнім Django EmailValidator, щоб визначити, що є, а що не є дійсним. Працює так само, як за замовчуванням CharField має максимальну довжину 254 символи, а також примушує рядок – це дійсний електронний лист.

Для аналізу ієрархії таблиць бази даних іноді використовуються діаграм зв'язку таблиць, це необов'язкові метод аналізу який потрібен тільки для наочності та полегшеного розуміння структури бази даних. Детально описуючи його можна виділити наступне.

Всередині діаграми бази даних у кожній зв'язку є три окремих елемента: кінцеві точки, стиль лінії і зв'язані таблиці.

Кінцеві точки. Кінцеві точки лінії показують вид зв'язку: "один до одного" або "один до багатьох".

Якщо на одній кінцевій точці зв'язку знаходиться ключ, а на іншій – цифра вісім, то це зв'язок «один до багатьох». Якщо в зв'язку по одному ключу на кожній кінцевій точці, то це зв'язок «один до одного».

Стиль лінії. Різновид лінії (не її кінцеві точки) показує, перевіряє чи СКБД посилальну цілісність для зв'язку при додаванні нових даних в таблицю, пов'язану з допомогою зовнішнього ключа.

Якщо зв'язок намальована у вигляді суцільної лінії, це означає, що СКБД перевіряє посилальну цілісність для зв'язку при додаванні або зміні рядків у таблиці, пов'язаної з допомогою зовнішнього ключа.

Якщо пунктирна лінія, це означає, що СКБД не перевіряє посилальну цілісність для зв'язку при додаванні або зміні рядків у таблиці, пов'язаної з допомогою зовнішнього ключа.

Зв'язані таблиці. Лінія зв'язку показує, що дві таблиці, зв'язані з допомогою зовнішнього ключа.

Для зв'язку «один до багатьох» таблиця, зв'язана з допомогою ключа, – це таблиця близько цифри на лінії.

Якщо обидві кінцеві точки лінії приєднані до однієї таблиці, це означає зворотну зв'язок.

Для будь-якої бази даних можна створити будь-яку необхідну кількість діаграм; кожна з таблиць бази даних може використовуватися у будь-якій кількості діаграм.

Таким чином, для візуалізації різних частин бази даних або для акцентування різних аспектів її конструювання можна створювати різні діаграми. Наприклад, можна створити велику діаграму, в якій будуть відображатися всі таблиці і стовпчики, а також меншу діаграму, в якій будуть відображатися всі таблиці, але не буде стовпців.

Всередині діаграми бази даних кожна таблиця має три окремих елементів: заголовка вікна, список вибору рядків і набір стовпців властивостей. Рядок заголовка в рядку заголовка відображається ім'я таблиці.

Якщо таблицю було змінено, але ще не збережено, то після імені таблиці з'являється зірочка (\*), що показує наявність незбережених змін. Область виділення рядка.

Щоб вибрати стовпець бази даних в таблиці, клацніть область виділення рядка. Якщо стовпець є первинним ключем таблиці, то в цьому списку відображається символ ключа.

Стовпці властивостей. Набір стовпців властивостей видно на поданні таблиці.

Таблицю можна переглянути в будь-якому з п'яти різних уявлень, що дозволяють підібрати підходящий розмір і розміщення елементів діаграми.

## Опис розробленого інтерфейсу до інформаційної системи

Інформаційну систему розроблено в кращих традиціях web-архітектури, за рахунок чого сайт складається з адміністраторської частини та клієнтської.

Адміністратор системи може додавати будь-яку інформацію на сайт, змінювати поточну, адмініструвати облікові записи та інше.

Панель адміністратора виконана у вигляді переліку, що спрощує навігацію, та дозволяє поступово виконувати різні дії з контентом.

Головне вікно панелі адміністратора зображено на рис. 3.8.

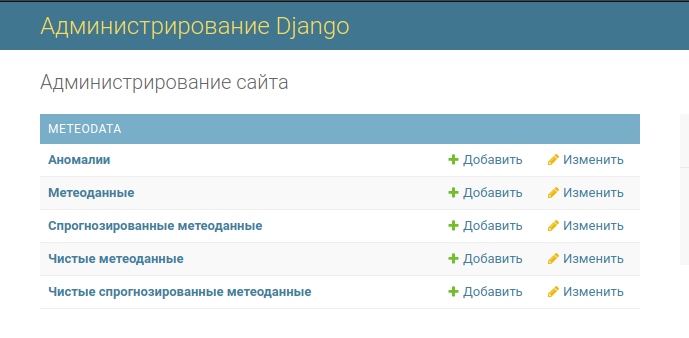


Рисунок 3.8 – Вікно панелі адміністратора

Головними складовими функціями адміністратора є додання нового контенту.

Для цього розроблено спеціалізовані форми, які дозволяють спрощувати цей процес та роботи його зрозумілим для кожного, рис. 3.9.

Це виконано з розрахунку на те, що в більшості метеорологічних центрів, адмініструванням контенту займаються прості метеорологи, які не мають спеціалізованого досвіду роботи з адміністрування сайтів.

Отже це було одним з головних критеріїв під час розробки панелі адміністратора.

Вона повинна бути як можливо спрощеною та відповідати усім вимогам до відповідного програмного засобу.

Форма додання нового запису метеоданих включає в себе:

* Id це поле целочисленном типу зберігає унікальний номер запису;
* Datetime це підлозі в якому зберігається дата і времм коли було зроблено запис;
* Place це поле в якому зберігається назва регіону в якому був зроблений запис;
* placeNumber це поле в якому зберігається номер регіону в якому був зроблений запис;
* temperature поле температури яка була зафіксована в момент запису;
* wind\_way поле зберігає напрямок вітру;
* wind\_speed поле зберігає значення швидкості вітру;
* air\_pressure поле зберігає значення атмосферного тиску;
* water\_pressure поле зберігає значення атмосферного тиску на рівні моря.

Основні джерела містять інформацію подібному форматі, різні інтернет ресурси закордонних спеціалізованих метеоцентри, що базуються в провідних університетах і великих компаніях. Це оптимальний набір даних для аналізу системою, який дає можливість легко збирати дані. Також різні архіви метеоданих містять всі вибрані для аналізу даної системою дані, що є важливим критерієм вибору, бо дає доступ до майже нескінченним ресурсів зборів Європейського метеоцентри.

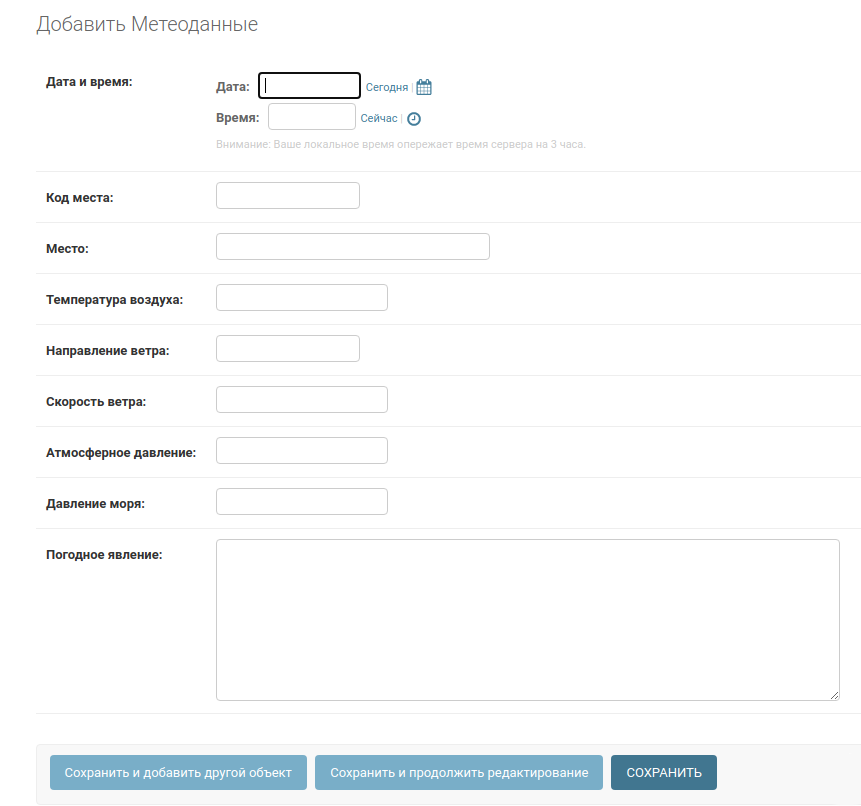


Рисунок 3.9 – Вікно додання нового запису метеоданих

Якщо необхідно відредагувати вже існуючий запис, то його форма буде максимально відповідати нормам змін. Усі додані до системи записи виводяться користувачам у вигляді переліків, що є нормою для перегляду великої кількості контенту схожого типу. Вікно форми редагування інформації зображено на рис. 3.10.

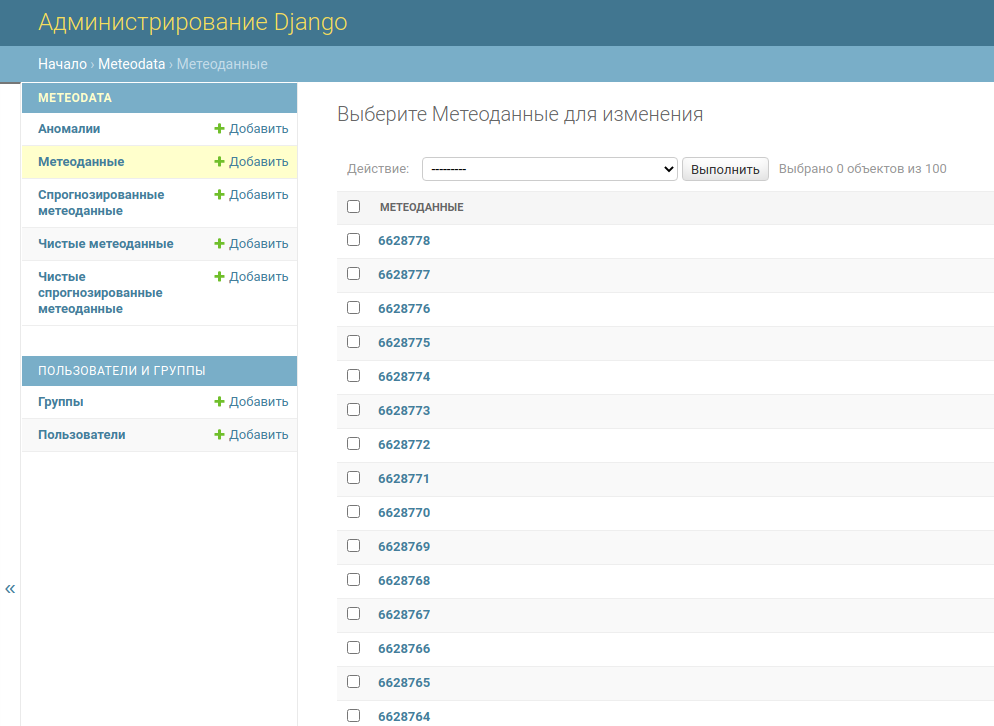


Рисунок 3.10 – Вікно редагування існуючих турів

Також не менш необхідною формою є форма для додавання, редагування та видалення аномалій. Дана форма складається зі спеціальних полів і в неї необхідно внести наступні дані:

* Id поле зберігає унікальний номер запису.
* Meteodata\_Id містить унікальний номер запису з таблиці метеоданнх.
* Fieldname містить назву поля в якому була обноружена аномалія.
* Value зберігає в собі значення поля в якому була виявлена аномалія.
* Anomaly містить кратне опис аномалії або її тип.

Усі інші форми редагування чи додання інформації розроблені аналогічним образом. Отже їх опис не є обов’язковим. Після розгляду функціональної складової панелі адміністратора, необхідно зробити опис користувацької частини.

Інтерфейс сайту виконаний в стилі подібному месенджер Telegram, який є еталонним з боку простоти дизайну і його приємності оці. Стиль спокійний і «однознанчний», його сособенності в тому що полььзователю зрозуміло як користуватися елементами інтерфейсу і якого «поведеднія» від них чекати.

