

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: Дослідження інструментів та ресурсів
для аналізу великих даних

Виконала: студент IV курсу, групи СНС-41
спеціальності 122 Комп'ютерні науки

(шифр і назва спеціальності)

Микитенко М.Ю.
(підпис) (прізвище та ініціали)

Керівник Мацюк О.В.
(підпис) (прізвище та ініціали)

Нормоконтроль Литвиненко Я.В.
(підпис) (прізвище та ініціали)

Завідувач кафедри Боднарчук І.О.
(підпис) (прізвище та ініціали)

Рецензент
(підпис) (прізвище та ініціали)

Тернопіль - 2023

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
Боднарчук І.О.
(підпис) (прізвище та ініціали)
«__» _____ 2023 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Бакалавр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Микитенку Максиму Юрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження інструментів та ресурсів
для аналізу великих даних

Керівник роботи Мацюк Олександр Васильович, к.т.н., доц.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «07» 02 2023 року № 4/7-133

2. Термін подання студентом завершеної роботи 21.06.2023р.

3. Вихідні дані до роботи наукові літературні джерела

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ

1. Аналіз предметної області.

2. Інструменти та ресурси для аналізу даних

3. Практична частина

4. Безпека життєдіяльності, основи охорони праці

Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Титулка. 2. Актуальність. 3. Мета, задачі, практична цінність дослідження. 4. Опис предметної області. 5. Схема аналізу даних. 6. Схема концепції Big Data

7. Мова програмування Python та її бібліотеки. 8. Приклад розрахунку із застосуванням Pandas. 9. Приклад використання NumPy. 10. Додаткові бібліотеки

11. Amazon Web Services та Kaggle. 12. Аналіз даних із застосуванням штучних неймереж

13. Реалізація аналізу даних в офтальмології

14. Висновки. Основні результати проведеного дослідження

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи хорони праці			

7. Дата видачі завдання _____ 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	08.02 – 12.02.23	<i>Виконано</i>
2.	Підбір джерел про інструменти та засоби аналізу великих даних	13.02 – 26.02.23	<i>Виконано</i>
3.	Опрацювання джерел про інструменти та засоби аналізу великих даних	27.02 – 15.03.23	<i>Виконано</i>
4.	Проведення дослідження інструментів та засобів аналізу великих даних	16.03 – 10.04.23	<i>Виконано</i>
5	Розроблення програмного коду	11.04 – 25.04.23	<i>Виконано</i>
6.	Оформлення розділу «Аналіз предметної області»	26.04 – 05.05.23	<i>Виконано</i>
7.	Оформлення розділу «Інструменти та ресурси для аналізу даних»	06.05 – 16.05.23	<i>Виконано</i>
8.	Оформлення розділу «Практична частина»	17.05 – 31.05.23	<i>Виконано</i>
9.	Виконання завдання до підрозділу «Безпека життєдіяльності, основи хорони праці»	12.05 – 25.05.23	<i>Виконано</i>
10.	Оформлення кваліфікаційної роботи	20.05 – 03.06.23	<i>Виконано</i>
11.	Нормоконтроль	01.06 – 05.06.23	<i>Виконано</i>
12.	Перевірка на плагіат	06.06 – 08.06.23	<i>Виконано</i>
13.	Попередній захист кваліфікаційної роботи	07.06 – 09.06.23	<i>Виконано</i>
14.	Захист кваліфікаційної роботи	22.06.23	

Студент

_____ (підпис)

Микитенко М.Ю.

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Мацюк О.В.

_____ (прізвище та ініціали)

АНОТАЦІЯ

Дослідження інструментів та ресурсів для аналізу великих даних // Микитенко Максим Юрійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем та програмної інженерії, кафедра комп'ютерних наук, група СНс-41 // Тернопіль, 2023 // С. – 56, рис. – 32, табл. –1, слайдів – 14, бібліогр. – 39.

Ключові слова: аналіз даних, база даних, медицина, прогнозування, big data, python, aws

Кваліфікаційна робота присвячена дослідженню програмних інструментів та спеціалізованих ресурсів для аналізу великих даних.

Описано особливості проведення аналізу даних, визначені головні поняття. Наведено види задач, котрі розв'язуються методами аналізу даних. Особлива увага приділена відносно новому напрямку - Big data, який може бути використаний для опрацювання «даних» у багатьох сферах: медицина, бізнес, політика, виробництво тощо.

Проаналізовано основні програмні засоби (мова програмування Python з спеціалізованими бібліотеками та надбудовами) та ресурси для аналізу даних (платформа хмарних сервісів Amazon та онлайн ресурс Kaggle). Описано їх основні характеристики і можливості. Також значна увага приділена аналізу даних із застосуванням штучних нейронних мереж.

У роботі було реалізовано практичне завдання аналізу даних в медицині, а саме в офтальмології. Проект створено на основі згорткової нейронної мережі, навченої на спеціалізованому наборі даних. Головними завданнями створеного програмного забезпечення є визначення кількості очних залоз на спеціально підготовлених фото. Було підібрано дані, проведено навчання нейронної мережі та написано необхідні скрипти для виконання аналізу даних.

ANNOTATION

Tools and Resources Research to Big Data Analysis // Mykytenko Maksym // Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science // Ternopil, 2023 // P. - 56, Fig. - 32, Table - 1, Slide - 14, References - 39.

Keywords: data analysis, database, medicine, forecasting, big data, python, aws

This thesis deals with the study of software tools and specialized resources for the analysis of big data.

Features of data analysis are described, main concepts are defined. Types of problems solved by data analysis methods are given. Special attention is paid to a relatively new direction - Big data, which can be used to process "data" in many areas: medicine, business, politics, production, etc.

The main software tools (Python programming language with specialized libraries and add-ons) and data analysis resources (Amazon cloud services platform and Kaggle online resource) were analyzed. Their main characteristics and capabilities are described. Considerable attention is also paid to data analysis using artificial neural networks.

The paper implemented a practical task of data analysis in medicine, namely in ophthalmology. The project is based on a convolutional neural network trained on a specialized data set. The main tasks of the created software are to determine the number of eye glands in specially prepared photos. The data was collected, the neural network was trained and the necessary scripts were written to perform the data analysis.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

AWS (Amazon Web Services) – платформа хмарних сервісів Amazon

DD (Deviation Detection) – визначення відхилень (викидів)

KDD (Knowledge Discovery in Data) – виявлення знань у базах даних

VVV (volume, velocity, variety) – обсяг, швидкість, різноманітність

АД (DM – Data Mining) – аналіз даних

БД – база даних

ЗПМ – зворотне поширення помилки

ПЗ – програмне забезпечення

ПК – персональний комп'ютер

ППМ – пряме поширення помилки

СД – сховище даних

ШІ – штучний інтелект

ШНМ – штучна нейронна мережа

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	10
1.1 Вступ до аналізу даних.....	10
1.2 Big data як новий рівень розвитку аналізу даних	12
1.3 Застосування Big data у реальному житті.....	15
РОЗДІЛ 2. ІНСТРУМЕНТИ ТА РЕСУРСИ ДЛЯ АНАЛІЗУ ДАНИХ.....	22
2.1 Мова програмування Python та її бібліотеки	22
2.2 Онлайн ресурси для аналізу даних.....	29
2.2.1 Amazon Web Servises	29
2.2.2 Kaggle	31
2.3 Аналіз даних із застосуванням штучних нейромереж	33
РОЗДІЛ 3. ПРАКТИЧНА ЧАСТИНА	37
3.1 Реалізація аналізу даних в офтальмології.....	37
3.2 Налаштування проекту	38
3.3 Розмітка зображення.....	39
3.4 Навчання та конфігурація мережі	40
РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ	46
4.1 Класифікація шкідливих та небезпечних виробничих факторів.....	45
4.2 Вплив вібрації на людину.	48
ВИСНОВКИ.....	52
ПЕРЕЛІК ДЖЕРЕЛ	53
ДОДАТКИ	

ВСТУП

Актуальність теми. 60 років тому американський вчений Джон Тьюкі запропонував реформувати академічну статистику. У статті «Майбутнє аналізу даних» [1] він зазначив існування науки, яку ще визнали. Предметом інтересу цієї науки було дослідження чи АД.

Інші вчені, зокрема Джон Чемберс, Білл Клівленд і Лео Брейман, незалежно один від одного, повторно запропонували академічній статистиці розширити свою область за межі класичної теоретичної статистики. Чемберс запропонував приділяти велику увагу підготовці та представленню даних, а не статистичному моделюванню. Брейман вважав за потрібне зробити великий наголос на передбачення, а не на висновок. Клівленд запропонував гучну назву "Наука про дані" для передбачуваної наукової галузі. Наприкінці 80х - початку 90х інший відомий вчений Григорій П'ятецький-Шапіро [2] зробив великий внесок у становлення нової науки АД, яка набула вже сучасного вигляду.

Сфера застосування АД немає обмежень - вона скрізь, де є якісь дані [3]. Насамперед методи АД дуже зацікавили комерційні організації, які розробляють проекти з урахуванням інформаційних СД. Саме досвід великої кількості таких фірм каже, що власне віддача від впровадження аналізу даних може сягати 1000%. Можна знайти відомості про проект у 20 мільйонів доларів США, який зміг окупитись за рекордні 4 місяці. Іншим прикладом є річна економія близько 700 тисяч доларів завдяки впровадженню аналізу даних у мережу універсамів у Великій Британії.

Актуальність даної роботи зумовлена тим, що АД представляє велику цінність не тільки для керівників та аналітиків у їхній щоденній діяльності, а ще й для простих працівників. Зокрема, у цій роботі буде розглянуто застосування АД у медичній сфері. Бурхливе зростання технологій і методів АД починає щільно впроваджуватися в медичну ціну. Медична галузь збирає величезну кількість різних даних, тому для їх обробки потрібно на високому рівні володіти комп'ютерними методами АД, щоб оптимізувати роботу лікарів та покращити якість лікування пацієнтів.

Мета роботи – проаналізувати засоби та онлайн-платформи для аналізу великих даних та створити програму мовою програмування Python для АД.

Для досягнення мети виділено ряд завдань:

- вивчення спеціалізованих бібліотек мови програмування Python;
- дослідження онлайн-курсів з машинного навчання з використанням web -сервісів Amazon;
- аналіз друкованих та інтернет ресурсів, присвячених АД;
- вибір завдання з аналізу даних та вирішення його за допомогою отриманих знань.

Практична цінність роботи обумовлена важливістю обраної тематики для АД. Даний аналіз ґрунтується на реальних завданнях медицини та справжніх вихідних даних.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Вступ до аналізу даних

Проблеми, у яких задіяний АД, постали значно раніше, ніж сам термін. Кінець 80х – початок 90х років ХХ ст. – це час, коли комп'ютерні технології та техніка розвивалися дуже швидко, з'явилася мережа Інтернет (у сучасному розумінні) - все це уможливило збирання, зберігання, керування, передавання та власне аналіз величезної кількості різноманітних даних. У ході всього перерахованого вище, АД перетворився на самостійний науковий напрям.

DM і близький до нього термін KDD в 1989 (за іншою версією в 1992) були запропоновані Григорієм П'ятецьким-Шапіро (президентом та головним редактором сайту KDnuggets.com). Фахівець визнаний одним із основоположників у цій області. Він спантеличив питанням: «Чи можна автоматично знайти закони, які б допомогли збільшити швидкість виконання запитів до об'ємних СД?».

DM - це процес отримання нових знань (у необроблених даних), які повинні бути [4]:

- невідомими раніше;
- оригінальними;
- легко інтерпретуються;
- корисними для різної людської діяльності;
- зрозумілими людям без спеціальної підготовки.

Визначення терміну DM пов'язане з класифікацією рівнів інформації [5].

Сирі дані (raw data) - стартові та неопрацьовані дані, котрі одержані при моніторингу будь-яких об'єктів та показують їх стан у різні моменти часу (к приклад, дані про ціни на автомобілі).

Інформація: систематизовані початкові дані, які наведені у зручному вигляді (як варіант, результати запиту у пошуковій системі); опрацьовані дані, які мають цінність для людини (наприклад, статистичні таблиці із середньою зарплатою в Україні по областях).

Знання включають: неявні взаємозв'язки між різноманітними ознаками об'єктів; нововведення, способи та алгоритми вирішення завдань.

