

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

МАЖИРІН ВАДИМ ОЛЕКСАНДРОВИЧ

Допускається до захисту:
завідувач кафедри
інформаційних технологій,
д. т. н., доцент
Т. В. Нескородева
«_____» _____ 2022р.

**ДОСЛІДЖЕННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ
ЦУКРОВОГО ЗАВОДУ**

Спеціальність 122 «Комп'ютерні науки»

Кваліфікаційна (магістерська) робота

Науковий керівник:
Нескородева Т.В.,
д-р техн. наук, доцент
завідувач кафедри
інформаційних технологій

_____ (підпис)

Оцінка: _____ / _____ / _____

(бали за шкалою ЄКТС/за національною
шкалою)

Голова ЕК: _____
(підпис)

Вінниця – 2022

Анотація

Мажирін В.О. Дослідження методів інтелектуального аналізу даних цукрового заводу. Спеціальність 122 «Комп'ютерні науки», Освітня програма «Комп'ютерні технології обробки даних (Data Science)». Донецький національний університет імені Василя Стуса, Вінниця, 2022.

У кваліфікаційній (магістерській) роботі досліджено методи інтелектуального аналізу даних компаній цукрової галузі. Розглянути об'єкти аналізу і визначені набори даних які характеризують діяльність компаній виробництва цукру. Сформовані набори можливих слабких та сильних властивостей виробництва, виявлення яких може бути метою аналізу. Визначені фактори, які впливають на появу даного виробництва та фактори, що впливають на продуктивність праці та родючість. Здійснений аналіз критеріїв для обрання застосунків для реалізації моніторингу на основі методів інтелектуального аналізу даних компаній цукрової галузі. Продемонстрована програмна реалізація елементів інтелектуального аналізу виробництва цукру. Розроблено модель для аналізу впливу факторів на обсяги виробництва, зокрема використання добрив на землях та наслідків їх використання у вигляді цукористості. Проведено аналіз динаміки виходу цукру та порівняння потужностей підприємства.

Ключові слова: інтелектуальний аналіз даних, набори даних, фактори, що впливають на продуктивність праці та родючість.

65 с., 18 табл., 22 рис., 1 дод., 43 джерел.

ABSTRACT

Mazhyrin V.O. Research of methods of intellectual analysis of sugar plant data. Specialty 122 "Computer science", educational program "Computer technologies of data processing (Data Science)". Vasyl' Stus Donetsk National University, Vinnytsia, 2022.

In the qualification (master's) thesis, the methods of intellectual analysis of the data of companies in the sugar industry were investigated. Consider object analysis and identify data sets that characterize the activities of sugar companies. Formed sets of possible weak and strong properties of production, the detection of which can be assigned to analysis. Factors affecting the emergence of this production and factors affecting labor productivity and fertility are defined. An analysis of the criteria for selecting applications for monitoring based on the methods of intellectual data analysis of sugar industry companies was carried out. software implementation of elements of intellectual analysis of sugar production is demonstrated. A model was developed to analyze the influence of factors on production volumes the use of fertilizers on land and the consequences of their use in the form of sugar content. An analysis of the dynamics of sugar output and a comparison of the company's capacities was carried out.

Keywords: data mining, datasets, factors affecting labor productivity and fertility.

65 pages, 18 tables, 22 pictures, 1 annex, 43 source.

ЗМІСТ

РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ.....	5
1 Дослідження предметної області аналізу	5
1.2 Область застосування технологій інтелектуального аналізу	9
1.3 Інструменти інтелектуального аналізу	12
Висновок до першого розділу	13
2 ДОСЛІДЖЕННЯ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ ОБРОБКИ ДАНИХ	15
2.2 Вибір платформ для аналізу	19
2.3 Результати аналізу програмних інструментів	23
Висновок до другого розділу	25
3. Обрання галузі та вибір підприємства для аналізу	26
3.1 Історія появи галузі	26
3.2 Характеристика підприємства	29
Висновок до третього розділу	45
4 Програмна реалізація аналізу виробництва	46
4.1 Обрання платформи та засобу реалізації аналізу	46
4.2 Аналіз обраної області	49
Висновок до четвертого розділу	57
Висновок до роботи	58
Список використаних джерел	60
Додаток	63

РОЗДІЛ 1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ

1 Дослідження предметної області аналізу

Сучасні компанії генерують великі обсяги даних. Швидкий розвиток технологій призводить до збільшення ролі пристроїв у буденному житті людей. Щоб роботизована система могла приймати вірне рішення їй необхідно мати певні навички та володіти деякою інформацією. Для цього використовується інформація про навколишнє середовище та про оточуючі об'єкти, їх вміст та зміну їх положення у просторі. Одним із таких напрямів є семантична сегментація зображення з відео потоку. На основі даних які отримуємо із камер робот здійснює певні дії. Важливою задачею є вірне розпізнання об'єкта та точно сформована поведінка машини для вирішення різних задач.

Data Mining (DM) – частина науки (інформатики) що пов'язана із дисциплінами: статистика, теорія ймовірності, штучний інтелект та машинне навчання.

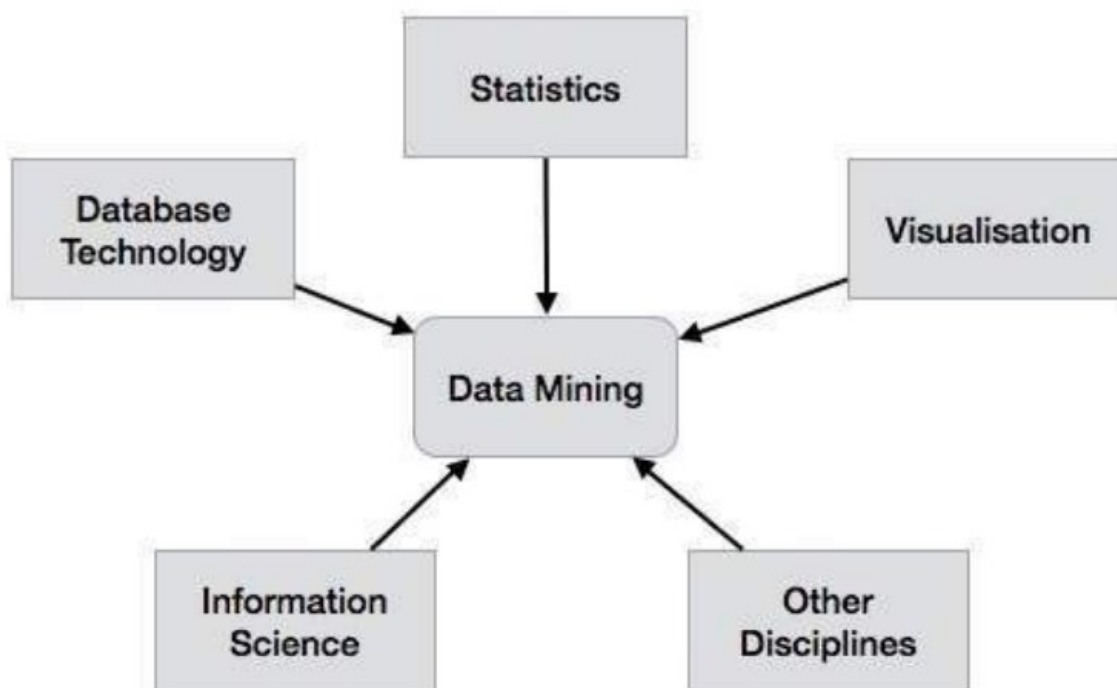


Рисунок 1 – Взаємодія Data Mining із різними дисциплінами

Завдяки своїм підходам до аналізу даних та виявленню інформації DM отримав велике признание в дослідницьких цілях. Окрім того він дає змогу отримати певні правила, процедури та алгоритми для створення гіпотез, отримання шаблонів з виявленням зв'язків в великих даних. А також включає в себе розширений пошук, обробку та моделювання даних кількома способами одночасно. Такий підхід дав змогу стати однією із найбільш впливових наук та технологією з різним застосуванням.