Базова сторінка є сторінкою перегляду метеоданих, вона складається з декількох соновном елементів. Перший елемент необхідний лоя роботи з системою це вершняя панель переходу між сторінками. На цій панелі розташовані три кнопки для переходу між сторінками перегляду метеоданих, перегляду прогнозів і перегляду аномалій шляхом натискання кнопок «перегляд метеоданих», «прогнозування метеоданих» і «Аналіз метеоданих» відповідно. Кнопка «Оновити дані» запускає основний алгоритм обробки даних, це основна кнопка по роботі з системою, рис. 3.11.

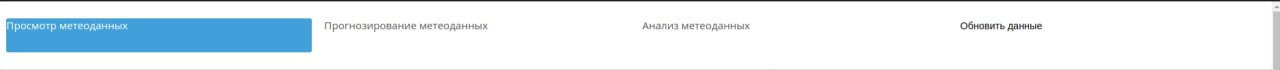


Рисунок 3.11 – Приклад зовнішнього вигляду верхній панелі управління

Другим важливим елементом на кожній сторінці явлется основний блок, в ньому може розташовуватися найрізноманітніша інформація, в даному випадку в ньому розташовані таблиці з даними для вільного їх перегляду користувачем. Використовувані таблиці мають панель управління, кожна робоча таблиця своя і саму таблицю в якій відображені дані. Дані в таблиці розбиті по стовпцях з іменами параметрів виведені в рядках нижче, при натисканні на назву парамтера проводиться сортування, а при подвійному натисканні зворотна сортування по параметру тексту, утримуючи кнопки заголовка. Інформація в таблиці виводитися в своїх окремих осередках, в центрі кожної з яких розміщений текст зі значенням параменту стовпця в цьому рядку записи, рис.3.12.

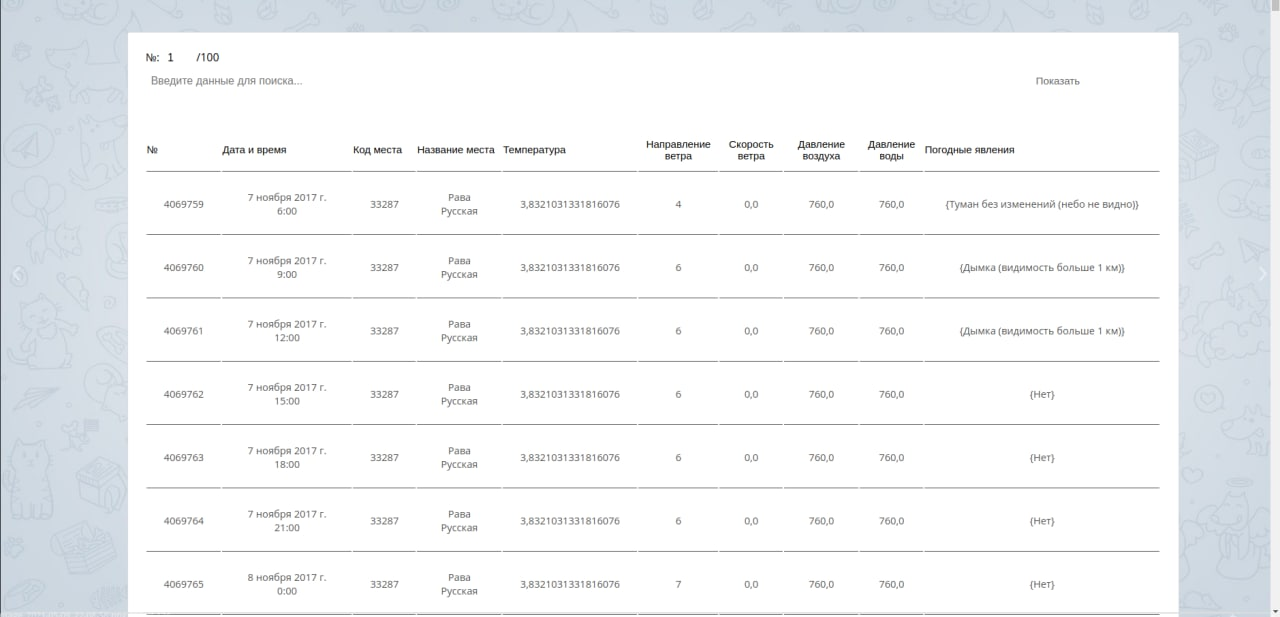


Рисунок 3.12 – Таблиця для відображеня даних.

Панель управління таблицею складається з рядка відображення поточної сторінки і максимального номера станиці. Даний елемент являє собою акуратну рядок тексту.

Також на панелі управління розташована рядок пошуку і фільтрації даних, що виводяться і кнопка застосувати.

Частково прозора рядок введення необхідна для пошуку в даних і їх фільтрації, для цього достатньо ввести якийсь текст і натиснути кнопку застосувати. Пошук осущетсвляется за всіма параметрами записи, рис. 3.13.



Рисунок 3.13 – Приклад зовнішнього вигляду панелі фільтрації таблиць.

Описані вище елементи інтерфейсу повторюються на кожній сторінці системи і в детальному описі не потребують. Інтерфейс виконаний в мінімалістичному дизайні щоб мінімально відволікати користувача від роботи з системою. Система розроблялася з ідеєю максимально швидкої роботи з нею [22].

## Практичне використання інформаційної системи

З метою перевірки роботоспосібності розробленої інформаційної системи, необхідно побудувати тестовий приклад, який дозволить проаналізувати ефективність роботи системи, а також забезпечить зворотній зв’язок щодо розробленого функціоналу. Спочатку необхідно обрати набір вхідних критерієв, на яких буде базуватися тестовий приклад, рис. 3.14.

Для базового прикладу функціональності програми досить перейти на сторінку перегляду метеоданих і натиснути на кнопку «Оновити дані». Після цього і будуть завантажені всі початкові дані необхідні для роботи алгоритмів і настройки всіх використовуваних нейромереж. З боку користувача потрібно тільки очікування, весь подальший процес польностью автоматизований і не вимагає додаткових втручань людини. Першим етапом буде то що система завантажить інформацію про погоду, в разі якщо дані частково були завантажені раніше то система оновить їх вибравши тільки ті що з'явилися з дати останнього запису в системі. Далі дані передаються на обробку нейромережі з пошуку аномалій, та записує їх в два списки «чистий», в якому знаходяться всі записи які пройшли перевірку і сирої, в нього входять всі отримані інформацію про погоду, також знайдені аномалії заносяться в окрему таблицю в базі.

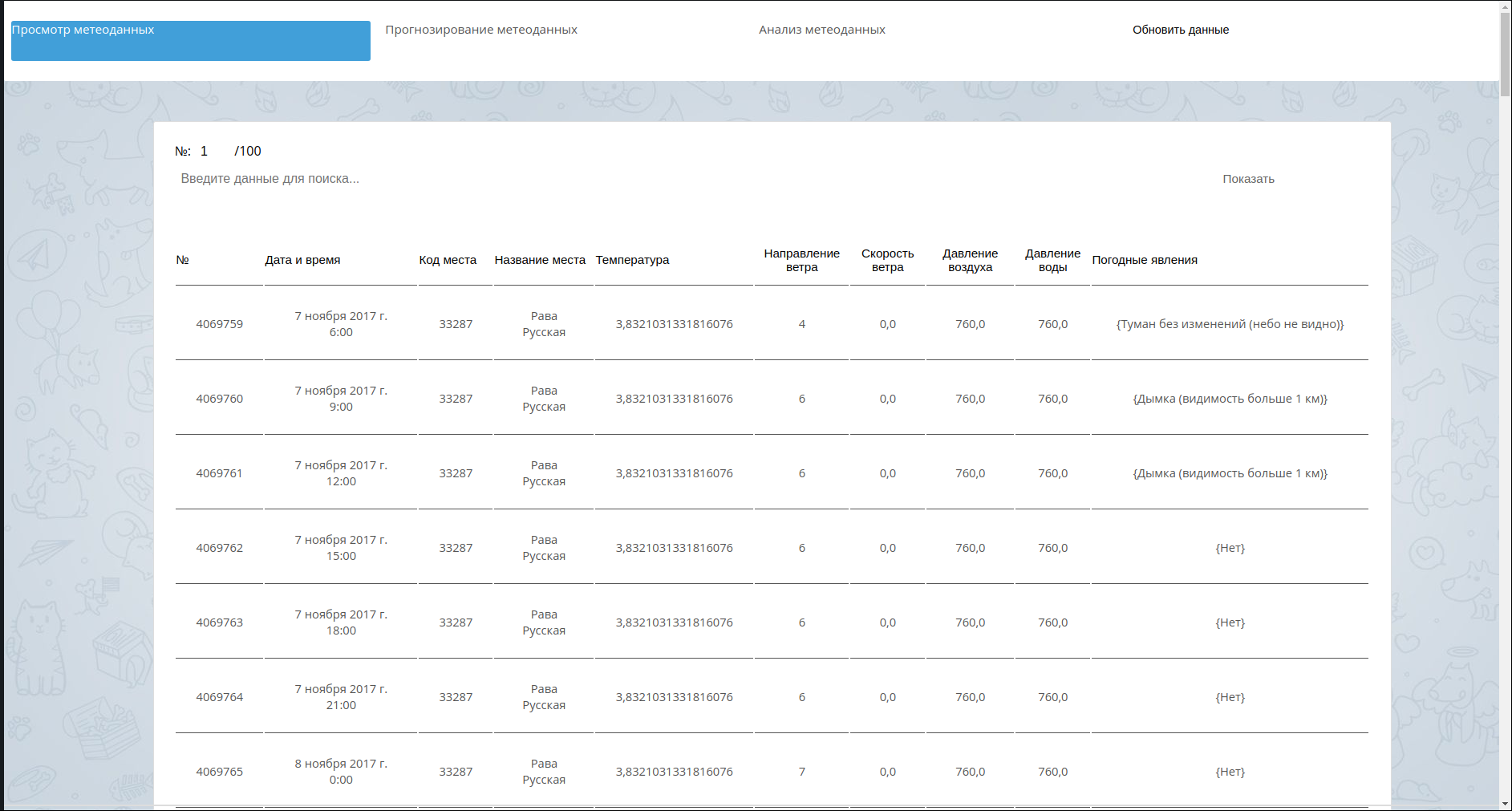


Рисунок 3.14 – Вікно перегляду метеоданих

Наступним етапоя вляется підготовка нейромереж для прогнозування, їх створюється дві копії, одна для повного списку а друга пада тільки для чистого списку. Після підготовки нейромереж обидві пари проходять свої списки і будують прогноз метеоданих і погодних явищ. В процесі система збирає статистику алгоритмів прогнозування для подальшого порівняння ефективності алгоритмів. В кінці цього процесу відбувається запість розрахованих прогнозів базу даних, прогноз пари робила розрахунок на чистих даних записується в список чистого прогнозва, прогноз другої пари в повний список прогнозів відповідно. Закінчення алгоритму в тому що користувача переносять на відповідну сторінку, на якій зображена стовпчаста діаграма з коефіцієнтами точності алгоритмів, кількість отриманих записів і число знайдених в них аномалій а також час скільки зайняв весь процес розрахунку.

Всі дані які були отримані в процесі записані у відповідні таблиці і можуть бути переглянуті користувачем пізніше. Головна перевага даної системи в тому що вона дозволяє легко перевірити алгоритми прогнозування і пошуку аномалій. При постановці подібних задачь користувачеві досить натиснути одну кнопку а для перевірки різних алгоритмів обробки і аналізу досить підмінити заздалегідь заготовлені файли нейромереж.

## Висновки до третього розділу

Проаналізувавши різні методи розробки інформаційних систем, а також засоби реалізації, було спроектовано сучасне програмне рішення, яке може впроваджуватися в метеорологічні центри та використовуватися як самостійна система, або як додатковий модуль з розширенням функціональних можливостей системи. При проектуванні інформаційної системи було визначено головні можливості системи, а також її функціональні особистості.