Властиво DM - це поєднання множини різних методів визначення нових знань. Власне задачі, котрі розв'язуються методами DM, можна розділити на наступні види [6]:

- класифікація (Classification) Одна з найпростіших задач АД. За підсумками її розв'язання визначаються атрибути, що класи. Завдяки виявленим атрибутам, власне новий об'єкт можна визначити у той чи інший клас;

- кластеризація є наступною стадією ідеї власне класифікації. Це трудомістка задача, її суть у цьому, що класи об'єктів наперед невідомі. Її результатом є поділ об'єктів на групи;

- асоціація (Associations) У завданнях, які виконують пошук асоціативних норм, виокремлюються логічні одиниці між пов'язаними подіями у СД. Головна особливість асоціації, яка відрізняє її від попередніх задач АД - знаходження закономірностей відбувається не на базі характеристик об'єкта, котрий піддається аналізу, а між набором подій, які проходять в один проміжок часу;

- послідовність (Sequence) - допомагає виявити непостійні закономірності між різними транзакціями. Ця задача схожа на попередню, але і них різна кінцева мета. Завдання послідовності встановлює закономірності не між подіями, що трапляються за один часовий проміжок, а між тими, які пов'язані між собою в часі (що відбуваються в деякому певному інтервалі часу);

- прогнозування (Forecasting) Результатом розв'язання такої задачі є оцінка пропущених чи майбутніх значень шуканих коефіцієнтів. Для вирішення подібних задач використовують методи статистики та нейронних мереж;

- DD - це знаходження та аналітика найбільш відмінних від усього масиву даних і визначення непритаманних шаблонів;

- оцінювання (Estimation), фактично є передбаченням таких значень характеристики, котрі є безперервними;

- аналіз зв'язків (Link Analysis) займається знаходженням залежностей у вибірці даних;

- візуалізація (Visualization або Graph Mining). Сенс цієї задачі полягає у побудові графічного варіанту даних, які застосовуються для аналізу;
- підбиття підсумків (Summarization). Метою є опис заданих груп об'єктів з того обсягу даних, котрий аналізується.

АД складається з елементів [7], показаних на рис. 1.1. Також включає математичну логіку, теорію інформації, ШІ та елементи машинного навчання.

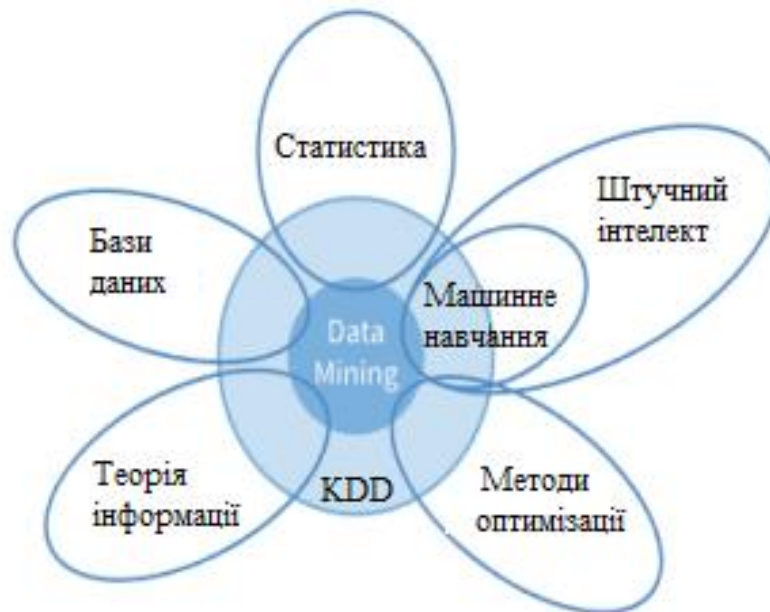


Рисунок 1.1 – Схема складу АД

1.2 Big data як новий рівень розвитку аналізу даних

Ще донедавна застосування принципу Big Data не було таким потрібним, як у наш час. Однак цифрові технології постійно розвиваються, і об'єми інформації стрімко ростуть. Тому зараз багато IT - фахівці активно вивчають «Big Data» та обговорюють свої досягнення у цій сфері у спеціалізованих журналах, газетах, на форумах та конференціях.

Що ж таке Big Data і як принцип застосовується практично? [8] Big Data або іншими словами «великі дані» містять у собі набір парадигм та методів, котрі застосовуються для обробки колосального числа структурованих і також неструктурованих даних. При допомозі цих парадигм спеціалісти виділяють

найбільш корисну та потрібну інформацію та отримують нові знання. Основними атрибутами великих даних є принцип VVV [9, 10].

За даними компанії IDC Digital Univers, до 2025 року обсяг загальної кількості відомостей планети Земля досягне 50 зеттабайтів чи 6000 Гб на кожного мешканця Землі. Ця фантастична цифра змусила ІТ- фахівців почати думати над систематизацією даних та виокремлюванням із загального потоку інформації найпотрібнішу та важливішу.



Рисунок 1.2 – Схема концепції Big data

Навіщо потрібно ретельно вивчати Big Data? Це дає змогу вирішити відразу кілька важливих проблем [12]:

- зберігання та керування гігантським масивом відомостей, що займають багато тисяч терабайтів. Звичайні бази даних не можуть допомогти;
- розподіл спонтанно розташованої інформації, що містить документи, зображення, відео, аудіо та інші типи файлів.

Обробка даних для їхнього подальшого структурування, формування зручних аналітичних панелей, що відображають результати АД та складають на

їх основі точні прогнози.

Успішність реалізації вищезазначених завдань залежить від наступних факторів:

- по-перше, важливим є якість власне роботи фахівців, які досліджують Big Data. Якщо якість знаходиться на максимально високому рівні, то інвестування в проект дослідження принципу буде вище, і процес відбудуватиметься швидше;

- по-друге, аналітики, що працюють з «Big Data», повинні перестати ґрунтуватися на старих технологіях та звернути свою увагу та зусилля на вирішення реальних бізнес-завдань.

У яких випадках слід використовувати Big Data? Організації, які застосовують їх, отримують додаткові бонуси, бо гігантські масиви даних обробляються, аналізуються та застосовуються ефективніше та раціональніше.

Переваги застосування Big Data:

- найбільш точне представлення про бізнес у цілому та використання цих знань на благо компанії;

- проведення більш якісного аналізу конкурентів, і на цій основі пропонуємо унікальні рішення для залучення нових клієнтів;

- отримання нової інформації про цільову аудиторію та створення більш вигідних пропозицій.

Таким чином, використання у компаніях «Big Data» є значним важелем для:

- підвищення кількості продажів;

- поліпшення рівня обслуговування;

- оптимізації витрат;

- покращення якості власне продуктів та послуг.

В яких сферах можуть стати в нагоді технології Big Data? При запуску в діяльність компанії Big Data потрібно розуміти, яких результатів допоможе досягти інформації, що знаходиться в наявності [13]. Наприклад, за допомогою досліджень британської компанії був створений алгоритм, який міг передбачити місце розташування конкретної людини протягом доби з максимальною

точністю близько 20 метрів. За допомогою цього компанія дала маркетологам можливість планування найбільш вигідних для розміщення реклами місць та захоплення більшої аудиторії покупців.

Крім реклами, існують ще кілька областей, у яких використання Big Data корисно найбільше:

- ритейл. У роздрібній торгівлі Big Data дозволяє зібрати відомості для створення довгострокових та дружніх взаємин із клієнтами та партнерами;
- банкінг. Фахівці в даній сфері щодня стикаються з величезною кількістю інформації, тому їм потрібна найбільш грамотна обробка даних з метою підвищення рівня задоволеності клієнтів, мінімізації ризиків та профілактики шахрайства;
- виробництво. Сильна конкуренція вимагає оптимізувати витрати на сировину та підвищити якість продукції. Предикативна аналітика з використанням Big Data дозволяє це зробити без великих витрат;
- освіта. Big Data може легко допомогти покращити систему освіти, мотивувати школярів та студентів до більш продуктивних занять. Викладачі зможуть легше знаходити відстаючих та оцінювати засвоєння навчальної програми;
- охорона здоров'я. Big Data дозволить об'єднати в єдиній БД всі необхідні дані - від історій хвороби до досліджень генетики. Це дозволяє покращити ефективність лікування пацієнтів та підбору відповідних методів терапії.

1.3 Застосування Big data у реальному житті

Великі ІТ компанії - саме те місце, де найбільше розвивається наука про дані, тому їх внутрішня будова у цій сфері дуже цікава [14]. Наприклад, компанія Google розробила Map Reduce (модель розподілених обчислень, яка застосовується для паралельних обчислень над гігантським об'ємом даних), створила свій внутрішній підрозділ, який займається навчанням власних програмістів технологіям та методам машинного навчання. Це дає компанії

конкурентну перевагу: після освоєння нових знань, працівники впроваджуватимуть новітні методи в ті проекти Google, де вони працюють на постійній основі. Список сфер, у яких компанія зможе здійснити революцію - дуже значний. Наприклад, компанія активно використовує нейронні мережі з метою оптимізування витрат за енергію у центрах обробки даних.

Інша велика корпорація Apple також використовує машинне навчання у всіх своїх продуктах. Головна її перевага полягає у наявності великої екосистеми, в якій знаходяться всі цифрові пристрої, випущені компанією, та які використовуються в реальному житті. Це допомагає Apple досягати неймовірних висот: у компанії найбільша кількість даних про своїх користувачів у порівнянні з конкурентами. При цьому політика конфіденційності компанії є дуже строгою, Apple завжди робила акцент на те, що не використовує дані про своїх користувачів з метою реклами. Інформація користувачів зашифровується таким чином, що а ні юристи самої Apple, а ні ФБР з ордером не зможуть нею скористатися.

З кожним роком автовиробники починають використовувати складніші методи збору даних, які застосовують Big Data. Це означатиме, що технології можуть сповіщати водіїв про погані дорожні умови шляхом автоматичного включення АБС та інших систем.

Інші концерни, наприклад, такі як BMW, застосовують парадигми Big Data разом з відомостями, зібраними з зразків, що тестуються, інтегрованою в авто системою «запам'ятовування помилок» і наріканнями клієнтів, для того щоб на початковій стадії виготовлення виявити вади автомобіля. Тепер не потрібне ручне оцінювання даних, яке потребує багато місяців, а використовується модерний алгоритм. Число помилок та витрат на їх усунення зменшується, що допомагає прискорити процес аналізу відомостей у BMW.

У 2022 році обіг ринку включених в єдину мережу автомобілів досяг приблизно \$140 млрд. Це цілком очікувано, якщо брати до уваги, якими темпами автоконцерни впроваджують нові технології, які стають важливою частиною автомобіля.

Застосування Big Data дозволяє зробити автомобіль безпечнішим та

функціональнішим. Наприклад, компанія Toyota вбудовує інформаційні комунікаційні модулі та прив'язує нові автомобілі до Toyota Big Data Center. Цей засіб, котрий функціонує на базі Big data, опрацьовує та аналізує різноманітну інформацію, зібрану модулем, щоб у майбутньому одержувати користь від неї.

Big data в медицині дає змогу лікарям ретельніше вивчати захворювання та підібрати найбільш ефективний курс лікування для кожного випадку. За допомогою АД, медикам стає простіше передбачати рецидиви та проводити профілактичні заходи. Результатом стає найбільш вірна постановка діагнозу та вдосконалення методів лікування.

Свіжа методика допомогла поглянути на недуги пацієнтів з нового боку, що призвело до виявлення раніше невиявлених джерел недуг. Наприклад, певні раси генетично схильні до серцевих хвороб більше, ніж інші етнічні групи. Тепер, коли пацієнта непокоїть конкретне захворювання, медики вивчають дані про представників раси пацієнта, які вже поводитися з подібною проблемою.

Збір та АД допомагає отримати багато нової інформації про хворих: переваги в їжі, стиль життя, генетична структура ДНК, метаболіти тканин, клітин та органів. У Центрі медицини у місті Канзас-Сіті використовують технології Big data для термінового розшифрування ДНК хворих та проведення розбору мутацій у генетичному коді, які провокують утворення ракових захворювань. Знаходження індивідуального підходу до кожного окремого пацієнта з огляду на його ДНК підвищує ефективність лікування на якісно новий рівень.

Розуміючи те, як використовувати Big data, впливає вкрай важлива зміна медицини. Під час проходження пацієнтом курсу лікування, лікарня або інший медичний заклад може отримувати багато вкрай корисної інформації про пацієнта. Всі зібрані дані використовуються для прогнозів рецидивів хвороб з певною точністю. Наприклад, якщо людина перенесла інсульт, медики розглядають дані про час відхилення у кровообізі мозку, досліджують проміжок між попередніми випадками (якщо такі були), наголошуючи на стресових випадках та сильних фізичних навантаженнях пацієнта. Лікарі, ґрунтуючись на цих даних, складають пацієнтові конкретний комплекс заходів, щоб спробувати

уникнути повторного інсульту.