Основними завданнями аналізу даних є :

- Класифікація – це визнання об'єкту за певними характеристиками.
- Регресія – має завдання подібне до класифікації та дає змогу визначити параметри та характеристику цілі. Відмінність від класифікації в кінцевому результаті, параметром якого є безліч дійсних чисел.
- Визначення залежностей між подіями, об'єктами, процесами та інших дій які впливають на кінцевий результат.

Усі завдання у свою чергу діляться на описові та очікуємі.

Описові це ті що дають покращення розуміння даних що аналізується. Основним моментом в таких моделях є легкість та прозорість для сприйняття інформації людиною.

Передбачувані це завдання які діляться на два етапи. На першому будується модель. На другому використовується як прогноз на основі отриманих даних.

Основні методи Data Mining є такі методи :

1. Древа рішень
2. Алгоритм кластеризації
3. Регресійний аналіз
4. Нейронні мережі
5. Часові ряди

Дерева рішення – дає можливість розглядати різні сценарії розвитку подій під впливом різних факторів ризику.

Область використання такого методу можна об'єднати в три класи:

I. Опис даних – дозволяє зберегти інформацію про вибірку у компактній та зручній формі для подальшої обробки даних, що містить в собі точний опис об'єкту.

II. Класифікація – віднести завдання до певного класу що б точніше вирішити поставлену проблему.

III. Регресія – якщо ми маємо недостатню кількість даних або недостовірні значення то дерево рішень дає змогу визначити залежність цільової змінної від вхідних змінних.

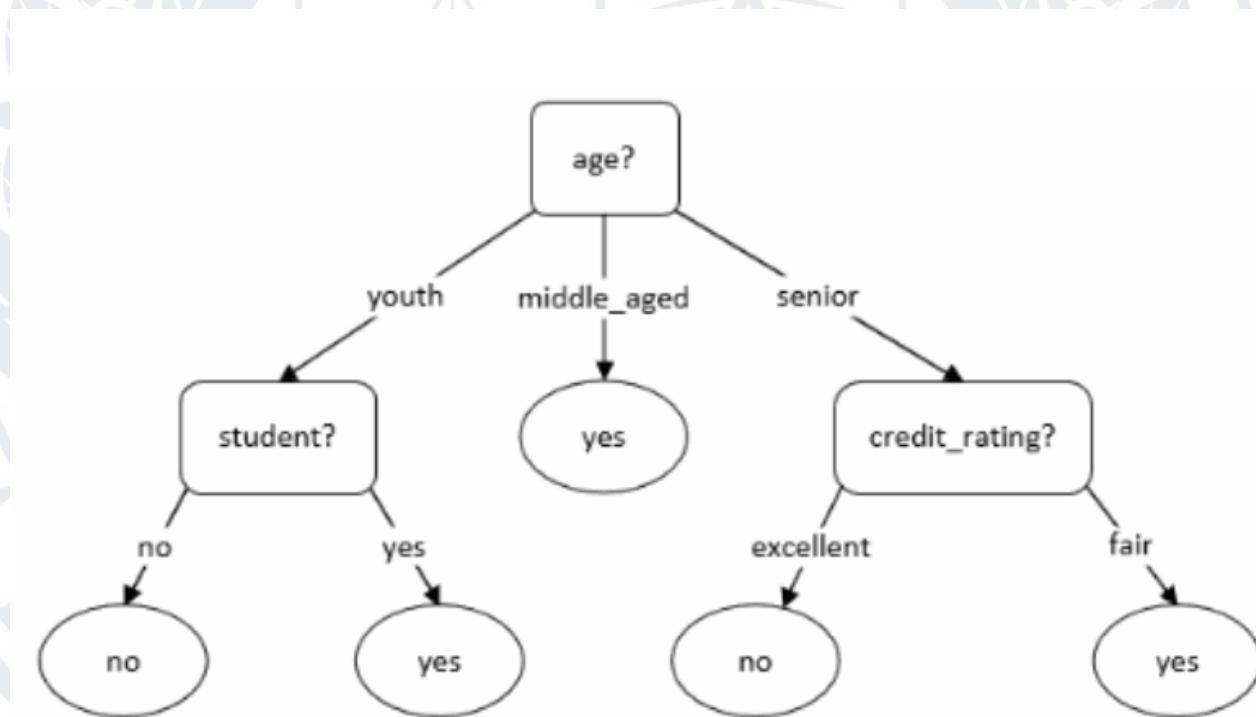


Рисунок 2 – приклад дерева рішень

Нейронні мережі – напрямок наукового дослідження в роботі створення штучного інтелекту, основою якого є прагнення створити нервову систему

людини, вивчити систему яка дасть змогу виправляти помилки та самонавчатися. Ідеальним завершенням роботи якої має бути відтворення роботи мозку людини. Найпоширенішою архітектурою є багатошаровий персептрон (рис3). Кожен нейрон вищого рівня на вході має інформацію із нейрона виходу нейрона нижчого. Таким чином це дає змогу покращити інформацію, що приблизить до нулевої ймовірності помилку таких систем.

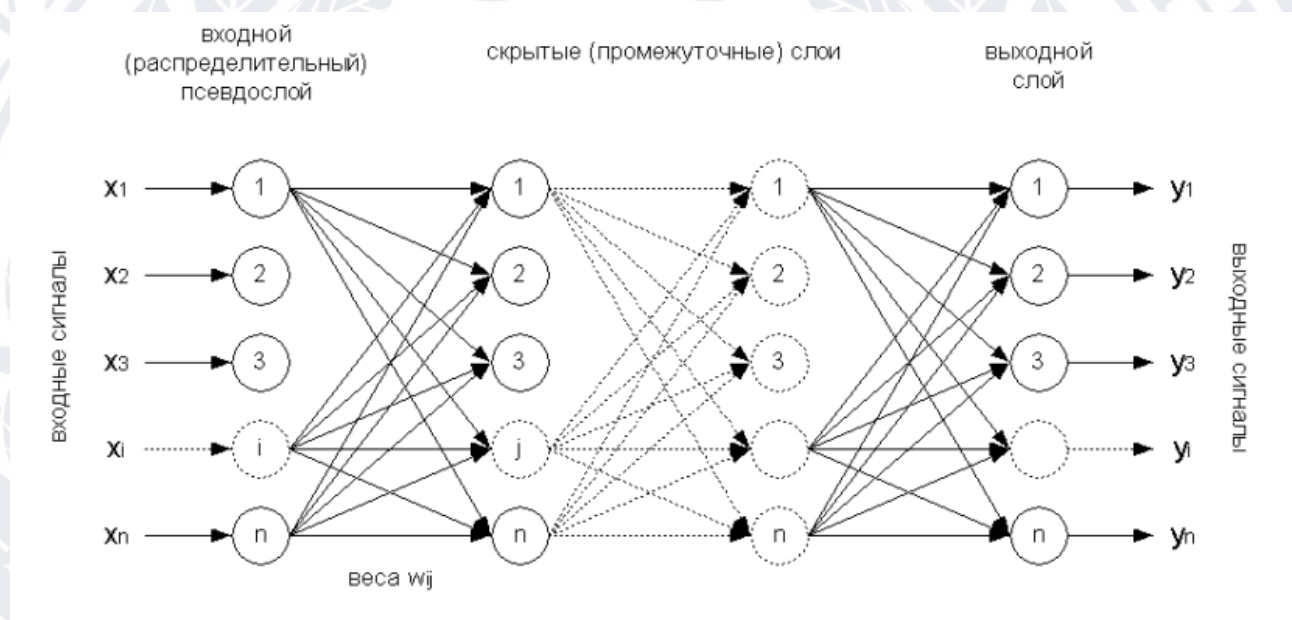


Рисунок 3 - Структура багатошарового персептрона

В результаті на виході нейрона верхнього рівня виробляється певне значення, яке рахується як відповідь усієї мережі на вхідні данні.