Головною складовою інформаційної системи є модуль обробки та аналізу даних, конкретно системи пошуку аномалій та побудудвання прогнозу. Система фільтрації та пошуку побудована з використанням різних пошукових критеріїв, які оброблюються різними спеціалізованими туристичними базами даних.

Додатковим програмним модулем є реалізація системи збору метеорологічних даних. Модуль опитує зовнішні джерела на предмет нових метеоданих та завантажуе їх по конкретним запитам.

Розроблена інформаційна система може використовуватися в якості самостійного програмного засобу аналізу та спрощення роботи метеорологів, а також як додатковий програмний модуль, який може бути впроваджено до існуючої інформаційної системи метеорологічних центрів.

# ВИСНОВКИ

В рамках виконання роботи були закріплені і розширені теоретичні знання, отримані при вивченні дисциплін, розвинені необхідні практичні вміння та навички відповідно до вимог рівня підготовки випускника та затвердженої тематики дослідження.

В результаті отриманих знань, і додатково вивчених матеріалів було проведене дослідження та розроблене програмне забезпечення що дозволяє оцінити ефективність нейромереж у задачах обробки та аналізу метеоданих.

Під час написання роботи, було виконано:

* аналіз того як машинне навчання допомогае вирішувати задачі метеорології;
* аналіз впливу інформаційних технологій на робочі процеси метеорологів;
* обґрунтувано впровадження інформаційних систем в сферу метеорологіх;
* визначення програмні засоби реалізації інформаційної системи;
* проектування інформаційної системи;
* розробку інформаційної системи для сфери метеорології;
* проведено дослідження результатів обробки та аналізу метеоданих на основі різних алгоритмів машинного навчання;
* проведено дослідження того які вхідні параметрі мали найбільший вплив на результат в різних алгоритмах машинного навчання.

Розробку програмного комплексу виконано з використанням мови програмування Python та спеціалізованого фреймворку для роботи з web-системами Django. В якості системи керування базами даних використано SQLite.

# ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. ДСТУ 8302:2015 «Інформація та документація. Бібліографічне посилання. Загальні положення та правила складання» – Чинний від 04.03.2016. – Київ: УкрНДНЦ, 2020. – 20 с.
2. Антоні С. Опануй самостійно програмування за 21 день / С. Антоні. – Москва: Вильямс, 2021. – 562 с.
3. Аронов І.З. Сучасні проблеми безпеки технічних систем і аналізу ризику / І.З. Аронов. – Москва: Інтуит. 2021 – 451 с.
4. Астелс Д. Керівництво з екстремального програмування / Д. Астелс. – Москва: Вільямc, 2022. – 468 с.
5. Багриновский К.А. Нові інформаційні технології / К.А. Багриновский, Є.Ю. Хрустальов. – Москва: ЕКО, 2019. – 212 с.
6. Гмурман В.Є. Теорія ймовірностей і математична статистика / В.Є. Гмурман. – Москва: Вища. Шк, 2022. – 479с.
7. Бахтізін В.В. Технології розробки програмного забезпечення / В.В. Бахтізін, Л.А. Глухова. – Мінськ: БДУІР, 2020. – 267 с.
8. Бенін Д.М. Системи підтримки прийняття рішень / В.Л. Сніжко, Д.М. Бенін – Москва: Тріада, 2022. – 165 с.
9. Бхаграва А. Грокаем алгоритми / А. Бхаграва – СПб .: ПИТЕР, 2020. – 288с
10. Вірт Н. Алгоритми і структури даних. / Н. Вірт – Москва: Свет, 2019. – 651 с.
11. Владимиров В.А. Управління ризиком / В.А. Владимиров, Ю.Л. Воробйов, Г.Г. Малинецкий. – Москва: ЮНИТИ, 2020. – 352 с.
12. Шаруп Ч. Х. Укус пітона / Ч. Х. Шаруп – Москва: ЮНИТИ, 2022 – 152с.
13. Дронов В. Django 3 практика створення веб-сайтів на python / В. Дронов – СПб.: «БХВ-Петербург», 2021 – 712с
14. Прохоренок Н.А. Python 3 и PyQt 5. Разработка приложений / Н.А. Прохоренок, В.А. Дронов. – СПб.: БХВ-Петербург, 2020. – 832c.
15. Padmanabhan T.R. Programming with Python / T.R. Padmanabhan. – СПб.: Springer, 2020. – 349p.
16. Эрик М. Изучаем Python. Программирование игр, визуализация данных, веб-приложения / М. Эрик. – СПб.: Питер, 2021. – 496c.
17. Коггзолл Д. PHP 5. Полное руководство / Д. Коггзолл. – Москва: Диалектика, 2021. – 752 с.
18. Бейлин Л. Изучаем MySQL / Л. Бейлин. – Москва: Эксмо, 2022. – 1060 с.
19. Lane D. Web Database Application with PHP and MySQL / D. Lane. – New Jersey: O'Reilly, 2021. – 816 p.
20. Прохоренок Н.А. HTML, JavaScript, PHP и MySQL. Джентльменский набор Web-мастера / Н.А. Пархоменко. – СПб.: Питер, 2019. – 768 с.
21. Роббинс Д. HTML5, CSS3 и JavaScript. Исчерпывающее руководство / Д. Роббинс. – Москва: Эксмо, 2019. – 528 с.
22. Шварц Б. MySQL. Оптимизация производительности / Б. Шварц. – Москва: Символ-Плю, 2019. – 483 с.
23. Гольцман В. MySQL 5.0 / В.Гольцман. – Москва: Питер, 2019. – 764 с.
24. Яргер Р.Д. MySQL и mSQL. Базы данных для небольших предприятий и Интернета / Р.Д. Яргер. – Москва: Символ-Плюс, 2021. – 929 с.
25. Уильман Л. MySQL. Руководство по изучению языка / Л.Уильман. – Москва: ДМК Пресс, 2021. – 764 с.
26. Аткинсон Л. MySQL. Библиотека профессионала / Л.Аткинсон. – Москва: Вильямc, 2021. – 1493 с.
27. Артеменко Ю. Н. MySQL. Справочник по язику / Ю.Н. Артеменко. –Москва: Вильямс, 2021. – 843 с.
28. Никсон Р. «Learning PHP, MySQL, JavaScript, CSS & HTML5 A Step-by-Step Guide to Creating Dynamic Websites» / Р. Никсон. – O'Reilly Media, 2020. – 730 с.
29. Rocha G.D. Learning SQLite for iOS / G.D. Rocha. – NY.: [Packt Publishing](https://www.packtpub.com/application-development/learning-sqlite-ios), 2022. – 579 р.
30. Owens M. The Definitive Guide to SQLite / M. Owens, G. Allen. – NY.: Apress, 2022. – 739 р.
31. Lary D.J. Machine learning in geosciences and remote sensing / D.J. Lary, A.H. Alavi, A.H. Gandomi, A.L. Walker // Geoscience Frontiers. – Vol. 7. – N 1. – 2022. – PP. 3–10.
32. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop. – NY: Springer. – 2021. – 311 p.
33. LeCun Y. Deep learning / Y. LeCun, B. Yoshua, H. Geoffrey // Nature. – Vol. 521. – N 7553. – 2022. – PP. 436–444.
34. Склярев О.М. Разработка программных продуктов на языках высокого уровня / О.М. Склярев. – СПб.: Глоунас, 2021. – 541 с.
35. Буйначев С.К. Основы программирования на языке Python / С.К. Буйначев. – М.: Издательство Уральского университета, 2021. – 92 с.

# ДОДАТОК А

Лістинг коду інформаційної системи

from ..models import Meteodata, ClearMeteodata, ForecastMeteodata, ClearForecastMeteodata, MeteodataAnomalies

from django.db.models import Q

class DataReader():

def \_\_init\_\_(self):

self.\_rows\_count = 50

self.\_meteodata\_page = 1

self.\_meteodata\_search = ''

self.\_meteodata = Meteodata.objects.all()

self.\_clear\_meteodata\_page = 1

self.\_clear\_meteodata\_search = ''

self.\_clear\_meteodata = ClearMeteodata.objects.all()

self.\_forecast\_page = 1

self.\_forecast\_search = ''

self.\_forecast = ForecastMeteodata.objects.all()

self.\_clear\_forecast\_page = 1

self.\_clear\_forecast\_search = ''

self.\_clear\_forecast = ClearForecastMeteodata.objects.all()

self.\_anomaly\_page = 1

self.\_anomaly\_search = ''

self.\_anomaly = MeteodataAnomalies.objects.all()

self.\_firstOrderName = ''

self.\_secondOrderName = ''

self.\_thirdOrderName = ''

self.\_fothOrderName = ''

self.\_fifthOrderName = ''

print(type(self.\_meteodata))

self.\_updateData(1)

self.\_updateData(2)

self.\_updateData(3)

self.\_updateData(4)

self.\_updateData(5)

print(type(self.\_meteodata))

def set\_page(self,tableId, num):

print('set\_page')

print(type(self.\_meteodata))

if tableId == 1:

self.\_meteodata\_page = num

self.\_updateData(1)

elif tableId == 2:

self.\_clear\_meteodata\_page = num

self.\_updateData(2)

elif tableId == 3:

self.\_forecast\_page = num

self.\_updateData(3)

elif tableId == 4:

self.\_clear\_forecast\_page = num

self.\_updateData(4)

elif tableId == 5:

self.\_anomaly\_page = num

self.\_updateData(5)

def search(self, tableId, search):

print('search')

print(type(self.\_meteodata))

temp = 0

if tableId == 1:

self.\_meteodata\_page = 1

if search != '':

self.\_meteodata\_search = search

self.\_meteodata = self.\_meteodata.filter(

Q(id\_\_contains=search) |

Q(datetime\_\_contains=search) |

Q(place\_\_contains=search) |

Q(placeName\_\_contains=search) |

Q(temperature\_\_contains=search) |

Q(wind\_way\_\_contains=search) |

Q(wind\_speed\_\_contains=search) |

Q(air\_pressure\_\_contains=search) |

Q(water\_pressure\_\_contains=search) |

Q(weather\_\_contains=search)

)

else:

self.\_meteodata\_search = ''

self.\_meteodata = Meteodata.objects.all()

self.\_updateData(1)

elif tableId == 2:

self.\_clear\_meteodata\_page = 1

if search != '':

self.\_clear\_meteodata\_search = search

self.\_clear\_meteodata = self.\_clear\_meteodata.filter(

Q(id\_\_contains=search) |

Q(datetime\_\_contains=search) |

Q(place\_\_contains=search) |

Q(placeName\_\_contains=search) |

Q(temperature\_\_contains=search) |

Q(wind\_way\_\_contains=search) |

Q(wind\_speed\_\_contains=search) |

Q(air\_pressure\_\_contains=search) |

Q(water\_pressure\_\_contains=search) |

Q(weather\_\_contains=search)

)

else:

self.\_clear\_meteodata\_search = ''

self.\_clear\_meteodata = ClearMeteodata.objects.all()

self.\_updateData(2)

elif tableId == 3:

self.\_forecast\_page = 1

if search != '':

self.\_forecast\_search = search

self.\_forecast = self.\_forecast.filter(

Q(id\_\_contains=search) |

Q(datetime\_\_contains=search) |

Q(place\_\_contains=search) |

Q(placeName\_\_contains=search) |

Q(temperature\_\_contains=search) |

Q(wind\_way\_\_contains=search) |

Q(wind\_speed\_\_contains=search) |

Q(air\_pressure\_\_contains=search) |

Q(water\_pressure\_\_contains=search) |

Q(weather\_\_contains=search)

)

else:

self.\_forecast\_search = ''

self.\_forecast = ForecastMeteodata.objects.all()

self.\_updateData(3)

elif tableId == 4:

self.\_clear\_forecast\_page = 1

if search != '':

self.\_clear\_forecast\_search = search

self.\_clear\_forecast = self.\_clear\_forecast.filter(

Q(id\_\_contains=search) |

Q(datetime\_\_contains=search) |

Q(place\_\_contains=search) |

Q(placeName\_\_contains=search) |

Q(temperature\_\_contains=search) |

Q(wind\_way\_\_contains=search) |

Q(wind\_speed\_\_contains=search) |

Q(air\_pressure\_\_contains=search) |

Q(water\_pressure\_\_contains=search) |

Q(weather\_\_contains=search)

