

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО–НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра Інженерії програмного забезпечення

Пояснювальна записка

до магістерської роботи
на ступінь вищої освіти магістр

на тему: «РОЗРОБКА МЕТОДУ ПРОГНОЗУВАННЯ ПОГОДИ З
ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ»

Виконав: студент 6 курсу, групи ПДМ–62

спеціальності 121 Інженерія програмного забезпечення
(шифр і назва спеціальності/спеціалізації)

Івченка Д.Є.
(прізвище та ініціали)

Керівник Трінтіна Н.А.
(прізвище та ініціали)

Рецензент _____
(прізвище та ініціали)

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ

Навчально–науковий інститут інформаційних технологій

Кафедра Інженерії програмного забезпечення

Ступінь вищої освіти – «Магістр»

Спеціальність підготовки – 121 «Інженерія програмного забезпечення»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Інженерії програмного забезпечення

Негоденко О.В.

“ ___ ” _____ 2022 року

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТА

Івченко Дмитро Євгенович

(прізвище, ім'я, по батькові)

1. Тема роботи: «Розробка методу прогнозування погоди з використанням штучного інтелекту»

Керівник роботи Трінтіна Наталія Альбертівна, доцент кафедри, к.т.н.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «12» жовтня 2022 року № 122.

2. Строк подання студентом роботи 31.12.2022

3. Вихідні дані до роботи:

3.1 Матеріали переддипломної практики, методи моделювання прогнозу погоди та погодних явищ;

3.2 Науково-технічна література з питань, пов'язаних з алгоритмами та методами прогнозування;

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно

розробити):

- 4.1 Огляд предметної області;
- 4.2 Опис моделювання методу;
- 4.3 Розробка методу прогнозування;
- 5. Перелік графічного матеріалу (презентація)
 - 5.1 Мета, об'єкт та предмет дослідження;
 - 5.2 Порівняльна характеристика існуючих моделей
 - 5.3 Принцип створення методу
 - 5.4 Практичний результат
 - 5.5 Покращення результату
 - 5.6 Висновки
- 6. Дата видачі завдання «14» жовтня 2022 року

КАЛЕНДАРНИЙ ПЛАН

№ з /п	Назва етапів бакалаврської роботи	Строк виконання етапів роботи	Примітка
1	Підбір науково-технічної літератури	14.10.2022	
2	Огляд предметної області	24.10.2022	
3	Створення моделі для прогнозування погоди	27.10.2022	
4	Аналіз методів прогнозування	15.11.2022	
5	Розробка моделей та методів	20.11.2022	
6	Розробка програмного забезпечення методу	23.11.2022	
7	Порівняльний аналіз результатів	25.11.2022	
8	Написання та оформлення пояснювальної записки	27.11.2022	
9	Розробка графічних та презентаційних матеріалів	11.12.2022	
10	Захист магістрескої роботи	17.01.2023	

Студент

Керівник роботи

_____ Івченко Д.Є
(підпис) (прізвище та ініціали)
_____ Трінтіна Н.А.
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Текстова частина магістерської роботи: 71 с., 35 рис., 2 дод., 55 джерел.

МЕТОД, ШТУЧНИЙ ІНТЕЛЕКТ, МАШИННЕ НАВЧАННЯ, ПРОГНОЗУВАННЯ, МОДЕЛЮВАННЯ МЕТОДУ.

Об'єкт дослідження - процес прогнозування погоди.

Мета дослідження - вдосконалення методу прогнозування погоди та погодних явищ з використанням штучного інтелекту.

Предмет дослідження - метод прогнозування погоди та погодних явищ за допомогою штучного інтелекту.

Методи дослідження – розробка та моделювання методу прогнозування погоди та погодних явищ за допомогою збору даних і їх аналізу для подальшого прогнозування за допомогою штучного інтелекту.

У роботі проведено аналіз сучасних методів прогнозування погоди та погодних явищ, типів та методів прогнозування стану погоди. Виконано огляд сучасних систем прогнозування та моделювання погоди. Оглянуто питання потрібності прогнозування погоди для людини та у сферах промислової діяльності. Здійснено розробку методу прогнозування погоди з використанням штучного інтелекту за допомогою машинного навчання.

Розроблено модель методу прогнозування погоди, яка дозволяє здійснювати прогнозування погоди за принципом штучного інтелекту, використовуючи дерева прийняття рішень та регресійну модель, аналізуючи попередні показники та надаючи короткостроковий та довгостроковий прогноз погоди.

Проведено аналіз отриманих результатів моделювання методу та порівняно з вже існуючими моделями прогнозування погоди, визначено їх переваги та недоліки у рамках обмежених параметрів вхідних даних від пристроїв збору інформації про навколишнє середовище

ABSTRACT

The text part of the master's work: 71p., 35 fig., 2 add., 55 source.

METHOD, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, FORECASTING, SIMULATION METHOD.

The object of research is the process of weather forecasting.

The purpose of the research is to improve the method of forecasting weather and weather phenomena using artificial intelligence.

The subject of research is the method of forecasting the weather and weather phenomena using artificial intelligence.

Research methods – development and modeling of a method of forecasting pogo and weather phenomena using data collection and their analysis for further forecasting using artificial intelligence.

The paper analyzes modern weather forecasting methods and weather phenomena, types and methods of weather forecasting. An overview of modern weather forecasting and modeling systems was performed. The issue of the necessity of weather forecasting for humans and in the spheres of industrial activity is reviewed. A method of weather forecasting using artificial intelligence with the help of machine learning has been developed.

A model of the weather forecasting method has been developed, which allows weather forecasting based on the principle of artificial intelligence, using decision trees and a regression model, analyzing previous indicators and providing short-term and long-term weather forecasts.

The obtained results of the simulation of the method were analyzed and compared with existing weather forecasting models, their advantages and disadvantages were determined within the limited parameters of input data from devices for collecting information about the enviro

ЗМІСТ

ВСТУП.....	11
1 ОСНОВНІ ПОНЯТТЯ ПРОГНОЗУВАННЯ ПОГОДИ	12
1.1 Процеси виникнення погодних явищ.....	12
1.2 Існуючі методи прогнозування погоди.....	13
1.3 Штучний інтелект в прогнозуванні.....	16
1.3.1 Інтелектуальні системи.....	19
1.3.2 Штучний нейрон	21
1.3.3. Машинне та глибоке навчання	26
1.4 Обробка даних для штучного інтелекту	29
1.5 Обробка великих даних	31
2. МОДЕЛЮВАННЯ МЕТОДУ ПРОГНОЗУВАННЯ ПОГОДНИХ ЯВИЩ.....	34
2.1 Математична модель прогнозування	34
2.2 Навчання штучного інтелекту	39
2.3 Прогнозування за допомогою множинної регресії.....	44
2.4 Байєсівські мережі.....	47
2.5 Метод опорних векторів	49
2.6 Випадковий ліс	51
3 Розробка методу прогнозування погоди.....	56
3.1 Вибір інструментів розробки	56
3.2 Оцінка та обробка вхідних даних	59
3.3 Реалізація методу прогнозування.	62
3.4 Розробка методу прогнозування погоди.....	64
3.5 Результат ефективності прогнозування.....	67

ВИСНОВКИ.....	70
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	72
ДОДАТКИ.....	77

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ШІ	Штучний інтелект
НМ	Нейронна мережа
BD	Big data
RF	Random forest
RBF	Радіальна базисна функція
ANN	штучним нейронним мережам
MSE	середньоквадратична помилка
MAE	середня абсолютна помилка
KNN	К-найближчих сусідів

ВСТУП

Концепція прогнозування погоди та погодних явищ за допомогою штучного інтелекту стає все більш актуальною з розвитком штучних інтелектуальних методів і використання машинного навчання. Прогнозування погоди потрібне як і людині, для розумінні погодних умов на найближче майбутнє, так і для більшості галузевих областей, які мають за основу планування робочого процесу у залежності від прогнозованих погодних умов.

Застосування штучного інтелекту у прогнозуванні дає можливість не тільки передбачати прогноз погоди, але і виявляти та передбачувати аномалії погодних явищ, які можуть виникати в наслідок катаклізмів чи техногенних загроз. Використання прогнозування погоди за методом створення синоптичних карт є точною, але потурбує багато ресурсів для аналізу отриманих даних, та довгого розрахунку отриманих даних.

Аналіз отриманих даних можна трактувати як точну зміну подальших умов стану атмосфери і аналізувати чинники впливу на навколишнє середовище. Об'єднання декількох методів прогнозування може дати неабиякий результат прогнозування, зменшивши похибку прогнозу в порівнянні з аналогічними існуючими системами. Моделювання параметрів погодних явищ, їх кореляцію між собою дає можливість не тільки оцінювати температуру чи вологість, а й надавати нетривіальні висновки про стан ґрунту або вірогідність дощу.

Таким чином створення методу прогнозування погоди є сучасною і актуальною роботою.

Об'єкт дослідження - процес прогнозування погоди.

Предмет дослідження - метод прогнозування погоди та погодних явищ за допомогою штучного інтелекту.

Мета дослідження - вдосконалення методу прогнозування погоди та погодних явищ з використанням штучного інтелекту.

1 ОСНОВНІ ПОНЯТТЯ ПРОГНОЗУВАННЯ ПОГОДИ .

1.1 Процеси виникнення погодних явищ

Погода – це стан атмосфери в певному місці, в певний час або в певний день, тиждень та місяць.

Погода має значний вплив на проживання людини, роботу над виробництвом їжі та особистий комфорт для життя. Людина залежить від стану та складу погодних умов, а саме: погоди та погодних явищ. Висока температура та вологість можуть викликати дискомфорт і призвести до хвороби; сильні дощі можуть спричинити повені, переміщення людей і порушення господарської діяльності; торнадо, грози, град та мокрий сніг можуть зіпсувати або знищити повністю врожай, будівлі, дороги та транспортні засоби. Бурі можуть навіть вбити або завдати шкоди людям та тваринам.

Атмосферні процеси можна класифікувати за критеріями, відмінними від просторово-часових властивостей. Підвищення якості прогнозів погоди відбувається за двома напрямками:

- Підвищення ефективності.
- Своєчасності прогнозу.

Формування атмосферних процесів і явищ у кожній конкретній місцевості відбувається під впливом самих процесів.

Можуть бути мікромасштабні погодні явища Мезомасштабні або середньорозмірні атмосферні явища.

Мікромасштабними погодними явищами є: невеликі пилові дияволи (ураган), терміки, турбулентні вихори. Вони відбуваються протягом дуже короткого періоду часу, від секунд до хвилин, і мають середній діаметр від менше метра до кількох кілометрів [1].

Мезомасштабні або середньорозмірними природними явищами є: грози , морські бризи та лінії шквалів. Вони належать до явищ між мікро-та

макромасштабами, які тривають від годин до днів. З точки зору динаміки, мезомасштабні явища — це не геострофічні вітри, а рухи, які зазвичай відповідають критеріям для припущення гідростатичного тиску. Мезомасштабні явища є результатом внутрішніх і зовнішніх механізмів впливу. Внутрішньо, масштабні системи створюють свою власну структуру і циркуляцію за рахунок виділення захищеної теплової енергії і динамічного руху, що залежить від місцевих температурних тиску.

Метеорологічні надзвичайні ситуації – це небезпечні природні процеси та явища, що відбуваються в атмосфері під впливом різноманітних природних факторів або їх комбінацій, що справляють або можуть мати істотний вплив на людей, сільськогосподарські рослини і тварин, економічні інтереси та навколишнє природне середовище [2].

1.2 Існуючі методи прогнозування погоди

Прогнозування називають найбільш розвиненою областю в метеорології. Сучасне прогнозування досить складне. Прийнято розрізняти три методи наукового прогнозування погоди, а саме: синоптичний метод прогнозування погоди, чисельний (також відомі як гідродинамічний) метод і статистичний метод прогнозування погоди.

Сучасні прогнози погоди – це результати багатьох процесів, що відбуваються по черзі і умовно створюються з декількох етапів. Тому це не просто.

Щоб зрозуміти, що станеться з погодою в найближчі дні, нам потрібно знати, як виглядає атмосфера і які зміни відбулися за останні кілька днів. Для цього майже 10 000 наземних метеорологічних станцій здійснюють моніторинг погоди кожні 3 години, і близько 1000 радіостанцій здійснюють радіомоніторинг верхньої атмосфери двічі на день [3].

Наука про прогнозування погоди стає все кращою завдяки запуску нових метеорологічних супутників і прогресу в технології. Метеорологи покладаються на

дані супутників, кораблів, літаків, метеостанцій і буїв, а також на обладнання, скинуте з літаків або повітряних куль [4].

Синоптичний метод - метод аналізу та прогнозування атмосферних процесів і умов у великих районах, використовуючи використання метеорологічних карт і відповідних погодних умов. Ці карти складають основу прогнозу погоди. Завдання створення синоптичної картина постійній основі передбачає збір і аналіз величезної кількості даних спостережень, отриманих в результаті багаторазових спостережень метеостанція. Ми можемо відчувати зміни в атмосфері, зокрема рух атмосферних змін, рух, трансформацію та взаємодію повітряних мас та тощо.

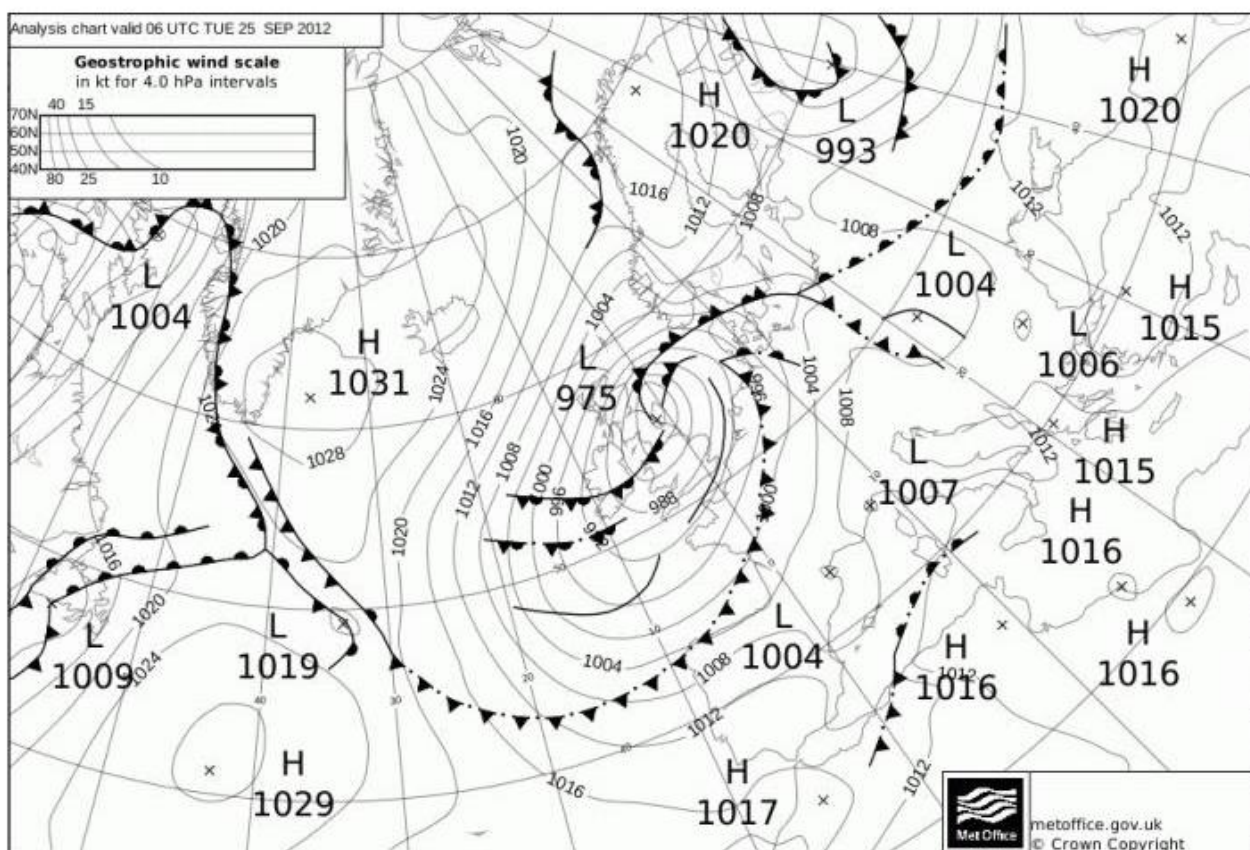


Рисунок 1.1 - Синоптична карта

Чисельний метод (гідродинамічний) – це метод атмосфери заснований на розробці математичних моделей атмосфери та взаємодії між атмосферою і океаном. Сучасні прогнози погоди зараз використовують методи чисельного прогнозу

погоди (NWP). Такий метод зроблений на тому факті, що гази атмосфери підпорядковуються ряду фізичних принципів.