Для того що б мережа працювала коректно їй потрібно провести навчання на основі попередніх даних, для яких є відповіді на вхідні параметри.

Часовий ряд – статистичні показники які мають деякі зміни у часі та мають певну послідовність у своїх процесах. Дають змогу проаналізувати дії та результати в момент певного часу. Метою такого аналізу є знаходження найвищого або найнижчого рівня даних.

Спектральний аналіз дає можливість знайти повторюванні складові часового ряду. Попередні данні зображені на рисунку 4.

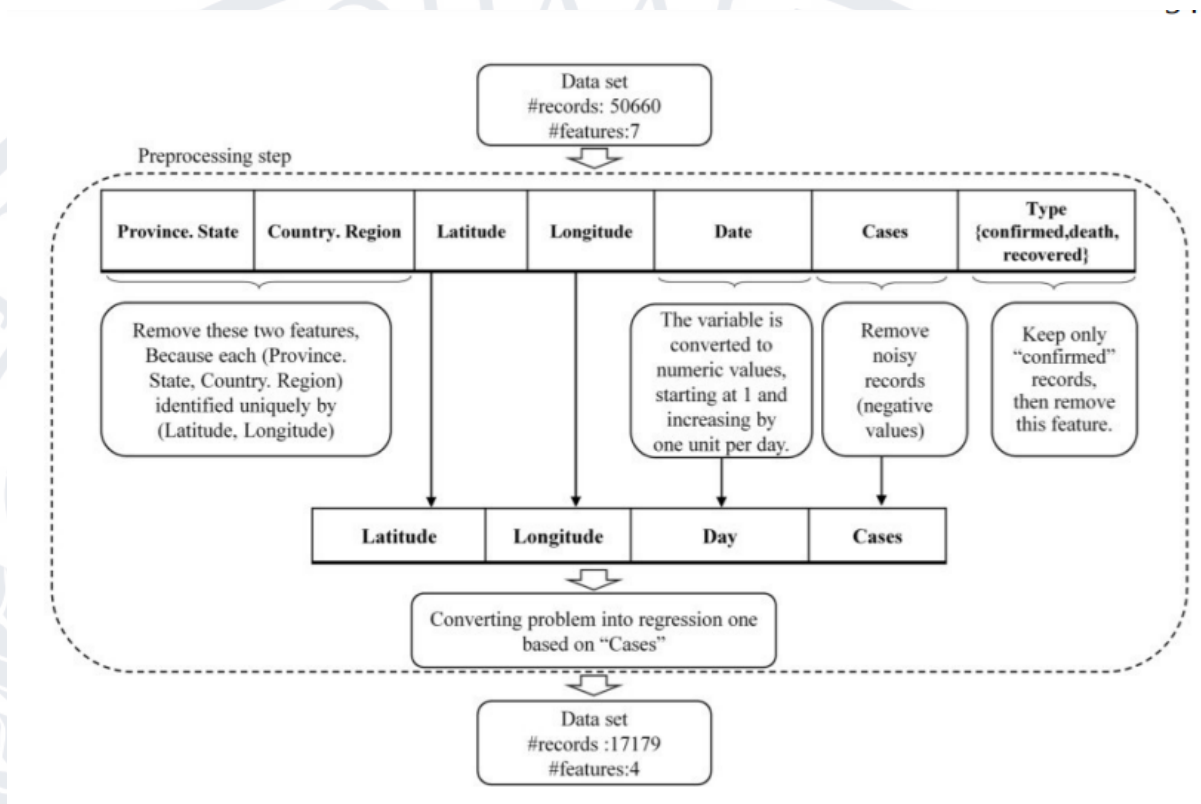


Рисунок 4 – структура попередніх даних

1.2 Область застосування технологій інтелектуального аналізу

Інтелектуальний аналіз даних (data mining) - це процес виявлення в "сирих" даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності.

Технології інтелектуального аналізу даних на сьогодні є найдосконалішим інструментом для вирішення складних аналітичних завдань. Необхідно відзначити, що технології data mining не мають власного унікального математичного апарату та програмного інструментарію, а поєднують різні математичні методи та системи штучного інтелекту.

Системи інтелектуального аналізу даних ґрунтуються на потужному математичному та статистичному апараті, грамотне застосування якого дозволяє досягти високих результатів у бізнесі.

Компанія BI Partner пропонує реалізацію повного спектру послуг у галузі інтелектуального аналізу даних, що передбачає проведення дослідження накопиченої статистики, виявлення закономірностей, створення моделі даних, її апробацію та впровадження в бізнес-процеси підприємства. ПЗ, яке ми використовуємо у своїх рішеннях, успішно використовується в багатьох російських компаніях.

Області застосування технологій інтелектуального аналізу даних.

Торгівля. Аналіз споживчого кошика, дослідження тимчасових шаблонів, створення прогнозуючих моделей, оптимізація складських запасів.

Банківська справа. Сегментація клієнтів, виявлення шахрайства із кредитними картками, прогнозування зміни клієнтури, аналіз фінансових ризиків.

Страховий бізнес. Сегментація клієнтів, виявлення шахрайства, аналіз страхових ризиків, розробка нових продуктів, розрахунок страхових премій.

Телекомунікації. Аналіз лояльності клієнтів, сегментування клієнтської бази та послуг, аналіз зовнішніх факторів на відмови обладнання, виявлення випадків несанкціонованого доступу до мережі.

Виробничі підприємства. Оптимізація закупівель, діагностика шлюбу ранніх стадіях, діагностика устаткування, маркетинг.

Нафтогазова галузь. Діагностика обладнання та нафто /газопроводів , прогнозування цін, розвідка родовищ, аналіз впливу зовнішніх та внутрішніх факторів на обсяги продажів.

Типи закономірностей, що виявляються

Основою систем data mining є виявлення різних закономірностей у даних.

Асоціація. Якщо факт А є частиною події, то з ймовірністю Х факт В буде частиною тієї самої події.

Послідовність. Якщо відбулася подія А, то з ймовірністю Х через період Т відбудеться подія В.

Класифікація. З інформації про властивості об'єкта присвоєння йому тієї чи іншої дискретного значення показника, яким проводиться класифікація.

Кластеризація. Кластеризація аналогічна до класифікації, але на відміну від останньої класи об'єктів заздалегідь не відомі.