Важливу роль грають гаджети, які покликані знаходити проблеми зі здоров'ям, навіть якщо людина не має виражених симптомів будь-якої хвороби. Замість того, щоб стежити за станом пацієнта, проводячи тривалі курси обстежень, лікар може зробити висновки, ґрунтуючись на зібраним фітнес-трекером або смарт годинником даними. У Лурдському медичному центрі Богоматері у місті Нью-Джерсі нещодавно стався цікавий випадок. Один пацієнт проходив обстеження після нападу судом, яке було наслідком пропущеного прийому лікарських засобів. Лікарі з'ясували, що у чоловіка набагато серйозніші проблеми зі здоров'ям. Цією проблемою стала фібриляція передсердь. У постановці діагнозу допоміг той факт, що лікарі змогли отримати доступ до смартфона пацієнта, а точніше до програми, яка працювала з фітнес-трекером пацієнта. Інформація з програми виявилася ключовою у визначенні правильного діагнозу, бо на період обстеження у пацієнта жодних серцевих відхилень не було знайдено. Цей випадок показує, що сьогодні використання Big data у медичній сфері є надзвичайно важливим.

Big data у торгівлі. Вивчення запитів користувачів та націлювання є максимально відомими громадськості областями використання парадигм Big Data. Ці технології допомагають у аналізі клієнтських навичок, щоб у майбутньому краще розуміти запити клієнтів. Компанії зацікавлені у розширенні традиційного набору даних із соцмереж та пошукових запитів для формування найбільш детальної картини. Найчастіше потужні компанії як пріоритетну мету обирають розробку своєї моделі для передбачень.

Приміром, мережа Target із застосуванням АД та своєї особистої системи, яка робить прогнози, може з величезною ймовірністю встановити вагітна жінка чи ні. За всіма клієнтами встановлюється унікальний ID номер, який має прив'язку до банківської картки, нікнейму або електронної пошти. ID є своєрідним кошиком для купівель, в котрому зберігаються дані про всі придбання людини за весь час. Фахівці мережі встановили, що вагітні жінки активно купують неароматизовані засоби до настання другого триместру вагітності, а упродовж перших двох триместрів часто купують спеціальні

добавки. На основі даних Target розсилає клієнтам картки дитячих товарів. При цьому разом зі знижками на дитячі речі надсилаються купони на інші для того, щоб пропозиції купити їх не здавалися б занадто нав'язливими.

Навіть урядові організації змогли знайти спосіб застосування Big Data для оптимізації проведення виборів. Багато хто вважає, що виграш Дж. Байдена на виборах у президенти США 2020 пояснюється продуктивною роботою його команди консультантів, які змогли грамотно розпорядитися величезною кількістю даних.

Приблизно близько 48% директорів з маркетингу вважають, що застосування парадигм Big Data простежується в оптимізації пошуку (SEO), електронній пошті та мобільному маркетингу, де АД є надзвичайно корисним у створенні маркетингових програм (рис. 1.3). Лише на чотири відсотки менше респондентів думають, що Big Data відіграватиме головну роль у кожній маркетинговій стратегії протягом багатьох років.



Рисунок 1.3 – Застосування Big Data у бізнесі

Big Data у масштабах нашої планети. Дуже цікаво те, як дані технології використовуються для зменшення впливу людини на планету Земля. Є

ймовірність, що машинне навчання в результаті стане єдиною можливістю, яка зможе підтримувати тендітну та важливу рівновагу. Тема про вплив людей на глобальне потепління є актуальною досі і викликає велику кількість суперечок [15]. Саме тому лише правдиві передбачувальні моделі, що ґрунтуються на аналізі величезних об'ємів даних, можуть видати точну відповідь (рис. 1.4). Зрештою, зниження рівня викидів допоможе людям: витрати на енергію стануть меншими.

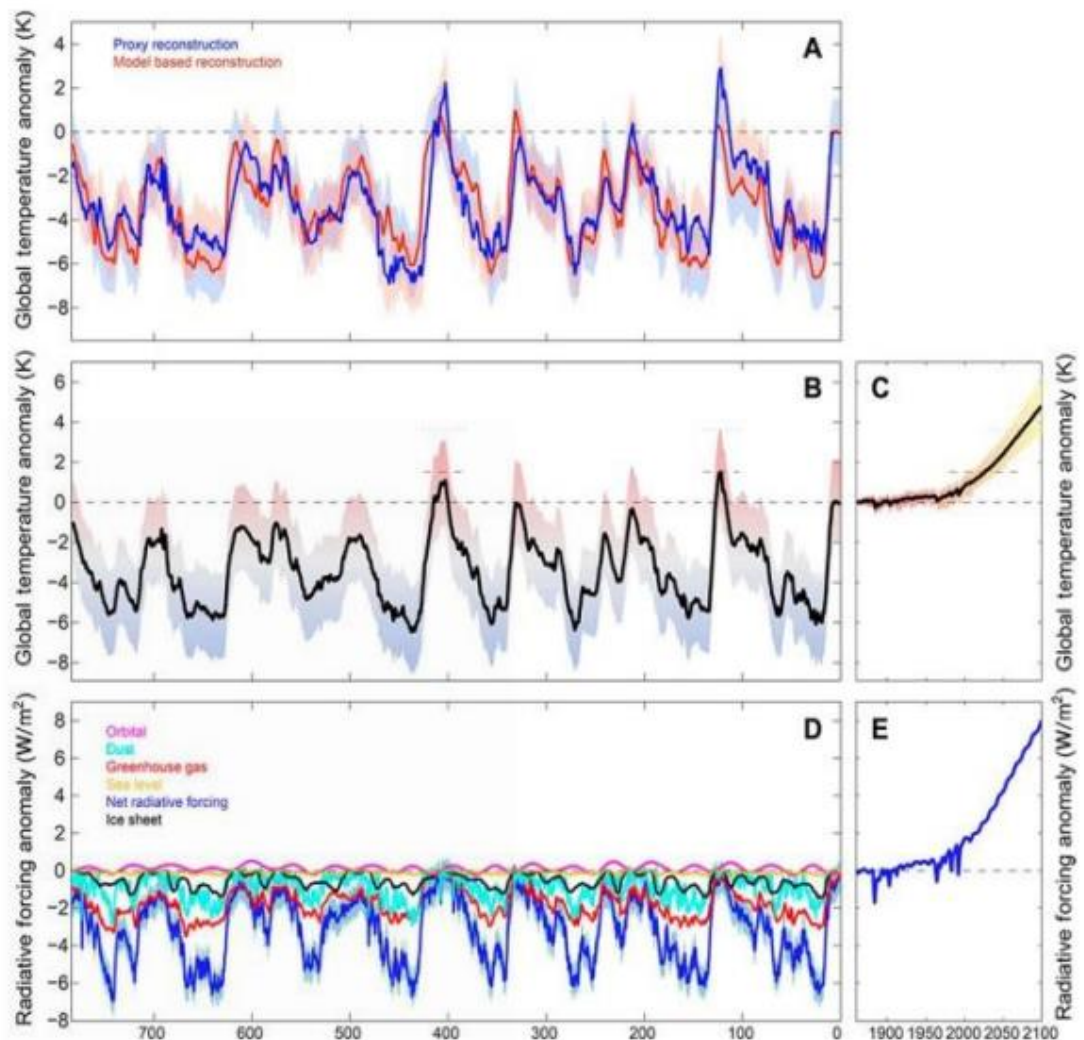


Рисунок 1.4 – Прогнозування глобального потепління

Зараз Big Data вже не є абстрактним судженням, яке використовуватиметься десь у майбутньому. Це діючий набір парадигм, котрий спроможний забезпечити велику користь майже у всіх галузях діяльності людини: від медичної та охорони правопорядку до маркетингу та продажу.

Активна інтеграція Big Data в наше життя почалася нещодавно, і хто достеменно зна, які висоти очікують на Big Data вже в найближчому майбутньому?

РОЗДІЛ 2. ІНСТРУМЕНТИ ТА РЕСУРСИ ДЛЯ АНАЛІЗУ ДАНИХ

2.1 Мова програмування Python та її бібліотеки

Python - улюблена багатьма мова програмування (рис. 2.1). Вона з'явилася в 1991 році і набрала величезної популярності серед інших динамічних мов програмування (наприклад, таких як Ruby та Perl) [16].



Рис. 2.1 – Значок Python

Python часто називають скриптовою мовою, натякаючи, що вона призначена лише для маленьких програм – скриптів. Але це не так, з безлічі інтерпретованих мов Python має величезне і активне співтовариство наукових розрахунків. Застосування цієї мови у сфері наукових досліджень почало розвиватися з початку 2000х років.

У тих випадках, коли ми намагаємося вирішити проблему з АД, то стикаємося зі значною кількістю питань, які допоможе вирішити Python [17].

Чому саме Python? Python – потужна мова програмування, але що це означає у реальному житті? Що робить Python такою потужною?

Python дуже простий. Статус універсальної мови Python отримав через просте вивчення. Синтаксис мови придуманий, щоб легко читати. Python має величезну популярність у наукових обчисленнях. Люди, які працюють у цій сфері по-перше вчені, а потім уже програмісти.

Python ефективний. Сучасний світ працює з гігантськими обсягами даних,

які називаються Big data. Чим більше даних потрібно обробити, тим пріоритетніше використання керованої пам'яті. У цьому ключі Python виявляє себе максимально ефективно.

Python швидкий. Хоча Python і є інтерпретованою мовою, багато хто думає, що вона повільна, але мова постійно розвивається і її продуктивність підвищується. Python є найкращим, якщо потрібно зробити високопродуктивні обчислення.

Мова програмування Python в останні роки все більше використовується для АД (у науці та комерційній сфері) [18]. Цьому сприяє широке розмаїття різноманітних відкритих бібліотек. Нижче наведено приклади базових бібліотек, які допоможуть зробити з мови Python потужний і корисний інструмент для аналізу та візуалізації даних [19, 20]. На цих базових бібліотеках ґрунтуються спеціалізовані бібліотеки (їх часто називають SciPy).

Jupyter є інтерактивною оболонкою для мови програмування Python (рис. 2.2). Дає розширену інтроспекцію і новий командний синтаксис [21].



Рис. 2.2 – Значок Jupyter

Jupyter також зберігає історію всіх введів для всіх сеансів, підсвічує та самостійно доповнює програмний код. Web- інтерфейс відмінно справляється: з дослідженнями та первинною обробкою даних, тестуванням початкових версій коду та покроковим налагодженням.

При використанні мови розмітки markdown та бібліотек для візуалізації можна сформулювати звіти з аналізом відразу в браузері. Або переробити звіт на

слайди презентації HTML. За допомогою Jupyterhub_налаштовується спільна робота команди на сервері з груповою авторизацією.

Pandas надає структури даних та корисні інструменти для аналізу (рис. 2.3). Ця бібліотека покликана допомогти у обробці неповних, невідсортованих та немаркованих даних (такі часто трапляються у житті) [22].



Рис. 2.3 – Значок Pandas

Pandas дозволяє замінити великі та складні операції з даними на парочку простих команд. Бібліотека має велику кількість готових методів групування, фільтрації, об'єднання даних та можливості розпізнавання різних видів джерел [23].

У цій бібліотеці можна об'єднати таблиці за прикладом із SQL JOIN. Дані черпаються прямо з файлів, тому йде необхідність організації БД.

Однією з головних рис Pandas є швидкість її роботи. Наприклад, для групування даних з таблиці в 10 млн. рядків, підрахунок середнього чека і сортування результатів за зменшенням, потрібний лише один рядок коду і менше п'яти секунд (рис. 2.4):

Розрахунок категорій з макимальним середнім чеком

```
%%time  
data.groupby('category').mean().sort_values('payment', ascending = False)['payment']
```

Wall time: 4.72 s

```
category  
6632f    7805.250000  
ce304    5978.881667  
e1291    5667.943333  
f4107    4834.665000  
38221    4788.160000  
ec052    3907.200000
```

Рисунок 2.4 – Приклад розрахунку із застосуванням Pandas

NumPy – це головна бібліотека мови Python, яка скорочує роботу з векторами та матрицями (рис. 2.5). Вона включає готові методи для різних операцій: створення, зміна форм, множення і розрахунок детермінанта матриці, навіть рішення лінійних рівнянь і сингулярного розкладання [24].



Рисунок 2.5 – NumPy лого

Приклад її використання наведено на рис. 2.6.