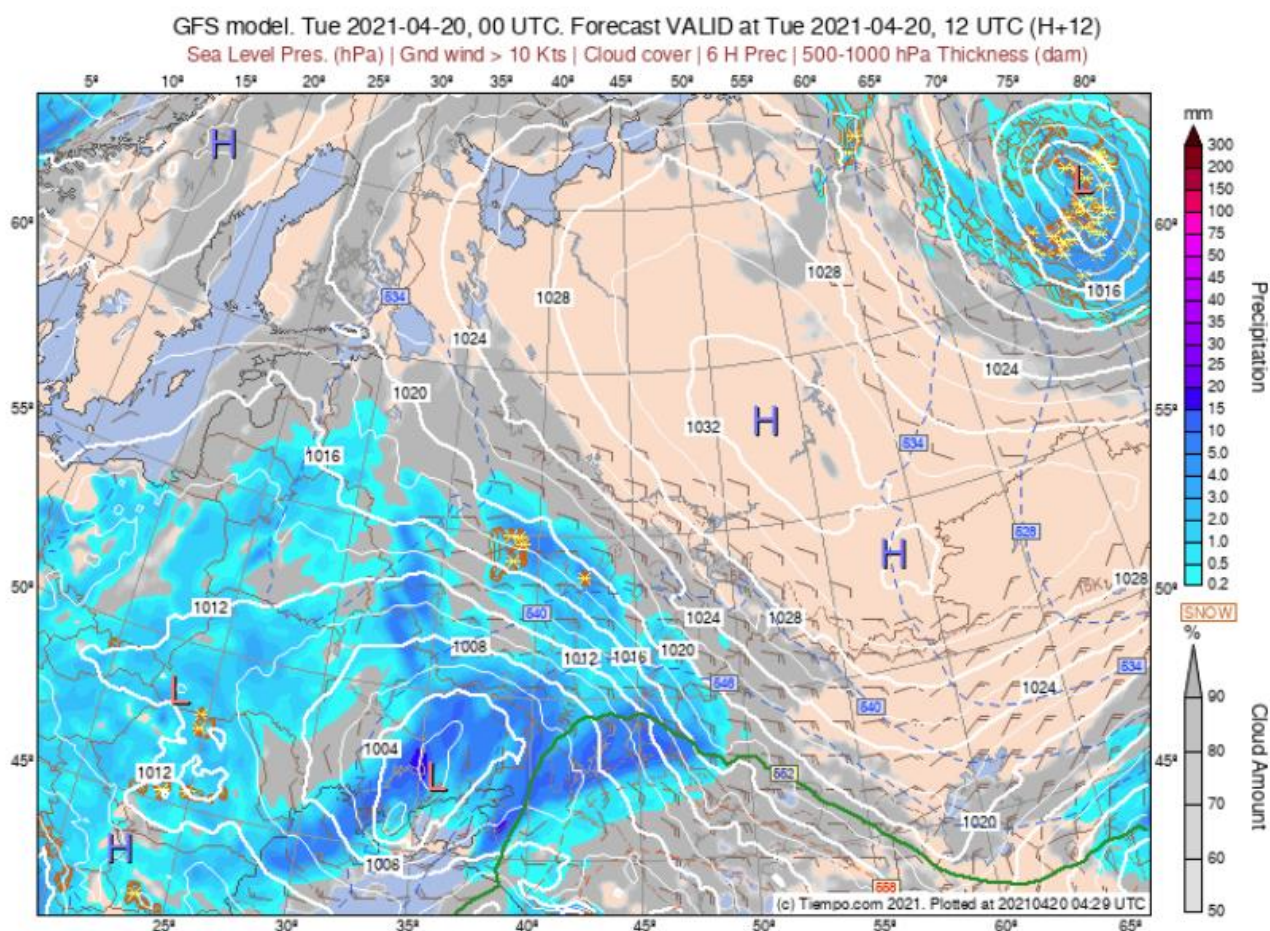


Рисунок 1.2 - Модель чисельного прогнозу погоди і статистична корекція

Завдяки отриманій статистиці моделі ми можемо виключити типові помилки в прогнозуванні. Крім того, модельну статистику для конкретних регіонів можна використовувати для більш точного прогнозування конкретних подій, особливо опадів, гроз тощо. Такий підхід значно підвищує точність прогнозів, дозволяючи врахувати особливості місцевого клімату [5].

Статистичний метод - цей метод часто доповнюється до гідродинамічного методу. Статистичні методи використовуються для запису минулих метеорологічних даних, залежно від припущення, що в майбутньому погода повториться.

Використання математичної моделі вимагає повного знання початкового стану атмосфери. Це досягається шляхом спостережень за атмосферою за допомогою радіозондових станцій по всьому світу.

Дані радіозонду доповнюються супутниковими спостереженнями радіації, які також надають дані про вологість і хмарний покрив.

Кожна з цих методів повинна не тільки передбачити місцезнаходження і можливе посилення існуючих систем тиску, але й визначати створення нових центрів шторма [6].

Насправді ці моделі представляють гіпотетичну атмосферу, але їх користь для прогнозування погоди безсумнівна.

Синоптики роблять прогнози погоди на основі даних з усієї мережі для моніторингу того, що відбувається на Землі.

Найкращі існуючий методи прогнозування використовуються в великому комп'ютерному моделюванні атмосферної фізики. Ці моделювання добре працюють саме для довгострокового прогнозування погоди, але менш ефективні в прогнозуванні того, що станеться в наступному році або про те, що називається поточним прогнозом.

1.3 Штучний інтелект в прогнозуванні

Штучний інтелект являє собою унікальну технологію технологічного прогресу у світі, який дозволяє машинам вчитися та розвиватися за допомогою людського та власного досвіду, може адаптуватися до нових розумів завдяки своїм додатків, можуть виконувати різні завдання, які могла зробити тільки людина, також може передбачати якісь події та робити кращим ресурси різноманітного характеру.

Штучний інтелект не замінить людину, але значно робить легшим життя в сучасний час, оскільки важливо прискорити обробку інформації і ефективно розподіляти ресурси [7].

Є чотири основних видів штучного інтелекту:

- Реактивні машини
- Обмежена пам'ять
- Теорія розуму
- Самосвідомість

До особливостей штучного інтелекту можна віднести: можливість реагувати на інформацію, доступну в середовищі; можливість вирішувати зазначені проблеми; пам'ять і навички на досвіді; можливість сприймати слухові, зорові та тактильні відчуття; може з легкістю обробляти великий обсяг інформації; адаптація; стійка здатність до оптимізації [8].

В ігровій індустрії штучний інтелект це зручно, а саме тим, що можна грати проти «ігрових роботів», коли людські гравці не доступні.

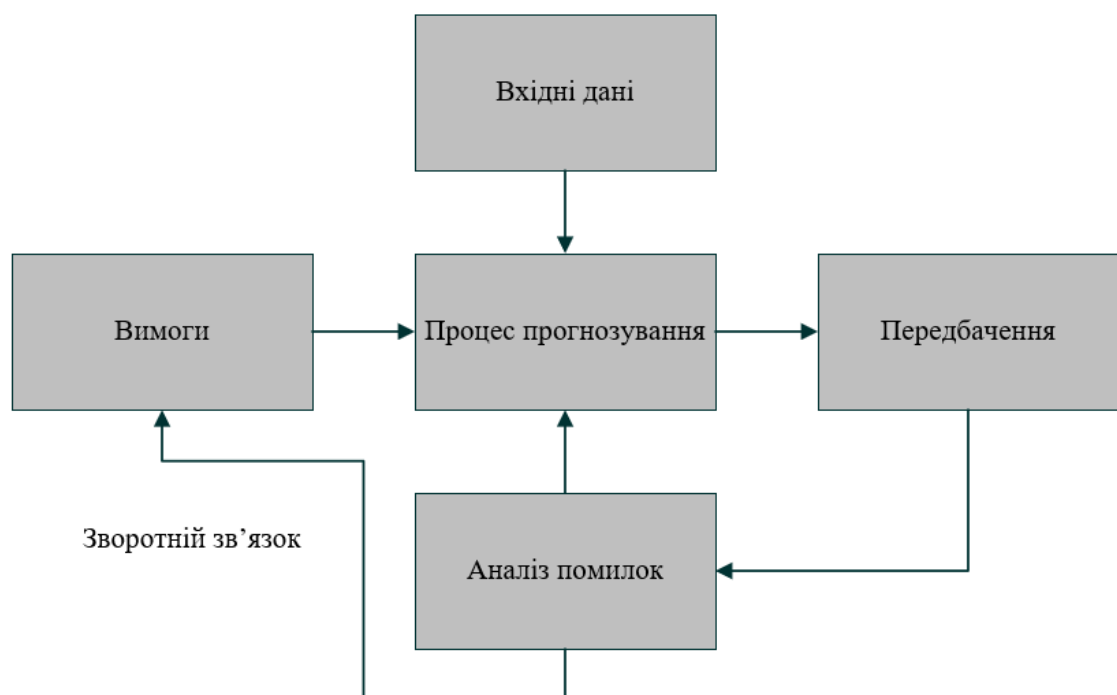


Рисунок 1.3 – принцип роботи прогнозування штучним інтелектом.

Військові та авіація можуть використовувати штучний інтелект для класифікації інформації, про це надає пілотам найкращий спосіб орієнтуватися в

повітряному просторі, щоб уникнути заторів та правильно планувати свої дії. Таким чином можна зробити висновок, що використання штучного інтелекту має актуальність і потреба цієї області знань вкрай важлива для сьогодення.

Кожна частина прогнозування за допомогою штучного інтелекту має свої переваги та недоліки, але кожний метод є унікальним. Вивчаючи кожен метод, можна зробити висновки стосовно того, який метод підходить для прогнозування, так як використання одного методу може не відповідати параметрам прогнозування в той чи іншій галузевій області прогнозування. Лінійна прогнозування або прогнозування за допомогою періодичності даних може використовуватися в різних сферах.

Планування методу прогнозування може бути основоположним для вирішення цілей та мети прогнозування.

Метеорологія часто стикалася з проблемою великих даних. Я навіть вважаю, що наука займалася створенням великих даних до того, як таке слово стало метафорою.

Використання штучного інтелекту для прогнозування погоди ще тільки є на початковому рівні. Одним зі способів, штучного інтелекту може революційно спрогнозувати сезонні та підсезонні часи – це додавання нових змінних в моделей прогнозування.

Обчислювальна потужність та розвиток технологій, таких як штучний інтелект, дозволили не тільки швидко і легко аналізувати дані, але і навчитися використовувати історичні дані для кращого розуміння ситуації і прийняття рішень [9].

Однією з цілей штучного інтелекту у прогнозуванні погоди для того, щоб краще прогнозувати погодні явища, які виникають у результаті змін навколишнього середовища, використовуючи певний універсальний алгоритм, який буде виконувати свою функцію, незалежно від вхідних отриманих параметрів і використанні при цьому обмеженого ресурсу комп'ютеру.

1.3.1 Інтелектуальні системи

Інтелектуальна інформаційна система керує інтеграцією сучасних технологій, обробкою та розподілом даних у багатьох обчислювальних середовищах.

Інтелектуальна інформаційна система (ІС) — це один із типів автоматизованої інформаційної системи, інколи ІС можуть називати як системою, яка була заснована на знаннях [10].

Логічна мова, яка використовується для розробки ІС, а в наш час використовують різні процедурні мови, а саме:

- Математичне
- Лінгвістичне
- Програмне
- Технічне
- Технологічне
- Кадрове

У загальному вигляді всі системи, засновані на знаннях, можна розділити на системи, що вирішують аналітичні задачі, і системи, що вирішують синтетичні задачі [11].

У технологіях прийняття рішень Інформаційної системи - це інформаційно-обчислювальна система з інтелектуальною підтримкою, яка вирішує проблеми без участі осіб, які приймають рішення, на відміну від інтелектуальної системи, в якій присутній оператор [12].

Інтелектуальні інформаційні системи дозволяють групам користувачів, пристроям і даним працювати разом, щоб приймати рішення без потреби участі третіх сторін.

Складні методи статистичного машинного навчання відображають інформацію так, як вам потрібно.

Інтелектуальні інформаційні системи можуть шукати та збирати, або навіть аналізувати дані, та спілкуватися з іншими системами навіть у мінливому середовищі [13].

Інформаційні системи зі штучним інтелектом вважаються кібернетичними системами, але вони є лише частиною кібернетики.

Хоча саме поняття інтелекту ще не було чітко визначено, його часто розуміють як здатність правильно реагувати на нові ситуації, особливо коли мова йде про штучний інтелект [14].

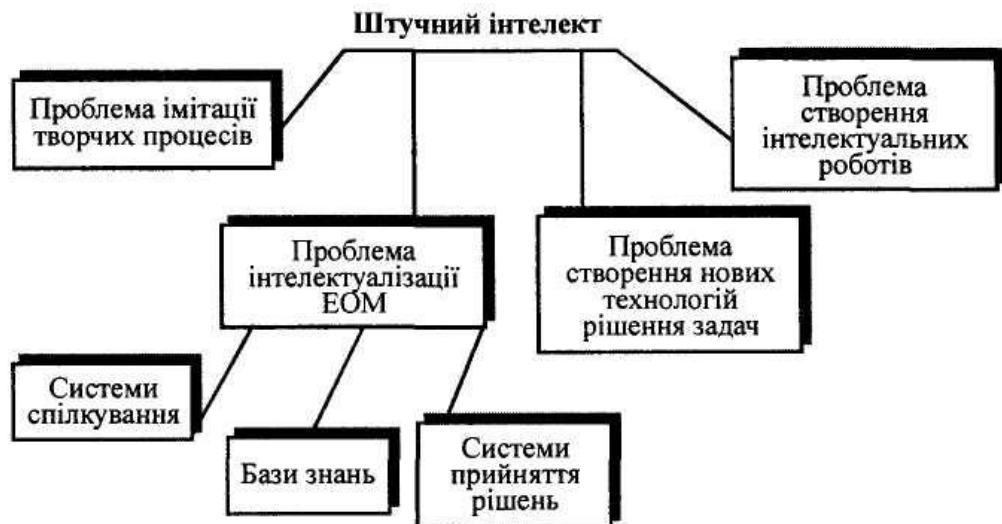


Рисунок 1.4 – Проблеми реалізації інтелектуальних систем

Відсутність інтелекту у комп'ютері це нормально, адже це набір даних, який може взаємодіяти лише за чітким алгоритмом взаємодії, і відповідати лише по завчасно запланованому плану, не маючи права відхилятися від поставлених цілей та задач.

Інтелектуальна система програмування система програмування, що виконує автоматичний синтез програм на основі написаного користувачем опису проблеми.

Особливості, які можуть бути в інтелектуальній системі та які стосуються баз знань:

- Машинне навчання
- Автоматичне доведення
- Інтроспекція
- Доведення висновку

Робота інформаційних систем пов'язана з обміном інформацією між інформаційними системами з компонентами та середовищем. Під час роботи персонал час від часу отримує інформацію про стан системи, про досягнення заданої мети з метою впливу забезпечувати виконання системних та управлінських рішень.

1.3.2 Штучний нейрон

Штучні нейронні мережі – це програмна реалізація нейронної структури людського мозку.

Штучні нейронні мережі імітують поведінку мозку більш простим способом. На рис. нейрон людського мозку.[15] Їх можна навчати контрольованими і не контрольованими способами. У контрольованій штучні нейронні мережі, мережа навчається, надаючи приклади від повідної вхідної та вихідної інформації [16].

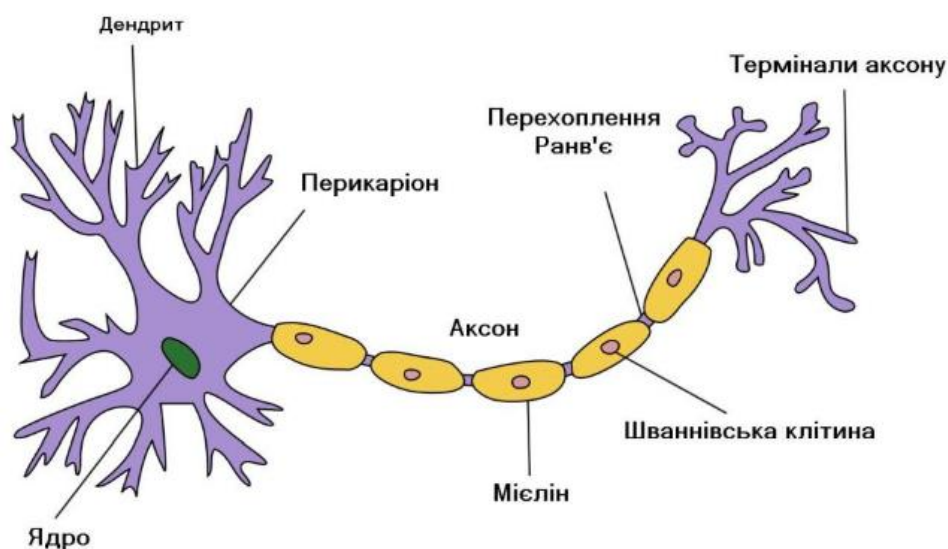


Рисунок 1.5 - Нейрон людського мозку

Неконтрольоване навчання штучні нейронні мережі «володіє» штучні нейронні мережі, намагаючись «зрозуміти» структуру на даної вхідної інформації.

На сьогоднішній день нейронні мережі використовуються як альтернатива усім алгоритмам машинного перекладу, розпізнавання мови та музики, обробки зображень та ідентифікації об'єктів на фотографіях і відео [17].

Біологічні нейрони моделюються в штучні нейронні мережі за допомогою функцій активації:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (1.1)$$

Штучна нейрона мережа є моделлю нейронної структури мозку, обробка, зберігання та формування інформації. Така модель реагує на зміну та видає результат або діяти або не рухатись. Ця система працює як базова булева алгебра, тобто нейрон обробляє інформацію, яку отримує та видає результат простих рішень, 0 або 1. Об'єднання великої множини таких нейронів дозволяють всебічно зважувати отриману інформацію та робити висновки, і за більшості результатів буде обрана правильна відповідь

Штучні нейронні мережі побудовані з базових одиниць, які називаються штучними нейронами. Нейронна мережа- це величезна кількість з'єднань, що з'єднують окремі нейрони. Групування нейронів в головного мозку люди забезпечують обробку інформації динамічним, інтерактивним, самоорганізованим способом [18].

Оригінальність нейронних мереж як біологічних аналогів мозку полягає в їх здатності навчатися.

Штучних нейронних мереж багато, вони складаються із простих нейронів, прості нейрони, які взаємодіють і дозволяють виконувати паралельні обчислення.

Нейронні мережі можуть бути реалізовані в швидкісних апаратних пристроїв, більшість досліджень воно виконується за допомогою програмного моделювання на звичайному комп'ютері.

Програмне моделювання пропонує досить дешеве та гнучке середовище щоб знайти та перевірити дослідницькі ідеї моделювання виявилось цілком адекватним та достатнім [19].

На малюнку показано типову структуру штучної нейронної мережі, багатосаровий тип. Є мережа, яка містить лише однорівневі або навіть

одноелементні, більшість програм вимагають мереж, які включають щонайменше три типи шарів: вхідний, прихований і вихідний.

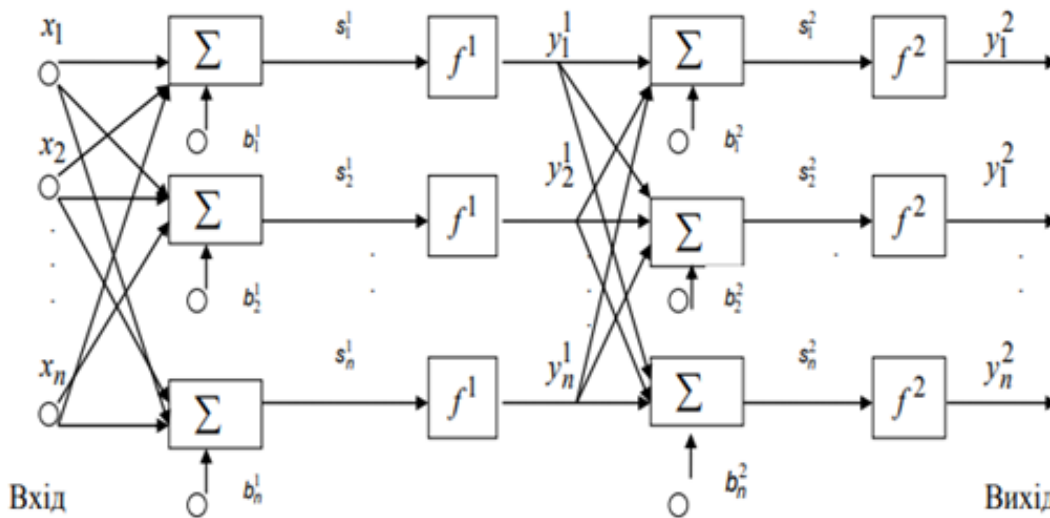


Рисунок 1.6 - Структура тришарової нейронної мережі

Ця схема відображає структуру тришарової нейронної мережі складові такої мережі мають такі параметри:

- 0-шар – вхідний;
- 1-шар прихований;
- 2-шар – вихідний;

Така система обробки структурної інформації дозволяє обробляти інформацію шляхом коефіцієнтів потрібності. Основні процеси вибору потрібних величин відбуваються саме у прихованому шарі.