Прогнозування. Використання наявних фактичних значень величин для прогнозування майбутніх виходячи з трендів і наявної статистики.

Математичний апарат

Основою систем data mining є методи виявлення різних закономірностей у даних: дерева рішень; алгоритми кластеризації; регресійний аналіз; нейронні сіті; часові ряди.

1.3 Інструменти інтелектуального аналізу

Програмне забезпечення таких галузей немає точного поняття та залежить від сфери на якій його застосовують. Основним призначенням інструментів data mining це пошук даних , їх необхідність, поширення та монетизація.

Основними вимогами до програмних забезпечень є:

I. Інструменти та програма мають мати простий графічний інтерфейс. Це пов'язано із тим що з такими системами працюють не тільки фахівці комп'ютерних наук , а і звичайні користувачі із різних галузей.

II. Попередня обробка даних має бути якісною – перетворення вихідних даних в зручну форму. Під обробку мається на увазі очищення даних від непотрібної інформації або доповнення відсутньої.

Інструменти Data Mining найчастіше використовуються

- спеціалістами з аналізу даних;
- маркетинговими аналітиками;
- спеціалістами з аналізу ризиків;
- маркетинговими аналітиками;
- спеціалістами із запобігання шахрайським діям.

Висновок до першого розділу

Отже інтелектуальний аналіз даних відіграє важливу роль в дослідженнях різних галузей виробництва. Використовуючи певні знання та правильно підібрані данні можна набагато спростити собі роботу. Аналіз даних має в собі багато способів для його опрацювання. Також у даному розділі ми дізнались що технологія інтелектуального аналізу даних використовується для великого обсягу даних. Такий метод дає можливість вирішити завдання класифікації, регресії, пошук асоціативних правил, кластеризації та тому подібне. Було вказано призначення та використання таких методів. Навіщо потрібний аналіз даних? Процес аналізу складається з кількох етапів. Аналітику слідуює:

Сформулювати цілі аналізу. Для цього необхідно визначитися з власними очікуваннями та питаннями, на які потрібно отримати відповіді. Слід точно розуміти, які дані мають бути зібрані (наприклад, про кількість відвідувачів, оформлені замовлення, відкриті листи і т.д.).

Зібрати інформацію з різних джерел, об'єднавши її, очистивши, упорядкувавши та систематизувавши.

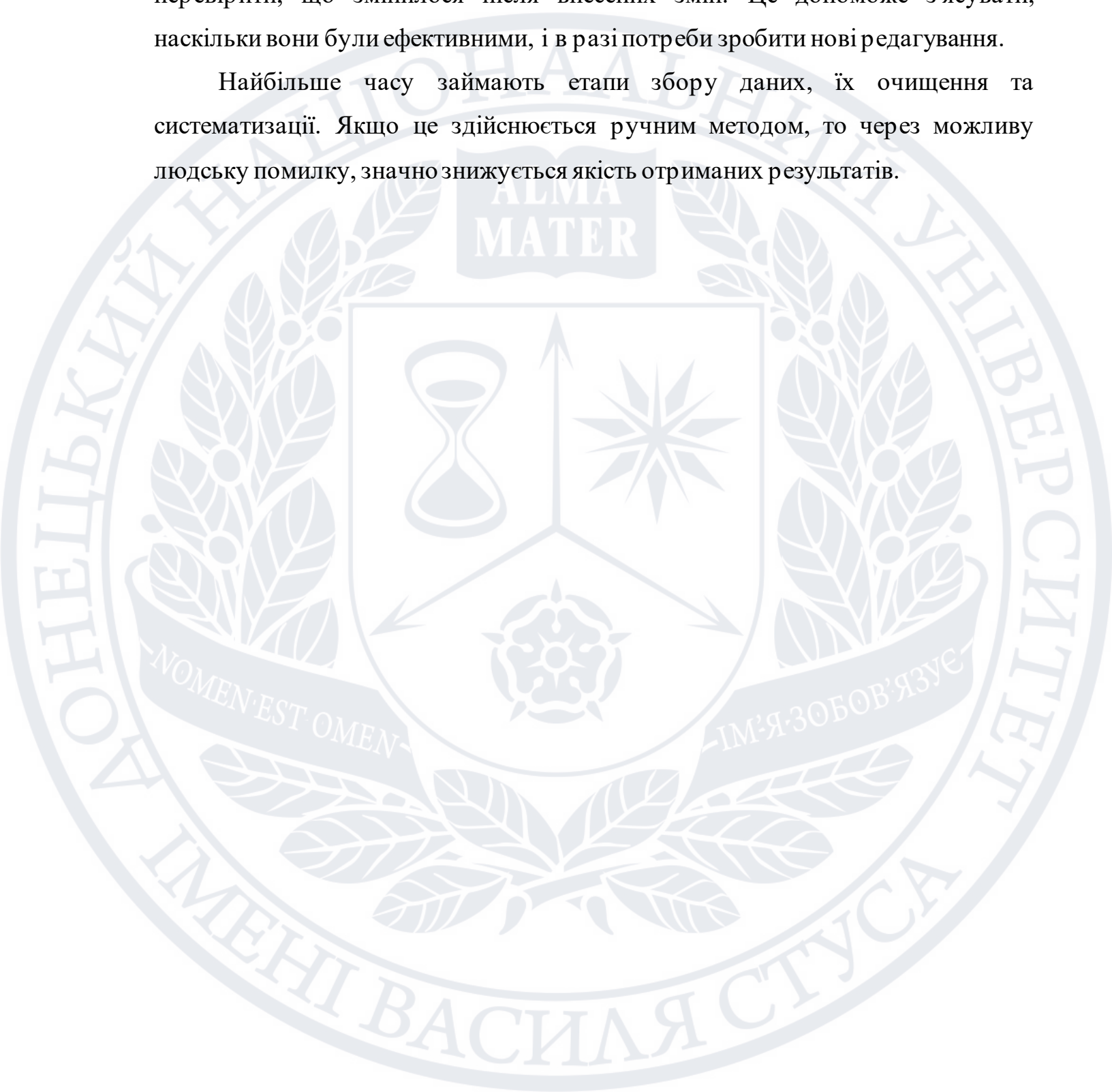
Подбати про якість зібраних даних, перевіривши їх на помилки та пропуски. Якщо пропустити цей етап, то подальше тлумачення буде недостовірним.

Провести роботу з результатами: проаналізувати та інтерпретувати їх. Полягає у складанні дашбордів, звітів чи діаграм, і навіть їх зіставленні з очікуваннями і сформульованими питаннями.

Виконати дії, що базуються на отриманих результатах. Виявлені інсайти слід використовувати як основу для подальшої роботи. Це допоможе зробити маркетингову стратегію ефективнішою.

Повторити весь цикл проведеного аналізу через деякий час, щоб перевірити, що змінилося після внесених змін. Це допоможе з'ясувати, наскільки вони були ефективними, і в разі потреби зробити нові редагування.

Найбільше часу займають етапи збору даних, їх очищення та систематизації. Якщо це здійснюється ручним методом, то через можливу людську помилку, значно знижується якість отриманих результатів.



2 ДОСЛІДЖЕННЯ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ ОБРОБКИ ДАНИХ

2.1 Дослідження методів аналізу

Засоби data mining призначені для пошуку прихованих, дійсних та всіх можливих шляхах серед набору великий даних. Цей метод, дає змогу досліднику виявляти неочікувані та приховані зв'язки між даними.