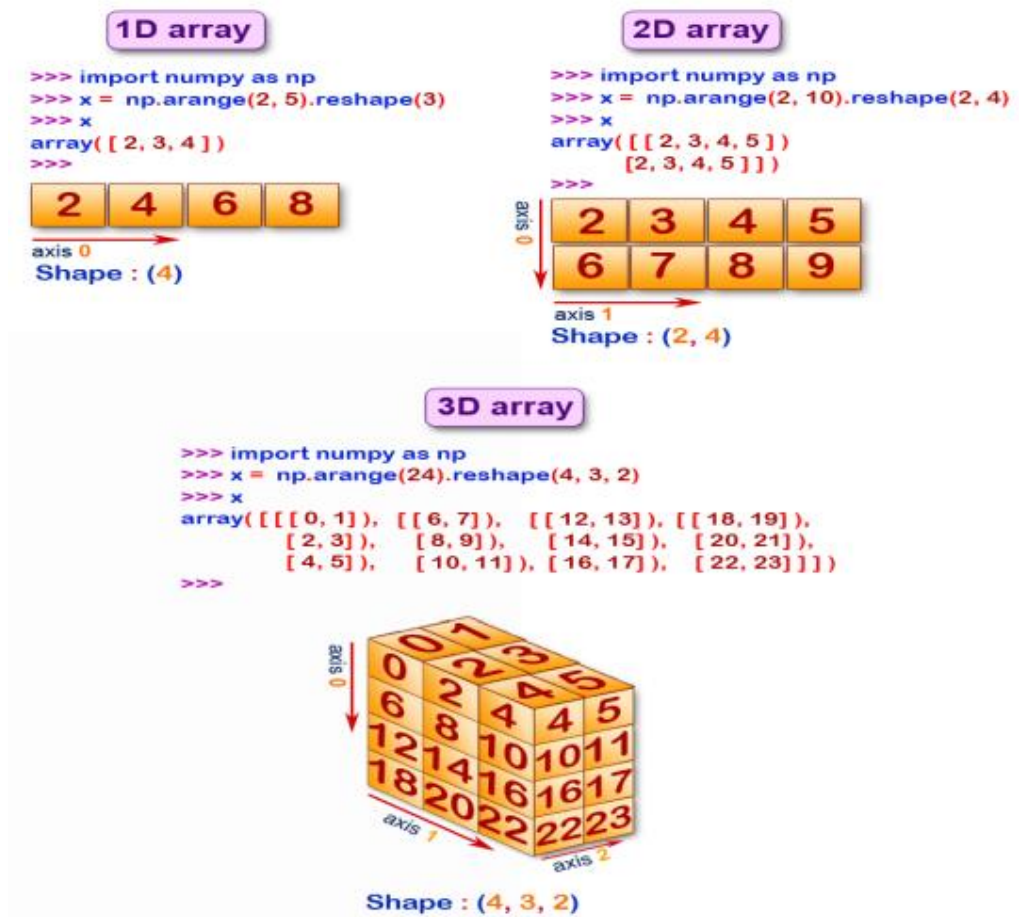


Рисунок 2.6 – Приклад використання NumPy

SciPy бібліотека базується на NumPy і відкриває нові можливості (рис. 2.7). Вона включає методи лінійної алгебри, методи для роботи теорією ймовірності, інтегральними обчисленнями та перетвореннями Фур'є [25].



Рисунок 2.7 – SciPy лого

Matplotlib [26] представляє низькорівневу бібліотеку для побудови двовимірних діаграм та графіків (рис. 2.8).



Рисунок 2.8 – Matplotlib лого

З її допомогою можна збудувати будь-який графік. Однак для більш складної візуалізації даних потрібно більше коду, ніж у найрозвиненіших бібліотеках. Приклад візуалізації із застосуванням можливостей Matplotlib представлено на рис. 2.9.

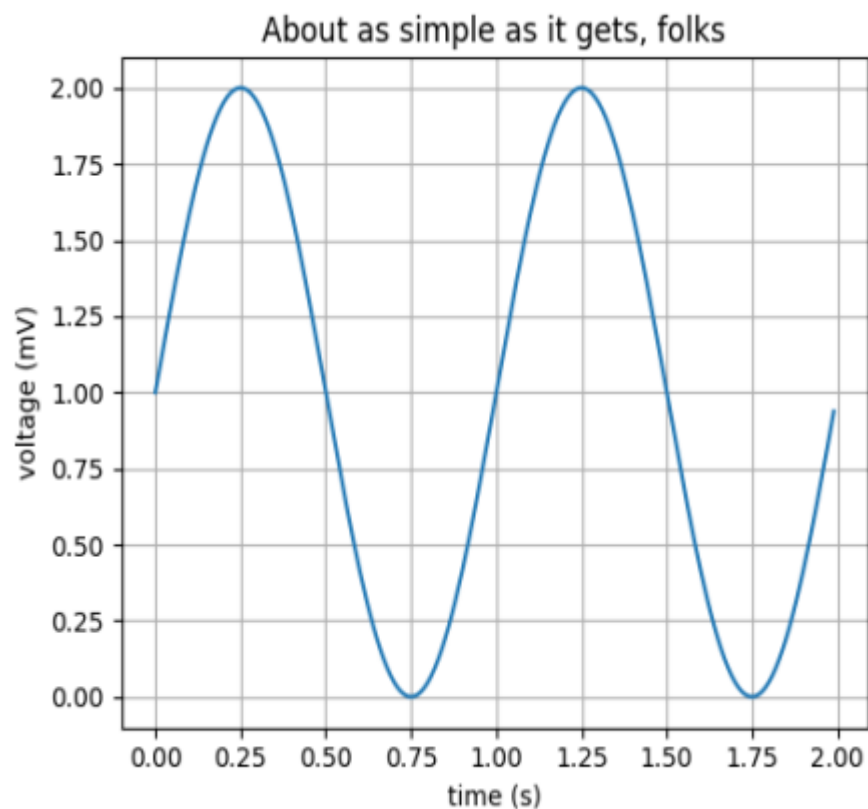


Рисунок 2.9 – Приклад використання Matplotlib

Scikit-learn (рис. 2.10) ґрунтується на бібліотеках NumPy та SciPy. Вона надає комплекс алгоритмів для машинного навчання та інтелектуального АД (кластеризація, регресія та класифікація) [27]. Є однією з найкращих бібліотек

для тих компаній, які працюють з величезним обсягом даних. Компанії, що її використовують: Evemote, OKCupid, Spotify, Birchbox at ін.



Рисунок 2.10 – Scikit-learn лого

Theano - це бібліотека, яка потрібна для оцінювання та покращення різних математичних виразів (рис. 2.11). Її синтаксис схожий на NumPy. Вона здійснює необхідні розрахунки з великим обсягом даних у сотні разів швидше, ніж CPU (центральний процесор), оскільки задіює GPU (відеокарти). За це її оцінили ті, хто займається глибоким навчанням та має справу з обчислювальними завданнями [28].



Рисунок 2.10 – Theano лого

TensorFlow – бібліотека, яку створила компанія Google для навчання нейронних мереж (рис. 2.11). Ця бібліотека використовується для налаштування, тренувань та використання ШНМ з різними масивами даних [29]. За допомогою цієї бібліотеки Google має можливість визначення об'єктів на фотографіях, а програма з розпізнаванням голосу може розуміти мову.



Рисунок 2.11 – Theano лого

Scrapy (рис. 2.12) часто використовується для створення ботів-павуків, які моніторять сторінки різних сайтів і накопичують структуровані дані: ціни, контактна інформація та URL -адреси. Також Scrapy може отримувати дані з API [30].



Рисунок 2.12 – Scrapy лого

Мова програмування Python дуже багата на бібліотеки. Вище були перераховані найбільш популярні, відомі та корисні. Все це переконливо говорить про те, що Python є серйозним і значним інструментом для АД.

2.2 Онлайн ресурси для аналізу даних

2.2.1 Amazon Web Services

АД щільно увійшов у наше життя. Багато великих компаній активно використовують технології АД у своїй діяльності. Це не дивно, бо без грамотного збору та АД компанія не зможе розраховувати на максимальну продуктивність, більший прибуток та ефективне управління. Також багато

компаній будують свою роботу на наданні послуг з АД. Одним із таких прикладів є компанія Amazon.

Amazon.com Inc. є однією з найбільших і відомих американських компаній, однією з перших інтернет магазинів, займає лідируючу позицію серед інших інтернет магазинів за обсягами продажів. Компанія Amazon одна з перших почала продавати реальні товари за допомогою мережі Інтернет. Найпершими товарами цієї компанії стали книжки. У новому інтернет магазині можна було розмістити необмежену кількість товарів та запропонувати їх колосальній кількості покупців. Дана бізнес-модель від Amazon добре прижилася і принесла величезний успіх і славу компанії.

Також компанія Amazon активно займається хмарними обчисленнями [31]. AWS - найпоширеніша хмарна платформа з великими можливостями у світі (рис. 2.13). Вона надає приблизно 165 повнофункціональних сервісів, які використовують центри обробки даних по всій планеті. Мільйони користувачів, стартапи, що розвиваються швидкими темпами, гігантські корпорації та різні урядові установи, використовують AWS для розміщення інфраструктури, підвищення гнучкості та зменшення витрат.



Рисунок 2.13 – AWS лого

Також Amazon відкрив вільний доступ до всіх своїх курсів по машинному навчанню, які спочатку призначалися інженерам компанії.

Навчальні матеріали поділені на блоки для розробників, менеджерів, науковців та аналітиків.

Топ менеджер зі ШІ в AWS Метт Вуд заявив, що місія компанії полягає в

тому, щоб машинне навчання, яке раніше було долею лише великим технологічним компаніям, могли вивчати будь-які розробники.

Програма навчальних курсів включає понад 30 навчальних блоків, розтягнутих на термін більше 45 годин, а також відеоматеріали, практичні завдання та необхідну документацію. Курси доступні всім користувачам через AWS.

Кожен окремий курс розповідає про фундаментальні основи машинного навчання і наводить приклади практичних завдань, які колись вирішувалися інженерами Amazon. Наприклад, учасникам курсів необхідно буде визначити придатність подарункової упаковки або оптимізувати маршрути доставок товарів.

Курси також можуть допомогти розпочати роботу з різними сервісами машинного навчання AWS, включаючи Amazon SageMaker, AWS DeepLens, Amazon (Rekognition / Polly / Comprehend) .

Крім цього, користувачі курсів зможуть скласти платний іспит та отримати сертифікат, який зможе підтвердити знання з машинного навчання та сприятиме визнанню в галузі. Вартість цього іспиту - 300\$.

2.2.2 Kaggle

Спочатку онлайн ресурс Kaggle (рис. 2.14), який має слоган «ваш будинок для науки про дані», був майданчиком для змагань з машинного навчання, але тепер там можна знайти ресурси для АД [32].



Рисунок 2.12 – Kaggle лого

Інтерфейс головної сторінки ресурсу Kaggle наведено на рис. 2.15.

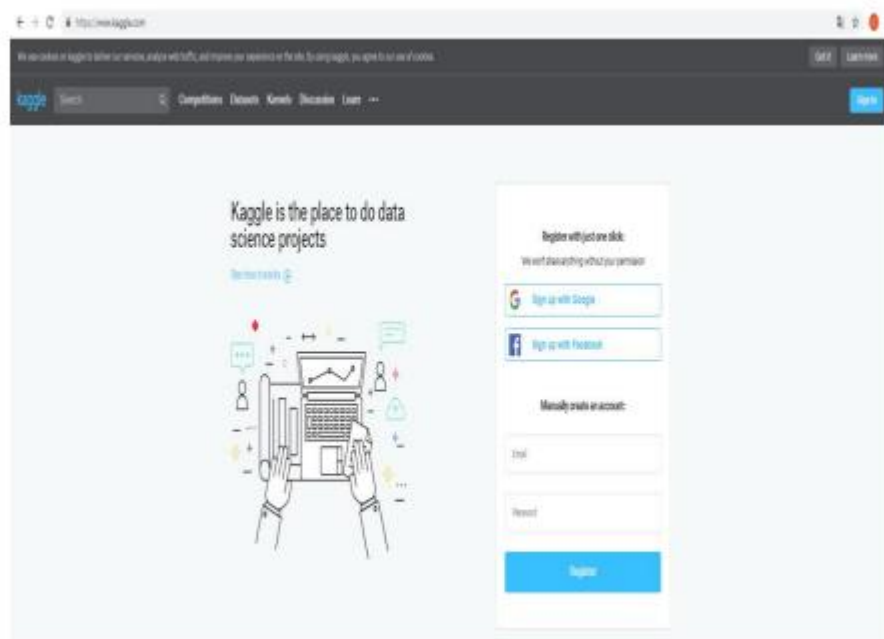


Рисунок 2.15 – Головна сторінка Kaggle

Головні особливості Kaggle [33]:

- **Datasets (набори даних):** наявність великого набору даних різних типів та розмірів, які знаходяться у відкритому доступі. Тут можна знайти цікаві дані для самостійного вивчення або тестування своїх навичок з моделювання;
- **Machine Learning Competitions (змагання з машинного навчання):** колись вони були основою Kaggle. Ці тести на моделювання є найкращим способом вивчення нових видів машинного навчання і тренування здібностей з допомогою оригінальних проблем, що з реальними даними;
- **Learn (Вивчення):** безліч навчальних матеріалів з вивчення даних, що стосуються SQL і глибоке навчання, які можуть потрапити в Jupyter Notebooks;
- **Discussion (обговорення):** існує форум, де є можливість поставити свої запитання та отримати відповіді та рекомендації від тисячі експертів з АД;
- **Kernels (ядра) –** це онлайн середовище розробки для програмування, яке функціонує на серверах Kaggle. У цьому середовищі можливе написання Python або R -скриптів і робота в Jupyter Notebooks. Це середовище абсолютно безкоштовне (можна навіть задіяти GPU) і не вимагає налаштувань свого або

навчальних комп'ютерів. Середовище відмінно підходить для аналізу будь-якого масиву даних, проведення змагань з машинного навчання або роботи над завданнями з розділу навчання. Доробками в середовищі можна ділитися з іншими користувачами.