До задач, складові які розв'язуються з використання штучних нейронних мереж являється:

- прогнозування режимних параметрів ;
- прогнозування електричного навантаження ;
- прогнозування економічного стану;
- прогнозування погодних явищ;
- прогнозування доходів бізнесу;

- відновлювальних джерел енергії;
- прогнозований пошук місць сейсмічної активності.

Вихідний нейрон може бути надісланий іншим нейронам нейрони над перерваними зв'язками. Залежно від значення ваги сигналу може збільшити чи зменшити значення.

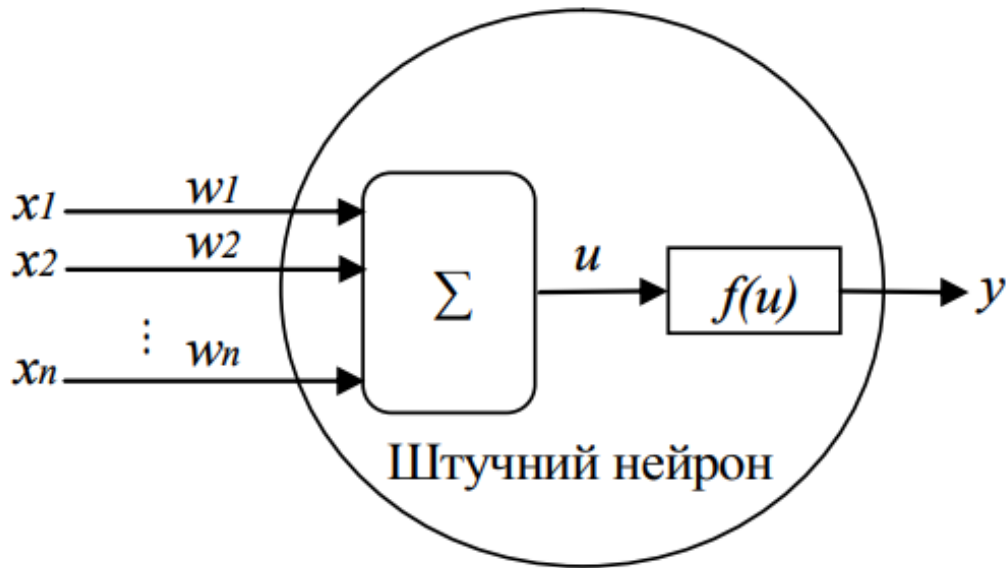


Рисунок 1.7 - Структура штучного нейрона

Функцією активації або передавальною функцією, математично штучний нейрон зазвичай представляється як нелінійна функція:

$$y = f(u) \quad (1.2)$$

Від одного аргументу u – усі лінійні комбінації вхідний сигнал:

$$u = \sum_{i=1}^n w_i x_i \quad (1.3)$$

де x_i і w_i – відповідно сигнали на входах нейрона і вагові коефіцієнти входів.

Лінійна передавальна функція розраховується за таким виразом, де k – коефіцієнт пропорційності:

$$f(x) = kx \quad (1.4)$$

Поки сигнал не досягне входу нейрона з нульового значення вихідний сигнал дорівнює нулю. Як тільки сигнал на вході нейрона перевищує нуль, стрибки вихідного сигналу змінюються на одиницю [20].

Сьогодні відома велика кількість нейронних структур та їх модифікацій, що орієнтовані на вирішення конкретного типу задач. Типи структур на рис. [21].

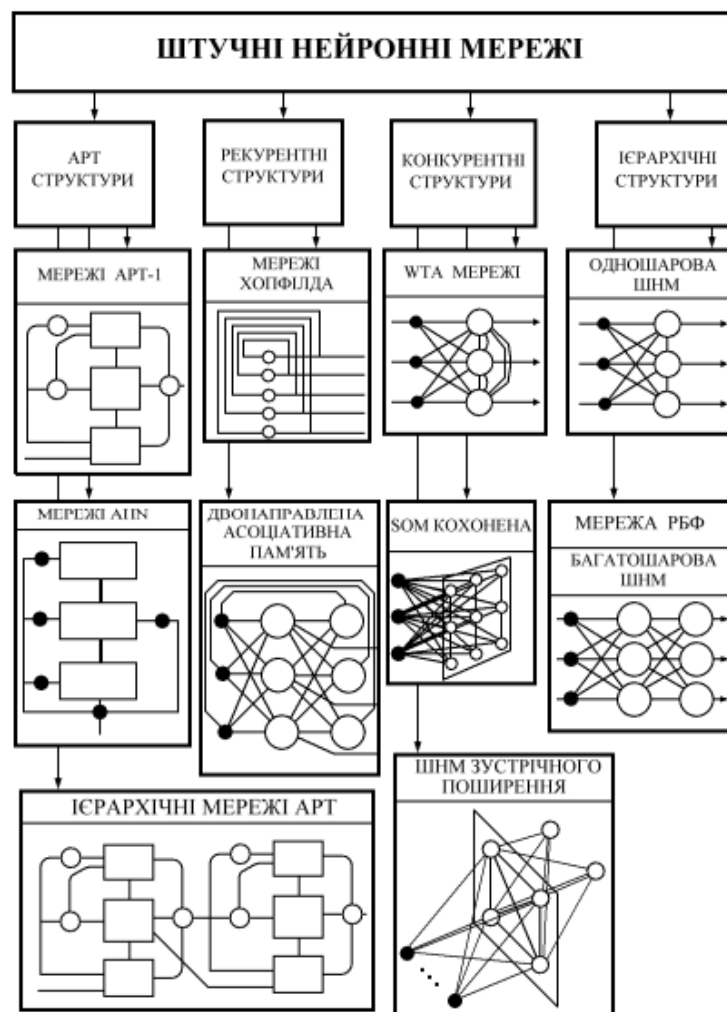


Рисунок 1.8 - Типи нейронних структур

Повністю зв'язана структура є узагальненою структурою, як і будь-яка комбінація інших нейронів вважається підмножиною цієї структури. Популяризація таких структур обмежується недоліком, значного зростання кількість міжнейронних зв'язків збільшується нейрон [22].

1.3.3. Машинне та глибоке навчання

Машинне навчання – відноситься до методів штучного інтелекту, які навчають комп'ютери самостійно вирішувати різні завдання.

Алгоритми визначаються відповідно до завдання, яке необхідно вирішити, і даних, які є у розробника [23].

Іншими словами, машини можуть знаходити шаблони в складних багато параметричних завданнях, завдяки цьому можна знаходити більш точні відповіді.

Машинне навчання розроблено для створення найточніших прогнозів на основі вхідних даних [24].

Є безліч безкоштовних і навчальних інструментів, які доступні розробникам програми з усього світу. Одним із найпопулярніших таких інструментів є бібліотека машинного навчання TensorFlow, розроблена Google. На основі бібліотеки Google було створено хмарні сервіси, які дозволяють використовувати машини в різних галузях [25].

Вхідні дані для машинного навчання зазвичай записується у формі матриці об'єкта, що залежить від певної залежності, але не є відомою. Суть машинного навчання полягає у пошуку цієї залежності, тобто необхідно зробити для початку алгоритм, який зможе на основі інформації про об'єкт добре класифікувати можливу ознаку.

Встановити вхідні дані для більш ефективного передбачення не обов'язково, тому що неправильне налаштування даних може призвести до результату, який не відповідає дійсності.

Є декілька типів машинного навчання [26]:

- навчання за прецедентами, чи індуктивне навчання, що базується на
- знаходженні деяких емпіричних закономірностей у вибірці;
- дедуктивне навчання означає певну формалізацію та оцінку експертних
- знань та перенесення їх у базу знань.

На зображенні показано загальна схема роботи алгоритму машинного навчання:



Рисунок 1.9 - Загальна схема роботи алгоритму машинного навчання

Дедуктивне навчання зазвичай стосується фахівців, тому «за прецедентами» та «машинного навчання» вважається ідентичними поняттями.

Одна з головних цілей машинного навчання полягає в тому, щоб усунути або частково автоматизувати в різних професійних роботах.

Машинне навчання можуть допомогти вам захистити інформацію, яка зберігається у хмарі, розкриваючи і аналізуючи підозрілі вхідні хмари та аналізи.

Існує багато моделей машинного навчання, але зазвичай вони стосуються одного з трьох видів [27]:

- навчання з учителем (supervised learning);
- навчання без учителя, або самонавчання (unsupervised learning);
- навчання з підкріпленням (reinforcement learning).

Кожен метод є унікальним, і може знадобитися для кожної конкретної ситуації. Машинне навчання з учителем це навчання машини певним алгоритмом,

який буде зауважувати та на етапі навчання буде сумніватися у своєму виборі, тобто буде тільки аналізувати правильні твердження і аналогічні з ним. Алгоритм самостійного навчання потребує дещо більше початкові параметрів для початку прогнозування. На етапі роботи, коли алгоритм зрозуміє початкову ціль, такий алгоритм буде самонавчатися довше, але результативніше. Мінусом використання такого алгоритму є те, що не має зовнішнього втручання, і отримання хибних результатів на початку тренування то не вийде правильного результату через накопичення помилки прогнозування з кожним етапом.



Рисунок 1.10 - Методи машинного навчання, їх класифікація

Машинне навчання та глибоке навчання - це усі підгалузі штучного інтелекту. Однак глибоке навчання насправді є підполем машинного навчання.

Різниця між глибоким навчанням і машинним навчанням полягає в тому як навчається алгоритм.

Поглиблене навчання автоматизує більшу частину процесу для усунення властивостей, усуває деякі необхідні людські процеси і дозволяє використання великих полів даних [28].

Глибоке навчання виявляє величезну кількість неструктурованих інформації, якої люди зазвичай потребують десятиліттями, щоб зрозуміти і виконати її. Це форма машинного навчання з можливістю працювати в нелінійних процесах прийняття рішень [29].

Як підмножина машинного навчання, для аналізу даних використовують ієрархічну мережу нейронів.

Глибоке машинне навчання може використовувати мічені збірки даних, які також відомі як контрольоване вивчення, щоб інформувати алгоритми, але цього не потрібно [30]. Вона не вимагає, щоб людина втрутилася в процес обробки даних, який дозволяє машинному навчанню зростати більш цікавим чином.

1.4 Обробка даних для штучного інтелекту

Попередня обробка даних – це інформація, яку використовують для перетворення необроблених даних у корисну і ефективну форму.

Штучний інтелект і обробка даних є динамічні частини інформації, які стають дедалі важливішими. За допомогою алгоритмів машин, математичних моделей та статистичних знань можна провести цілий процес автоматизації.

Метою програмування є опис процесів обробки даних. Дані – це поняття фактів і ідей у формі формул, придатній для передавання та переробки у певний процес, і інформація надає їм значення, яке надає інформація [31].

Щоб перевірити якість даних, слід оцінити наступне:

- кількість записів;
- кількість атрибутів (або компонентів);
- типи даних атрибуту (номінальні, порядкові чи безперервні);
- кількість пропущених значень;

- правильність формату даних;
- неузгоджені записи даних.

Надання послуг обробки даних такі [32]:

1.Збір даних:

Такі дані можна отримати з перевірених і надійних джерел. Завдяки цьому збір інформації буде кращим та простішим. Вони дадуть найкращий результат під час перевірки моделі штучного інтелекту.

2. Приготування:

Ці дані будуть неопрацьовані і не матимуть відповідного формату. Ця інформація не подається безпосередньо у систему. Ви можете приготуватися двома способами — вручну чи автоматично.

3. Введення:

Підготування даних не завжди є читаним форматом машин. Для перетворення даних в систему, яку легко читати, потрібні алгоритм, щоб перетворитися. Для того щоб ефективніше виконувати роботу, потрібно неабиякої функціональності й точності.

4. Обробка даних:

Це фаза, на якій алгоритми і методи машинного навчання виконують вказівки з найточнішими точністю та точним аналізом. Але основним етапом для створення Системи штучного інтелекту це очищення даних. Великі дані, які отримує алгоритм, повинні бути очищеними від:

- Повторюваних значень;
- Фіктивних значень;
- Відсутність даних;
- Суперечливих даних.

5.Вихід:

На цій стадії результати отримуються в режимі, який користувач легко розуміє. Вихідні дані можна використовувати як звіти, електронні записи, графіки або відео.

6.Зберігання:

На цьому останньому етапі результати, модель інформації та всі цінні дані зберігаються, щоб продовжити використання.

У наш час більша кількість інформації, більша частина праці, ґрунтується на самих фактах, і тому дедалі більше інформації з'являється в різних цілях, наприклад, у наукових дослідженнях, наукових дослідженнях, приватному та особистому використанні, комерційному використанні, інституціональному використанні тощо [33].

1.5 Обробка великих даних

- це термін, який стосується широкого спектру структурованої та неструктурованої інформації, а також методів обробки та аналізу цієї інформації.

Системи, які обробляють і зберігають Big Data, а також інструменти, які підтримують аналіз Big Data, є загальною частиною архітектури керування даними в організаціях [34].

Big Data надходить із безлічі джерел, можна навести такі приклади, як: системи обробки транзакцій, бази даних про клієнтів, документи, електронна пошта, мед записи, веб-журнали кліків, мобільні програми та соціальні мережі.

Сам алгоритм Big Data впливає з реалізації Перші високопродуктивні сервери (мейнфрейми) мали Достатні ресурси для обробки інформації, придатні для Комп'ютерні розрахунки та подальший аналіз [35].

У сучасному світі більшість найбільш відомих постачальників інформаційні технології в організаційній бізнес-стратегії, використання поняття «Big Data » та провідні аналітики інформаційного ринку окреме дослідження присвячено поняттю технології.

Використання технологій Big Data в наш час дуже велике. Так, ви можете дізнатися з них про інтереси клієнтів, про те, наскільки ефективним є маркетинг, чи провести аналіз ризиків.

Запропоновано 5 виміри визначення великих даних, а саме: об'єм, різноманітність, швидкість, достовірність та мінливість [36].

Об'єм «Volume» - це велика кількість даних, які утворюються з додатків. Різноманітність «Variety» визначає, чи наявні у визначеному виборі різноманітні форми даних. Швидкість «Velocity» відноситься до генерації швидкості та аналізу даних. Достовірність «Veracity» Вибір надійних даних. Якість записаних даних може сильно відрізнятись, із-за цього погіршується точність аналізу. Мінливість «Variability» Неузгоджена інформація ускладнює, а іноді й заплутує процеси обробки та управління даними.

Вважається, що ці «вміння» дають їм нагоду виявити приховані факти, які виходять з обмеженого людського розуміння.



Рисунок 1.11 - виміри визначення великих даних

Проекти, засновані на великих даних, вже допомогли: лікувати захворювання, відкривати нові далекі планети, прогнозувати надзвичайні ситуації, запобігти злочинам, оптимізувати прибутки та тощо.

Big Data збирає неймовірну кількість інформації про наше приватне життя, але ми хочемо зберігати її в таємниці. Тому виникає питання про баланс між тим, чим ми готові поділитися, і тим, наскільки комфортнішим стає наше життя, якщо ми відкриті [37].

Навіть якщо припустити, що Big Data використовує лише шляхетні наміри, це не гарантує, що особисті дані будуть захищені від злочинців і хакерів.

Методами Big Data є на рис.

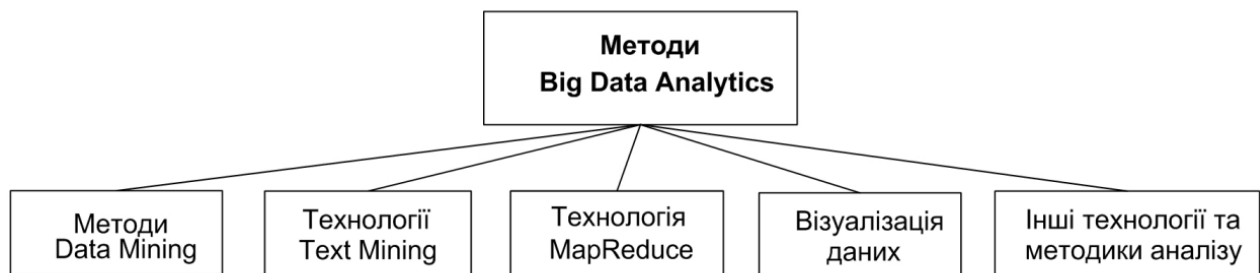


Рисунок 1.12 - Групи методів аналітики Великих даних

Велика інформація дуже корисна, як технологія, що допомагає розв'язувати сьогоденні щоденні проблеми, але створює ще більше нових. Велика інформація може змінити наше життя, роботу та мислення.

Складність аналізу великої кількості даних знаходиться у спеціальній збірці, поділі, збереженні, переносі, візуалізації та збереженні конфіденційної інформації [38].

Під розрахунком великої інформації, часто можна зрозуміти застосування уявних аналітиків або інших передових водоростей та методів, щоб усунути якомога більше корисних даних.

2. МОДЕЛЮВАННЯ МЕТОДУ ПРОГНОЗУВАННЯ ПОГОДНИХ ЯВИЩ

2.1 Математична модель прогнозування

Математичний метод прогнозування є базовим методом для створення математичної моделі та алгоритму прогнозування. Для того щоб зрозуміти які можливі варіанти створення прогнозу потрібно знати підхід до математичної частини цього процесу. Використання математики - це основний параметр для створення систем.

Для розумних базових потреб математичної моделі прогнозування, потрібно усвідомити основну модель, за якої вже йде розгалуження і модифікація інших моделей. Базову модель прогнозування впливає із такої формули:

$$y = mx + n \quad (2.1)$$

Де Y кінцевий результат прогнозування. При прогнозуванні погоди таких кінцевих результатів буде декілька, тому що за основу береться багато параметрів погоди. Тому для початку потрібно відокремити кожний параметр, а тільки після обробки вже можна об'єднувати їх[39]. X це певний проміжок часу, це можуть бути сезони чи навіть дні. У цьому випадку є чіткі параметри прогнозованого алгоритму. M параметр нахилу прогнозу.