Є безліч інструментів для інтелектуального аналізу даних. Найпопулярніші та найкращі ми розглянемо нижче.

Sas Data Mining – система статистичного аналізу, розробник продукту SAS. Таку програму було зроблено для аналізу та управління даними (рис. 4).

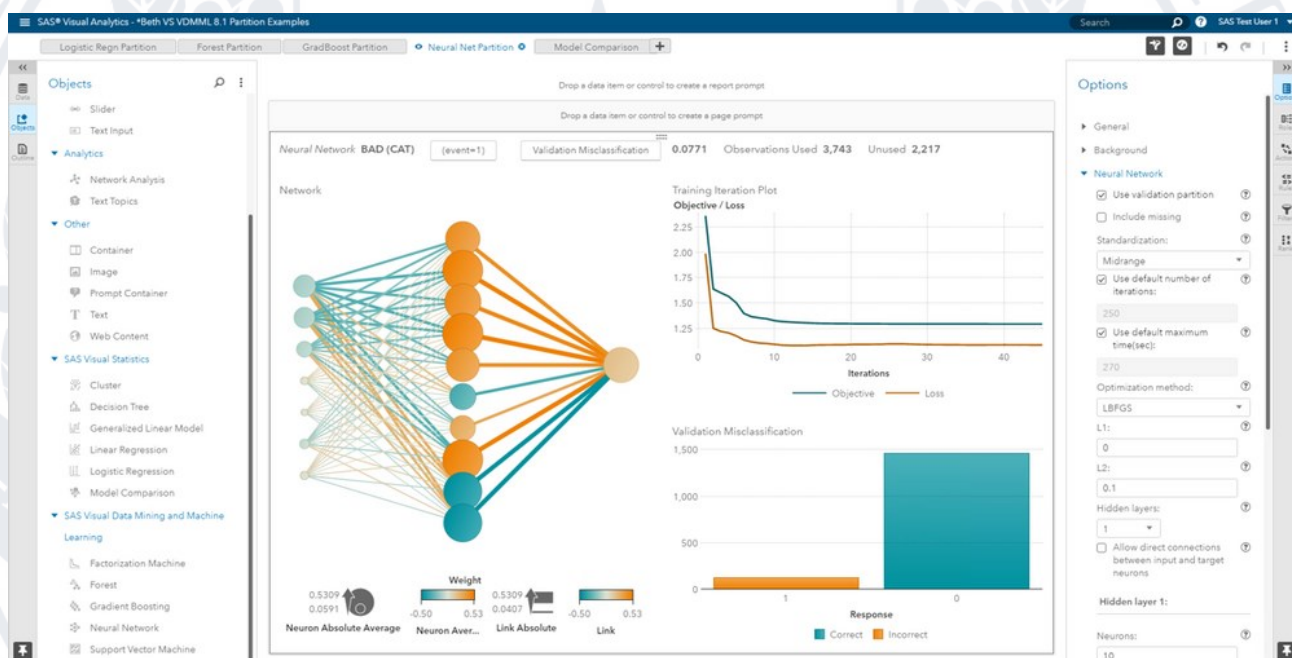


Рисунок 4 – зовнішній вигляд інтерфейсу

Інструменти програми дають змогу аналізувати текст, робити великий аналіз та оптимізувати його.

Наступна програма це Teradata – являє собою відкриту паралельну систему обробки та розробки великих додатків. Інтерфейс на рисунку 5. Така система працює на всі популярних платформах. Оптимізатор підтримує до 64 з'єднань у запиті. Дані мають найнижчу вартість володіння. Що дає змогу налаштовувати, проводити певні дії з адмініструванням та підтримку.

Основною особливістю є те що вона дає змогу працювати із SQL так само як і з таблицями.

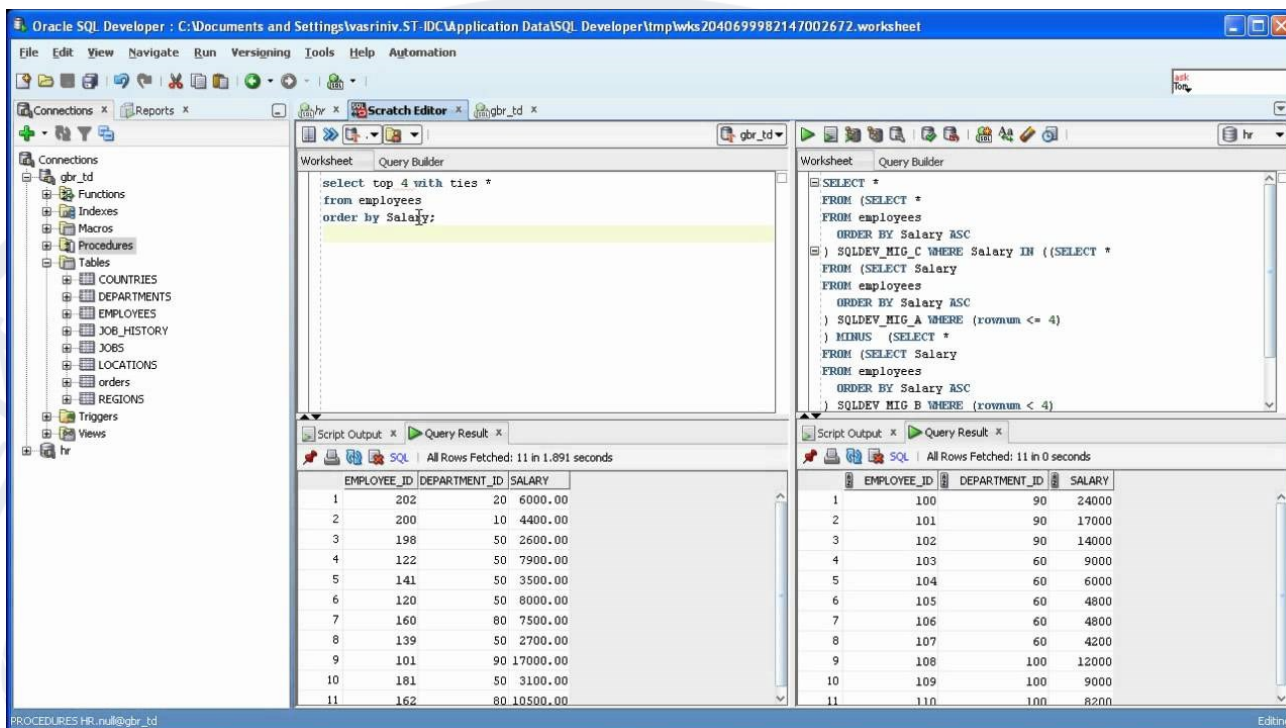


Рисунок 5 – інтерфейс Teradata

R-програмування – мова для обчислень та створення графіків. Вона використовується для великих обсягів даних. Особливість даної мови є те що вона пропонує великий спектр тестів приклад на рисунку 6.

Ефективний засіб обробки даних та їх зберігання, надає перебір операторів для розрахунку на матрицях, масивах, а також є в наявності інтегрований набір інструментів для великих даних, які будуть відображатись на екрані або в друкованому вигляді.