)

else:

self.\_clear\_forecast\_search = ''

self.\_clear\_forecast = ClearForecastMeteodata.objects.all()

self.\_updateData(4)

elif tableId == 5:

self.\_anomaly\_page = 1

if search != '':

self.\_anomaly\_search = search

self.\_anomaly = self.\_anomaly.filter(

Q(id\_\_contains=search) |

Q(meteodata\_id\_\_contains=search) |

Q(fieldname\_\_contains=search) |

Q(value\_\_contains=search) |

Q(anomaly\_\_contains=search)

)

else:

self.\_anomaly\_search = ''

self.\_anomaly = MeteodataAnomalies.objects.all()

self.\_updateData(5)

def get\_max\_pages(self, tableId):

print('get\_max\_pages')

print(type(self.\_meteodata))

size = 0

if tableId == 1:

size = self.\_meteodata.values('id').count()

elif tableId == 2:

size = self.\_clear\_meteodata.values('id').count()

elif tableId == 3:

size = self.\_forecast.values('id').count()

elif tableId == 4:

size = self.\_clear\_forecast.values('id').count()

elif tableId == 5:

size = self.\_anomaly.values('id').count()

val = int(size // self.\_rows\_count)

if size % self.\_rows\_count != 0:

val += 1

return val

def sort(self, tableId, dataName):

print('sort data')

print(type(self.\_meteodata))

if tableId == 1:

if self.\_firstOrderName == '' or self.\_firstOrderName != dataName:

self.\_meteodata = self.\_meteodata.order\_by(dataName)

self.\_firstOrderName = dataName

else:

self.\_meteodata = self.\_meteodata.order\_by(str('-' + self.\_firstOrderName))

elif tableId == 2:

if self.\_secondOrderName == '' or self.\_secondOrderName != dataName:

self.\_clear\_meteodata = self.\_clear\_meteodata.order\_by(dataName)

self.\_secondOrderName = dataName

else:

self.\_meteodata = self.\_meteodata.order\_by(str('-' + self.\_secondOrderName))

elif tableId == 3:

if self.\_thirdOrderName == '' or self.\_thirdOrderName != dataName:

self.\_forecast = self.\_forecast.order\_by(dataName)

self.\_thirdOrderName = dataName

else:

self.\_forecast = self.\_forecast.order\_by(str('-' + self.\_thirdOrderName))

elif tableId == 4:

if self.\_fothOrderName == '' or self.\_fothOrderName != dataName:

self.\_clear\_forecast = self.\_clear\_forecast.order\_by(dataName)

self.\_fothOrderName = dataName

else:

self.\_clear\_forecast = self.\_clear\_forecast.order\_by(str('-' + self.\_fothOrderName))

elif tableId == 5:

if self.\_fifthOrderName == '' or self.\_fifthOrderName != dataName:

self.\_anomaly = self.\_anomaly.order\_by(dataName)

self.\_fifthOrderName = dataName

else:

self.\_anomaly = self.\_anomaly.order\_by(str('-' + self.\_fifthOrderName))

def \_updateData(self, tableId):

print('updateData')

print(type(self.\_meteodata))

if tableId == 1:

self.\_meteodata = self.\_meteodata[(self.\_meteodata\_page \* self.\_rows\_count):(self.\_meteodata\_page \* self.\_rows\_count) + self.\_rows\_count]

elif tableId == 2:

self.\_clear\_meteodata = self.\_clear\_meteodata[(self.\_clear\_meteodata\_page \* self.\_rows\_count):(self.\_clear\_meteodata\_page \* self.\_rows\_count) + self.\_rows\_count]

elif tableId == 3:

self.\_forecast = self.\_forecast[(self.\_forecast\_page \* self.\_rows\_count):(self.\_forecast\_page \* self.\_rows\_count) + self.\_rows\_count]

elif tableId == 4:

self.\_clear\_forecast = self.\_clear\_forecast[(self.\_clear\_forecast\_page \* self.\_rows\_count):(self.\_clear\_forecast\_page \* self.\_rows\_count) + self.\_rows\_count]

elif tableId == 5:

self.\_anomaly = self.\_anomaly[(self.\_anomaly\_page \* self.\_rows\_count):(self.\_anomaly\_page \* self.\_rows\_count) + self.\_rows\_count]

print(type(self.\_meteodata))

from django.db import models

class Meteodata(models.Model):

id = models.BigAutoField("id", primary\_key=True)

datetime = models.DateTimeField("Дата и время")

place = models.BigIntegerField("Код места")

placeName = models.CharField("Место", max\_length=80)

temperature = models.FloatField("Температура воздуха")

wind\_way = models.BigIntegerField("Направление ветра")

wind\_speed = models.FloatField("Скорость ветра")

air\_pressure = models.FloatField("Атмосферное давление")

water\_pressure = models.FloatField("Давление моря")

weather = models.TextField("Погодное явление")

def \_\_str\_\_(self):

return self.id

class Meta:

verbose\_name = "Метеоданные"

verbose\_name\_plural = "Метеоданные"

class ClearMeteodata(models.Model):

id = models.BigAutoField("id", primary\_key=True)

datetime = models.DateTimeField("Дата и время")

place = models.BigIntegerField("Код места")

placeName = models.CharField("Место", max\_length=80)

temperature = models.FloatField("Температура воздуха")

wind\_way = models.BigIntegerField("Направление ветра")

wind\_speed = models.FloatField("Скорость ветра")

air\_pressure = models.FloatField("Атмосферное давление")

water\_pressure = models.FloatField("Давление моря")

weather = models.TextField("Погодное явление")

def \_\_str\_\_(self):

return self.id

class Meta:

verbose\_name = "Чистые метеоданные"

verbose\_name\_plural = "Чистые метеоданные"

class ForecastMeteodata(models.Model):

id = models.BigAutoField("id", primary\_key=True)

datetime = models.DateTimeField("Дата и время")

place = models.BigIntegerField("Код места")

placeName = models.CharField("Место", max\_length=80)

temperature = models.FloatField("Температура воздуха")

wind\_way = models.BigIntegerField("Направление ветра")

wind\_speed = models.FloatField("Скорость ветра")

air\_pressure = models.FloatField("Атмосферное давление")

water\_pressure = models.FloatField("Давление моря")

weather = models.TextField("Погодное явление")

def \_\_str\_\_(self):

return self.id

class Meta:

verbose\_name = "Спрогнозированные метеоданные"

verbose\_name\_plural = "Спрогнозированные метеоданные"

class ClearForecastMeteodata(models.Model):

id = models.BigAutoField("id", primary\_key=True)

datetime = models.DateTimeField("Дата и время")

place = models.BigIntegerField("Код места")

placeName = models.CharField("Место", max\_length=80)

temperature = models.FloatField("Температура воздуха")

wind\_way = models.BigIntegerField("Направление ветра")

wind\_speed = models.FloatField("Скорость ветра")

air\_pressure = models.FloatField("Атмосферное давление")

water\_pressure = models.FloatField("Давление моря")

weather = models.TextField("Погодное явление")

def \_\_str\_\_(self):

return self.id

class Meta:

verbose\_name = "Чистые спрогнозированные метеоданные"

verbose\_name\_plural = "Чистые спрогнозированные метеоданные"

class MeteodataAnomalies(models.Model):

id = models.BigAutoField("id", primary\_key=True)

meteodata\_id = models.BigIntegerField("Код записи")

fieldname = models.TextField("Название поля")

value = models.TextField("Значение")

anomaly = models.TextField("Аномалия")

class Meta:

verbose\_name = "Аномалии"

verbose\_name\_plural = "Аномалии"

def \_\_str\_\_(self):

return self.name

def get\_absolute\_url(self):

return reverse("\_detail", kwargs={"pk": self.pk})

import psycopg2

from datetime import date

class DBControl():

def \_\_init\_\_(self):

self.\_connect()

self.\_tables = ('meteodata', 'forecastmeteodata', 'clearmeteodata', 'clearforecastmeteodata', 'meteodataanomalies')

self.\_curTable = 0

self.\_lastQuere = ''

# removers

def removeMeteodata(self, id):

self.\_curs.execute('DELETE from public.meteodata\_meteodata WHERE id ={}'.format(str(id)))

def removeForecast(self, id):

self.\_curs.execute('DELETE from public.meteodata\_forecastmeteodata WHERE id ={}'.format(str(id)))

def removeClearMeteodata(self, id):

self.\_curs.execute('DELETE from public.meteodata\_clearmeteodata WHERE id ={}'.format(str(id)))

def removeClearForecast(self, id):

self.\_curs.execute('DELETE from public.meteodata\_clearforecastmeteodata WHERE id ={}'.format(str(id)))

def removeAnomalies(self, id):

self.\_curs.execute('DELETE from public.meteodata\_meteodataanomalies WHERE id ={}'.format(str(id)))

# inserters

def insertMeteodata(self, data):

\_id = int(self.getMaxId(self.\_tables[2]) + 1)

self.\_curs.execute("INSERT INTO public.meteodata\_meteodata VALUES(DEFAULT, '" + data[0] + "', " + data[1] + ", '" + data[2] + "', " + data[3] + ", " + data[4] + ", " + data[5] + ", " + data[6] + ", " + data[7] + ", '{" + data[8] + "}')")

def insertForecast(self, data):

\_id = int(self.getMaxId(self.\_tables[3]) + 1)

self.\_curs.execute("INSERT INTO public.meteodata\_forecastmeteodata VALUES(DEFAULT, '" + data[0] + "', " + data[1] + ", '" + data[2] + "', " + data[3] + ", " + data[4] + ", " + data[5] + ", " + data[6] + ", " + data[7] + ", '{" + data[8] + "}')")

def insertClearMeteodata(self, data):

\_id = int(self.getMaxId(self.\_tables[2]) + 1)

self.\_curs.execute("INSERT INTO public.meteodata\_clearmeteodata VALUES(DEFAULT, '" + data[0] + "', " + data[1] + ", '" + data[2] + "', " + data[3] + ", " + data[4] + ", " + data[5] + ", " + data[6] + ", " + data[7] + ", '{" + data[8] + "}')")

def insertClearForecast(self, data):

\_id = int(self.getMaxId(self.\_tables[3]) + 1)

self.\_curs.execute("INSERT INTO public.meteodata\_clearforecastmeteodata VALUES(DEFAULT, '" + data[0] + "', " + data[1] + ", '" + data[2] + "', " + data[3] + ", " + data[4] + ", " + data[5] + ", " + data[6] + ", " + data[7] + ", '{" + data[8] + "}')")

def insertAnomalies(self, data):

\_id = int(self.getMaxId(self.\_tables[0]) + 1)

self.\_curs.execute("INSERT INTO public.meteodata\_meteodataanomalies VALUES(DEFAULT," + data[0] + ", '{" + data[1] + "}', '{" + data[2] + "}', '{" + data[3] + "}')")