Для оцінки і обробки даних отриманих від машин потрібно не тільки збирати, а і обробляти отриману інформацію.

Існують чотири техніки математичного прогнозування:

- Пряма лінія.
- Ковзке середнє.
- Проста лінійна регресія
- Множина лінійна регресія.

Прямолінійний метод прогнозування - це один із простих методів, який має лінійну залежність від вхідних параметрів. Цей метод підходить виключно для тих нових значень, які мають накопичувальний характер.

Наприклад, прогнозування з попередніми даними отриманий від опадів. Лінійне передбачення складає, те що є змінна яка збільшується на деякі відсоток протягом певного періоду. Це відноситься до статистичних даних але не для прогнозування в цілому. Цей метод доцільно можна використовувати, якщо потрібні приблизні значення в подальшому майбутньому, не використовуючи параметри у реальному часі і вважати за дійсними тільки минулі результати.



Рисунок 2.1 – існуючі методи математичного прогнозування

Ковзка середня - це техніка зважування деяких періодів. Аналіз проводиться за попередніми показниками. Дивлячись на те, що для прогнозування погоди є сталі факти, наприклад, якщо зимою холодно а влітку тепло, і із цього можна зробити висновок що таке реляційна модель кожного наступного року буде схожа.

Методів прогнозування досить багато, але вони мають основну мету - це створювати прогноз. Методи прогнозування поділяються:

- Інтуїтивні.
- Формалізовані.

У свою чергу ці два основоположних принципа прогнозування поділяються на різні критерії прогнозування.

Не всі класифікації підходять саме для прогнозування погоди, але основоположними є наступні види моделей:

- Факторні.
- Структуровані.
- Комбіновані.
- Імітаційні.
- Найменших квадратів.
- Дисперсії.
- Регресійного аналізу.
- Кореляції.

Перераховані методи математичного моделювання мають в собі початкове розуміння для складання моделі прогнозування за допомогою штучного інтелекту.

Також одним з основоположним методом [40] побудови моделі прогнозування є така формула:

$$S_n = a \cdot \sum_{j=0}^{n-1} (1-a)^j \cdot x_n - j \quad (2.2)$$

де x це ряд аналізованих даних, а це коефіцієнт згладжування, а S - це злагожене значення, яке є прогнозом.

Аналіз даних погодних параметрів мають не фіксовані показники та можуть коливатися в залежності від часу. Тому, проста лінійна регресія може виконувати

доцільне прогнозування за параметрами. Отримуючи лінію прогнозу, можна відтворити подальшу дію інших параметрів.

Множинна лінійна регресія для прогнозування погоди доцільно створює прогноз завдяки використанню двох або більше незалежних змінних.

Методи оперують певною статистикою параметрів і найчастіше використовується середній квадрат:

$$\sigma = \sqrt{\frac{\sum_{j=1}^n (x_i - x)^2}{n}} \quad (2.3)$$

Ми можемо використовувати статистичні таблиці, якщо знаємо певний розподіл значень. Знання послідовності ймовірності коефіцієнтів k , визначаємо ймовірність прогнозування.

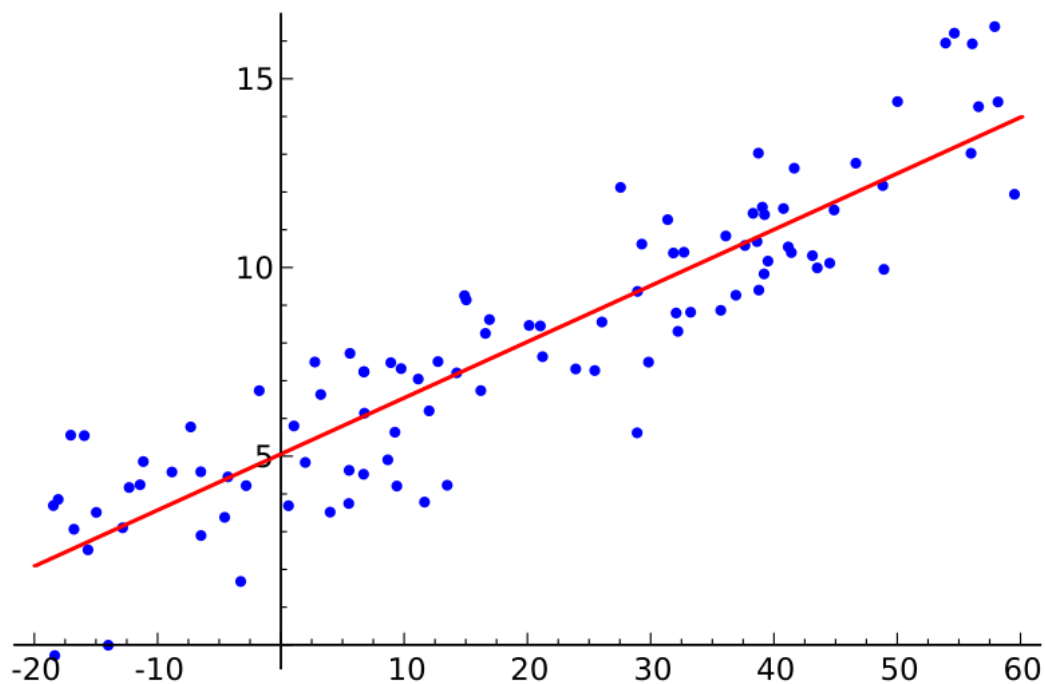


Рисунок 2.2 – вигляд аналізу лінійної регресії

Аналізуючи рисунок, можна зробити висновки що цей метод влучно відображає усереднення результату прогнозування.

Такий прогноз можна вважати лінійним, але потрібно взяти до уваги те, що цей прогноз не відтворює реальний стан у кожному проміжку часу. Параметри погодних явищ не можуть бути лінійними, вони змінюються і мають сезонні зміни, такі як день і ніч, зима та літо.

Бувають випадки, коли розширення середнього значення корисно. Наприклад, можливо, що означення не були знайдено в області досконалості змінної під дослідження. Це може бути проблемою для розпізнавання хибних і ймовірних значень.

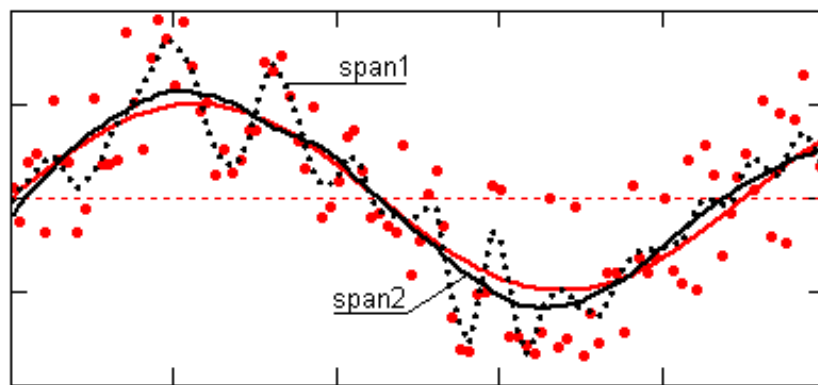


Рисунок 2.3 – Криволінійна зональна регресія

Для отримання результату прогнозування у непостійних змінних, потрібно використовувати побудову кривих, які більш точно відображають середнє квадратичне значення у кожний момент часу. Це дуже важливо для прогнозування. Лінійне прогнозування не може дати реальне значення якщо воно має результат лише прямої лінії. Наприклад 2 високі значення може перекрити одне низьке значення, які стоять поряд, але це призведе до коливань графіку. Для зональної регресії даних використовують таку формулу:

$$D = \sum_i (f1(i) - y_{o_i})^2 \quad (2.4)$$

Для детального аналізу потрібно знайти середню вагу, але якщо частота градацій є симетричною, враховуємо ваги значущі. Для цього потрібно вивити

максимальне і мінімальне значення для обох випадків прогнозування, та виявити середню похибку усіх результатів.

$$I = \frac{x_{max} - x_{min}}{n} \quad (2.5)$$

У цій формулі N виступає як показник кількості результатів. Розвиток потужних комп'ютерів та більш досконаліх методів моделювання повернуло чисельне прогнозування погоди до моделей, які дуже схожі на попередні моделі. Численні прогнози погоди використовують прості базові формули, як формули прогнозу.

Він враховує метеорологічні дані на конкретній станції на невеликій відстані в конкретному регіоні. Такі параметри, як мінімальна температура, максимальна температура та відносна вологість, розраховуються на основі математичних взаємозв'язків різних періодів часу із зібраними даними часових рядів. Отримані результати показують, що може точніше оцінювати погодні умови.

2.2 Навчання штучного інтелекту

Штучний інтелект - це наука яка змушує машини думати та приймати рішення, як роблять це люди. Тобто це алгоритм який може відтворювати дії та робити вибір у той чи іншій ситуації. Шляхом порівняння аналізу та співвідношення варіантів правильно або неправильно відповіді алгоритм може приймати рішення.

ШІ має величезний вплив на сучасну освітню ситуацію, і його вплив є незмірним. Це допомагає ідентифікувати людей за ознаками поведінки та спрямовувати їх у правильному напрямку. Це також допомагає педагогам сприяти інклюзії. Наприклад, мовний переклад за допомогою штучного інтелекту дозволить більше студентам навчатися.

Машинне навчання - це один з розділів ШІ, алгоритм якого розгортають роботу комп'ютера та запроваджують його робити висновки на основі вхідних заданих параметрів і вихідних даних. Але головною метою машинного навчання, що використані параметри і дані можуть бути проігноровані або відредаговані і не слідувати жорстко заданими параметрами або правилами[41].

Основою якісного машинного навчання іде базова схема і три основоположні параметри, і максимально ефективно взаємодія між ними. Основа в машинному навчанні та штучному інтелекті базується на трьох показниках:

- Дані
- Ознаки
- Алгоритми.

Стосовно даних, то це базова інформація або вхідні дані за якими надалі буде навчатися алгоритм.

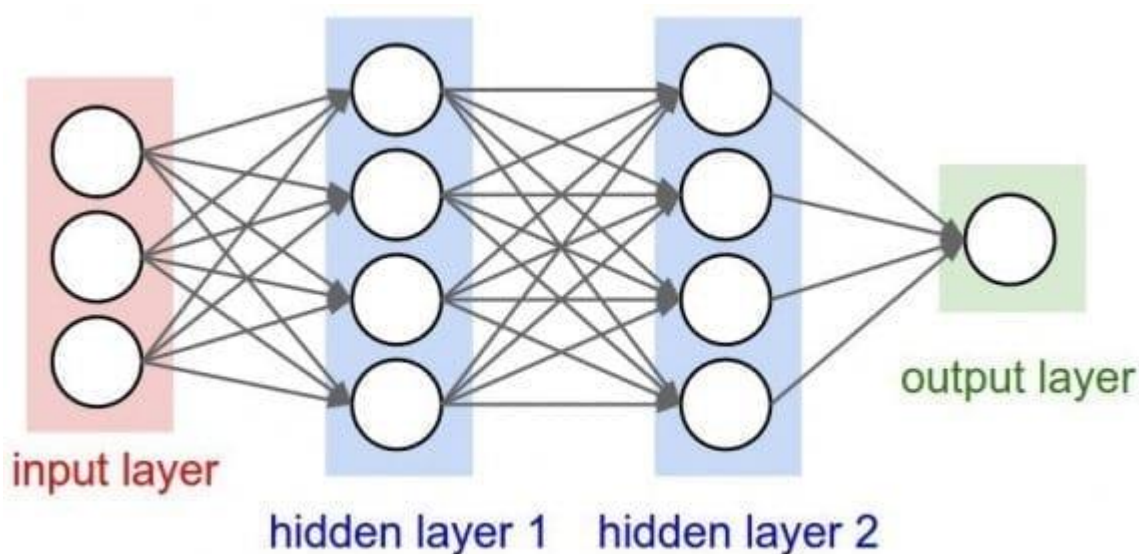


Рисунок 2.4 – принцип роботи глибоко навчання

На рисунку зображено кульки, які відображають взаємодію між всіма компонентами системи навчання штучного інтелекту. Гочаткове навчання починається від зовнішнього прошарку, це прошарок вхідних даних. Для прогнозування погоди потрібні вхідні дані стану погодних явищ в конкретний період часу.

Після вхідних даних йде прихований шар, який включає математичні рішення конкретної задачі прогнозування. Лінії, які видно між цими шарами, називаються «вагами». Ваги показують, наскільки один сегмент рішення переважає над іншим сегментом. Тобто є або правильним або точним за отриманими попередніми результатами навчання.

Ознаки для машинного навчання - це процес створення певних умов за яким буде працювати сам алгоритм машинного навчання. Це можуть бути параметри обмеження, формули, вхідні розрахункові параметри. Обмеження в ознаках може бути як ключовим фактором навчання, так і фактором неефективності процесу навчання.

Вихідний рівень дає нам окремі результати. Після того, як шар просумую усі ваги, він визначити, чи є відтворювані дані реальними чи ні.

В принципі прогнозування погоди це порівняння буде між попередніми даними та прогнозу майбутнього. Тобто при використанні попередніх даних, ми будемо знати вже наявні нові дані.

$$y = f \left(\sum_{i=1}^n x_i * w_i + b \right) \quad (2.6)$$

Рішення цієї формули дає результат роботи такого методу штучного навчання. N- це кількість вхідних параметрів, x як вхідні значення, w як ваги, які відповідають за значущість результату, а b – зміщення, або просто похибка відхилення.

Якщо штучний інтелект відтворить правильно послідовність. І прогнозу буде відповідати результатам реально часу, то такий алгоритм можна використовувати в подальшому для навчання, адже від вже буде розуміти. За яким принципом проходить навчання та кореляція між попередніми показниками атмосфери та реальними.

Щоб машинне навчання працювало коректно, потрібно чітко та формульовано складати параметри впливу на алгоритм. Хибні твердження або неправдиві формули можуть порушити структуру та результат навчання.

Процес машинного навчання використовує навчальні дані як алгоритм вибору.

Стосовно алгоритмів, набір методів для вирішення поставленого завдання. На основі створення алгоритмів взаємодії і починається навчання ШІ.

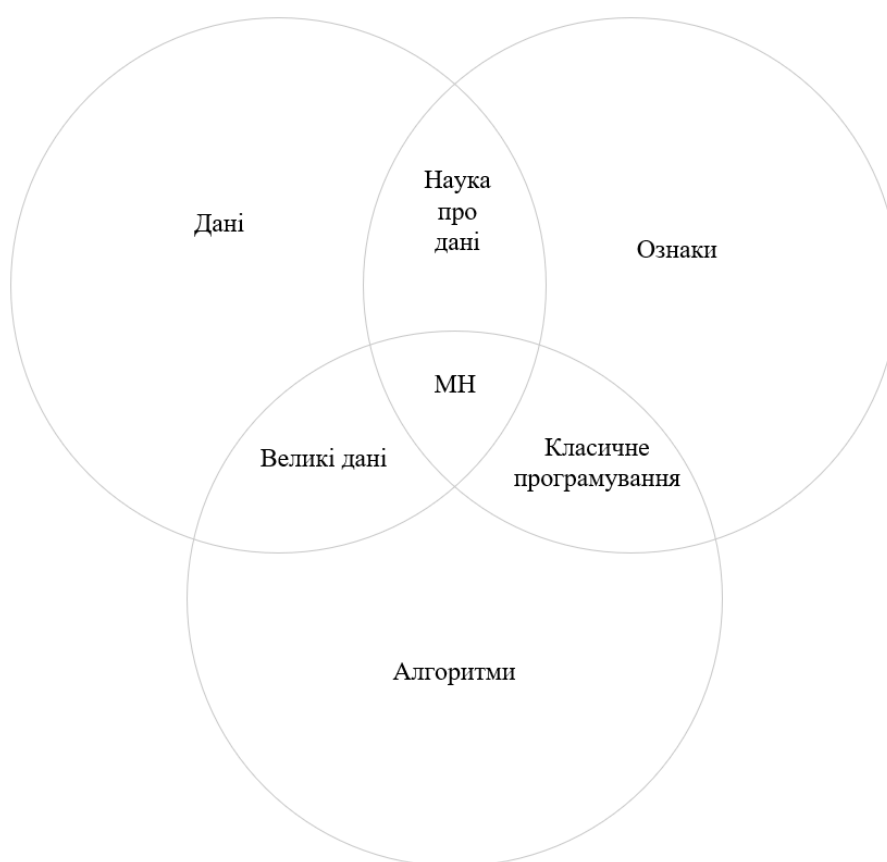


Рисунок 2.5 – схема компонентів машинного навчання

Для того щоб програма навчилася аналізувати та відтворювати модель інтелекту її потрібно перш за все навчити та надати параметри за якими вона буде робити вибірку рішень. Модель роботи машинного навчання має наступний характер. Програма отримує інформацію, потім вона проводить обробку отриманої інформації за певними алгоритмами, далі алгоритм виділяє більше значуще або потрібні матеріали та дані і видає результат який потрібен.

Глибокий (глибкість) навчання можна застосовувати лише з огляду на складніші показники ANN, які містять кілька схованих рівнів. На цьому етапі нейрони можуть виконувати шари, які важко збагнути. Кожна наступна стадія мережі шукає з'єднання за попереднім рівнем. Це ANN, який може знайти не лише прості послання, але й послання між відношеннями

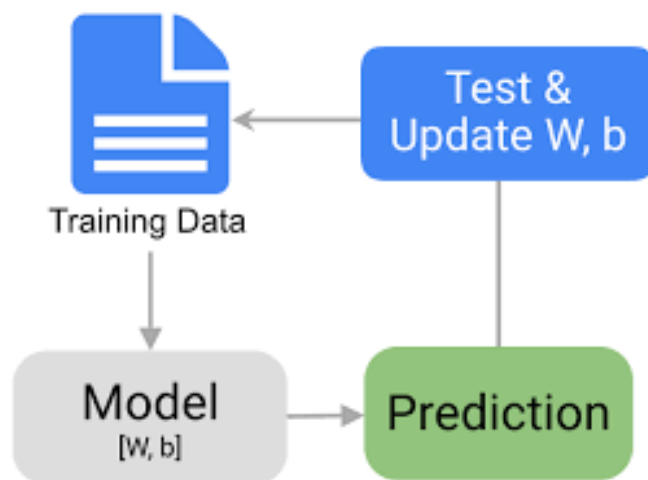


Рисунок 2.6 – Процес навчання штучного інтелекту

На даному рисунку представлено просту модель машинного навчання. Алгоритму дається метод невеликий датасет інформації, тобто тренувальні дані, які підготовлені для початку тренування. У таких датасетах дані чітко структуровані, та не мають містити складні варіації даних. Наступним етапом дані проходять до моделі або способу машинного алгоритму, який обробляє дані та кінцевим результатом видає прогнозування, що саме потрібно робити чи як діяти. Цей прогноз може бути чисельним або відповідати на питання булевого варіанту, тобто так чи ні.