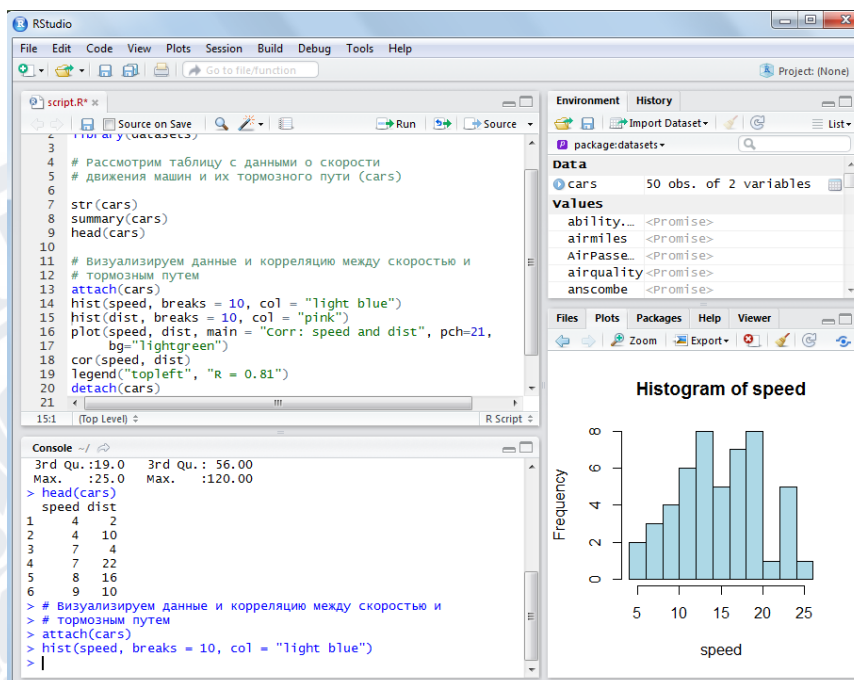


Рисунок 6 – зовнішній вигляд програми

Наступною програмою, яка використовується для аналізу даних, це Python, згідно із загальним визначенням це мова яка є високорівневою та орієнтована на підвищення продуктивності та зручності читання коду. За час свого існування він набув великої популярності через свою простоту та використання безлічі бібліотек. Найбільшу зацікавленість для аналізу даних набули такі бібліотеки як :

- I. Pandas – бібліотека яка відповідає за обробку даних
- II. NumPy – працює із матрицями та багатовимірними масивами даних.

Використання цих бібліотек як одне ціле дає змогу зробити аналіз не тільки на рівні програми а і візуально. Таким чином ми можемо взяти файл із певними даними, взяти відповідний індекс або комірку та проводити із ними різні операції а після чого створити графік. Приклад на рисунку 7


```
In [31]: coefs = np.polyfit(x,y,1)
         predict = np.polyld(coefs)
```

```
In [32]: x_test = np.linspace(0,16)
         y_pred = predict(x_test[:,None])
         plt.scatter(x,y)
         plt.plot(x_test,y_pred,c='r')
         plt.show()
```

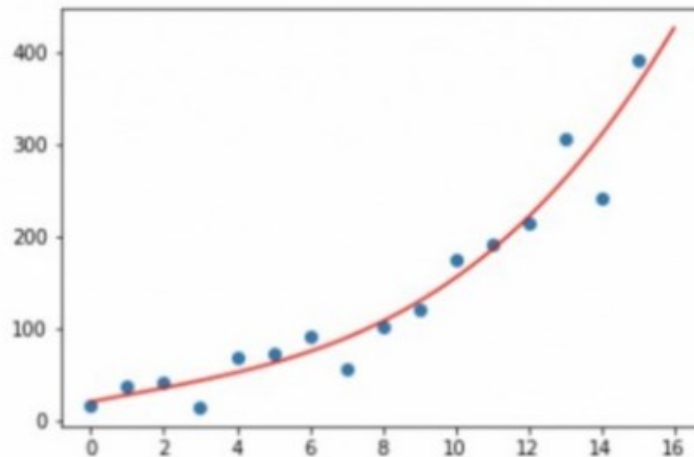


Рисунок 7- Приклад прогнозування за допомогою polyfit

Поліноміальні функції — це функції, які містять цілі невід’ємні степені x . До них входять такі функції, як квадратичні функції, кубічні функції та решта степеневих прогресій. Процес підгонки поліноміальної кривої — це процес побудови математичної функції, яка найкраще підходить для ряду точок даних, таким чином, щоб крива відповідала більшості наявних точок даних. Головне — визначити «гладку» криву, яка найкраще підходить. Ця бібліотека дає змогу читачеві визначити таку криву за попередніми даними. Для визначення цього результату використовується програмне забезпечення Matlab.

2.2 Вибір платформ для аналізу

Перед вибором платформи для аналізу слід врахувати характер аналізу та данні які будуть підлягати подальшим діям. Оскільки різні бібліотеки можуть проводити аналіз та взаємодіяти із ними по різному. Такі критерії включають у себе оцінку фахівця, тобто його вміння, можливість працювати із великими даними, ступінь аналізу та перетворення даних.

Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 1 для зручності порівняння.

Якість обробки даних	Python	R	Weka	Knime	Rapid Miner
Практичні навички	4	2	1	1	1
Можливість проводити складні перетворення	3	3	1	1	1
Якість простих перетворювань	4	4	4	4	4
Загальна оцінка	11	9	6	6	6

Таблиця 1 – результат оцінки платформ

Наступним критерієм буде «візуалізація даних» вона складається із двох оцінок, гнучкості та естетичного вигляду. Метою візуалізації є уявлення спрощеної інформації, що б користувач дивлячись на таблицю міг легко знайти дані які є необхідними. Слід знати що ідеальна візуалізація вважається та при якій втрачаючи весь текст вона збереже свій сенс. Користувачі що працюють на високому рівні не мають часу перечитувати усі рядки вибираючи тільки те що

їм необхідне. Найбільша частина інформації приходить до людини саме через очі, тому якість візуалізації це важливий показник для сприйняття.

Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 2 для зручності порівняння. Отож результат такої оцінки зображено в таблиці 2

Візуалізація	Python	R	Weka	Knime	Rapid Miner
Гнучкість	4	4	2	2	2
Естетичний вигляд	2	4	2	2	2
Загальна оцінка	6	8	4	4	4

Таблиця 2 – загальна оцінка «візуалізації»

Третім критерієм буде «машинне навчання» складається із оцінки кількості методів для навчання та можливості налаштувань певних параметрів. Алгоритм такого навчання дає змогу користувачу аналізувати та досліджувати складні набори даних та знаходити в них залежності. В моделях машинного алгоритму використовуються параметри що залежать від початкових даних. Чим більший обсяг початкових даних тим більш правильно буде проведено аналіз. В різних алгоритмах застосовуються різні підходи та методи аналізу. Нерідко вони групуються за методами в рамках яких виконується: контрольоване навчання, не контрольоване та навчання із підкріпленням.

Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 1 для зручності порівняння. У таблиці 3 загальна оцінка та критерії оцінювання

Машинне навчання	Python	R	Weka	Knime	Rapid miner
Кількість методів	4	4	4	3	3
Гнучкість налаштування	4	4	2	2	2
Загальна оцінка	8	8	6	5	5

Таблиця 3- загальна оцінка «машинного навчання»

Ще одним важливим критерієм є «Представлення результатів роботи»

Вона складається із оцінки витрати на обробку результату та на можливість налаштування. Він є важливим оскільки при великому обсягу даних швидкість є важливою. Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 4 для зручності порівняння. Загальна оцінка та порівняння зображено в таблиці 4.