# updaters

def updateMeteodata(self, data):

self.\_curs.execute("UPDATE public.meteodata\_meteodata SET datetime='" + data[0] + "', place=" + str(data[1]) + ", \"placeName\"" + data[2] + "', temperature=" + str(data[3]) + ", wind\_way=" + str(data[4]) + ", wind\_speed=" + str(data[5]) + ", air\_pressure=" + str(data[6]) + ", water\_pressure=" + str(data[7]) + ", weather='{" + data[8] + "}' WHERE id=" + str(data[0]))

def updateForecast(self, data):

self.\_curs.execute("UPDATE public.meteodata\_forecastmeteodata SET datetime='" + data[0] + "', place=" + str(data[1]) + ", \"placeName\"" + data[2] + "', temperature=" + str(data[3]) + ", wind\_way=" + str(data[4]) + ", wind\_speed=" + str(data[5]) + ", air\_pressure=" + str(data[6]) + ", water\_pressure=" + str(data[7]) + ", weather='{" + data[8] + "}' WHERE id=" + str(data[0]))

def updateClearMeteodata(self, data):

self.\_curs.execute("UPDATE public.meteodata\_clearmeteodata SET datetime='" + data[0] + "', place=" + str(data[1]) + ", \"placeName\"" + data[2] + "', temperature=" + str(data[3]) + ", wind\_way=" + str(data[4]) + ", wind\_speed=" + str(data[5]) + ", air\_pressure=" + str(data[6]) + ", water\_pressure=" + str(data[7]) + ", weather='{" + data[8] + "}' WHERE id=" + str(data[0]))

def updateClearForecast(self, data):

self.\_curs.execute("UPDATE public.meteodata\_clearforecastmeteodata SET datetime='" + data[0] + "', place=" + str(data[1]) + ", \"placeName\"" + data[2] + "', temperature=" + str(data[3]) + ", wind\_way=" + str(data[4]) + ", wind\_speed=" + str(data[5]) + ", air\_pressure=" + str(data[6]) + ", water\_pressure=" + str(data[7]) + ", weather='{" + data[8] + "}' WHERE id=" + str(data[0]))

def updateAnomalies(self, data):

self.\_curs.execute("UPDATE public.meteodata\_meteodataanomalies SET meteodata\_id=" + str(data[1]) + ", fieldname='{" + data[2] + "}', value='{" + data[3] + "}', anomaly='{" + data[4] + "}' WHERE id=" + str(data[0]))

# getters

def getMeteodata(self, filter='id, datetime, place, \"placeName\", temperature, wind\_way, wind\_speed, air\_pressure, water\_pressure, weather', where=''):

self.\_curTable = 0

tmp = 'SELECT {} FROM public.meteodata\_meteodata {}'.format(filter, where)

self.\_lastQuere = tmp

self.\_curs.execute(tmp)

records = self.\_refactorRecords(self.\_curs.fetchall())

return records

def getForecast(self, filter='id, datetime, place, \"placeName\", temperature, wind\_way, wind\_speed, air\_pressure, water\_pressure, weather', where=''):

self.\_curTable = 1

tmp = 'SELECT {} FROM public.meteodata\_forecastmeteodata {}'.format(filter, where)

self.\_lastQuere = tmp

self.\_curs.execute(tmp)

records = self.\_refactorRecords(self.\_curs.fetchall())

return records

def getClearMeteodata(self, filter='id, datetime, place, \"placeName\", temperature, wind\_way, wind\_speed, air\_pressure, water\_pressure, weather', where=''):

self.\_curTable = 3

tmp = 'SELECT {} FROM public.meteodata\_clearmeteodata {}'.format(filter, where)

self.\_lastQuere = tmp

self.\_curs.execute(tmp)

records = self.\_refactorRecords(self.\_curs.fetchall())

return records

def getClearForecast(self, filter='id, datetime, place, \"placeName\", temperature, wind\_way, wind\_speed, air\_pressure, water\_pressure, weather', where=''):

self.\_curTable = 4

tmp = 'SELECT {} FROM public.meteodata\_clearforecastmeteodata {}'.format(filter, where)

self.\_lastQuere = tmp

self.\_curs.execute(tmp)

records = self.\_refactorRecords(self.\_curs.fetchall())

return records

def getAnomalies(self, filter='id, meteodata\_id, fieldname, value, anomaly', where=''):

self.\_curTable = 2

tmp = 'SELECT {} FROM public.meteodata\_meteodataanomalies {}'.format(filter, where)

self.\_lastQuere = tmp

self.\_curs.execute(tmp)

records = self.\_refactorRecords(self.\_curs.fetchall())

return records

# Additional functions

def getTableNames(self):

return self.\_tables

def getMaxId(self, table):

if table in self.\_tables:

self.\_curs.execute('SELECT MAX(id) FROM public.{}'.format(table))

elif isinstance(table, int):

if table >= 0 and table < len(self.\_tables):

self.\_curs.execute('SELECT MAX(id) FROM public.{}'.format(self.\_tables[table]))

else:

return 0

records = self.\_curs.fetchall()

if records[0][0] is None:

return 1

else:

return int(records[0][0])

def getHederLabels(self):

if self.\_curTable == 0:

return ['№', 'Дата и время', 'Код места', 'Название места', 'Температура', 'Направление ветра', 'Скорость ветра', 'Давление воздуха', 'Давление воды', 'Погодные явления']# meteodata\_meteodataanomalies

elif self.\_curTable == 1:

return ['№', 'Дата и время', 'Код места', 'Название места', 'Температура', 'Направление ветра', 'Скорость ветра', 'Давление воздуха', 'Давление воды', 'Погодные явления']# payment\_table

elif self.\_curTable == 2:

return ['№', '№ метеоданных', 'Имя поля', 'Значение', 'Описание аномалии']# MeteodataTable

def \_refactorRecords(self, records):

result = list()

for row in records:

result\_row = list()

for element in row:

if isinstance(element, datetime):

result\_row.append(str(element))

elif isinstance(element, str):

if element.isdigit():

result\_row.append(float(element))

else:

if element[0] == '{':

result\_row.append(element[1: -1])

else:

result\_row.append(element)

elif isinstance(element, list):

result\_row.append(element[0])

elif isinstance(element, int):

result\_row.append(element)

elif isinstance(element, float):

result\_row.append(float(element))

result.append(result\_row)

return result

def save(self):

self.\_conn.commit()

def \_connect(self):

self.\_conn = psycopg2.connect(dbname='Weather', user='postgres', password='12345', host='localhost')

self.\_curs = self.\_conn.cursor()

def \_\_del\_\_(self):

self.\_conn.commit()

self.\_curs.close()

self.\_conn.close()

import requests

from bs4 import BeautifulSoup as bs

from bs4 import NavigableString

from calendar import monthrange

import psycopg2

import time

import re

from datetime import date

class MyParser:

def \_\_init\_\_(self):

self.\_rawData = []

self.\_urls = ['https://meteopost.com/weather/archive/']

def StartParse(self):

citySet, data = self.raw\_update1()

return self.prepare\_data1(citySet, data)

def raw\_update1(self, url='https://meteopost.com/weather/archive/', startParseRequest='', endParseRequest=''):

resultList = []

monthSet = {'value': 0, 'text': 0}

yearSet = {'value': 0, 'text': 0}

citySet = {'value': 0, 'text': 0}

badCitySet = {'value': 0, 'text': 0}

# get city

r = requests.get(url)

html = bs(r.content, 'html.parser')

badCitySet['value'], badCitySet['text'] = self.\_getSetData(name='city', html=html, part=0)

citySet['value'], citySet['text'] = self.\_getSetData(name='city', html=html, part=1)

# get month and year

yearSet['value'], yearSet['text'] = self.\_getSetData(name='y', html=html, part=0)

monthSet['value'], monthSet['text'] = self.\_getSetData(name='m', html=html, part=0)

r.close()

startParseFlag = False

for year in yearSet['value']:

y = year

for month in monthSet['value']:

m = month

days = str(monthrange(int(y), int(m))[1])

for city in citySet['value']:

city = city

req = 'd=1&m={}&y={}&city={}&arc=2&days={}'.format(m, y, city, days)

print(req)

hd = {

"Host": "meteopost.com",

"User-Agent": "Mozilla/5.0 (X11; Ubuntu; Linux x86\_64; rv:84.0) Gecko/20100101 Firefox/84.0",

"Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,\*/\*;q=0.8",

"Accept-Language": "en-US,en;q=0.5",

"Accept-Encoding": "gzip, deflate, br",

"Content-Type": "application/x-www-form-urlencoded",

"Content-Length": "39",

"Origin": "https://meteopost.com",

"DNT": "1",

"Connection": "keep-alive",

"Referer": "https://meteopost.com/weather/archive/",

"Cookie": "\_ga=GA1.2.1952077172.1608990993; \_gid=GA1.2.2059205127.1609852737; \_gat=1",

"Upgrade-Insecure-Requests": "1",

"TE": "Trailers"

}

if req == startParseRequest or startParseRequest == '':

startParseFlag = True

if not startParseFlag:

continue

if req == endParseRequest:

return

time.sleep(0.15) # not to ddos timer)

r = requests.post(url, data=req.encode(), headers=hd, stream=True)

html = bs(r.content, 'html5lib')

items = html.find\_all("table")

html = items[3]

items = html.find\_all("tr")

for i in range(len(items)):

s = ''

tr = items[i]

if len(tr) > 5:

for child in tr:

x = child.center

try:

img = x.img

num = 0

if img['src'] == '/pic/180.png':

num = 0

elif img['src'] == '/pic/225.png':

num = 1

elif img['src'] == '/pic/270.png':

num = 2

elif img['src'] == '/pic/360.png':

num = 3

elif img['src'] == '/pic/0.png':

num = 4

elif img['src'] == '/pic/45.png':

num = 5

elif img['src'] == '/pic/90.png':

num = 6

elif img['src'] == '/pic/135.png':

num = 7

elif img['src'] == '/pic/00.png':

num = 8

s = s + '\_' + str(num)

except:

try:

b = x.b

s = s + '\_' + b.string

except:

# set \_ symbol for delimiter

if not isinstance(x, type(None)):

s = s + '\_' + x.get\_text()

else:

s = s + '\_=' # if data missing = symbol is placeholder

print('SSS - ', s)

resultList.append(s + '&&' + req + '\n')

r.close()

return citySet, resultList

def \_getSetData(self, name='', html='', part=0):

items = html.find\_all(attrs={'name': name})

temp1 = []

temp2 = []

for child in items[part].children:

if not isinstance(child, NavigableString):

temp1.append(child['value'])

temp2.append(child.string)

return temp1, temp2

def prepare\_data1(self, citySet, data):

resultList = []

r = re.compile('^days=[0-9](2)[\n]$')

i = 0

while True:

line = data[i]

nextLine = data[i+1]

temp = str(line[-1:-9:-1])

if temp[::-1] not in ('days=28\n', 'days=29\n', 'days=30\n', 'days=31\n'):

print(list(temp[::-1]))

data[i] = data[i].rstrip() + nextLine

data.pop(i+1)

i = i + 1

if i + 1 == len(data):

break

i = 0

headers = []

while True:

line = data[i]

baseList = line.split('&&')

if line[0] == '&':

print(baseList[1])

i = i + 1

continue

dataList = baseList[0].split('\_')

querryList = baseList[1].split('&')

if dataList[0] == '':

dataList.pop(0)

if line[:3] == '\_Де':

headers = dataList

i = i + 1

continue

d = dict()

for j in range(len(dataList)):

if dataList[j] != '=':

d.update({headers[j]: dataList[j]})

else:

d.update({headers[j]: 0})

for item in querryList:

temp = item.split('=')

d.update({temp[0]: temp[1]})

if isinstance(d['Темп. Возд'], str):

d['Темп. Возд'] = d['Темп. Возд'][:-1]

if isinstance(d['Скор ветра'], str):

d['Скор ветра'] = d['Скор ветра'].split(' ')

d['Скор ветра'] = d['Скор ветра'][0]

if d['День'] == '0':

d['День'] = '26'

xx = d['Час'].split(':')

if int(xx[0]) > 24:

d['Час'] = '12:00'

resultList.append(d)

i = i + 1

#11-3

#4-10

if i == len(data):

break

return citySet, resultList

from .Deamon.ForecastDeamon import ForecastDeamon

from .MeteodataMiner import MeteodataMiner

from .DataReader import DataReader

class MainMenu:

Deamon = ForecastDeamon()

Miner = MeteodataMiner()