Після отримання результату прогнозування всі отримані дані або параметри проходять тестування на відповідність, якщо ці дані були правильними, то вони знову записуються до основного датасету як основоположні.

Ці дані можна використовувати як базу для повторного та подальшого навчання. Все інше або оновлюється, змінюючи параметри прогнозування, або видаляється, в залежності від критерій оцінювання даного алгоритму.

Машинне навчання набагато легше узгодити з оновленням часу, оскільки це дозволяє більше інформації. Прогнозна аналітика зазвичай працює з стаціонарним набором даних і потребує постійного відсвіження.

Машинне навчання можна поділити та реалізовувати наступними способами:

- Контрольоване навчання.
- Навчання без контролю.
- Навчання з підкріпленням
- Навчання за допомогою ансамблевих методів.

Найефективнішим із методів навчання є навчання з підкріпленням чи так званим допомогою людини.

Тобто навчання штучного інтелекту іде самостійно за певним заданим алгоритмом але у алгоритм може втручатися людина та редагувати вибір програми.

Це дуже зручний інструмент на початку створення прогнозу і дії самого алгоритму.

Якщо прогноз не відповідає очікуваним прогнозам, алгоритм повторюється декілька разів, доки не знайде потрібні результати. Це дає можливість алгоритму постійно розвиватися і дати найбільш оптимістичну відповідь, яка поступово зростає пунктуальністю.

Коли програма помиляється, алгоритм пропонує людині або людина сама втручається в процес навчання та вказує на необхідний або більш правильний параметр. Після висновків програма редагує свої першочергові параметри та приймає схильність робити висновки самостійна використовуючи власний результат.

2.3 Прогнозування за допомогою множинної регресії

Метод прогнозування не завжди можна визначати як регресію. Але вона може виявляти залежності та кореляції між отриманими даними та прогнозованими.

Метод множинної лінійна регресія впроваджує оцінювання вихідних даних при отриманих даних незалежних змінних. Тобто при отриманні інформації у реальному часі про погоду ці параметри вважаються як незалежні змінні і вони входять у рівняння регресії.

На цьому етапі потрібно використовувати точковий прогноз. Якщо до рівняння множинної лінійної регресії додати значення змінних в які залежать від наступних або попередніх даних, то можна отримати прогнозовано значення або точковий прогноз. Цей прогноз є доцільним тільки у тому випадку коли є залежні змінні.

Прогнозування за моделлю множинної регресії дає оцінити очікуване значення певних змінних, які визначаються за допомогою значень незалежних від змінних, які є в регресії.

На цьому етапі прогнозування моделлю множинної регресії може виступати безперервне значення параметрів X_1, X_2, \dots, X_n , то значення прогнозованої [42] регресії точкового прогнозу виглядатиме наступним чином:

$$f(x) = n_0 + S_1x_1 + S_1x_1 + \dots S_nx_n \quad (2.7)$$

У цій формул S вступає як коефіцієнти регресії. Але ця формула може допомогти тільки якщо є точні кореляційні дані. Які мають деяку залежність від попередніх даних. Отримання даних для прогнозування погоди можуть бути не реляційними і навіть неочікуваними.

Прилади, які фіксують зміни мають власну похибку, тому доцільно для отримання більш точних показників враховувати похибку у прогнозуванні таким методом. Видозміна попередньої формули потрібна для урахування похибки і додається похибка під час регресійного процесу обчислення.

$$Y_i = n_0 + S_1x_1 + S_1x_1 + \dots S_nx_n + \varepsilon_i \quad (2.8)$$

У цьому випадку використання множинної регресії потрібно перевірити те, що значення двох чи більше незалежних змінних є корельованими. Якщо такі незалежні змінні є корельовані то вони підпадають під параметр мультиколінеарності. Якщо ця подія виявляється, то можна зробити висновки що метод не є ефективним і має у собі велику похибку прорахування прогнозу.

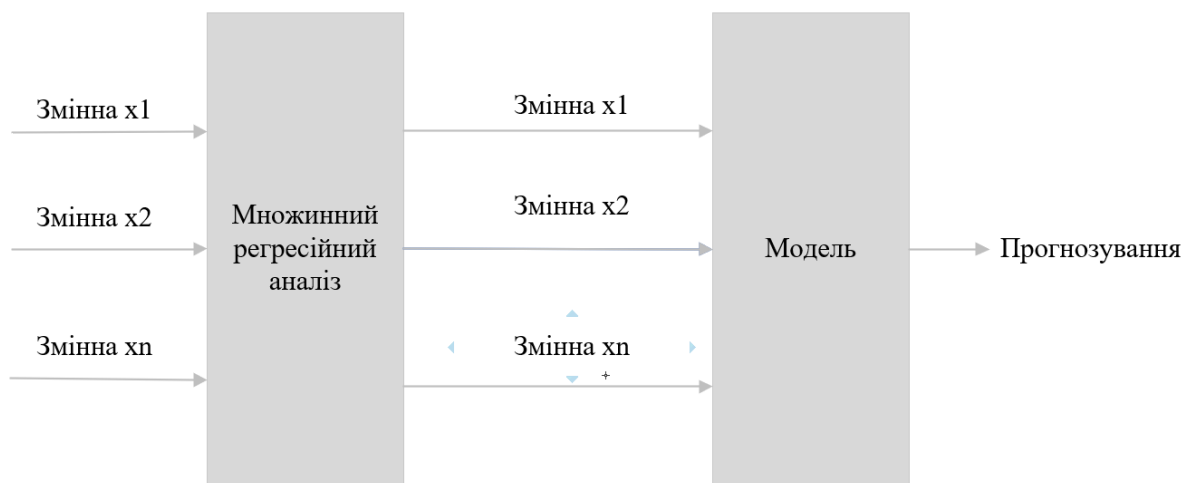


Рисунок 2.7 - Схема роботи множинного регресійного аналізу

Щоб запобігти негативного впливу прогнозування та усунути проблему не чіткого прогнозування використовують різні методики. Одна із технік це додавання чи видалення деяких вибірових змінних ітерацій. Якщо є декілька моделей з рівнем значусті $p < 0.05$, то вибір припадає на одну із моделей шляхом оцінювання вихідних даних. Кореляція даних у двох методах неможлива, навіть якщо вони мають спільні ознаки та принципи. Для вибору включення правильної моделі беруть значення та порівнюють їх методом оцінки значення суми квадратів результату до передбачення.

Більш поширене прогнозування використовуються інтервальним прогнозом. Інтервальний прогноз використовує технологію вибору мінімального та

максимального значення для регресії прогнозування і проміжок у якому може бути параметри подальшого прогнозування. Значення між двома точками та фактичним значенням вважається похибкою. Чим більше параметрів використовувати в інтервальному прогнозі тим менше буде похибка а прогноз точним.

2.4 Байєсівські мережі

Байєсівські мережі - це графи з деякими особливостями. Ідентичність БМ вказує на причини інтеграції, що є характерними для процесу графства [43].

Томас Байєс - англійський математик і пресвітеріанський священнослужитель. Він був першим, хто запропонував і пояснив правила, які дозволили б новоотриманій інформації змінити попередні переконання людини.

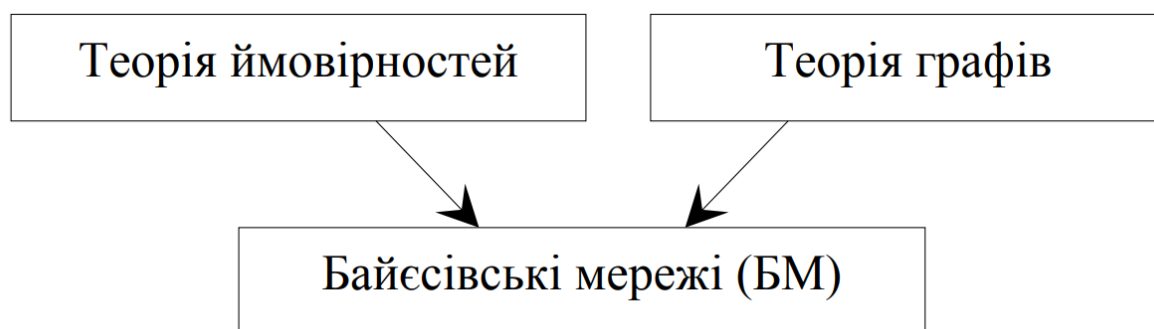


Рисунок 2.8 - БМ на стику двох наук

Це відносно молодий підхід до розвитку науки, який виник на основі теорії ймовірності і теорії графів.

Змінні можуть належати до будь-якого типу, довільного параметра, схованого у змінних або гіпотезах. Вони є двох частин: структури та параметрів [44].

Структури — це ациклічна графіка, яка виражає умовну незалежність і залежить від випадкових змінних, пов'язаних з вузлами. Параметри складаються з умовного розподілу параметрів, пов'язаних з будь-яким вузлом.

Байєсівські мережі - це стандартна графічна модель, що приймає форму графіку ациклічної графіки (DAG) разом з розподілом ймовірностей [45].

Теоремою Байєса.

$$p(H_k | X) = \frac{p(X | H_k) \cdot p(H_k)}{\sum_{i=1}^n p(X | H_i) \cdot p(H_i)}, \quad k = \overline{1, n} \quad (2.9)$$

Сила байєсівських методів полягає в здатності уточнювати попередні ймовірності відповідно до фактичних реалій досліджуваного процесу. Це дозволяє нам вказати ймовірність події, коли надходить додаткова інформація.

Повна спільна ймовірність БМ обчислюється за формулою [46]:

$$P_B(X^{(1)}, \dots, X^{(N)}) = \prod_{i=1}^N P_B(X^{(i)} | Pa(X^{(i)})) \quad (2.10)$$

З математичної точки зору БМ є моделлю, яка представляє існуючі та відсутні ймовірнісні залежності.

Насправді байєсівська методологія є набагато більш комплексною, ніж кластер ліків маніпулювання умовними ймовірностями в спрямованих графах [47].

Цей параметр використовується у разі, якщо відомо про початкову інформацію щодо параметрів, визначених як апіорного розподілу параметрів.

У простому випадку Байєсова мережа визначається як експерт, а потім використовуються для пояснення. В інших додатках для людини надто багато визначень. У цьому випадку має бути структура бітової нейронної мережі і можливість локального поділу між даними [48].

Щоб повністю визначити Байєсова мережу, яка повністю розділяє параметри розподілу, кожен X повинен визначити параметри X відповідно до його основного типу. Розділення X залежить від батьків. Оскільки це спрощення, зазвичай це обмежений чи анемічний розподіл. Їх іноді називають обмеженнями розподілу і до них застосовується принцип максимальної ендотропії. Це визначає спільний

розподіл, який максимізує обсяг формування з урахуванням обмежень. (У певних контекстах логічно, що біологічні мережі, що відповідають умовним розподілам тимчасового розгортання прихованих станів, звичайно, побудовані для розсіювання рівнів ентропії.) Функції часто мають невідомі значення, які необхідно класифікувати за даними за допомогою таких методів, як максимальна ймовірність. Регулярне заповнення ймовірностей (або аптометричних ймовірностей) є складним завданням, часто через відсутність змінних.

Ця проблема пов'язана з алгоритмами очікуваних значень, які обчислюють очікувані значення на основі того, як визначено інформацію, за умови, що значення були обчислені правильно. Це може бути максимальне значення. Цей процес відповідає максимальному (або максимальному апотеролеону) значення параметра.

Багатофункціональний біосферний підхід до параметрів полягає в тому, щоб розглядати їх як додаткові недосліджені змінні, обчислювати загальну теологію розподілу всіх вузлів за даними, що спостерігаються, і інтегрувати їх. Цей підхід може бути дуже дорогим і може призвести до розмірів моделей, які зроблять традиційні підказки кориснішими.

Баєсівські підходи чудово підходять для машинного навчання. До байєсівських відносять група методів, за допомогою яких ми можемо показати, чи цінуємо дані, створені шляхом створення і пошуку оптимального розв'язання завдань, які аналізують поведінку процесів та систем різних природи.

2.5 Метод опорних векторів

Алгоритм опорний векторної машини - це контрольовані алгоритми машинного навчання, які використовуються для регресії а також класифікації. Цей метод використовують більше для класифікації. Професійне використання методу не є доцільним але є можливим. В залежності від кількості ознак розмірність гіперплощини варіативна.

На прикладі можемо розглянути площину, яка має наступний вигляд:

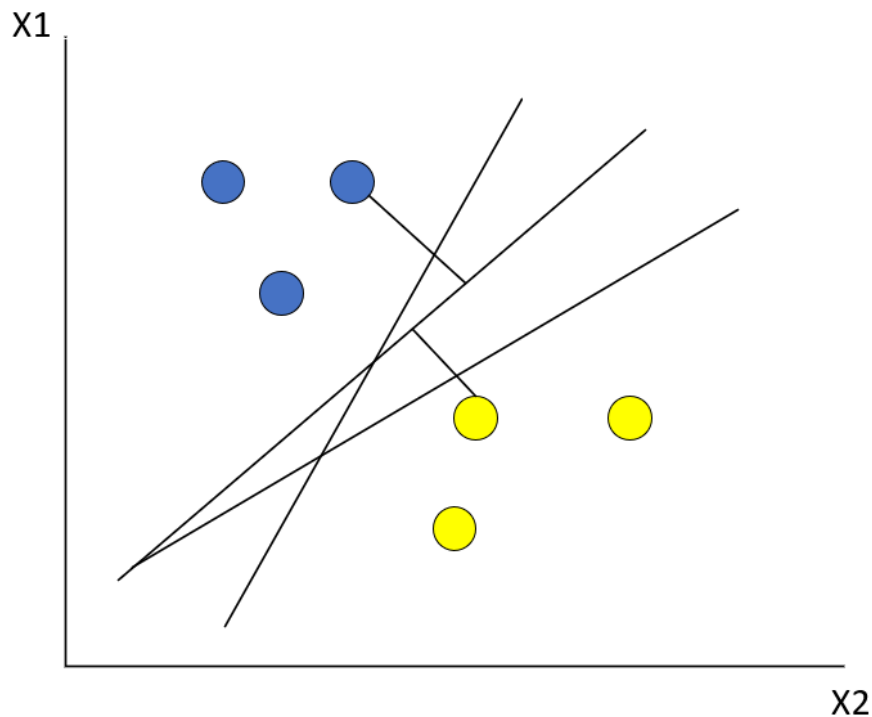


Рисунок 2.9 – Площина класифікації даних

$X1$ та $X2$ не є залежними змінними між собою, які мають тільки спільне поле обрахунку. Кольорові точки на рисунку позначають деякий набір даних, який відображає сутність кожного набору даних. Векторні лінії, які їх поділяють це є умови, які знаходяться на площині та розділяють їх за класами.

На прикладі видно, що вектори виконують класифікацію між синіми та жовтими колами. Наступним етапом потрібно вибрати найкращу гіперплощину у представленій схемі, щоб отримати площину, яка розділяє точки даних.

Для класифікацій площини необхідно задати параметри для кожної області класифікації та створити комбінацію різних типів класів. Таким чином безмежна кількість класів у вибірці шуканого і прогнозованого значення є комбінація одиночних k класів. Значення такої вибірки відображає формула:

$$L_n = \{1, \dots, n\} \quad (2.11)$$

де L це всі можливі класифікаційні варіанти, а від 1 до n це порядкові номери в класі інтервалів.

У цьому випадку необхідно створити безліч варіантів класифікацій, це допоможе збільшити площину вибірки, і збільшити регресійне прогнозування за допомогою опорних векторів. Також, при використанні функцій залежності, коли спостереження за роботою класу відноситься до k , функція приналежності виглядає наступним чином:

$$m_{ki}(x) = \min(1, D_i(x)), \text{ for } i \in L_k \quad (2.12)$$

З цієї формули можна визначити, що функція належності опорних векторів до вибраного регіону площини вимірюється методом пошуку найближчої гіперплощини до даного спостереження. Якщо значення пошукових даних є від'ємними, то висновок цього результату є те, що діюча належність не відноситься до жодного з класу. Якщо це трапляється, то спостереження відноситься до найближчого класу, який може відповідати параметрам класу.

Ядро опорних точок на площині - це функція, яка перетворює вхідний простір низької розмірності у простір великої розмірності N . Перетворення нероздільних задач у зрозумілі. Це в основному корисно для задач нелінійного розділення. Простіше кажучи, ядро знаходить процес, який виконує дуже складне перетворення даних і розділяє дані на основі заданої мітки або виведення.

2.6 Випадковий ліс

Алгоритми для прогнозування бувають дуже не типові для використання в тих чи інших системах. Зокрема використання алгоритму випадковий ліс - це контрольований алгоритм машинного навчання, який розкриває потенціал

алгоритму створення умов для подальшого навчання. Цей алгоритм бере за основу створення нейронів, які поєднуються між собою та створюють масштабне навчання, виконуючи певні задані умови з параметрами обмеження, які впливають на результат навчання.

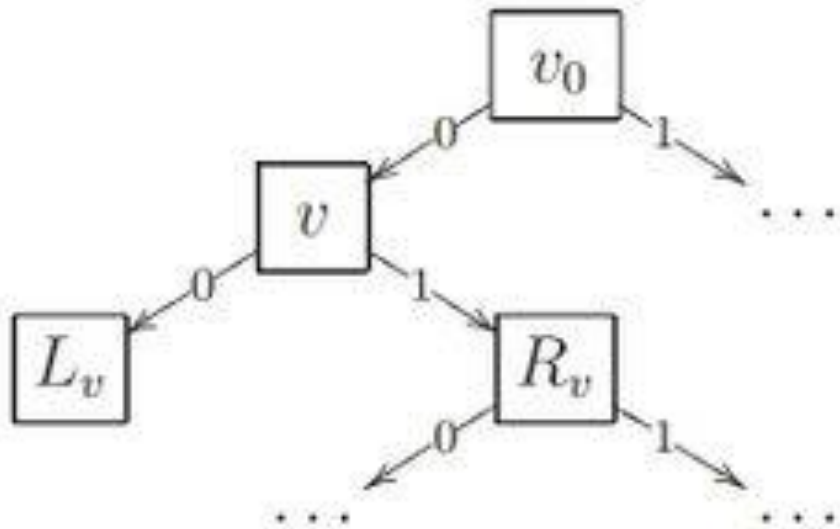


Рисунок 2.10 – Схема дерева рішень

Випадковий ліс використовується не тільки для завдань класифікації машинного навчання так і для регресії. Кероване машина навчання це коли алгоритм або модель алгоритму створюється за певним набором вхідних та вихідних даних, таким чином алгоритм навчається та класифікує будь-які на вихідні дані які вона отримує. Цей алгоритм доцільно використовують для систем яких потрібна прогнозування майбутні результатів. Використання алгоритму на базі машинного навчання випадкового лісу досить вдала, реалізував тему магістерської роботи а саме прогнозування погоди за допомогою штучного інтелекту.