Представлення результатів роботи	Python	R	Weka	Knime	Rapid Miner
	4	4	1	2	2
	1	2	4	3	3
	5	6	5	5	5

Таблиця4 – загальна оцінка «Представлення результатів»

Критерій «швидкість отримання попередніх результатів» містить в собі оцінку витрати часу на написання коду та його виконання. Критерій впливає на оперативність та швидкість прийняття рішень, та має важливе значення для деяких задач.

Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 5 для зручності порівняння.

Швидкість отримання попередніх даних	Python	R	Weka	Knime	Rapid Miner
Витрати часу на написання коду	1	1	4	4	4
Швидкість вводу результатів	2	3	4	4	4
Загальна оцінка	3	4	8	8	8

Таблиця 5- загальна оцінка та порівняння «Швидкість отримання попередніх результатів»

Наступний критерій «Наочність процесу аналізу даних» демонструє можливість платформи відображати процес аналізу даних в реальному часі. Більшість платформ дають таку змогу що робить аналіз більш зрозумілим та гнучким до налаштувань.

Оцінка кожного критерія буде зроблена за шкалою від 1 до 4, де 1 це мало відповідає а 4 відповідає повністю. Дані наведено в табличному вигляді, у таблиці 6 для зручності порівняння.

Наочність процесу аналізу даних	Python	R	Weka	Knime	Rapid Miner
Наочність	1	1	2	4	4
Загальна оцінка	1	1	2	4	4

Таблиця 6 – загальна оцінка та порівняння «Наочність процесу аналізу»

2.3 Результати аналізу програмних інструментів

В результаті зробленого аналізу обраних платформ за даними критеріями отримали загальну оцінку для кожної із систем, оцінки представлені в таблиці 7. Загальна оцінка, представлена у вигляді суми оцінок із попереднього розділу.

Таблиця 7
Результати проведеного порівняння аналізу платформи за критеріями

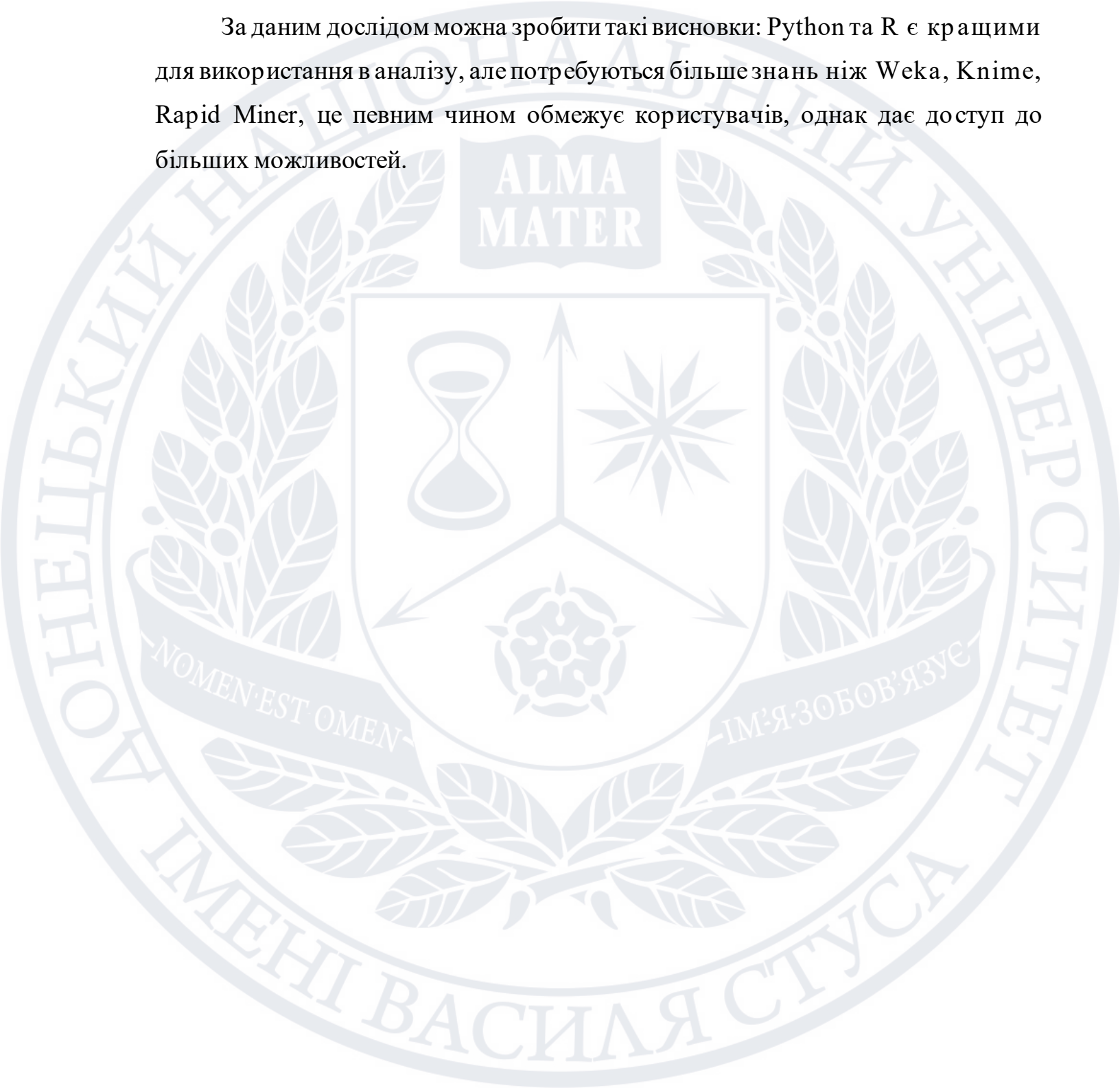
Критерій	Python	R	Weka	Knime	Rapid Miner
Якість обробки	11	9	6	6	6
Візуалізація	6	8	4	4	4
Машинне навчання	8	8	6	5	5
Представлення результатів	5	6	5	5	5
Швидкість отримання попередніх результатів	3	4	8	8	8
Наочність процесу	1	1	2	4	4
Загальна оцінка	34	36	31	32	32

Загальні висновки які можна зробити із проведеного дослідження такі:

1. Найбільшу кількість балів набрали Python та R, 34 та 36 балів.
2. Відмінність оцінок свідчить про те що при виборі платформи слід чітко розуміти поставлену задачу та принципи її виконання.
3. Якщо результати необхідно отримати швидко то краще використовувати Weka.Knime, Rapid Miner.
4. Для перегляду дій краще використати Knime або Rapid Miner оскільки ці системи краще демонструють наочність.

5. Задачі із машинним навчанням краще притаманні Python та R , оскільки в них більше можливостей для створення реалізацій алгоритмів.

За даним дослідом можна зробити такі висновки: Python та R є кращими для використання в аналізі, але потребуються більше знань ніж Weka, Knime, Rapid Miner, це певним чином обмежує користувачів, однак дає доступ до більших можливостей.



Висновок до другого розділу

У даному розділі було проаналізовано усі сучасні програми та додатки аналізу даних. Було розглянуто різні підходи та різні приклади для зручного дослідження. Також проведено дослід на порівняння між мовами та їх бібліотеками. Оскільки кожна мова та бібліотека мають щось спільне а щось ні потрібно чітко розуміти поставлену задачу та шляхи її вирішення. Таким чином виконуючи аналіз програмно ми наочно можемо дослідити зміни в режимі реального часу. Побачити приховані дані під час їх дослідження. Інколи бувають такі завдання що без втручання програм справитись із ними неможливо. Швидкий та простий для розуміння аналіз набув великого попиту, та застосовується у різних галузях. Так із збільшенням попиту збільшились і можливості бібліотек. Наприклад машинне навчання стало реальністю і чим більше обробка попередніх даних тим точніше буде розрахунок прогнозних або оціночних значень відгука.