Reader = DataReader()

currentContext = 1

def get\_context():

if MainMenu.currentContext == 1:

context = {

"meteodata\_page": MainMenu.Reader.\_meteodata\_page,

"meteodata\_search": MainMenu.Reader.\_meteodata\_search,

"meteodata\_max\_pages": MainMenu.get\_max\_pages(1),

"meteodata\_top\_labels": MainMenu.get\_top\_labels(),

"meteodata": MainMenu.Reader.\_meteodata,

"clear\_meteodata\_page": MainMenu.Reader.\_clear\_meteodata\_page,

"clear\_meteodata\_search": MainMenu.Reader.\_clear\_meteodata\_search,

"clear\_meteodata\_max\_pages": MainMenu.get\_max\_pages(2),

"clear\_meteodata": MainMenu.Reader.\_clear\_meteodata

}

elif MainMenu.currentContext == 2:

context = {

"forecast": MainMenu.Reader.\_forecast,

"forecast\_page": MainMenu.Reader.\_forecast\_page,

"forecast\_search": MainMenu.Reader.\_forecast\_search,

"forecast\_max\_pages": MainMenu.get\_max\_pages(3),

"forecast\_top\_labels": MainMenu.get\_top\_labels(),

"clear\_forecast": MainMenu.Reader.\_clear\_forecast,

"clear\_forecast\_page": MainMenu.Reader.\_clear\_forecast\_page,

"clear\_forecast\_search": MainMenu.Reader.\_clear\_forecast\_search,

"clear\_forecast\_max\_pages": MainMenu.get\_max\_pages(4),

}

elif MainMenu.currentContext == 3:

context = {

"anomaly": MainMenu.Reader.\_anomaly,

"anomaly\_page": MainMenu.Reader.\_anomaly\_page,

"anomaly\_search": MainMenu.Reader.\_anomaly\_search,

"anomaly\_max\_pages": MainMenu.get\_max\_pages(5),

"anomaly\_top\_labels": MainMenu.get\_top\_labels()

}

context.update(

{

"rows\_count": MainMenu.Reader.\_rows\_count,

"menu\_status": MainMenu.get\_menu\_status(),

"submenu\_status": MainMenu.get\_submenu\_status()

}

)

return context

def set\_page(num, tableId):

MainMenu.Reader.set\_page(num, tableId)

def search(search):

MainMenu.Reader.search(search)

def set\_context(num):

if num >= 1 and num <= 3:

MainMenu.currentContext = num

def data\_update():

print('data\_update')

MainMenu.Miner.updateMeteodata()

def forecast\_update():

print('forecast\_update')

MainMenu.Deamon.update()

def anomaly\_update():

print('anomaly\_update')

MainMenu.Deamon.scanForAnomalies()

def get\_menu\_status():

print('get\_menu\_status')

if MainMenu.currentContext == 1:

return ['button-active', 'button', 'button']

elif MainMenu.currentContext == 2:

return ['button', 'button-active', 'button']

elif MainMenu.currentContext == 3:

return ['button', 'button', 'button-active']

def get\_submenu\_status():

print('get\_submenu\_status')

if MainMenu.currentContext == 1:

if MainMenu.Miner.meteodataUpdateStatus():

return 'hide'

else:

return 'active'

elif MainMenu.currentContext == 2:

if MainMenu.Deamon.forecastUpdateStatus():

return 'hide'

else:

return 'active'

elif MainMenu.currentContext == 3:

if MainMenu.Deamon.anomaliesScanStatus():

return 'hide'

else:

return 'active'

def get\_top\_labels():

print('get\_top\_labels')

if MainMenu.currentContext == 1:

return ['№', 'Дата и время', 'Код места', 'Название места', 'Температура', 'Направление ветра', 'Скорость ветра', 'Давление воздуха', 'Давление воды', 'Погодные явления']

elif MainMenu.currentContext == 2:

return ['№', 'Дата и время', 'Код места', 'Название места', 'Температура', 'Направление ветра', 'Скорость ветра', 'Давление воздуха', 'Давление воды', 'Погодные явления']

elif MainMenu.currentContext == 3:

return ['№', '№ метеоданных', 'Имя поля', 'Значение', 'Описание аномалии']

def get\_max\_pages(tableId):

return MainMenu.Reader.get\_max\_pages(tableId)

def \_make\_forecast\_test():

print('make\_forecast\_test')

MainMenu.Deamon.\_make\_test()

def sort(dataName):

print('sort data')

MainMenu.Reader.sort(MainMenu.currentContext, dataName)

from django.shortcuts import render, redirect

from django.views.generic import ListView

from django.views.generic.base import View

from .models import Meteodata, ForecastMeteodata, MeteodataAnomalies

from django.db.models import Q

from .moduls import MainMenu

class MeteodataView(ListView):

def get(self, request):

MainMenu.set\_context(1)

context = MainMenu.get\_context()

return render(request, "meteodata/meteodataview.html", context=context)

class MeteodataChangePage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(1)

MainMenu.set\_page(int(d['page']))

MainMenu.search(d['search'])

try:

context = MainMenu.get\_context()

return render(request, "meteodata/meteodataview.html", context=context)

except:

return redirect('/meteodata/')

class MeteodataUpdateDataPage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(1)

if not MainMenu.\_is\_data\_updated:

MainMenu.\_data\_update()

return redirect('/meteodata/')

class ForecastView(ListView):

def get(self, request):

MainMenu.set\_context(2)

context = MainMenu.get\_context()

return render(request, "meteodata/forecastview.html", context=context)

class ForecastChangePage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(2)

MainMenu.set\_page(int(d['page']))

MainMenu.search(d['search'])

try:

context = MainMenu.get\_context()

return render(request, "meteodata/forecastview.html", context=context)

except:

return redirect('meteodata/forecast/')

class ForecastUpdateDataPage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(2)

if not MainMenu.\_is\_forecast\_updated:

MainMenu.\_make\_forecast()

return redirect('meteodata/forecast/')

class ForecastUpdateTestResultPage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(2)

MainMenu.\_make\_forecast\_test()

return redirect('meteodata/forecast/')

class AnomalyView(ListView):

def get(self, request):

MainMenu.set\_context(3)

context = MainMenu.get\_context()

return render(request, "meteodata/analysisview.html", context=context)

class AnomalyChangePage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(3)

MainMenu.set\_page(int(d['page']))

MainMenu.search(d['search'])

try:

context = MainMenu.get\_context()

return render(request, "meteodata/analysisview.html", context=context)

except:

return redirect('meteodata/anoomaly/')

class AnomalyUpdateDataPage(View):

def post(self, request):

d = request.POST

print(d)

MainMenu.set\_context(3)

if not MainMenu.is\_anomaly\_updated:

MainMenu.\_make\_anomaly()

return redirect('meteodata/anoomaly/')