Загалом Kaggle є чудовим місцем для навчання: як через звичні навчальні завдання, так і через змагання з розділу Competitions. Щоб дізнатися про нові методи машинного навчання, можна прочитати книгу або подивитися на Kaggle приклади того, як люди використовують ці методи на практиці. Друге є найцікавішим та ефективнішим. Також не варто забувати про те, що спільнота даного ресурсу вкрай товариська і завжди готова відповісти на будь-які запитання та прийти на допомогу.

2.3 Аналіз даних із застосуванням штучних нейромереж

ШНМ є спробою, за допомогою якої математичні моделі відтворюють діяльність мозку людини для побудови машин, котрі володіють ШІ [34].

ШНМ, найчастіше, навчається разом із учителем. Це свідчить про необхідність навчального властиво набору даних, що має приклади з еталонними значеннями, зорема тегами, класами, показниками тощо.

Для прикладу, якщо потрібно створити ШНМ для оцінювання тональності тексту, набором початкових даних стане перелік пропозицій із шаблонними емоційними оцінками. Власне саму тональність тексту характеризують різні ознаки: слова, фрази, структури речень, які мають негативний чи позитивний настрій. Вага ознак у кінцевому оцінюванні тональності тексту, який може бути позитивним, нейтральним чи негативним, безпосередньо залежить від тієї математичної функції, котра обраховується упродовж навчання мережі.

Колись люди генерували самостійно ці ознаки. Чим їх більше і грамотніше підібрані ваги, тим відповідь точніша. Нейронна мережа змогла автоматизувати цей процес.

ШНМ включає три основні компоненти (шари) (див. рис. 2.16):

– входу;

- прихованих обчислень;
- виходу.

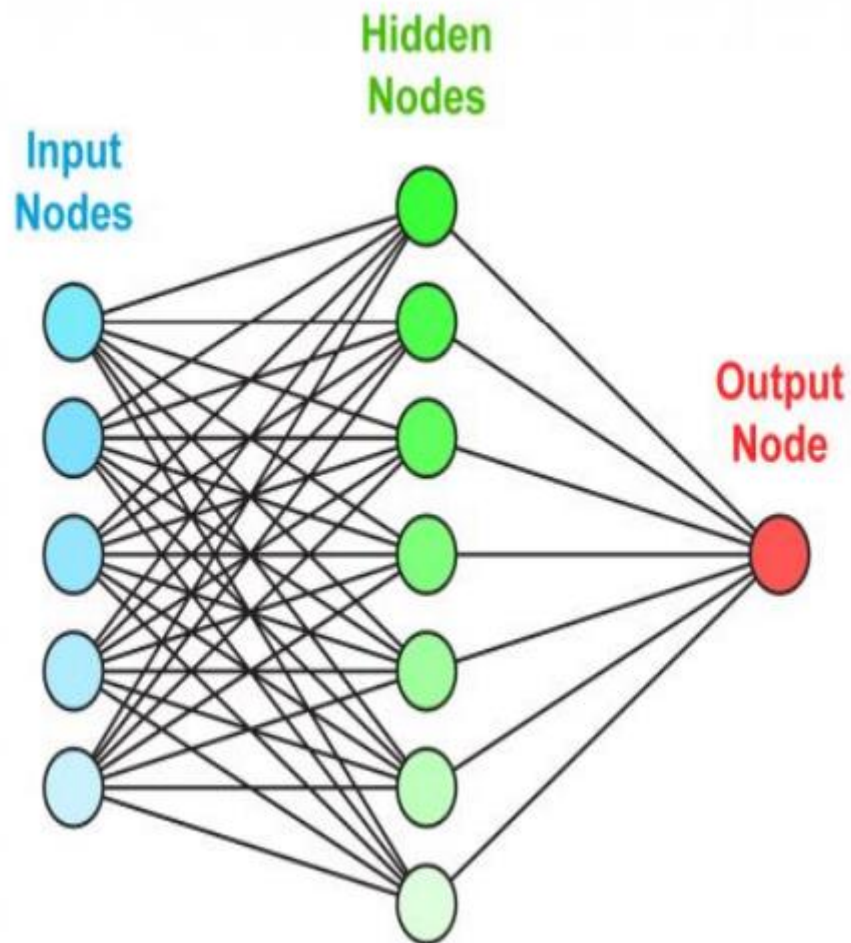


Рисунок 2.16 – Компоненти ШНМ

Навчання ШНМ містить 2 частини:

- ППМ (рис. 2.17);
- ЗПМ (рис. 2.18).

У періоді ППМ робиться прогноз результату. При ЗПМ помилка мінімізується між фактичною та передбаченою відповідями.

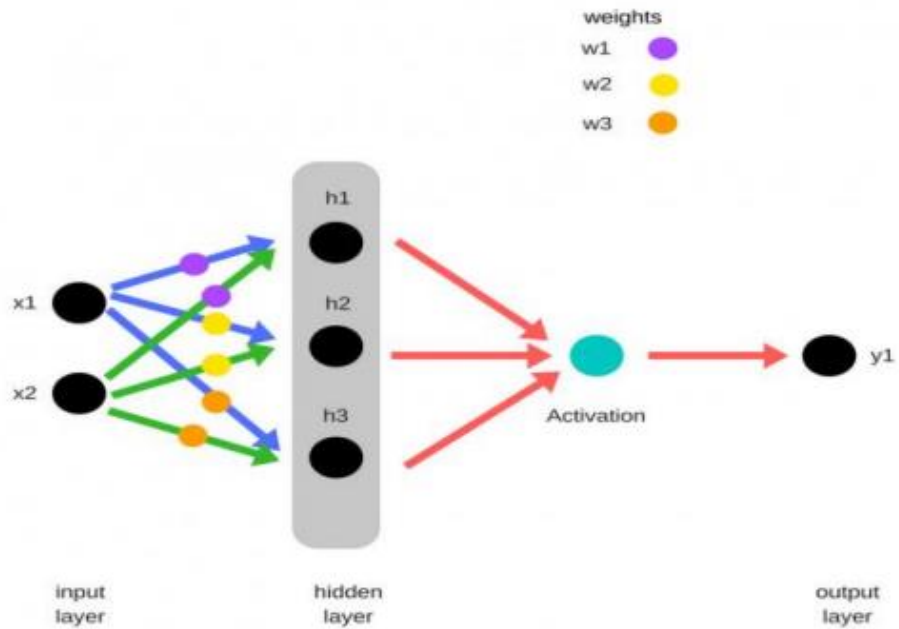


Рисунок 2.17– ППМ

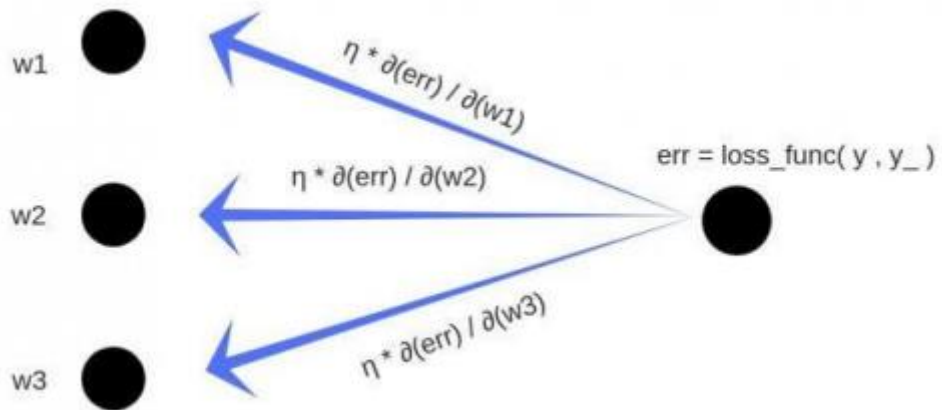


Рисунок 2.18 – ЗПМ

Нейронні мережі можна застосувати практично кожній сфері діяльності [35]. Области, в яких застосування нейронних мереж є актуальним вже сьогодні, описані нижче по тексту.

Охорона здоров'я та медицина. У цій галузі нейронні мережі можуть стати в нагоді при постановці діагнозів, обробці різних даних, при отриманні від

медичної апаратури зображень. З їх допомогою можна виконати моніторинг стану здоров'я пацієнтів, оцінити ефективність поточних методів лікування.

Сфера зв'язку. Зменшення обсягів пакетів даних, оптимізація процесів маршрутизації, передача даних у стільникових мережах.

Робототехніка. Нейронні мережі підвищують ефективність розпізнавання різних об'єктів, перешкод, покращують координацію рухів роботів. З їх допомогою вдасться вдосконалювати системи автопілотування.

Фінансова галузь. Дана технологія дозволяє прораховувати ризики, пов'язані з обігом коштів, виставляти об'єктивні та повні оцінки для компаній. Також вона дуже перспективна для трейдингу.

Виробництво. У цій сфері технологія дозволяє зробити автоматизацію виробничих процесів, підвищити контроль якості на новий рівень, знизити ймовірність аварійних ситуацій.

Соціологія та політологія. Нейронні мережі допомагають збільшити точність та ефективність соціологічних опитувань. Їх також логічно використовувати для передбачення результатів виборів.

Нейронні мережі активно використовують у ігрових додатках, у сучасній індустрії гаджетів, у сфері безпеки тощо.

РОЗДІЛ 3. ПРАКТИЧНА ЧАСТИНА

3.1 Реалізація аналізу даних в офтальмології

У рамках цієї роботи було виконано тренувальний проект з АД знімків людських очей. Основне завдання цього аналізу полягає у виявленні очних залоз. Ці дані можуть бути корисними лікарям - офтальмологам.

Усі необхідні скрипти та папки з даними розташовані на Google Disc. Він дозволяє зручно працювати з усіма речами, які на ньому розташовані та проводити необхідні обчислення за допомогою сервісу Colaboratory, що дає доступ до GPU самого Google (рис. 3.1).

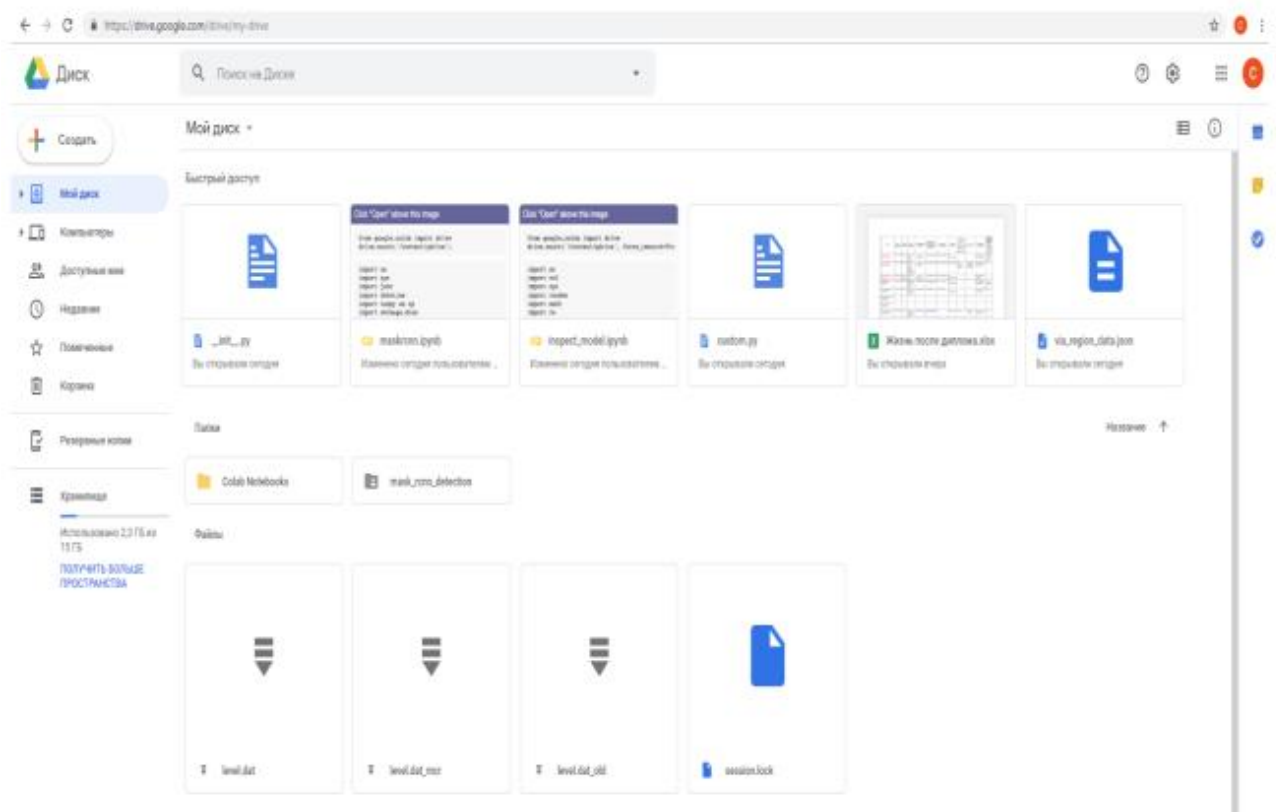


Рисунок 3.1 – Скріншот початкової сторінки Google Disc

Проект створено на основі згорткової нейронної мережі, навченої на даних Microsoft COCO (Common Objects in Context) (рис. 3.2) [36]. Дані містять знімки, де вже виконано розмітку, але також можна зробити розмітку вручну.