Математична модель регресійної помилки, в які дані розгалужуються від вузла, який знаходиться в системі, знаходиться по такій формулі:

$$M = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (2.12)$$

Для даної формули N - це кількість точок даних, вона може бути необмежена, f_i – значення яке повертає модель, тобто результат обробки, y_i – фактичне значення для кожної i .

Цей алгоритм працює в двох варіантах використання, а саме класифікаційна аналізі та регресійному аналізі. У класифікаційному аналізі базова ознака залежна. Завдання на класифікація й спроба визначити мітку класу для предметної області навчання, тобто навчання йде типовим за алгоритмом який завдає людина. У регресійного аналізі залежно атрибут має бути тільки числовим. Регресія використовується коли вихід на змінна є реальна або безперервним значенням. Впроваджуючи випадкові ліси на основі класифікації, потрібно використовувати індекс Джіні, або формулу визначення вузлів у початковій гілці дерев.

$$Gini = 1 - \sum_{i=1}^c (p_i)^2 \quad (2.13)$$

Формула бере клас i ймовірність прогнозу для визначення Джіні кожної гілки у вузлі, визначаючи яке дерево має більше ймовірність з'явитися.

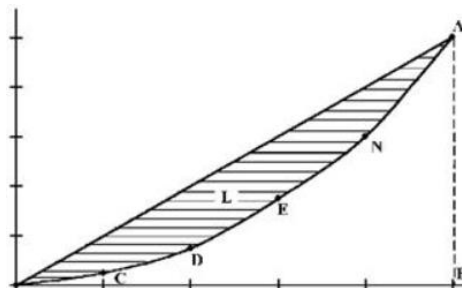


Рисунок 2.11 – крива Лоренса

Для розуміння що таке дерево рішень можна навести приклад який буде виглядати наступним чином.

Короткий опис алгоритму полягає в наступному якщо ми маємо деякі питання в ми можемо на нього відповісти або так або ні і залежності від відповіді ми маємо отримати результат або числовий або у відсотках та продовжити вирішення питання тим самим способом на наступних етапах. Після завершення етапу питання так чи ні можна дійти до кінцевого результату.

Алгоритм дерева рішень створює декілька дерев рішень, щоб об'єднати їх для точного прогнозування логіка такої моделі полягає в тому, що кілька або більше неконтрольованих моделей тобто окремих рішень які не сумісні між собою. Використовуючи цей алгоритм ви повинні внести початковий набір даних з функціями обмеження чи розширення і повинні сформулювати певний набір правил які використовуватиме алгоритм для прогнозування. Таким чином кожне дерево вибери спостереження за певним типом даних для побудови певних дерево рішень. Результатам роботи декількох дерев буде у середньому результат, що дасть більш правильне рішення у прогнозуванні.

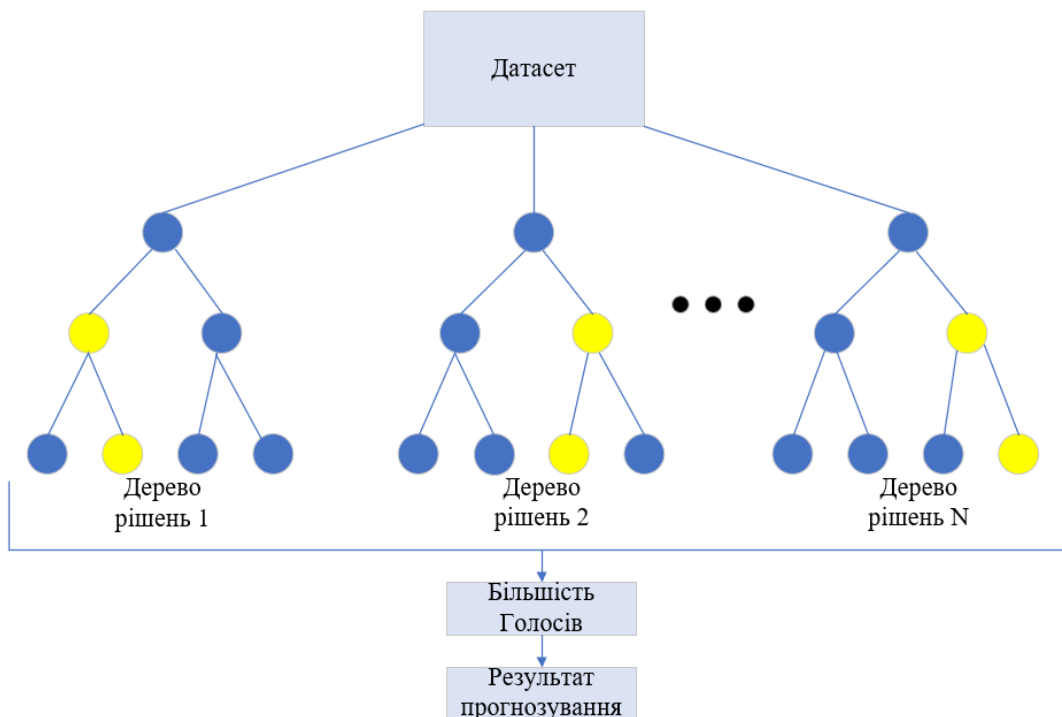


Рисунок 2.12 – Структура випадкового лісу

Тренування дерева у випадковому лісі починається з-за допомогою методу анкетування. Метод анкетування це тип алгоритму ансамбль машинного навчання під назвою *bootstrap aggregation*. Методооб'єднання прогнозу з декількох алгоритмів та об'єднує їх результати.

Дисперсія може виникати через чутливість методу до великих коливань у наборі даних які використовуються для навчання. Велика дисперсія може призвести до того, що алгоритм буде відображати не релевантні дані які можуть посилюватися з кожним наступним прогнозом. У такому випадку потрібно змінювати та допрацьовувати модель під час навчання. Це допоможе уникнути великих похибок під час фактичного тестування та роботи моделі в цілому.

Також використовують ентропію, це дозволяє визначати як саме розгалужуються взаємозв'язки між вузлами дерева рішень.

$$Entropy = \sum_{i=1}^c -p_i * \log_2(p_i) \quad (2.14)$$

Найбільшу перевагу цього методу інтерпретує в тому, що цей алгоритм має здатність виявляти похибки на етапі навчання та коригувати їх. Також під час навчання випадковий ліс, який має багато параметрів, дає значно нижчу похибку ніж інші методи. Чим більше методів включено в цей алгоритм, тим похибка буде знижуватись а прогнозування становитись значно точним. Отже випадковий ліс - це контрольований алгоритм у машинному навчанні, що складається з дерев рішень.

3 Розробка методу прогнозування погоди

3.1 Вибір інструментів розробки

Python – це мова програм, яка підтримує декілька парадигм програм, у тому числі програмне забезпечення та функціональне забезпечення. Об'єктивна, високо синхронна, добре розшифрована програма. Він також відомий за динамічним семантичним і величезними можливими можливостями швидкого розвитку додатків, і він також широко застосовується для створення скриптів, а Python досить універсальний і тому щоденно стає популярною мовою програмного забезпечення [49].

Він є широким спектром завдань, і його зазвичай використовують у багатьох галузях, у тому числі в різних сферах :

- веб-розробка,
- аналіз даних,
- машинне навчання ,
- DevOps і системне адміністрування,
- автоматизоване тестування,
- прототипування програмного забезпечення та багато інших.

Крім того, Python надає змогу розробникам програми та запуском прототипи, що значно пришвидшує розвиток. Після того як проект знаходиться на шляху до аналітичних інструментів або добавок, ми можемо перевести його на складніші мови, як - от Java чи C, якщо необхідно [50].

На сьогоднішній день в пакетах Python налічується понад 70 000 бібліотек, кількість їх продовжує зростати досі.

Python придатний для пришвидшення копіювання програмного обладнання, тому що він просто використовується і підтримує TDD (test Driven development) для вилучення деяких компонентів програми відразу після її створення [51].

Однією з характерних особливостей розвитку в середовищі Python є його модуль, в якому не всі функції і фреймворки доступні відразу, але ви повинні бути імпортовані з інших моделей.

Ключеві переваги Python:

- чистий синтаксис;
- портативність програм;
- стандартне розповсюдження містить багато корисних модулів;
- з легкістю вирішує задачі.
- відкритий код

Програмне забезпечення на Python створені як модулі, які можуть бути зібрані у пакетах.

TensorFlow – це друзі з Python, він з відкритим кодом розробки цифрових підрахунків, який дозволяє користувачам навчатися швидше і простіше [52].

Це платформа, на якому ви можете обробляти дані, створювати і тренувати моделі технічного навчання, доглядати за моделями з різних мов програмування.

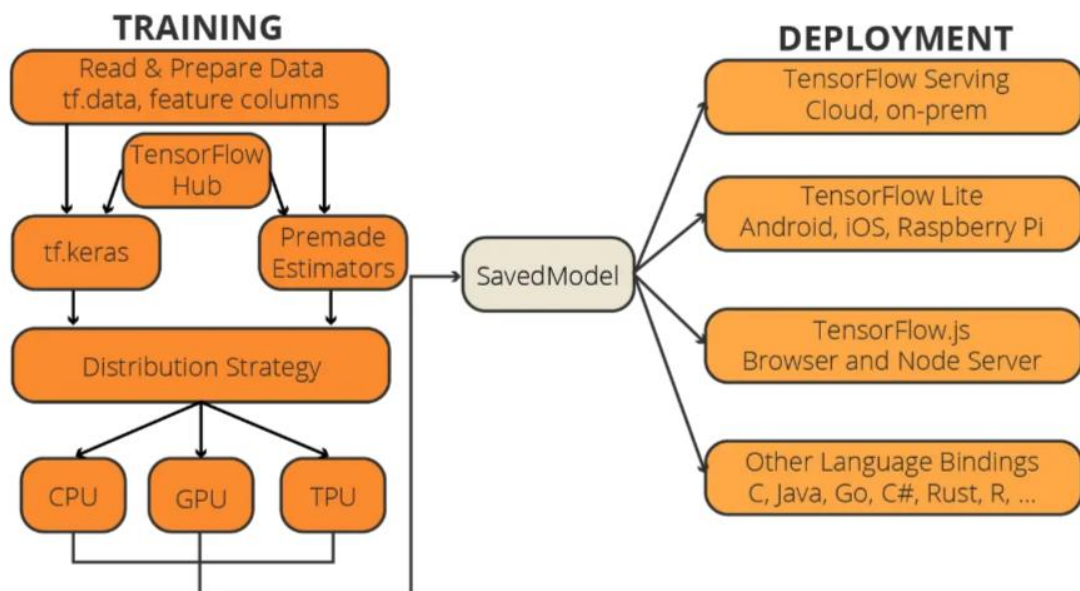


Рисунок 3.1 Поточна діаграма платформи TensorFlow

TensorFlow може вчитися і виконувати глибокі нейронні мережі, які виконують класифікацію цифр, розпізнавання малюнків, вставлення слів, рекреаційні мережі, моделі від послідовності до послідовного перекладу, редагування природної мови і дизайн процесів на основі PDE [53].

Завдяки цим розробникам можна створити віджети даних — структуру, яка описує, як вони пересуваються через граф або декілька вузлів. Завдяки мови програмування Python, TensorFlow надає все для програміста.

Colab, — це продукт від Google Research, безкоштовне інтерактивне середовище хмар для кодів програм. Код працює на віртуальному комп'ютері, який є вашим обліковим записом. Якщо віртуальна машина, після певного часу простою не запускалась, то віртуальні машини видаляються, а також мають максимальний проміжок життя, який встановлений Colab службою [54].

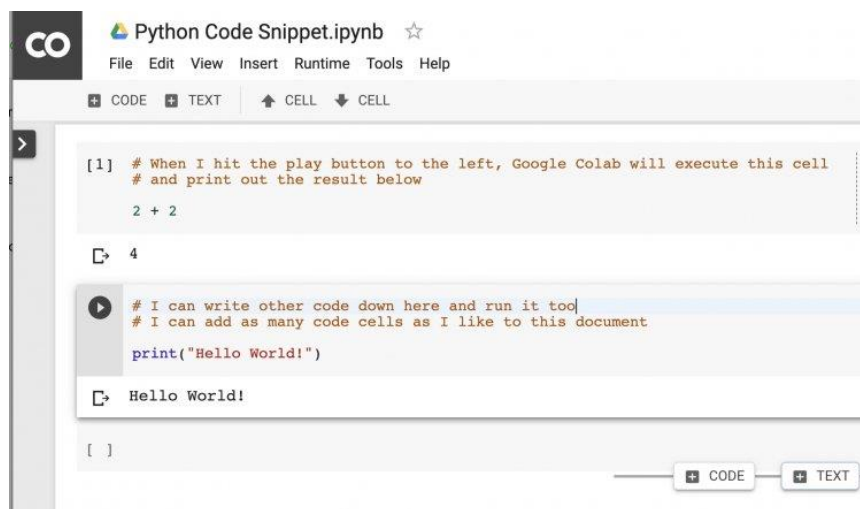


Рисунок 3.2 – інтерфейс Google Colab

Це дуже швидкий режим для перевірки коду без потреби встановити обладнання і увійти в систему машинного навчання, глибоке навчання, штучний інтелект та наука про дані [55].

Google Colab це гарний інструмент та підійде всім, хто працює з Big Data, а саме:

1. інженерам даних;
2. аналітикам даних;

3. дослідникам даних.

Можна безкоштовно користуватися із браузера за допомогою Google Colab, можна одразу написати або виконати код на Python, якщо код має математичне рівняння, то можна його прокоментувати, також можна створити завантажити або надати доступ до блокноту та тощо.

3.2 Оцінка та обробка вхідних даних

Високоякісні дані є передумовою для аналізу та використання великих даних, а також для забезпечення цінності даних. Наразі немає всебічного аналізу чи дослідження стандартів якості та методології для оцінки якості великих даних. По-перше, у статті підсумовуються огляди досліджень щодо якості даних. По-друге, у статті аналізуються характеристики даних у середовищі великих даних, окреслюються проблеми якості, з якими стикаються великі дані, і створюється ієрархічна структура якості даних з точки зору користувача даних.

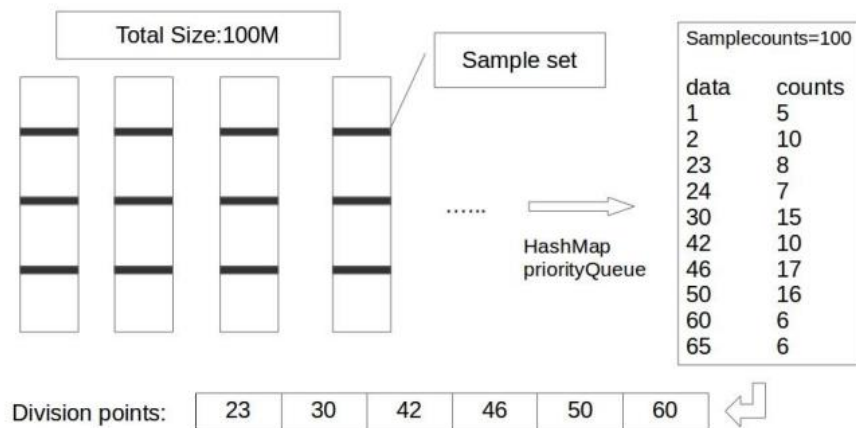


Рисунок 3.3- Приклад сортування даних

Ця структура складається з параметрів якості великих даних, характеристик якості та індексів якості. Нарешті, документ будує на цій основі процес динамічного оцінювання якості даних. Цей процес є високомасштабованим і адаптованим і може задовольнити вимоги оцінки якості великих даних.

Для того щоб оцінити характеристику вхідних даних потрібно для початку їх отримати. Отримання може бути або шляхом ручного заміру або використання спеціальних технічних датчиків або метеостанцій, які отримують стан характеристик складових погоди. Використання офіційних джерел погодних умов у реальному часі має в основі впевненість у тому що дані в реальному часі будуть достовірними та правдивими. Якщо використовувати власні датчики або використовувати параметри які можна отримати самостійно, наприклад температуру з градусника, то такі параметри можна вважати неточним.



Рисунок 3.4 – схема отримання релевантних даних

Якщо використовувати сукупність вимірювальних приладів то середня арифметичне значень буде більш точно відображати стан атмосфери у конкретній області прогнозування.

За періодом отримання даних у відкритих джерелах є наступні показники:

- Кожну хвилину
- Кожні 10 хвилин
- Кожну годину.

Для більш точного прогнозування краще використовувати дані з проміжком часу хвилини або менше. Чим більше показників виміру тим складніше

обчислювати але точність прогнозу збільшується завдяки аналізу отриманих даних. Найкращим варіантом для такої системи прогнозування погоди та погодніх явищ це вимір та отримання даних в реальному часі щосекунди або менше. Середнє відхилення таких систем буде мінімальним. Для прогнозування в дикій місцевості прийнято вважати що інформація про стан погоди в реальному часі може мати однакові показники в радіусі 5 км.

Тобто для покриття міста потрібно використовувати багато обладнання для виміру стану атмосфери. Якщо аналізувати багато таких даних, та використовувати більш обширний системи в радіусі більше 20 км.

Хто система буде у середньому погоду. Це призведе до того що стан погоди буде відхилятися від нормально осі прогнозування. Отримані дані будуть похибкою та недостовірними.

Чим більше вхідних даних з конкретної точки тим менші показник похибки прогнозування. Взаємодія між різними показниками є дуже корильованими, тому що кожний показник впливає в той чи інший момент часу на інший. Із прикладу вологість де є обмеження по нижній точці. Але якщо додати вплив переміщення повітряних мас, тобто вітру, то ці показники не будуть обмежувати а збільшувати коефіцієнт нижньої точки температури. Після отримання даних вони розподіляються, сортують ся та доводяться до стану великих величин.

Після отримання даних у відкритому доступі з різних метеорологічних станцій процес обробки даних полягає в наступному. Інформацію Від data аналізує, займаючись процесами зберігання й управління великими обсягами даних та оновлення інформації.

Метеорологи використовують датчики та збирають великі обсяги даних від них для того щоб відстежувати стан навколишнього середовища для вивчення :

- Розуміння впливу глобального потепління
- Створення прогнозу погоди.
- Вивчення стихійних лих.
- Рання прогнозування про небезпечні ситуації.