# ДОДАТОК Б

Слайди презентації



Рисунок Б1 – Титульний слайд презентації



Рисунок Б2 – Слайд з визначенням предметної області

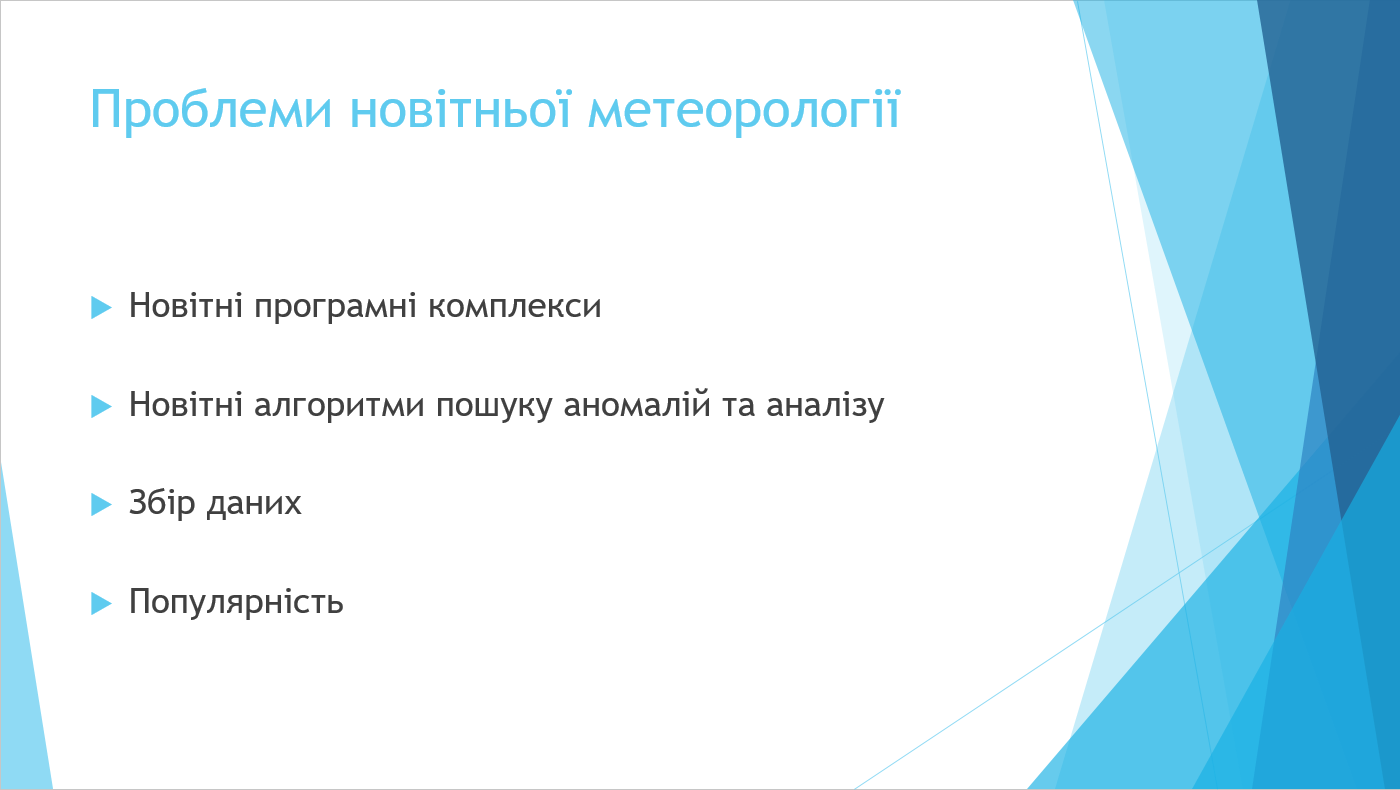


Рисунок Б3 – Слайд з визначеними проблемами предметної області

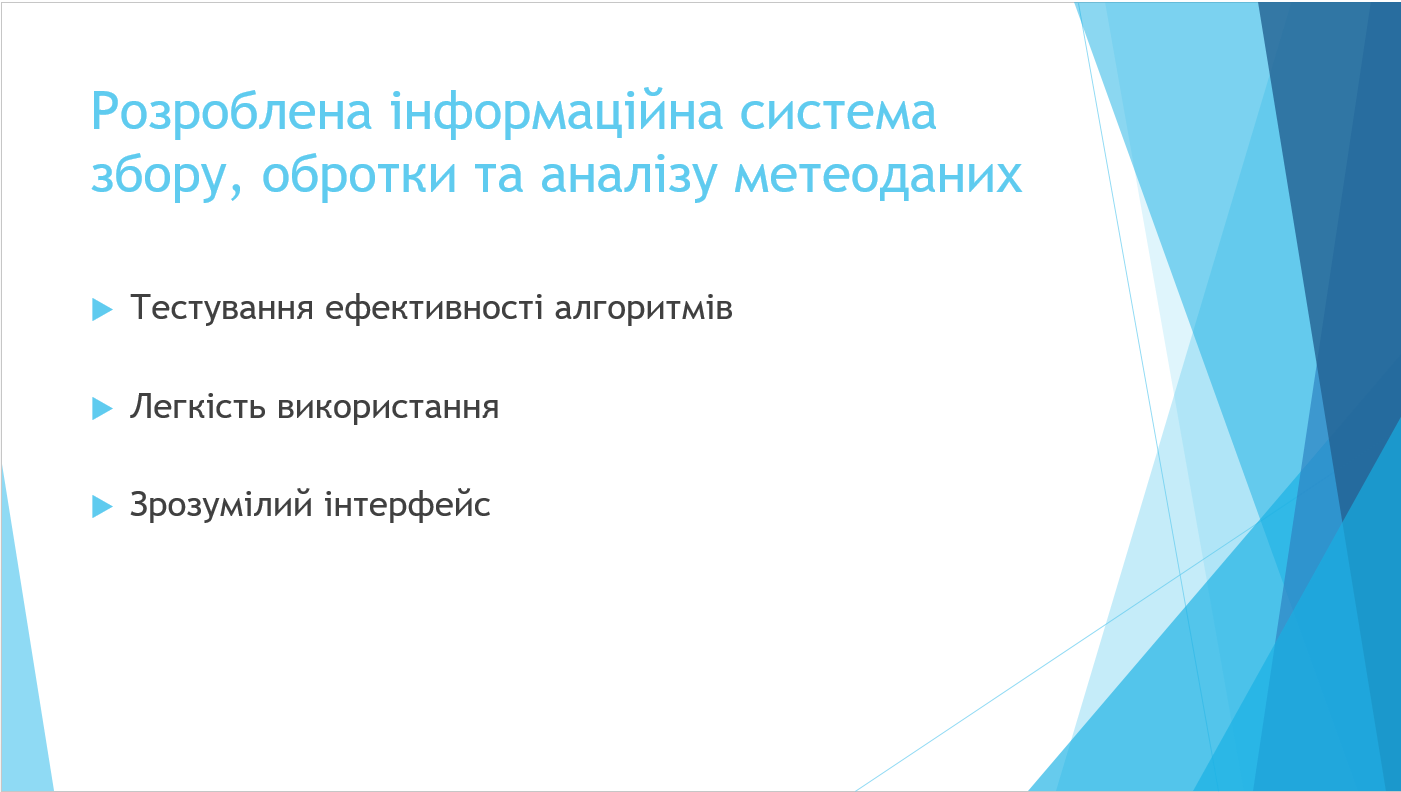


Рисунок Б4 = Слайд з описом функціоналу розробленої системи

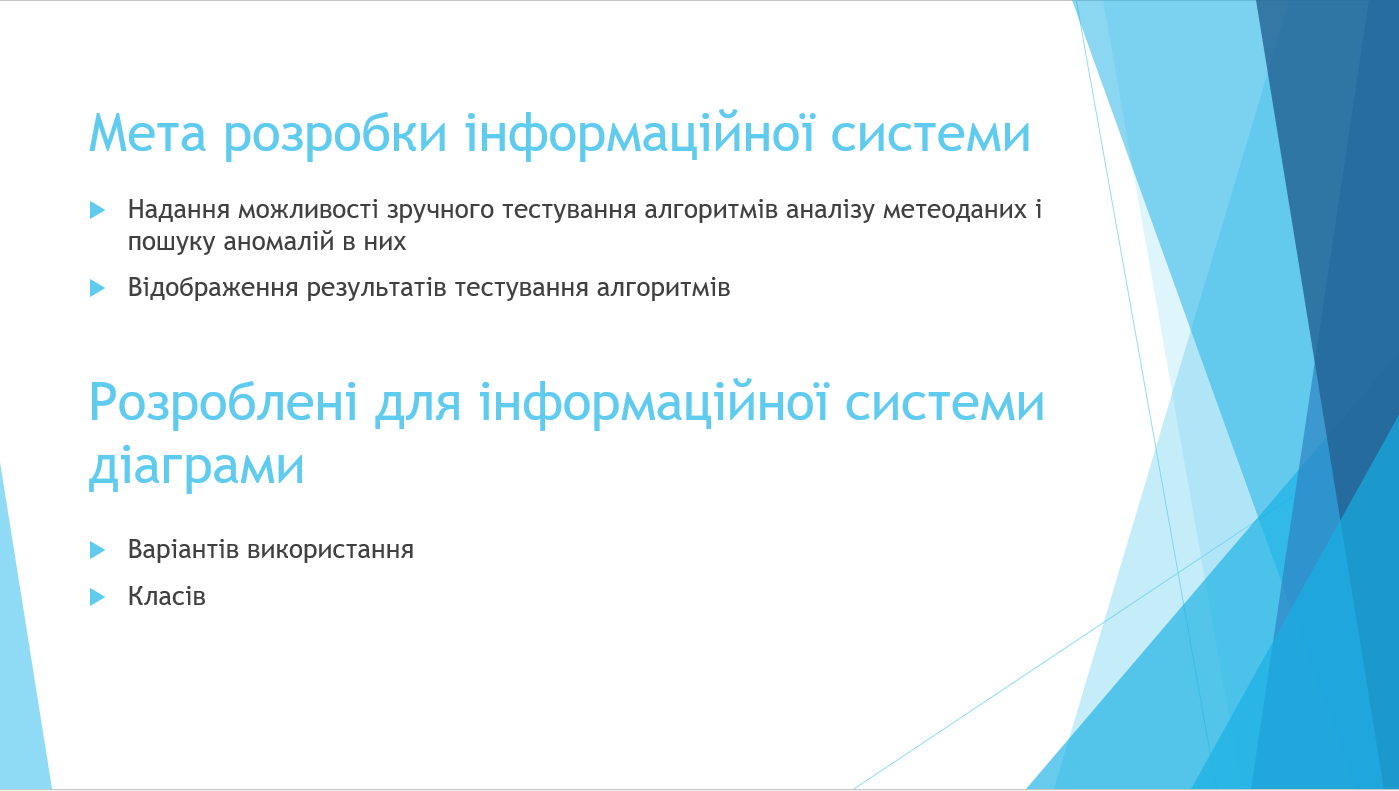


Рисунок Б5 – Слайд описом мети та описом розроблених діаграм

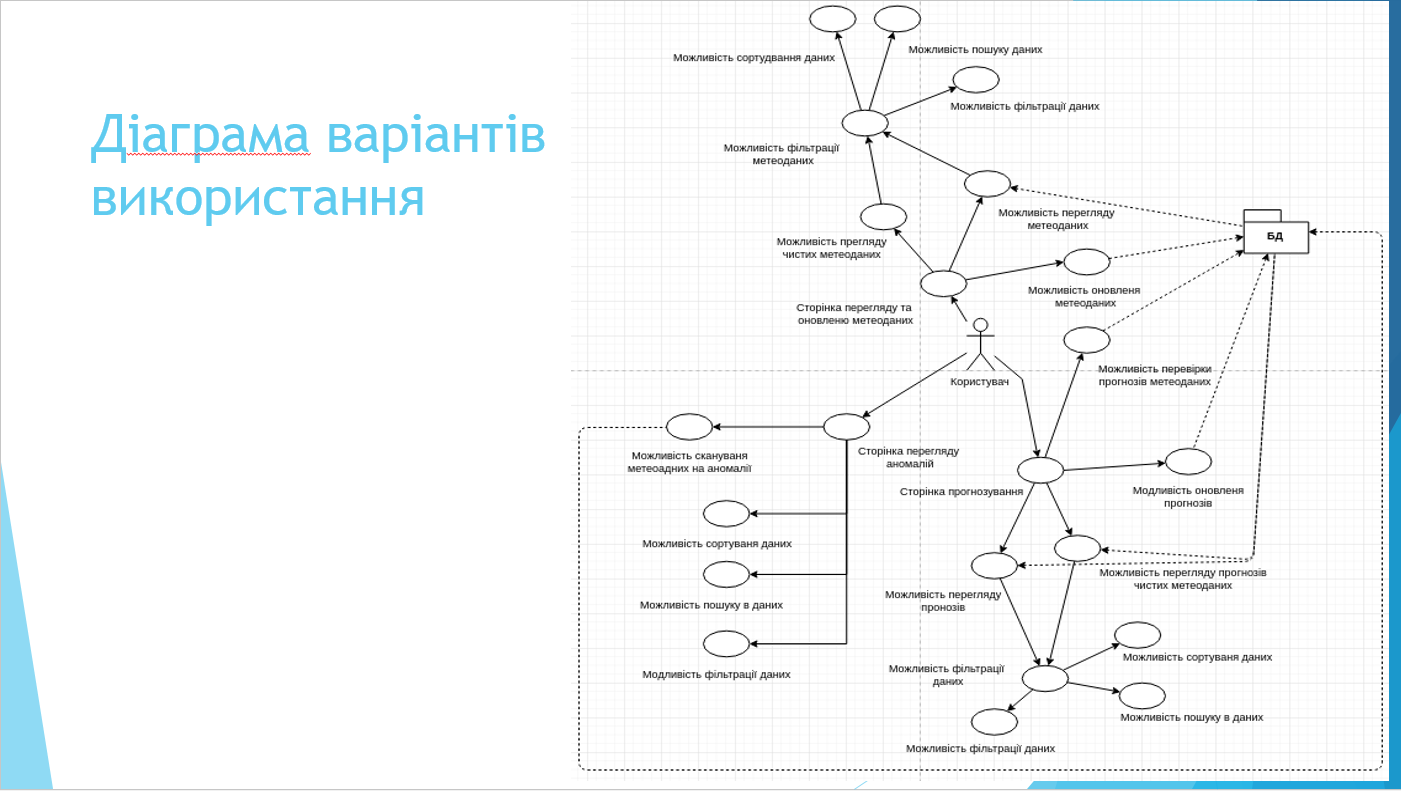


Рисунок Б6 – Слайд з діаграмою варіантів використання

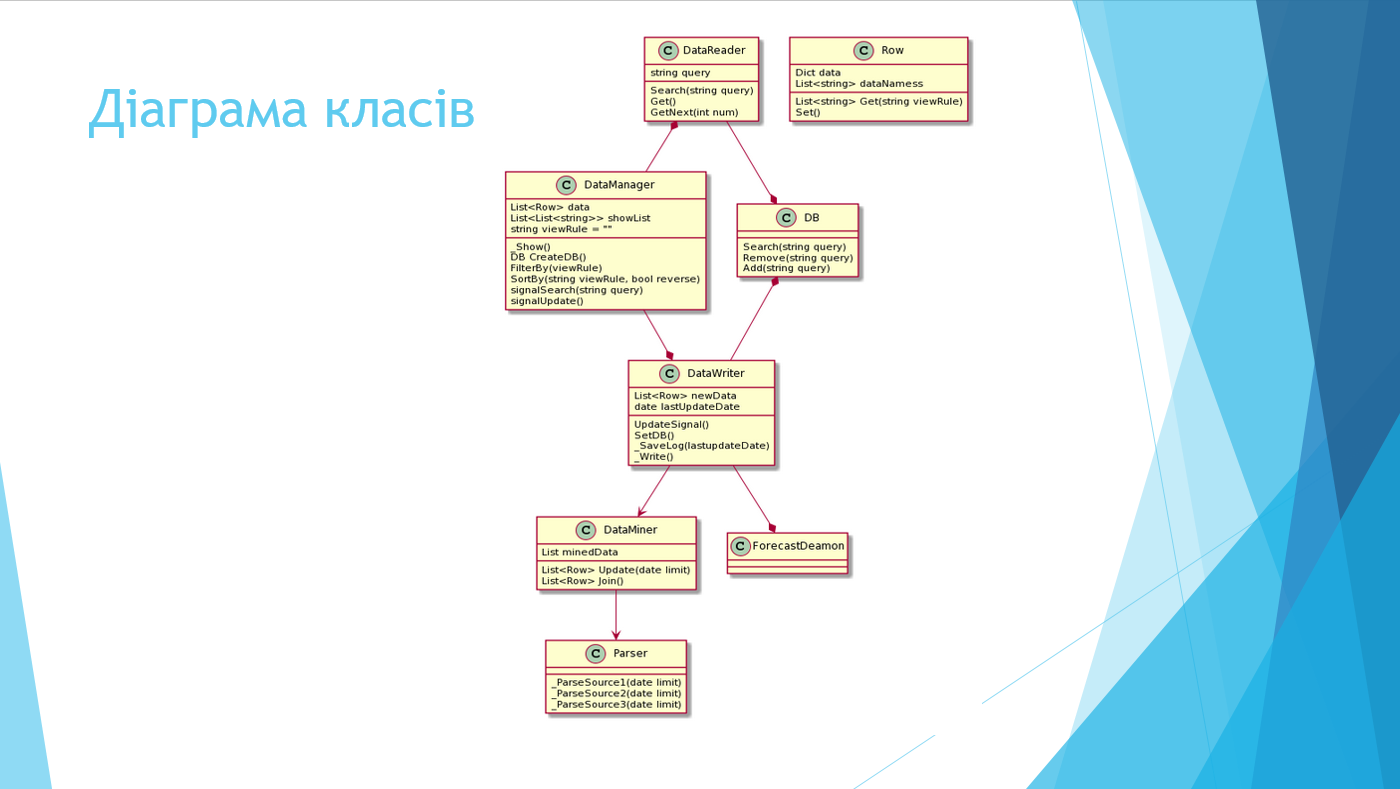


Рисунок Б7 – Слайд з діаграмою класів

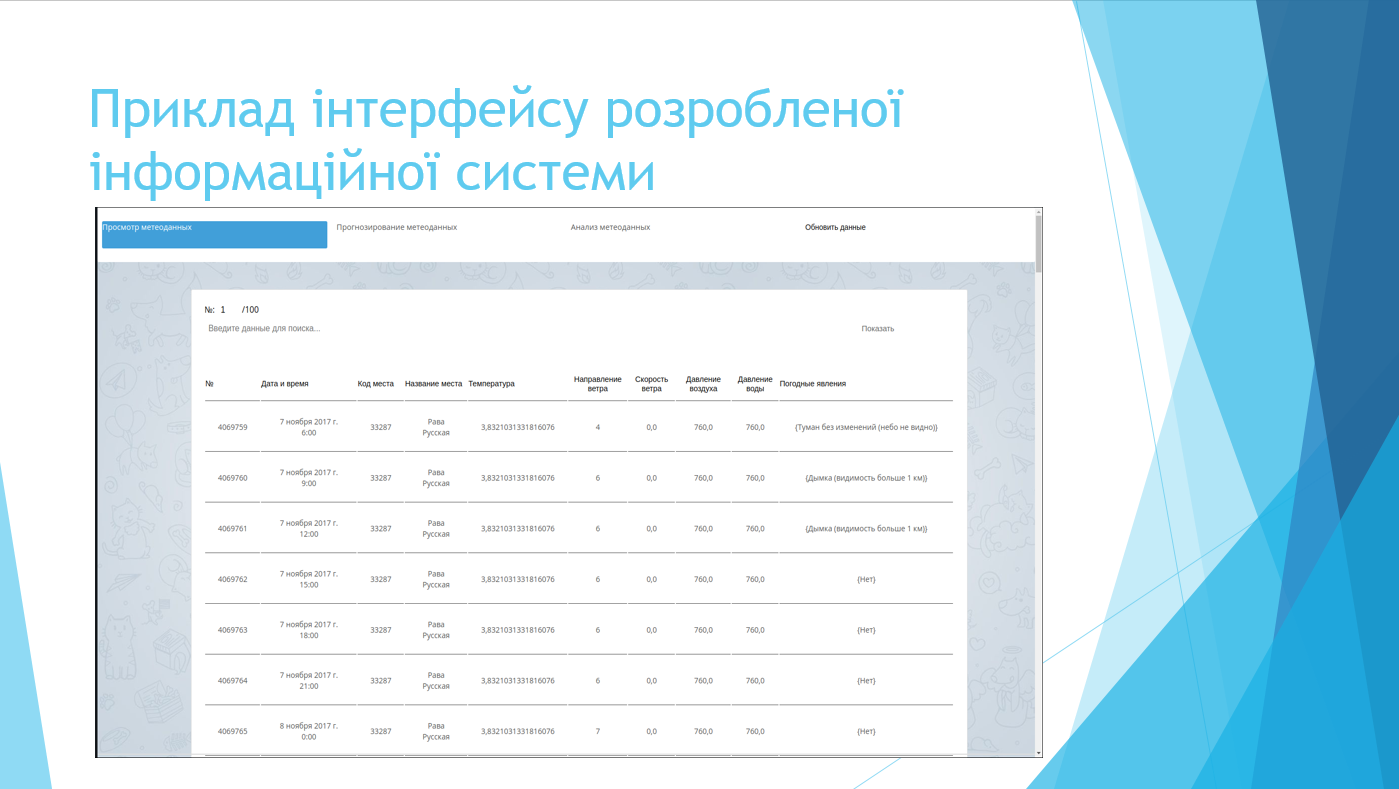