Рисунок 3.2 – COCO лого

3.2 Налаштування проекту

Насамперед відбувається налаштування проекту. На цьому етапі імпортуються всі залежності - бібліотеки та файли. Далі задається шлях до файлу ваг, тут існують два варіанти, або завантаження збереженого файлу передбаченого на даних COCO dataset, або завантаження власного файлу для навчання мережі (він також до навчання на початковому файлі COCO dataset). Скрипт наведено на рис. 3.3.

```
import os
import sys
import json
import datetime
import numpy as np
import skimage.draw
import cv2

# Root directory of the project
ROOT_DIR = os.path.abspath("/content/gdrive/My Drive/mask_rcnn_detection/")

# Import Mask RCNN
sys.path.append(ROOT_DIR) # To find local version of the library
from mrcnn.config import Config
from mrcnn import model as modellib, utils

# Path to trained weights file
# COCO_WEIGHTS_PATH = "/content/gdrive/My Drive/Colab Notebooks/mask_rcnn_coco.h5"
COCO_WEIGHTS_PATH = os.path.join(ROOT_DIR, "logs", "gland20190402T1459", "mask_rcnn_gland_0178.h5")

DATASET_PATH = os.path.join(ROOT_DIR, "dataset")

# Directory to save logs and model checkpoints, if not provided
# through the command line argument --logs
DEFAULT_LOGS_DIR = os.path.join(ROOT_DIR, "logs")
```

Рисунок 3.3 – Скрипт налаштування проекту

У блоці з конфігурацією відбувається налаштування основних параметрів для навчання мережі - лейбл об'єкта, що досліджується, кількість зображень, що подаються на GPU, кількість кроків в епісі навчання, мінімальне значення передбачення (рис. 3.4).

```
[ ] class CustomConfig(Config):
    """Configuration for training on the toy dataset.
    Derives from the base Config class and overrides some values.
    """
    # Give the configuration a recognizable name
    NAME = "gland"

    # We use a GPU with 12GB memory, which can fit two images.
    # Adjust down if you use a smaller GPU.
    IMAGES_PER_GPU = 2

    # Number of classes (including background)
    NUM_CLASSES = 1 + 1 # Background + toy

    # Number of training steps per epoch
    STEPS_PER_EPOCH = 100

    # Skip detections with < 90% confidence
    DETECTION_MIN_CONFIDENCE = 0.9
```

Рисунок 3.4 – Скрипт налаштування ШНМ

3.3 Розмітка зображення

У проєкті використовуються навчальний та валідаційний набори даних у відповідних папках. Розмітка проводилася інструментом VGG Image Annotator (VIA).

VIA — це просте та автономне ПЗ для ручного анування зображень, аудіо та відео. VIA працює у веб-браузері та не потребує встановлення чи налаштування. Повне ПЗ вміщується в одну самостійну HTML-сторінку розміром менше 400 кілобайт, яка працює як автономна програма в більшості сучасних веб-браузерів. Є проєктом із відкритим кодом, який базується виключно на HTML, Javascript і CSS (не залежить від зовнішніх бібліотек). VIA

розроблено і випущено згідно з ліцензією BSD-2, що дозволяє використовувати його як для академічних проектів, так і для комерційних програм.

У проекті виділяються необхідні області на зображенні та підписуються. Потім зображення з цими контурами-анотаціями передаються у мережу для навчання. По координатах контурів створюється маска кожного зображення (рис. 3.5).

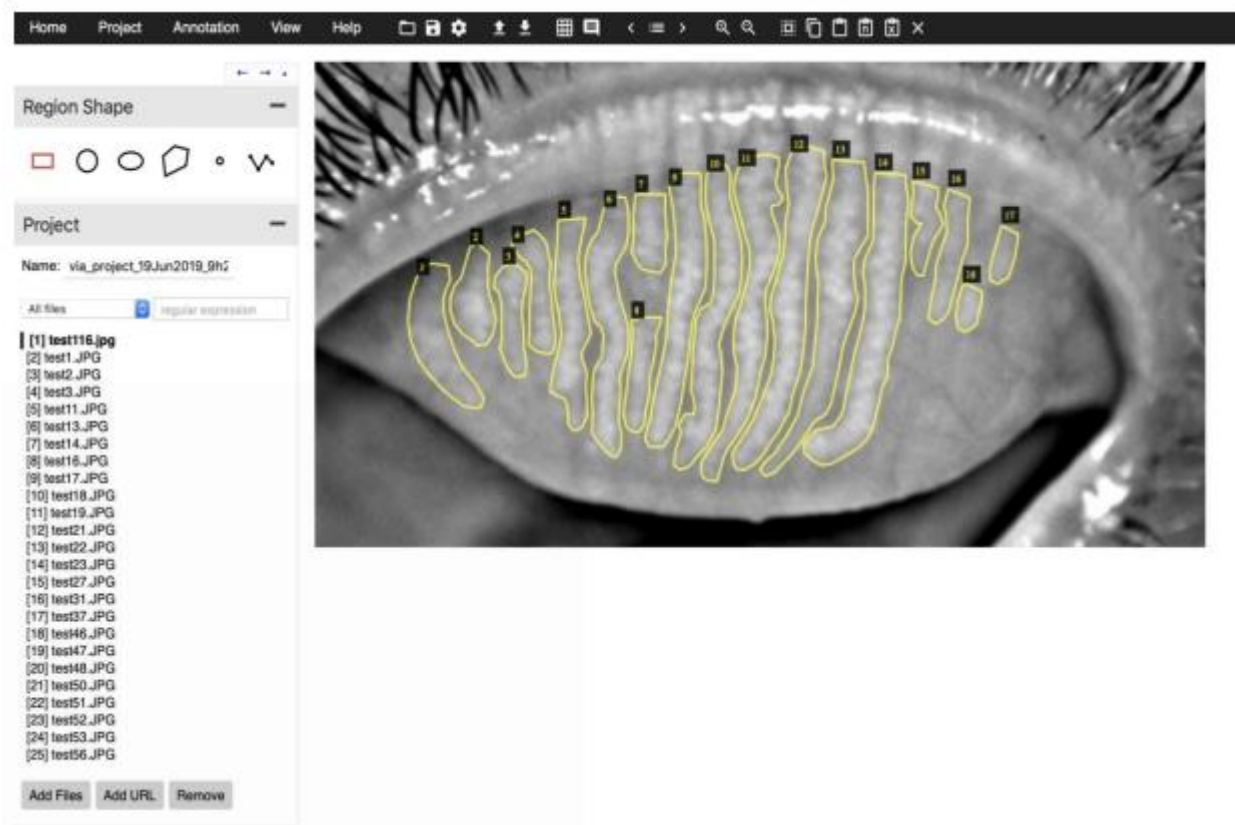


Рисунок 3.5 – Приклад розмітки зображення

3.4 Навчання та конфігурація мережі

У цьому проекті використовується дуже маленький набір даних для навчання та передвибраний файл ваг COCO dataset, тому навчати всі шари мережі немає сенсу, навчаємо лише останній вихідний шар.

Частина скрипта для навчання останнього шару мережі представлена на рис. 3.6 та 3.7.


```

def load_custom(self, dataset_dir, subset):
    """Load a subset of the Balloon dataset.
    dataset_dir: Root directory of the dataset.
    subset: Subset to load: train or val
    """
    # Add classes. We have only one class to add.
    self.add_class("gland", 1, "gland")

    # Train or validation dataset?
    assert subset in ["train", "val"]
    dataset_dir = os.path.join(dataset_dir, subset)

    # Load annotations
    # VGG Image Annotator saves each image in the form:
    # { 'filename': '28503151_5b5b7ec140_b.jpg',
    #   'regions': {
    #     '0': {
    #       'region_attributes': {},
    #       'shape_attributes': {
    #         'all_points_x': [...],
    #         'all_points_y': [...],
    #         'name': 'polygon'}},
    #     ... more regions ...
    #   },
    #   'size': 100202
    # }
    # We mostly care about the x and y coordinates of each region
    annotations1 = json.load(open(os.path.join(dataset_dir, "via_region_data.json")))
    # print(annotations1)
    annotations = list(annotations1.values()) # don't need the dict keys

    # The VIA tool saves images in the JSON even if they don't have any
    # annotations. Skip unannotated images.
    annotations = [a for a in annotations if a['regions']]

```

Рисунок 3.6 – Фрагмент скрипту навчання останнього шару

Проміжні файли ваги зберігаються в папці logs. Далі є можливість дослідити кожен файл, наскільки добре мережа справляється з завданням аналізу об'єктів на зображенні. Це потрібно робити за допомогою файла inspect_model.

```

[ ] config = CustomConfig()
    config.display()

# Create model
model = modellib.MaskRCNN(mode="training", config=config,
                           model_dir=DEFAULT_LOGS_DIR)

weights_path = COCO_WEIGHTS_PATH
# Download weights file
if not os.path.exists(weights_path):
    utils.download_trained_weights(weights_path)

# Load weights
# Exclude the last layers because they require a matching
# number of classes
model.load_weights(weights_path, by_name=True, exclude=[
    "mrcnn_class_logits", "mrcnn_bbox_fc",
    "mrcnn_bbox", "mrcnn_mask"])

train(model)

```

Рисунок 3.7 – Скрипт продовження навчання мережі

Також існує конфігурація нейронної мережі, де зібрані всі відомості про неї (назва, розміри аналізованих зображень, кількість епох, кроки кожної епохи і т.д.). Вона потрібна для того, щоб можна було змінювати налаштування та бачити поточні параметри (рис. 3.8).

```

[ ]
+ Configurations:
BACKBONE                resnet101
BACKBONE_STRIDES        [4, 8, 16, 32, 64]
BATCH_SIZE              1
BOX_STD_DEV             [0.1 0.1 0.2 0.2]
COMPUTE_BACKBONE_SHAPE None
DETECTION_MAX_INSTANCES 100
DETECTION_MIN_CONFIDENCE 0.9
DETECTION_NMS_THRESHOLD 0.3
FPN_CLASSIF_FC_LAYERS_SIZE 1024
GPU_COUNT              1
GRADIENT_CLIP_NORM     5.0
IMAGES_PER_GPU         1
IMAGE_MAX_DIM          1024
IMAGE_META_SIZE        14
IMAGE_MIN_DIM          800
IMAGE_MIN_SCALE        0
IMAGE_RESIZE_MODE      square
IMAGE_SHAPE            [1024 1024  3]
LEARNING_MOMENTUM      0.9
LEARNING_RATE          0.001
LOSS_WEIGHTS           {'rpn_class_loss': 1.0, 'rpn_bbox_loss': 1.0, 'mrcnn_class_loss': 1.0, 'mrcnn_bbox_loss': 1.0, 'mrcnn_mask_loss': 1.0}
MASK_POOL_SIZE         14
MASK_SHAPE             [28, 28]
MAX_GT_INSTANCES       100
MEAN_PIXEL             [123.7 116.8 103.0]
MINI_MASK_SHAPE        (56, 56)
NAME                   gland
NUM_CLASSES            2
POOL_SIZE              7
POST_NMS_ROIS_INFERENCE 1000
POST_NMS_ROIS_TRAINING 2000
ROI_POSITIVE_RATIO     0.33
RPN_ANCHOR_RATIOS     [0.5, 1, 2]
RPN_ANCHOR_SCALES     (32, 64, 128, 256, 512)
RPN_ANCHOR_STRIDE     1
RPN_BOX_STD_DEV       [0.1 0.1 0.2 0.2]
RPN_NMS_THRESHOLD     0.7
RPN_TRAIN_ANCHORS_PER_IMAGE 256
STEPS_PER_EPOCH       100
TOP_DOWN_PYRAMID_SIZE 256
TRAIN_BN              False
TRAIN_ROIS_PER_IMAGE  200
USE_MINI_MASK         True
USE_RPN_ROIS         True
VALIDATION_STEPS      8
WEIGHT_DECAY          0.0001

```

Рисунок 3.8 – Скріншот конфігурації ШНМ

У файлі `inspect_model` ми знову імпортуємо необхідні бібліотеки і вказуємо шлях до папки з даними для проведення аналізу. Ця програма працює з конкретним зображенням. Вказуємо шлях потрібного зображення та пропускаємо через нейронну мережу (рис. 3.9).