Після отримання даних потрібно використовувати методи сортування великих даних.

Це потрібно щоб правильно задати показники в реальному часі та отримувати нові дані. Сортування і обробка даних проходить виключено тільки за допомогою релевантних зв'язків. Після повної сортування виключення всіх похибкових даних, очищення даних від не коректних даних починається оцінка характеристик вхідних даних.

3.3 Реалізація методу прогнозування.

Реалізація методу прогнозування починається з отримання датасета зовнішніх даних параметрів погоди метеорологічних датчиків.

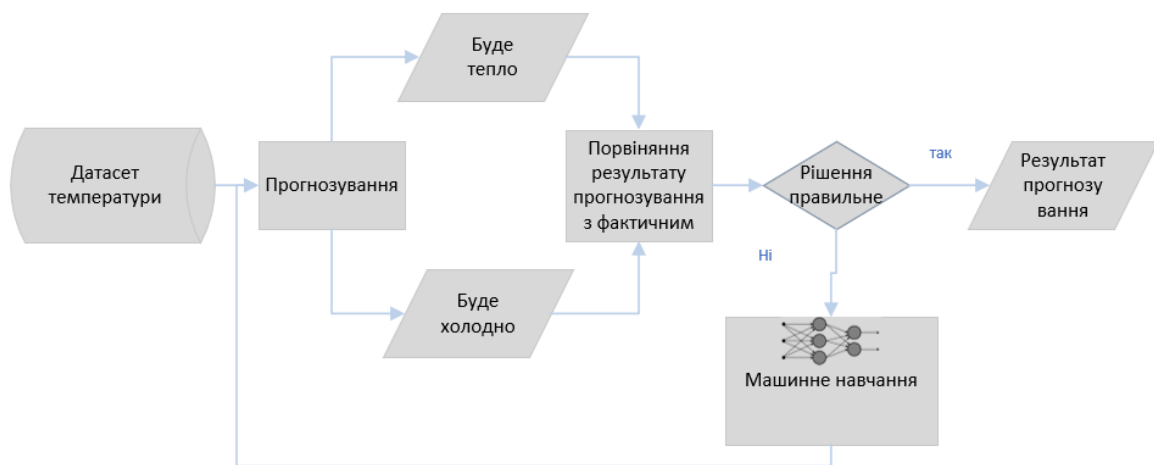


Рисунок 3.5 – схема роботи методу прогнозування

За допомогою Big data всі дані обробляються щоб отримати табличні дані показників попереднього періоду та в режимі реального часу. Далі всі дані передаються на метод штучного інтелекту та аналізуються попередній період даних до поточного стану навколишнього середовища. На цьому етапі чим більше програма отримує попередніх даних, точніше буде прогнозування. Машина

починає своє навчання. Модель штучного інтелекту у прогнозуванні полягає в тому щоб використати машиною навчання за декількома алгоритми машинного навчання. Об'єднуючи результат тренування машинного навчання за декількома алгоритмами дає перевагу у прогнозуванні.

В методі використовуються такі алгоритми тренування передбачень:

- Регресії.
- Випадкового лісу.
- Регресія найближчого сусіда
- Лінійної регресії.

Навчання полягає в тому щоб брати перші значення та робити прогнозування на наступні значення.

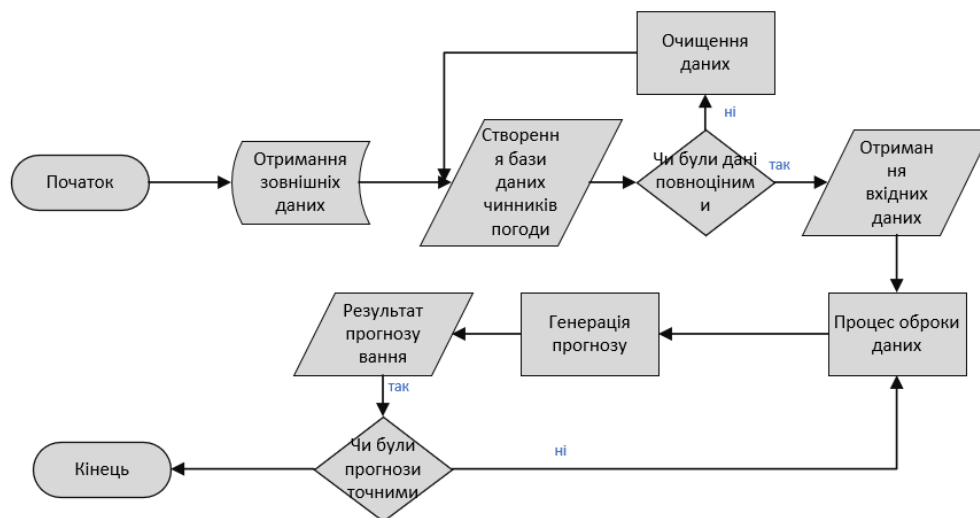


Рисунок 3.6 - Повний цикл методу прогнозування

Наприклад брати показники протягом двох днів робити прогноз якщо програма правильно навела прогноз на наступні дві години то алгоритм записується як правильне рішення. Таким чином крок за кроком програма аналізує поетапно кожний наступний день збільшуючи свою точність у прогнозі та зменшуючи похибку прогнозування.

Такий метод називається навчання з підкріпленням. Кріплення навчання іде за допомогою попередніх даних.

Наступним кроком ми називаємо значення та формули взаємодії між всіма атрибутами даних. Після прогнозування ми повинні отримати результат прогнозування на потрібний проміжок часу.

3.4 Розробка методу прогнозування погоди

Для розробки методу використовується Python та Colaboratory.

Отримаємо попередні дані погоди в форматі CSV. Формат отриманих даних включає:

- Температура.
- Вологість.
- Швидкість вітру.
- Напрямок руху вітру.
- Сонячна радіація.

02.09.2022 14:30		37	0	84	767	0,3	0
02.09.2022 14:40	19,4	37	2,5	82	767,2	0,3	0
02.09.2022 15:10	20,6	36	1,1	81	767,1	0,3	0
02.09.2022 15:30		36	6,1		767	0,3	0
02.09.2022 15:50	21,6	36	1,1	67	766,9	0,3	0
02.09.2022 16:00	21,9	36	1,1	156	766,8	0,3	0
02.09.2022 16:20	21,4		1,1			0,3	0
02.09.2022 16:50	20,7	35	2,5	45	766,7	0,3	0
02.09.2022 17:10	20,9	36	1,1	106	766,6	0,3	0
02.09.2022 17:20	20	36	1,1	308	766,7	0,3	0
02.09.2022 17:30	19,2	36	4,7	89	766,7	0,3	0
02.09.2022 17:40	19,1	35	1,1	13	766,6	0,3	0
02.09.2022 18:00	18,1	35	4,7	288	766,6	0,3	0
02.09.2022 18:30	17,8	35	0	288	766,7	0,3	0
02.09.2022 18:50	17,2	36	1,1	145	766,7	0,3	0
02.09.2022 19:10	16,7	37	0	284	766,9	0,3	0
02.09.2022 19:20	15,9	36	0	252	766,9	0,3	0
02.09.2022 19:40	14,4	37	0	256	766,9	0,3	0
02.09.2022 20:10	13,4	37	0	243	766,9	0,3	0
02.09.2022 20:30	13,7	37	0	58	767,2	0,3	0
02.09.2022 21:00	13	37	1,1	234	767,2	0,3	0
02.09.2022 21:20	13,4	37	0	351	767,2	0,3	0
02.09.2022 21:40	13,5	37	0	351	767,3	0,3	0
02.09.2022 22:00	13,9	37	0	321	767,4	0,3	0
02.09.2022 22:10	14	37	0	314	767,6	0,3	0
02.09.2022 22:40	14,1	37	1,1	50	767,3	0,3	0
02.09.2022 22:50					767,4	0,3	0
02.09.2022 23:00	13,9	37	0	19	767,4	0,3	0

Рисунок 3.7 – Датасет атрибутів погоди

Як можна побачити дані не очищені та не заповнені. Вони потребують певного налагодження, заповнення та очищення. Метрологічні датчики зчитують атрибути показників з датчиків кожні 10 хвилин.

```

body = soup.find("body").text
body = body.split(",")
# weather.append(temp)
day = body[-27][4:-1]
if (day[0] == '0'):
    day_2 = day[-1]
else:
    day_2 = day
weather.append(day_2)
month = body[0][-1]
weather.append(month)
weather.append(year)

temp = body[-24][1:-1]
weather.append(temp)

high = body[-25][1:-1]

low = body[-23][1:-1]
weather.append(high)
weather.append(low)

Humidity = body[-21][1:-1]
weather.append(Humidity)
dew = body[-22][1:-1]
weather.append(dew)
pressure = body[-26][1:-1]

```

Рисунок 3.8 – Базове отримання та обробка вхідних даних

Для короткотривалого і довготривалого прогнозу це достатньо, щоб мати невелику похибку відхилення від прогнозу. Незаповнені клітинки заповнюються середнім арифметичним між попередніми та наступними даними. Використання середніх значень можна записати за формулою:

$$n_i = \frac{n_{i-1} + n_{i+1}}{2} \quad (3.1)$$

Після заповнення всіх значень весь датасет обробляє кожен алгоритм, починається машинне навчання.

Паралельно працюючи декілька алгоритмів можуть через усереднене значення видавати більш точний прогноз. У цьому варіанті перевага надається алгоритму випадкового лісу, його ваги мають більш точний прогноз, але все одно

результат всього навчання залежить від всіх алгоритмів в цілому. Після успішної обробки та отримання датасету починається машинне навчання та аналіз.

Програма аналізує значення попередніх показників і створює власний прогноз. Якщо параметрів прогнозу замало, то процес навчання почнеться з базового отримання нових параметрів кожні 10 хвилин, за які оновлюється датасет.

```
def OnClicked_2(self, event):
    btn = event.GetEventObject().GetLabel()
    os.system('py -3.6 -m Main_Prediction.py')
    print ("Finished Prediction by Button = ",btn)

def OnToggle(self, event):
    state = event.GetEventObject().GetValue()

    if state == True:
        event.GetEventObject().SetLabel("Deactivate Auto-Prediction")
        while(state==True):
            print ("Enabled Auto-Prediction (daily update)")
            os.system('py -3.6 -m Main_Forecast.py')
            print("Finished daily update. Wait for next day.")
            i = 1
            while(i<20):
                print("Please wait "+str(20-i)+" seconds")
                time.sleep(1)
                i+=1
    else:
        print (" Disabled Auto-Prediction")
        event.GetEventObject().SetLabel("Activate Auto-Prediction")

app = wx.App()
Mywin(None, 'Weather Prediction using Machine Learning')
```

Рисунок 3.9 – Оновлення прогнозу

Чим більше попередніх показників, тим точніше буде прогнозування. Дані прогнозу записуються та порівнюються з наступними даними. Цей процес також може відбуватися у режимі реального часу.

Алгоритм штучного інтелекту повинен навчитися розуміти зміни, і видаючи прогнозування погодних явищ. Задані параметри дають можливість робити правильний прогноз, і відслідковувати нетипові зміни.

Якщо є порушення між прогнозованим значенням, яке отримала програма і значенням, яке буде відповідати дійсності, то можна зробити висновок, що є якісь порушення або в роботі датчиків метеорологічної станції, або це може бути ознакою різкої зміни погодних явищ або виникнення нетипового катаклізму.

```

# In[12]:
pred_rf = rf.predict(X_test)
pred_rf
y_pred = rf.predict(X_test)

# In[16]:
import seaborn as sns

g = sns.regplot(y_pred, y=y_test, fit_reg=True)
g.set(xlabel='Predicted Mean Temperature', ylabel='Actual Mean Temperature', title='Regression Plot for Actual vs Predicted Values')

# In[17]:
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df1 = df.head(25)
df1

# In[18]:
df1.plot(kind='bar', figsize=(10, 8))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.savefig("statistics/random_forest_comparison.png")
plt.show()

```

Рисунок 3.10 – метод навчання випадковим лісом.

Після отримання результату можна запросити побудувати графік на проміжному етапі прогнозування. Це може бути виведення графіку прогнозування чи співвідношення відсотків вірогідності між отриманим або не отриманим результатом. Якщо результат прогнозування задається графіком, то на виході ми отримаємо результат графіку кожного компонента змінної, а саме окремо температури, вологості, напрямку вітру.

3.5 Результат ефективності прогнозування

Результативність прогнозування можна порівняти лише з достовірними та наявними результатами прогнозування інших систем. За основу було взято популярний сервіс прогнозування, за яким було спостереження. Протягом 3 днів створено запис даних прогнозування, їх можливе відхилення, та порівняльний аналіз даних між методом прогнозування погоди за допомогою штучного інтелекту та аналізом прогнозування за допомогою синоптичних карт.

Таблиця 3.1 – порівняння результатів прогнозування

	Синоптичний Прогноз	Прогноз Штучного інтелекту	Покращення прогнозування погоди
Відносна похибка температури за 24 години, %	4,77	4,31	10,6
Відносна похибка вологості за 24 години, %	6,35	5,88	8
Відхилення Прогнозу за 24 години, %	6,20	5,54	11,9
Відхилення Прогнозу за 120 години, %	15,21	12,35	23,15

За даними результатами можна підтвердити ефективність роботи штучного інтелекту у прогнозуванні погоди, використання обмеженої кількості параметрів погодних ознак не зменшує рівень показників прогнозу.

Відносна та статистична похибка прогнозування є суттєво різною та значущою між системами прогнозування.

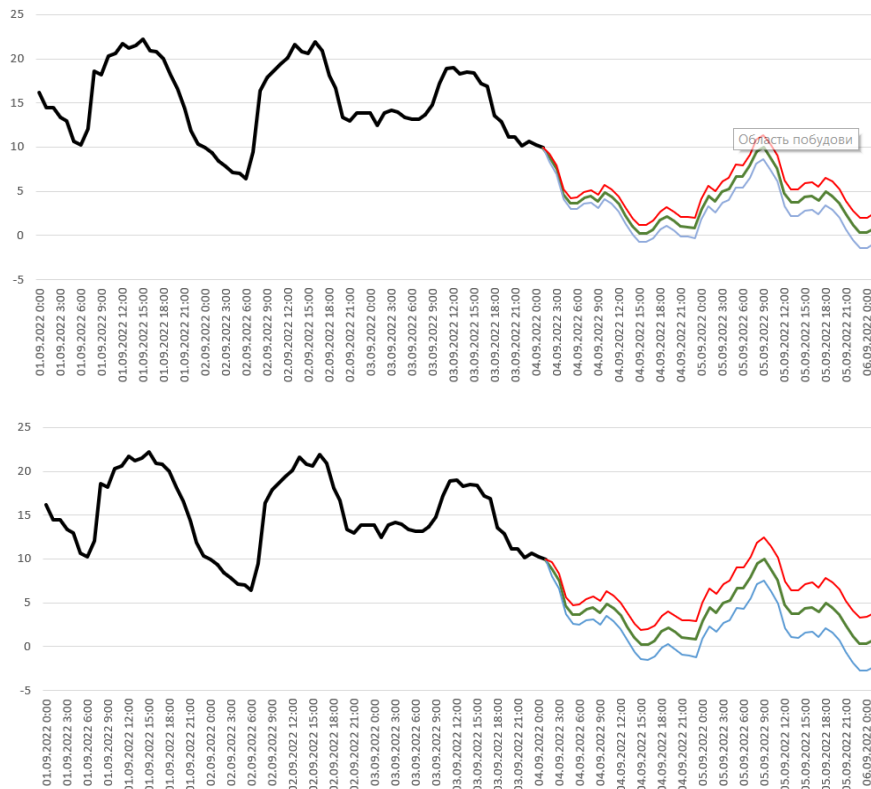


Рисунок 3.11 – Порівняльні графіки прогнозування.

За представленим рисунком можна побачити різницю між результатами прогнозування наочно. Зверху графік відображає результат прогнозування за допомогою штучного інтелекту. Інший графік показує прогнозування синоптичного характеру. Чорна частина графіку це період спостереження та навчання системи. Червоним та синім відображаються граничні точки прогнозування і можливе відхилення. На нижньому синоптичному графіку можна помітити, що розбіжність в цілці прогнозування сягає значного відхилення

Створений метод прогнозування погоди на основі штучного інтелекту має переваги, а саме масштабованість, точність, довготривалість. За результатами прогнозування можна побачити прогнозування точне на короткостроковій дистанції на 8 відостків, довготривалому прогнозі різниця між відхиленням у системах синоптичного прогнозу та прогнозу штучного інтелекту складає майже 23.15 відсотків.

ВИСНОВКИ

В результаті виконання магістерської роботи була здійснена розробка та моделювання методу прогнозування погоди та погодних явищ з використанням штучного інтелекту.

У роботі було детально проведено аналіз типів та методів сучасного прогнозування погоди та погодних явищ. Здійснено розробку методу прогнозування погоди з використанням штучного інтелекту за допомогою машинного навчання.

Використання штучного інтелекту для прогнозування погоди та погодних явищ дає поняття про майбутній стан навколишнього середовища. Це дозволить дізнаватися не тільки прогноз погоди, а створювати плани на майбутнє не тільки для людини та багатьох галузей використання, які залежать від актуального та майбутнього стану погоди. Завдяки цьому методу можна дізнатись та прогнозувати погоду в конкретній місцевості без використання складного та коштовного обладнання, і за даними, які є в наявності. Одна з причин штучного інтелекту полягає в тому, що вони використовують універсальні алгоритми, які виконуватимуть свою функцію незалежно від налаштувань, які буде використано у цьому обмеженому ресурсі комп'ютера.

З розвитком технологій штучного інтелекту зростає зацікавленість людей у використанні прогнозування погоди та передбачені погодних катаклізмів та явищ. Погода дуже сильно впливає на людину. Щоб зрозуміти, що станеться з погодою в найближчі дні, нам потрібно знати, як виглядає атмосфера і які зміни відбулися за останні кілька днів. Такий спосіб підвищує точність прогнозування, так як штучний інтелект вміє обробляти дуже багато даних та проаналізувати всі можливі випадки погоди.

Було розроблено моделі прогнозів погоди, які дають можливість прогнозувати погоду за принципом штучних розумових здібностей, за допомогою дерева прийняття рішень і регресійної моделі, аналізу попередніх показники, короткими та тривалими прогнозами погоди.