Рисунок Б8 – Слайд з прикладом інтерфейсу дороти

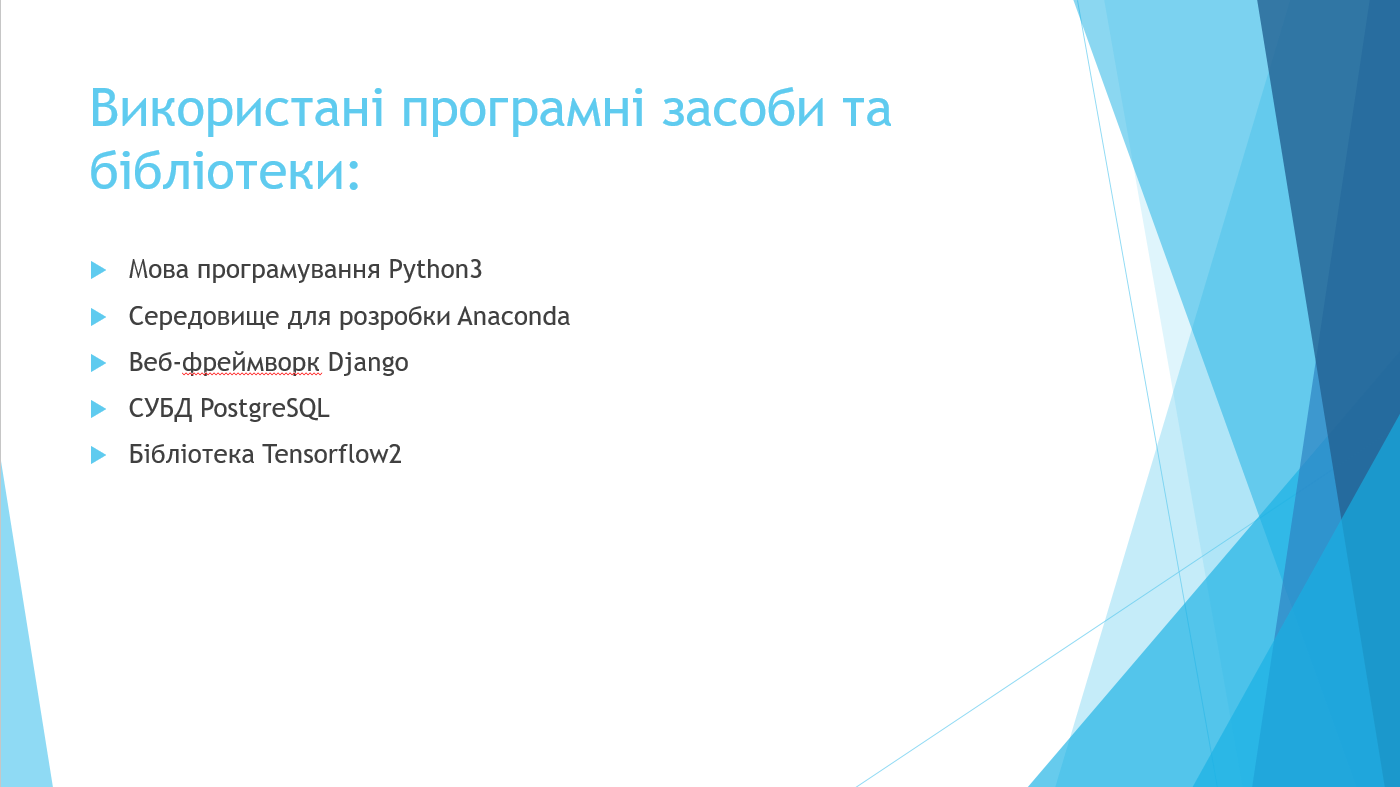


Рисунок Б9 – Слайд з визначенням засобів реалізації роботи

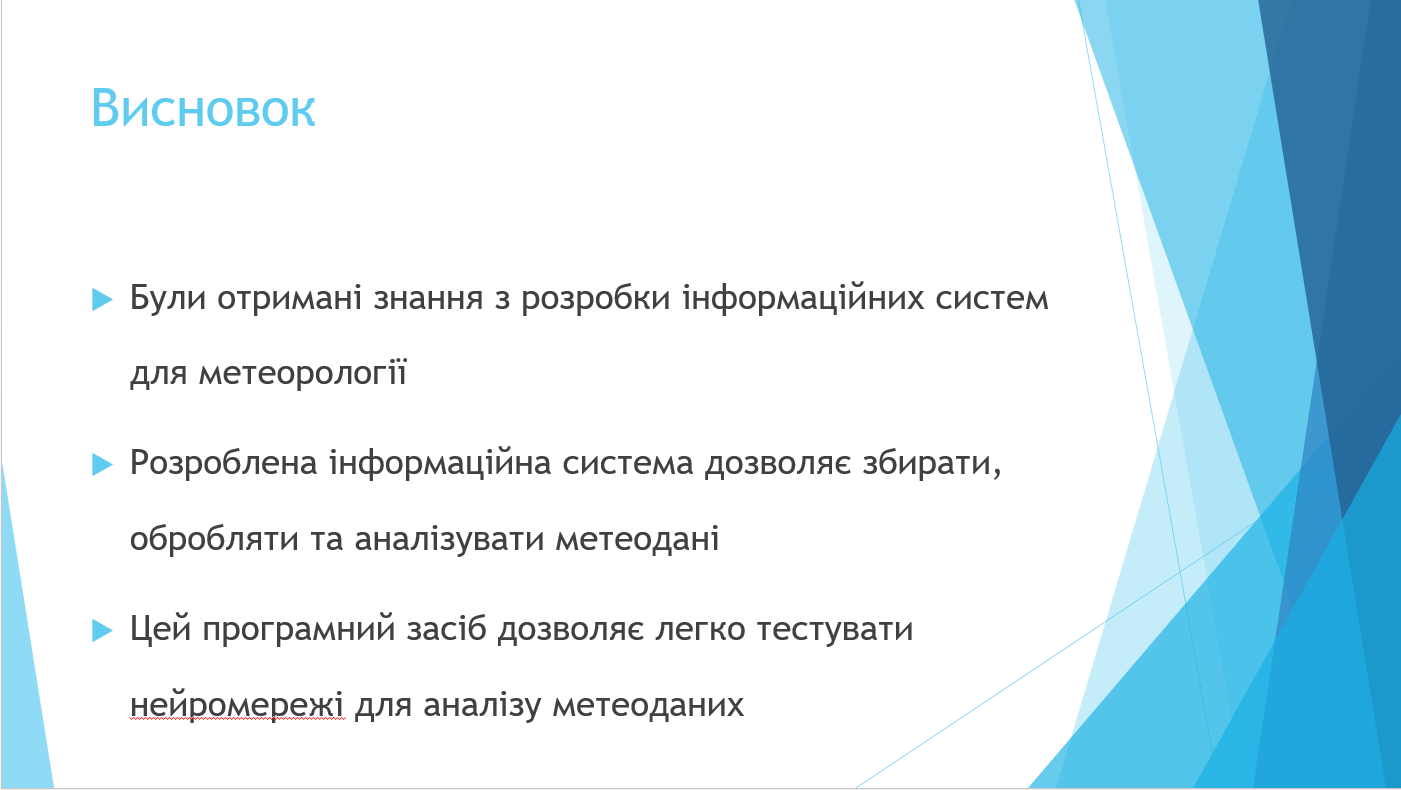


Рисунок Б10 – Слайд з загальними висновками



Рисунок Б11 – Слайд з висновками до роботи

ДОДАТОК В

Електронні матеріали

|  |  |  |
| --- | --- | --- |
| Зміст | Папка | Ім'я файла |
| Пояснювальна записка до роботи бакалавра |  | diploma.docх |
| Вихідні дані до роботи бакалавра |  | readme.txt |
| Докладний проект програми з кодами і поясненнями у середовищі проектування програмного продукту |  | project.zip |
| Демонстраційний ролик програмної системи (слайди, анімація тощо) |  | demo.pptх |