▼ Load Validation Dataset

```
[ ] # Load validation dataset
dataset = custom.CustomDataset()
dataset.load_custom(custom_DIR, "val")

# Must call before using the dataset
dataset.prepare()

print("Images: {}\nClasses: {}".format(len(dataset.image_ids), dataset.class_names))
```

```
Images: 8
Classes: ['BG', 'gland']
```

Рисунок 3.9 – Скрипт завантаження зображення

В результаті роботи програми отримується вихідне зображення, але вже із зафарбованими областями. Нейронна мережа змогла знайти необхідні об'єкти та зафарбувала області їхнього знаходження, а також вказала ймовірність істини (рис. 3.10).

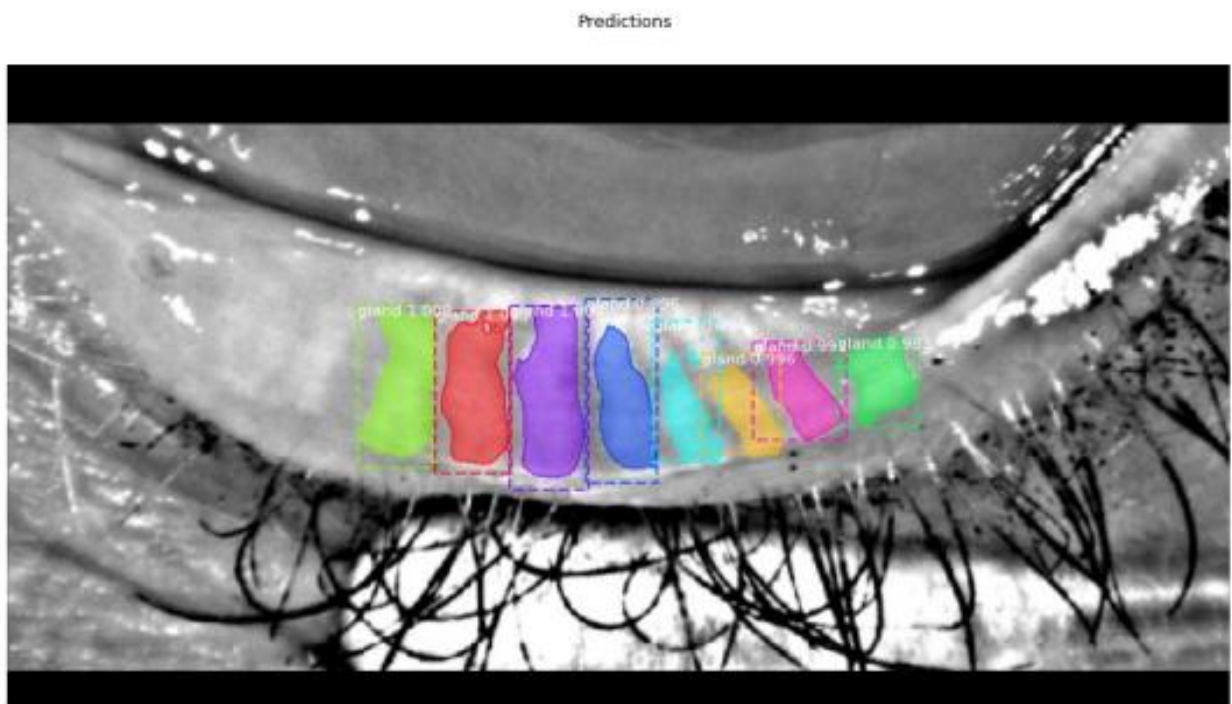


Рисунок 3.10 – Готовий результат

Одержаний результат є найточнішим для даної нейронної мережі. Вона

була навчена при допомозі близько 150 зображень. Варто зауважити, що для якісної її роботи потрібно використати значно більше навчальних даних, деє близько 100 тисяч.

РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ

4.1 Класифікація шкідливих та небезпечних виробничих факторів

Шкідливий виробничий фактор – небажане явище, яке супроводжує виробничий процес і вплив якого на працюючого може призвести до погіршення самопочуття, зниження працездатності, захворювання, виробничо зумовленого чи професійного, і навіть смерті, як результату захворювання. Небезпечний виробничий фактор – небажане явище, яке супроводжує виробничий процес і дія якого за певних умов може призвести до травми або іншого раптового погіршення здоров'я працівника (гострого отруєння, гострого захворювання) і навіть до раптової смерті [37].

Поділ несприятливих чинників виробничого середовища на шкідливі та небезпечні зумовлене різним характером їх дії на людський організм, тим, що вони потребують різних заходів та засобів для боротьби з ними та профілактики викликаних ними ушкоджень, а також рядом причин організаційного характеру. В той же час між шкідливими та небезпечними виробничими факторами інколи важко провести чітку межу. Один і той же чинник може викликати травму і профзахворювання (наприклад, високий рівень іонізуючого або теплового випромінювання може викликати опік або навіть призвести до миттєвої смерті, а довготривала дія порівняно невисокого рівня цих же факторів – до хвороби; пилинки, що потрапили в око, спричиняє травму, а пил, що осідає в легенях, – захворювання, що зветься пневмоконіоз). Через це всі несприятливі виробничі чинники часто розглядаються як єдине поняття – небезпечний та шкідливий виробничий фактор (НШВФ) [38]. За своїм походженням та природою дії всі НШВФ можна поділити на 5 груп: фізичні, хімічні, біологічні, психофізіологічні та соціальні. До фізичних НШВФ відносяться машини та механізми або їх елементи, а також вироби, матеріали, заготовки тощо, які рухаються або обертаються; конструкції, які руйнуються; системи, устаткування або елементи обладнання, які знаходяться під підвищеним тиском; підвищена запиленість та

загазованість повітря; підвищена або понижена температура повітря, поверхонь приміщення, обладнання, матеріалів; підвищені рівні шуму, вібрації, ультразвуку, інфразвуку; підвищений або понижений барометричний тиск та його різкі коливання; підвищена та понижена вологість; підвищена швидкість руху та підвищена іонізація повітря; підвищений рівень іонізуючих випромінювань; підвищене значення напруги в електричній мережі; підвищені рівні статичної електрики, електромагнітних випромінювань; підвищена напруженість електричного, магнітного полів; відсутність або нестача світла; недостатня освітленість робочої зони; підвищена яскравість світла; понижена контрастність; прямий та віддзеркалений блиск; підвищена пульсація світлового потоку; підвищені рівні ультрафіолетової та інфрачервоної радіації; гострі крайки, зачипки, шершавість на поверхні заготовок, інструментів та обладнання; розташування робочого місця на значній висоті відносно землі (підлоги); слизька підлога; невагомість.

Хімічні НШВФ:

- за характером дії на організм людини поділяються на токсичні, задушливі, наркотичні, подразнюючі, сенсibiliзуючі, канцерогенні, мутагенні та такі, що впливають на репродуктивну функцію;

- за шляхами проникнення в організм людини поділяються на такі, що потрапляють через: 1) органи дихання; 2) шлунково-кишковий тракт; 3) шкіряні покриви та слизова оболонка;

- які перебувають у різному агрегатному стані: 1) твердому 2) газоподібному 3) рідкому.

Біологічні НШВФ – це: - патогенні мікроорганізми (бактерії, віруси, рикетсії, спірохети, грибки, найпростіші) та продукти їхньої життєдіяльності; - макроорганізми (тварини та рослини) та продукти їхньої життєдіяльності. До психофізіологічних НШВФ відносяться фізичні (статичні та динамічні) перевантаження і нервово-психічні перевантаження (розумове перенапруження, перенапруження аналізаторів, монотонність праці, емоційні перевантаження). 6 Соціальні НШВФ – це неякісна організація роботи, понаднормова робота, змушеність праці в колективі з поганими відносинами між його членами,

соціальна ізоляція з відривом від сім'ї, зміна біоритмів, незадоволеність роботою, фізична та/або словесна образа та її ризик, насильство та його ризик. Один і той же НШВФ за природою своєї дії може належати водночас до різних груп.

4.2 Вплив вібрації на людину

Вібрація - це механічні коливання пружних тіл або коливальні рухи механічних систем. Для людини вібрація є видом механічного впливу, який має негативні наслідки для організму [39].

Причиною появи вібрації є неврівноважені сили та ударні процеси в діючих механізмах. Створення високопродуктивних потужних машин і швидкісних транспортних засобів при одночасному зниженні їх матеріалоемності неминуче призводить до збільшення інтенсивності і розширення спектру вібраційних та віброакустичних полів. Цьому сприяє також широке використання в промисловості і будівництві високоефективних механізмів вібраційної та віброударної дії.

Дія вібрації може приводити до трансформування внутрішньої структури і поверхневих шарів матеріалів, зміни умов тертя і зносу на контактних поверхнях деталей машин, нагрівання конструкцій. Через вібрацію збільшуються динамічні навантаження в елементах конструкцій, стиках і сполученнях, знижується несуча здатність деталей, ініціюються тріщини, виникає руйнування обладнання. Усе це приводить до зниження строку служби устаткування, зростання імовірності аварійних ситуацій і зростання економічних витрат. Вважають, що 80% аварій в машинах і механізмах здійснюється внаслідок вібрації. Крім того, коливання конструкцій часто є джерелом небажаного шуму. Захист від вібрації є складною і багатоплановою в науково-технічному та важливою у соціально-економічному відношеннях проблемою нашого суспільства [37].

Вплив вібрації на людину залежить від її спектрального складу, напрямку дії, прикладення, тривалості впливу, а також від індивідуальних особливостей людини. При оцінці вібраційного впливу потрібно враховувати, що коливальні

процеси притаманні живому організму. В основі серцевої діяльності і кровообігу та біоелектричних процесів мозку лежать ритмічні коливання. Внутрішні органи людини можна розглядати як коливальні системи з пружними зв'язками. Частоти їх власних коливань лежать у діапазоні 3..6 Гц. Частоти власних коливань плечового пояса, стегон і голови щодо опорної поверхні (положення стоячи) складають 4..6 Гц, голови щодо плечей (положення сидячи) 25...30 Гц.

При впливі на людину зовнішніх коливань (хитавиці, струсів, вібрації) відбувається їхня взаємодія з внутрішніми хвильовими процесами, виникнення резонансних явищ. Так, зовнішні коливання частотою менш 0,7 Гц утворюють хитавицю і порушують у людини нормальну діяльність вестибулярного апарата. Інфразвукові коливання (менш 16 Гц), впливаючи на людину, пригнічують центральну нервову систему, викликаючи почуття тривоги, страху. При певній інтенсивності на частоті 6..7 Гц інфразвукові коливання, втягуючи у резонанс внутрішні органи і систему кровообігу, здатні викликати травми, розриви артерій, тощо [38].

Вібрація, що діє на людину, має широкий діапазон – від десятків частот одного до декількох тисяч Гц. Характерними ознаками шкідливого впливу вібрації на людину є можливі зміни у функціональному стані: підвищена втома, збільшення часу моторної реакції, порушення вестибулярної реакції. Медичними дослідженнями встановлено, що вібрація є подразником периферичних нервових закінчень, розташованих на ділянках тіла людини, що сприймають зовнішні коливання. Адекватним фізичним критерієм оцінки її впливу на організм людини є коливальна енергія, що виникає на поверхні контакту, а також енергія, поглинена тканинами і передана опорно-руховому апарату та іншим органам. У результаті впливу вібрації виникають нервовосудинні розлади, ураження кістково-суглобної та інших систем організму. Відзначаються, наприклад, зміни функції щитовидної залози, сечостатевої системи, шлунково-кишкового тракту. Так, медичні дослідження показали, що у працюючих в умовах вібрації відбуваються значні зміни кістковосуглобної системи, які виражаються у функціональній перебудові кісткової тканини, регіональному остеопорозі, кістковидних утвореннях у кістках, асептичному некрозі кісток, хронічних

переломах. Відзначається, що терміни виникнення змін у кістках у працівників вібраційних професій коливається в межах від 6-8 місяців до 2-5 років.

Шкідливість вібрації збільшується при одночасному впливі на людину таких факторів, як знижена температура, підвищений шум, запиленість повітря, тривала статична напруга тощо. Сучасна медицина розглядає виробничу вібрацію як могутній стрес-фактор, що має негативний вплив на психомоторну працездатність, емоційну сферу і розумову діяльність, підвищує ймовірність виникнення різних захворювань і нещасних випадків. Особливо небезпечний тривалий вплив вібрації для жіночого організму. Широкий комплекс патологічних відхилень, викликаний впливом вібрації на організм людини, кваліфікується як віброзахворювання [39].