За результатами проведено аналіз моделювання методу та порівняно з вже існуючими моделями прогнозування погоди, визначено їх сильні та слабкі сторони у рамках обмежених параметрів вхідних даних від пристроїв збору інформації про навколишнє середовище.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Мікромасштабними погодніми явищами, J.S.M. Coleman, K.T. Law, in Reference Module in Earth Systems and Environmental Sciences, 2015 <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/meteorological-phenomenon> (дата звернення: 20.10.2022)
2. Метеорологічні надзвичайні ситуації, <https://e-kolosok.org/naukova-robotanespryatlyvi-pohodno-klimatychni-umovy-mista-chernivtsi/>(дата звернення: 22.10.2022)
3. Врублевська О. О. Кліматологія : підручник . упоряд.:О. О. Врублевська, Г. П. Катеруша, Л. Д. Гончарова ; МОН України ; Одес. держ. еколог. ун-т. Одеса : Екологія, 2013. С. 249–256.
4. Путренко В.В., Назаренко С.Ю. Визначення якості повітря на основі інтелектуального аналізу даних дистанційного зондування / В.В. Путренко, С.Ю. Назаренко // Математичне моделювання в економіці – 2016. –№3-4 – с.176-187.
5. Основні методи для складання прогнозів <https://energyline.com.ua/uk/tarif-elektroenerg/metodi-skladannya-prognoziv-viroblennya-energi%D1%97-ses/> (дата звернення: 28.10.2022)
6. Методи прогнозування погоди <https://www.preservearticles.com/essay/what-are-the-methods-used-for-weather-forecasting/15385> (дата звернення: 29.10.2022)
7. Morris R. Workshop Summary Kasparov vs. Big Blue: The Significance for Artificial Intelligence. ICGA Journal. 1997. Vol. 20.
8. Bobbie Johnson 2021, Штучний інтелект <https://uk.warbletoncouncil.org/inteligencia-artificial-2346> (дата звернення: 01.11.2022)
9. Artificial Intelligence In Weather Makes The Case For Human Input <https://www.forbes.com/sites/rennyvandewege/2022/02/17/artificial-intelligence-in-weather-makes-the-case-for-human-input/?sh=ecc7a615c477> (дата звернення: 01.11.2022)

10. Зайченко Ю.В. Основи проектування інтелектуальних систем. Навчальний посібник. К: Видавничий Дім «Слово», 2004. – 352 с 119 с
11. Логіко-інформаційна система // / В. І. Шинкарук (гол. редкол.) та ін. — Київ : Інститут філософії імені Григорія Сковороди НАН України : Абрис, 2002. Ст.21-25.
12. Інформаційна система у технологіях прийняття рішень <https://www.hitechnectar.com/blogs/intelligent-information-system/> (дата звернення: 02.11.2022)
13. Intelligent information system <https://www.hitechnectar.com/blogs/intelligent-information-system/> (дата звернення: 04.11.2022)
14. Загальні відомості про системи із штучним інтелектом <https://library.if.ua/book/100/6874.html> (дата звернення: 05.11.2022)
15. Рисунок нейрону людського мозку, <https://znai.com.ua/budova-neyrona-ta-yogo-funkc/> рис. (дата звернення: 05.11.2022)
16. Paugam-Moisy H. Computing with Spiking Neuron Networks / H. PaugamMoisy, S. Bohte // Handbook of Natural Computing / H. Paugam-Moisy, S. Bohte., 2012.
17. Нейронні мережі та глибоке навчання <https://evergreens.com.ua/ua/articles/neural-network.html> (дата звернення: 06.11.2022)
18. Krenker A. Introduction to the Artificial Neural Networks / A. Krenker, J. Bester, A. Kos // Artificial Neural Networks - Methodological Advances and Biomedical Applications / A. Krenker, J. Bester, A. Kos., 2011.
19. Rajesh Bordawekar, Bob Blainey, Ruchir Puri, Analyzing Analytics. – Morgan & Claypool Publishers, 2015. – 124 - 127 p.
20. Kriesel D. A Brief Introduction to Neural Networks, 2007, http://www.dkriesel.com/en/science/neural_networks
21. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення. – К.: Інститут Математики НАН України, 2004. – 408с. http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnNejronMeregNester2004.pdf.

22. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення. – К.: Інститут Математики НАН України, 2004. – 40с.
http://www.immsp.kiev.ua/postgraduate/Biblioteka_trudy/ShtuchnNejronMeregNester2004.pdf

23.Стаття про машинне навчання
<https://aiconference.com.ua/uk/news/mashinnoe-obuchenie-prostimislovami-oslognoy-tehnologii-97834> (дата звернення: 06.11.2022)

24. M. Swamynathan, Mastering Machine Learning with Python in Six Steps, DOI 10.1007/978-1-4842-2866-1

25. Grigorov O.V., Svirgun V.P. Improving the productivity of utility cranes through optimum motion control. Soviet machine science. 1986. P.110-115

26. Stephen Russo. IBM WW Dir Cognitive City solutions and public safety «Adapting to the Changing Threat Landscape: Security vs Convenience». Матеріали конференції IT Weekend Ukraine 2017-10-09, Київ, Україна

27. Types of Machine Learning Algorithms You Should Know. URL :
<https://towardssdatascience.com/types-of-machine-learning-algorithms-you-shouldknow-953a08248861>

28. Raschka S., Mirjalili V. Python Machine Learning. Livery Place: Packt Publishing Ltd, 2019. 296 с.

29. P. Kim. MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence. Apress. 151 p.

30.Nikhil Ketkar. Deep Learning with Python: A Hands-on Introduction. Apres, 2017, 226 p.

31 Н. Вірт. Систематичне програмування. - М.: Світ, 1977. - С. 94-164.

32.Data Processing for Artificial Intelligence (AI)
<https://www.edatamine.com/how-crucial-is-data-processing-for-machine-learning-ml-artificial-intelligence-ai/> (дата звернення: 07.11.2022)

33. Е. Дейкстра. Нотатки з структурному програмуванню / / У. Дал, Е. Дейкстра, К. Хоор. Структурне програмування. - М.: Світ, 1975. - С. 24-97

34. Khan Z, Anjum A, Kiani SL. Cloud Based Big Data Analytics for Smart Future Cities. In Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. IEEE Computer Society; 2013. pp. 381–386.
35. Batty M. Big data, smart cities and city planning. *Dialogues Hum Geog.* 2013;3(3):274–9
36. D. Laney, “3D Data Management: Controlling Data Volume, Velocity, and Variety,” *Application Delivery Strategies* by META Group Inc., vol. 949, p. 4, 2001.
37. Horrigan, M.W. (2013), Big Data: A Perspective from the BLS, *Amstat News*, January 2013, 25-27.
38. Складність аналізу, <https://new.minfin.com.ua/ua/kyivstar/bigdata> (дата звернення 07.12.2022).
39. K. Hornick, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* 2 (1989), 359–366.
40. *Machine Learning Algorithms: A reference guide to popular algorithms for data science and machine learning* / Giuseppe Bonaccorso., 2017. – 360 с.
41. Глинський Я.М. Штучний інтелект. Інтелектуальні роботи /Я.М.Глинський, В.А. Ряжська В.А. – Львів: Деол, 2002. - 168 с.
42. Siyu Wu, Jian-Qiao Sun, “Multi-stage Regression Linear Parametric Models of Room Temperature in Office Building,” *Elsevier’s Building and Environment*, 56, 2012, pp. 69–77
43. Neapolitan R. E. *Learning Bayesian networks*/ R. E. Neapolitan. – Prentice Hall Series in Artificial Intelligence, 2003. – 703 p.
44. Lam W. *Learning Bayesian belief networks. An approach based on the MDL principle* / W. Lam, F. Bacchu // *Computational intelligence*. – 2004. – №10. – P. 104–127.
45. Tsamardinos I. The max-min hill-climbing Bayesian network structure learning algorithm / I. Tsamardinos, L. E. Brow, C. F. Alifer // *Journal of Machine Learning*. – 2006. – № 65 (1). – P. 31–78.

46. Chickering D. M. Large-Sample Learning of Bayesian Networks is NP-Hard / D. M. Chickering, D. Heckerman, C. Meek // Journal of Machine Learning Research. – 2004. – № 5. – P. 1287–1330.
47. Scutari M., Learning Bayesian Networks with the bnlearn R // Journal of Statistical Software. - 2010. - Vol. 35. - P. 1-22.
48. Jouffe L. New search strategies for learning Bayesian networks / Jouffe L. and Munteanu P. // Proc. of tenth international symposium on applied stochastic models and data analysis (ASMDA 2001). – Compiègne (France). 12 – 15 June 2001. – Vol. 2. – P. 591-596.
49. What is Python? Executive Summary. Available: <https://www.python.org/doc/essays/blurb/> (дата звернення 08.12.2022)
50. A. Sweigart Invent Your Own Computer Games with Python. - 2008-2010. - 436 с.
51. Mckinney, W. In Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference; IEEE, 2010; pp 56– 61.
52. Lafuente, D.; Cohen, B.; Fiorini, G.; García, A. A.; Bringas, M.; Morzan, E.; Onna, D. A Gentle Introduction to Machine Learning for Chemists: An Undergraduate Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling. J. Chem. Educ. 2021, 98, 2892– 2898
53. Journal of Educational and Behavioral Statistics, v45 n2 p227-248 Apr 2020.
54. Johnston, B. Google Colab | Educational & Classroom Technologies. <https://mcgrawect.princeton.edu/tool/google-colab/> (дата звернення 10.12.2022)
55. Carneiro, T.; Nóbrega, R.V.M.D.; Nepomuceno, T.; Bian, G.B.; Albuquerque, V.H.C.D.; Filho, P.P.R. Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. IEEE Access 2018, <https://ieeexplore.ieee.org/document/8485684> (дата звернення 10.12.2022)

ДОДАТКИ



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ



Кафедра інженерії програмного забезпечення

МАГІСТЕРСЬКА РОБОТА

«РОЗРОБКА ТА МОДЕЛЮВАННЯ МЕТОДУ ПРОГНОЗУВАННЯ ПОГОДИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ»

Виконав: Студент групи ПДМ-62 Івченко Дмитро Євгенович

Керівник: к.т.н., доц., доцент кафедри ІПЗ, Трінтіна Н.А.

Київ - 2022

МЕТА, ОБ'ЄКТА ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

Мета дослідження: вдосконалення методу прогнозування погоди та погодних явищ з використанням штучного інтелекту.

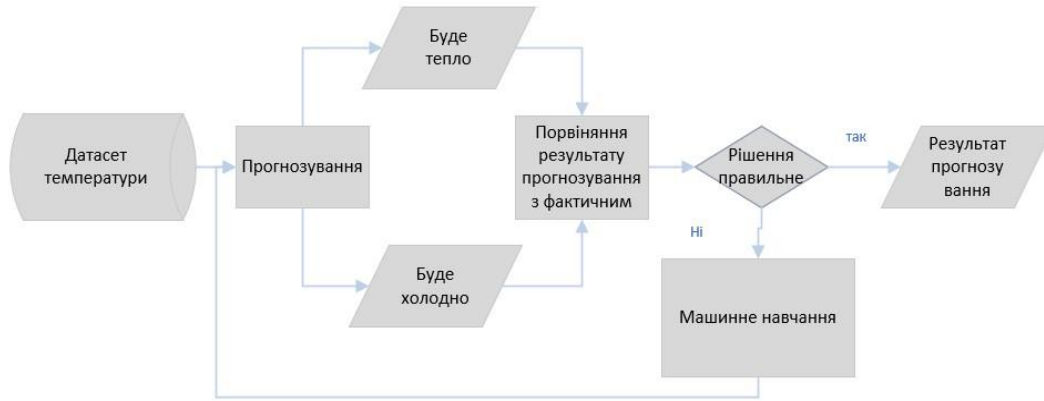
Об'єкт дослідження: процес прогнозування погоди

Предмет дослідження: метод прогнозування погоди та погодних явищ за допомогою штучного інтелекту.

Порівняння похибки методів прогнозування

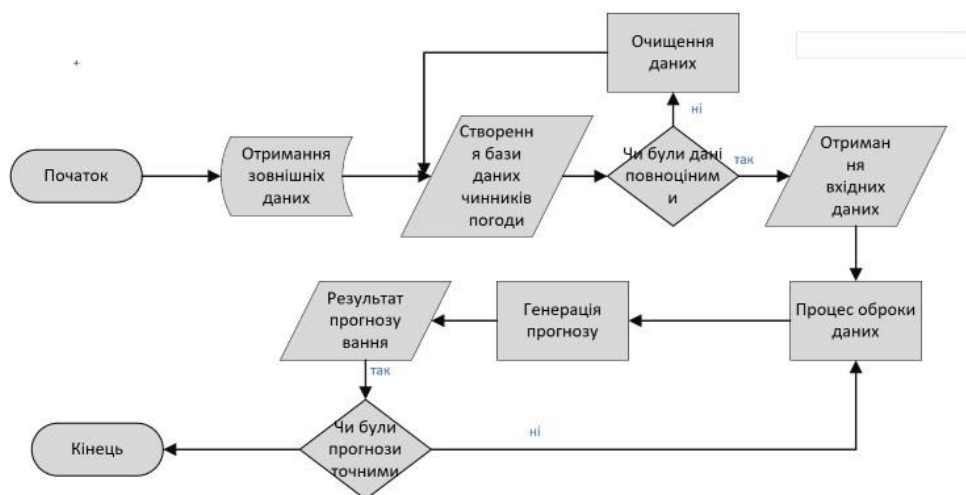
Технології (методи) прогнозування	Період аналізу	Похибка прогнозування Δa 24 год, %	середня відносна похибка прогнозу, MAPE 24 год, %
Стохастичне прогнозування	2016	8	10,2
Математичне прогнозування	2016	11,2	7,2
Статистичного висновку	2016	8,4	6,6
Синоптичні карти	2022	5,2	6,4

МОДИФІКОВАНИЙ МЕТОД ПРОГНОЗУВАННЯ



4

СХЕМА РОБОТИ МЕТОДУ ПРОГНОЗУВАННЯ



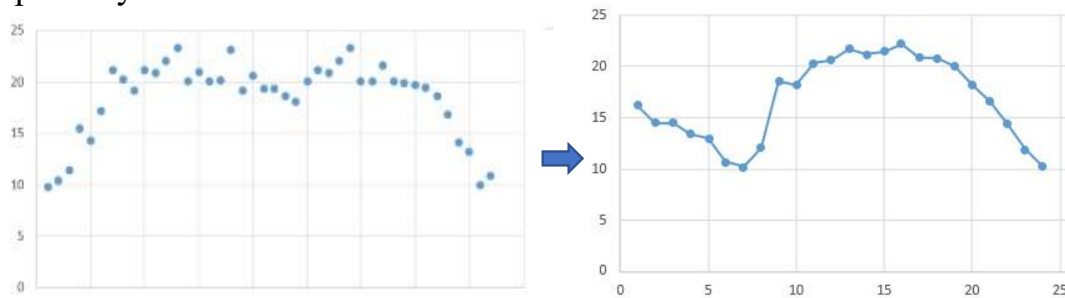
5

ПАРАМЕТРИ ВХІДНИХ ДАНИХ

Метод використовує методи використовується аналіз даних як деревного ухвалень рішень і параметри записується як регресійні моделі.

$$(x, Y) = (x_1, x_2, x_3 \dots x_k, Y)$$

Подальше обчислення йде за основу побудови ймовірностей. Дискретизація даних з використанням лінійного прогнозування необхідне для подальшої оброки даних. При отриманні хибних результатів алгоритм аналізувати дані та завдавати корективи при прогнозуванні.



Дискретизація даних температури

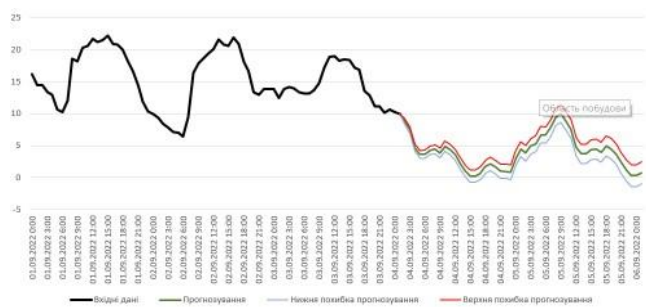
6

ПРАКТИЧНИЙ РЕЗУЛЬТАТ

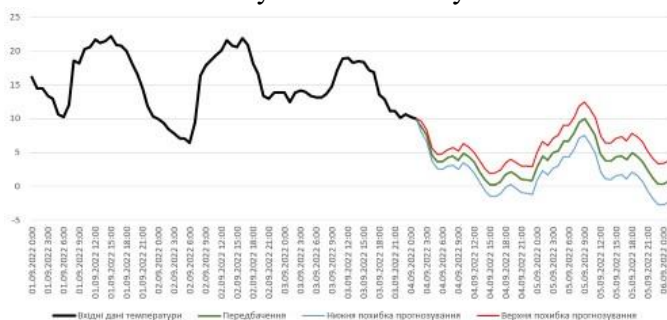
	Синоптичний Прогноз	Прогноз Штучного інтелекту	Покращення прогнозування погоди
Відносна похибка температури за 24 години, %	4,77	4,31	10,6
Відносна похибка вологості за 24 години, %	6,35	5,88	8
Відхилення Прогнозу за 24 години, %	6,20	5,54	11,9
Відхилення Прогнозу за 120 години, %	15,21	12,35	23,15

7

ПРАКТИЧНИЙ РЕЗУЛЬТАТ



Результат прогнозування із використанням штучного інтелекту



Результат синоптичного прогнозу

8

ВИСНОВКИ

1. Проаналізовано предметну область та існуючі методи прогнозування погоди.
2. Досліджено методи обробки отриманих даних, а саме збір даних про стан погодних характеристик в реальному часі та подальшому використанні цих даних з метою створення прогнозу погоди та погодних явищ за допомогою штучного інтелекту.
3. Проаналізовано алгоритми для обробки даних та прогнозування погоди а також доцільності використання штучного інтелекту в прогнозуванні.
4. Удосконалено метод прогнозування на основі штучного інтелекту, який дає меншу похибку в прогнозуванні у короткостроковий і довгостроковий період.

9

ПУБЛІКАЦІЇ ТА АПРОБАЦІЯ РОБОТИ

Статті:

1. Івченко Д.Є., Розробка та моделювання методу прогнозування погоди з використанням штучного інтелекту// Зв'язок. Прийнята до друку.

Тези доповідей:

1. Івченко Д.Є., Трінтіна Н.А. Створення методу прогнозування погоди за допомогою штучного інтелекту//ІІІ Науково-практична конференція «Проблеми комп'ютерної інженерії». – Київ: ДУТ, 2022

2. Івченко Д.Є. Трінтіна Н.А. Розробка та моделювання методу прогнозування погоди з використанням штучного інтелекту // XV Науково-технічна конференція «сучасні інфокомунікаційні технології» – Київ: ДУТ, 2022.

ДЯКУЮ ЗА УВАГУ!