Вібрація як фізичний чинник виробничого середовища спостерігається в металообробній, гірничодобувній, металобудівній, машинобудівній, авіаційній та інших галузях народного господарства. Джерелом вібрації можуть бути різні механізми, вібраційне устаткування, віброінструменти, акустичні системи, транспортні та сільськогосподарські машини.

Загальна вібрація поділяється на транспортну вібрацію, яка діє на людину на робочих місцях в транспортних засобах (трактори сільськогосподарські та промислові, самохідні сільськогосподарські машини (комбайни), тягачі, грейдери ті інші); транспортно-технологічну вібрацію, яка діє на людину на робочих місцях машин з обмеженою рухливістю (екскаватори, крани промислові та будівельні, гірничі комбайни, транспорт виробничих приміщень та інші) та технологічну вібрацію, яка діє на людину на робочих місцях стаціонарних машин чи передається на робочі, де немає джерел вібрації (верстати та метало-деревообробне, пресувально-ковальське обладнання, ливарні машини, електричні машини, насосні агрегати та вентилятори, обладнання для буріння свердловин, бурові верстати, машини для тваринництва, очищення та сортування зерна (у тому числі сушарні), обладнання промисловості будматеріалів (крім бетоноукладачів), установки хімічної та нафтохімічної промисловості та інші.

Оператори машин, які зазнають у процесі трудової діяльності впливу вібрації, підлягають попереднім та періодичним медичним оглядам відповідно

до Порядку проведення медичних оглядів працівників певних категорій, затвердженого Наказом МОЗ України від 21.05.2007 р. №246. Обов'язкові попередні (під час прийняття на роботу) та періодичні (протягом трудової діяльності) медичні огляди дозволять визначити стан здоров'я працівника та можливість виконання без погіршення стану здоров'я професійних обов'язків, своєчасно виявити ранні ознаки хронічного професійного захворювання, забезпечує динамічне спостереження за станом здоров'я в умовах дії шкідливих та небезпечних факторів і трудового процесу, вирішує питання щодо можливості продовжувати роботу в умовах дії шкідливих та небезпечних факторів і трудового процесу [39].

За результатами періодичних медичних оглядів роботодавець забезпечує проведення відповідних оздоровчих заходів Заключного акта у повному обсязі та усуває причини, що призводять до професійних захворювань. Організовує проведення лабораторних досліджень умов праці на робочих місцях та вживає заходів до усунення небезпечних і шкідливих для здоров'я виробничих факторів.

До роботи операторами машин допускаються особи не молодші 18 років, які пройшли попередній медичний огляд, мають відповідну кваліфікацію та ознайомлені з характером впливу вібрації на організм.

ВИСНОВКИ

З кожним роком АД дедалі більше впливає на оточуюче середовище. Темпи розвитку технологій дуже швидкі, але АД від них не відстає. Нещодавно з'явився новий напрямок аналізу даних - Big data. Цей напрямок може застосовуватися у будь-яких галузях: медицина, бізнес, політика, виробництво тощо. Технології Big data приходять на допомогу, коли йдеться про колосальні обсяги різних сирих даних. Також зросла потреба у кваліфікованих фахівцях з АД (data scientist) і найближчим часом лише зростатиме. Адже мало мати лише потужну техніку та нові передові технології, треба мати також фахівців, які грамотно працюватимуть з усім цим.

Основні результати, отримані в роботі:

- вивчені спеціалізовані бібліотеки мови програмування Python, котрі використовуються для здійснення аналізу великих даних;
- досліджено онлайн-платформи з використанням хмарних web - сервісів Amazon та машинного навчання Kaggle;
- описано особливості застосування ШНМ при АД у різних сферах;
- реалізовано практичне завдання АД у медичній сфері. Було підібрано дані, навчено нейронну мережу та написано необхідні скрипти для виконання АД.

Головними завданнями цієї програми є виявлення та підрахунок кількості очних залоз на спеціально підготовлених фотографіях. Дане завдання залишається поки більше навчальним, але при відповідному доопрацюванні може бути корисним в офтальмології.

ПЕРЕЛІК ДЖЕРЕЛ

1. John W. Tukey. The Future of Data Analysis // The Annals of Mathematical Statistics - Vol. 33, No. 1 (Mar., 1962), pp. 1-67.
2. Gregory Piatetsky-Shapiro. [Електронний ресурс]. – Режим доступу: https://en.wikipedia.org/wiki/Gregory_Piatetsky-Shapiro (Дата звертання: 28.04.2023)
3. Черняк О.І. Інтелектуальний аналіз даних: підручник. – К: Знання, 2014. – 599 с.
4. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020. – 298 p.
5. Данильченко О.М., Данильченко А.О. Інтелектуальний аналіз даних: Навч. посібник. – Житомир: ЖДТУ, 2009. – 405 с.
6. Задачі Data Mining та їх класифікація. Інформація та знання [Електронний ресурс]. – Режим доступу: https://moodle.znu.edu.ua/pluginfile.php/486125/mod_resource/content/1/Лекція%205.pdf (Дата звертання: 28.04.2023).
7. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. — К.: КН ЕУ, 2007. — 376 с.
8. Технології Big Data: ключові характеристики, особливості та переваги. [Електронний ресурс]. – Режим доступу: <https://aiconference.com.ua/uk/news/tehnologii-big-data-klyuchevie-harakteristiki-osobennosti-i-preimushchestva-97883> (Дата звертання: 28.04.2023).
9. Шумейко А. А. Интеллектуальный анализ данных (Введение в Data Mining): учеб. пособ. / А. А. Шумейко, С. Л. Сотник. – Днепропетровск: Белая Е.А., 2012. – 212 с.
10. Вайгенд А. BIG DATA. Вся технология в одной книге. М.: Эксмо, 2018. – 384 с.
11. Що таке Big Data: все що вам слід знати про великі дані? [Електронний ресурс]. – Режим доступу: <https://futurenow.com.ua/shho-take-big-data-velyki-dani/> (Дата звертання: 28.04.2023).

12. Волкова С. Просто BIG DATA. СПб.: Страта, 2019. – 148 с.
13. Фрэнк Билл. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики М.: Альпина Паблицер, 2014. – 430 с.
14. Методи аналізу даних: навчальний посібник для студентів /. В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
15. Зміни клімату: як глобальне потепління вплине на Землю. [Електронний ресурс]. – Режим доступу: <https://www.slovoidilo.ua/2021/10/04/video/suspilstvo/zminy-klimatu-yak-hlobalne-poteplinna-vplyne-zemlyu> (Дата звертання: 28.04.2023).
16. Python – Object Oriented. [Електронний ресурс]. – Режим доступу: https://www.tutorialspoint.com/python/python_classes_objects.htm (Дата звертання: 28.04.2023).
17. Маккинни У. Python и анализ данных – М.: ДМК Пресс, 2015. 482 с.
18. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.
19. Свейгарт Э. Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих. – М.: Вильямс, 2017. – 573 с.
20. Плас Дж.В. Python для сложных задач. Наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.
21. Jupyter Notebook [Електронний ресурс]. – Режим доступу: <https://jupyter.org/> (Дата звертання: 28.04.2023).
22. Pandas [Електронний ресурс]. – Режим доступу: <https://github.com/pandas-dev/pandas> (Дата звертання: 28.04.2023).
23. Pandas [Електронний ресурс]. – Режим доступу: https://www.w3schools.com/python/pandas/pandas_intro.asp (Дата звертання: 28.04.2023).
24. NumPy Introduction. [Електронний ресурс]. – Режим доступу: https://www.w3schools.com/python/numpy/numpy_intro.asp (Дата звертання: 28.04.2023).
25. SciPy. Електронний ресурс]. – Режим доступу: <https://scipy.org/> (Дата звертання: 28.04.2023).

звертання: 28.04.2023).

26. Matplotlib: Visualization with Python. [Електронний ресурс]. – Режим доступу: <https://matplotlib.org/> (Дата звертання: 28.04.2023).

27. An introduction to machine learning with scikit-learn [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/tutorial/basic/tutorial.html> (Дата звертання: 28.04.2023).

28. Theano. [Електронний ресурс]. – Режим доступу: <https://theanopublic.readthedocs.io/en/latest/> (Дата звертання: 28.04.2023).

29. TensorFlow. [Електронний ресурс]. – Режим доступу: <https://www.tensorflow.org/> (Дата звертання: 28.04.2023).

30. Scrapy. [Електронний ресурс]. – Режим доступу: <https://scrapy.org/> (Дата звертання: 28.04.2023).

31. AWS. [Електронний ресурс]. – Режим доступу: https://aws.amazon.com/?nc1=h_ls (Дата звертання: 28.04.2023).

32. Kaggle. [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/> (Дата звертання: 28.04.2023).

33. Open Knowledge Foundation [Електронний ресурс]. – Режим доступу: <https://okfn.org> (Дата звертання: 28.04.2023).

34. Що таке нейронні мережі та як вони працюють? Класифікація штучних нейромереж [Електронний ресурс]. – Режим доступу: <https://livingfo.com/shcho-take-nejronni-merezhi-ta-iaak-vony-pratsiuiut/> (Дата звертання: 28.04.2023).

35. Нейронні мережі: їх застосування, робота. [Електронний ресурс]. – Режим доступу: [https://www.poznavayka.org/uk/nauka-i-tehnika-2/neyronni-merezhi-yih-zastosuvannya-robota/](https://www.poznavayka.org/uk/nauka-i-tehnika-2/neyronni-merezhi-yih-zastosuvannya-robot/) (Дата звертання: 28.04.2023).

36. Big Data: the 3 VS explained [Електронний ресурс]. – Режим доступу: <https://bigdataidn.com/intelligence/big-data-the-3-vs-explained/> (Дата звертання: 28.04.2023).

37. Зеркалов Д.В. Безпека життєдіяльності та основи охорони праці. Навчальний посібник. К.: «Основа». 2016. – 267 с.

38. Яремко З. М. Безпека життєдіяльності: Навч. посіб. — Львів., 2005. –

301 с.

39. Желібо Є. П. Заверуха Н.М., Зацарний В.В. Безпека життєдіяльності. Навчальний посібник. – К.; Каравела, 2004. -328 с.

ДОДАТКИ

Фрагмент лістингу розробленого програмного забезпечення

Файл `maskrcnn`:

```

import os import
sys import json
import datetime
import numpy as
np import
skimage.draw
import cv2

# Root directory of the project
ROOT_DIR = os.path.abspath(7content/gdrive/My
Drive/mask_rcnn_detection/")

# Import Mask RCNN
sys.path.append(ROOT_DIR) # To find local version of the library
from mrcnn.config import Config from mrcnn import model as
modellib, utils

# Path to trained weights файл
# COCO_WEIGHTS_PATH = "/content/gdrive/My Drive/Colab Note-
books/mask_rcnn_coco.h5"
COCO_WEIGHTS_PATH = os.path.join(ROOT_DIR, "logs",
"gland20190402T 1459", "mask_rcnn_gland_0178.h5")
DATASET_PATH = os.path.join(ROOT_DIR, "dataset")

# Directory to save logs and model checkpoints, if not provided
# through the command line argument --logs
DEFAULT_LOGS_DIR = os.path.join(ROOT_DIR, "logs")

class CustomConfig(Config):
    Configuration for training on the toy dataset.
    Derives from the base Config class and overrides some
    values.
    MMM

# Give the configuration a
recognizable name NAME = "gland"

# We use a GPU with 12GB memory,
which can fit two images. # Adjust
down if you use a smaller GPU.
IMAGE_S_PER_GPU = 2

# Number of classes (including background)
NUM_CLASSES = 1 + 1 # Background + toy

# Number of training steps
.....

```

Файл inspect_model:

```
import os
import cv2
import sys
import random
import math
import re
import time
import numpy as np
import tensorflow as tf
import matplotlib
import matplotlib.pyplot as plt
import matplotlib.patches as patches
import
skimage
import glob

# Root directory of the project
ROOT_DIR = '/content/gdrive/My Drive/mask_rcnn_detection'

# Import Mask RCNN
sys.path.append(ROOT_DIR) # To find the local version of library
from mrcnn import utils
від mrcnn import visualize
від mrcnn.visualize import display_images
import mrcnn.model as modellib
від mrcnn.model import log
import custom
%matplotlib inline

# Directory to save logs i trained
model MODEL_DIR =
os.path.join(ROOT_DIR, "logs")
CUSTOM_WEIGHTS_PATH = os.path.join(ROOT_DIR, "logs",
"gland20190402T1459", "mask_rcnn_gland_0200.h5")n(model)
config = custom.CustomConfig()
custom_DIR = os.path.join(ROOT_DIR, "dataset")
config = custom.CustomConfigO
custom_DIR = os.path.join(ROOT_DIR, "dataset")# Override the
training conconfig = custom.CustomConfig()
custom_DIR = os.path.join(ROOT_DIR, "dataset")figurations with few
# changes for inferencing.
class InferenceConfig(config. class ) :
    # Run detection on image at a
    time GPU_COUNT = 1 IMAGES PER
    GPU = 1
.....
```