3. Обрання галузі та вибір підприємства для аналізу

3.1 Історія появи галузі

Для розвитку економічного становища, Україна розпочала пошук швидкого та ефективного економічного механізму господарства, якби дав би змогу змінити негативні ефекти розвитку в економіці.

Науково-технічний прогрес дав змогу великий поштовх для повного використання ресурсів країни. Найбільш широкого використання набуло землеробство, одне із яких вирощування цукрових буряків. Рівень буряківництва деякою мірою впливає на стан економіки. Швидкий розвиток бурякоцукрової галузі став зміцненням вітчизняної економіки, оскільки ця промисловість забезпечує робочі місця для сільського населення та збільшує бюджет держави через податки, а також зростання внутрішнього валового доходу.

Розміщення України та її клімат є досить сприятливими для вирощування буряків, що і стало наслідком того що вона залишається однією із провідних країн світу за площею посіву та обсягами вирощування, а також переробки на цукор.

Під час аналізу даних виробництва продукції рослинництва необхідно вивчити:

- 1) виконання плану волових зборів продукцій в розділі на різні групи та культури, а також вплив на продаж державі.
- 2) Виконання планів що до посіву площ та їх структури.
- 3) Рівень врожайності від всіх культур сільськогосподарських культур, фактичний та очікуємий результат, перевірка відповідності дотримання комплексу агротехнічних заходів.
- 4) Фактори які можуть або вплинули на відхилення від запланованого збору, втрати чи приріст фактичного збору.
- 5) Слабкі місця та їх причини, план що до усунення недоліків.

б) показники передових відділків, ферм, бригад, ланок порівняно з середніми показниками господарства, в результаті чого їх досягнуто та як це позначилось на загальних показниках господарства. Крім того визначають місце господарства в районі за основним показниками рослинництва, проводять порівняльний аналіз цих показників і показників передових господарств та виявляють невикористані резерви.

В аналізі абсолютні величини використано для порівняння плану та виконаних робіт збору, площ для посіву та врожаю. Відносні величини будуть характеризувати структуру площ для посіву, стадію завдання по врожайності та валових зборах культур. Для того щоб перевірити вживанні заходи використовується метод порівняння, усі заходи для підвищення врожайності записуються у журнал заходів, після чого робиться перевірка на змінні, позитивний чи негативний вплив конспектується. Із розвитком біологічних добавок є необхідність у перевірці певних препаратів на ділянках посіву. Кожного року виділяється окрема ділянка для тестування таких добрив. Після чого буряки із таких площ відправляють на перевірку у лабораторію для визначення сахаристості та вмісту негативних речовин.

Із всієї кількості виробництва світу майже 40% припадає на цукрові буряки, в деяких країнах вони являються джерелом єдиного добування цукру. В Україні наприклад, цукор добувають тільки із цукрового буряка. Коренеплоди сортів та гібридів буряків містять 17-18% цукру, у найкращий врожайний період може досягати до 20%. Під час переробки коренеплодів із 350-400ц/га можна отримати по 50-60ц цукру з 1га.

Під час переробки цукрових отримується також інша продукція, наприклад гичка, жом, меляс що є кормом для великої рогатої худоби. Використання таких добавок забезпечує високий рівень продуктивності тварин, а саме збільшення виробництва продуктів тваринного походження.

Останнім часом в Україні зменшили виробництво цукру що призвело до зниження рівня продажу за межі країни. Основними областями де засівають буряк є: Вінницька, Черкаська, Хмельницька, Сумська та Полтавська, вони забезпечують близько 70% валового виробництва буряків.

Виробництво пов'язане із великими витратами грошей і трудовими ресурсами. Порівняти витрати навіть за два роки важко, оскільки інфляційні процеси не дають можливості це зробити.

Відношення між результатами праці та витрата для досягнення результату або використаними ресурсами визначає рівень грошової ефективності виробництва. До важливих характеристик у буряківництві відносять: врожайність, цукровість буряка із 1 га посіву, витрати праці на одиницю продукції, вартість 1 ц цукрових буряків, ціна продажу та рівень рентабельності з рахунку на 1 га посіву.

Напрямок підвищення економічної ефективності буряківництва є послідовна інтенсифікація галузі. Декілька років все більше використовують інтенсивні технології виробництва цукрових буряків. Етап інтенсифікації є головною складовою, яка впливає на ефективне виробництво цієї культури. Її головною характеристикою є виробничі витрати в розрахунок на 1 га посіву. У цей показник враховують усі витрати, а саме, вартість добрив які використовувались, насіння, паливо, амортизація основних засобів, витрати на оплату працівникам.

На практиці ж ми можемо побачити що не завжди підвищення рівню інтенсифікації має позитивний вплив. Наприклад коли виростати надмірну кількість добрив то зростає врожайність але зменшується сахаристика, від чого ми можемо вдосконалити тільки кількість вхідного матеріалу, тобто буряка, а вихідний продукт (сахар) залишиться не змінним або навпаки зменшиться.

Праця із буряками є трудомісткою галуззю сільськогосподарського виробництва. Технології що використовуються при переробці не дають

можливості повністю виключити ручну працю. При доглядах за густоти насаджень, огляду посівних площ, збиранні коренеплодів першочергове значення має продуктивність праці.

Збільшення виробництва цукрових та збільшення економічності пов'язане із індустріальними та інтенсивними технологіями. Які включають агротехнічні заходи вирощування буряків які базуються на попередніх досвідах та розвитку технологій. При введенні індустріальних технологій велике значення мають добрива які вносяться відповідно до плану, густота посіву насіння, використання агротехнічних та хімічних засобів для збільшення маси коренеплоду.

Головним завданням аналізу є: систематичний контроль плану виробництва товару, підрозділу господарства та визначення дії різних факторів на обсяги продукції, виявлення резервів, збільшення виробництва та покращення її якості. Оцінка перспектив збільшення товару виробництва з урахуванням різних факторів, розробка плану що до застосування виявлених резервів збільшення продукції.

3.2 Характеристика підприємства

Земельні ділянки та їх місце знаходження зображено в таблиці 8

Таблиця 8

Дані про землі та розміщення компанії

Компанія	Місцезнаходження	Площа, га
Товариство з обмеженою відповідальністю «Продовольча компанія «Зоря поділля»	Вінницька область, Гайсинський район	49 540
Приватне акціонерне товариство «Продовольча компанія «Поділля»	Вінницька область, Тульчинський район	47 381
Товариство з обмеженою відповідальністю «Правобережне»	Дніпропетровська область, Верхньодніпровський район	2 224
Сільськогосподарське товариство з обмеженою відповідальністю «Агрофірма» кряж-агро-1»	Запорізька область, Запорізький район	-
Товариство з обмеженою відповідальністю агрофірма «Іванківці»	Кіровоградська область, Знам'янський район	4 274
Товариство з обмеженою відповідальністю «Брассіка-2010»	Кіровоградська область, Знам'янський район	29
Товариство з обмеженою відповідальністю «Мас – агро»	Кіровоградська область, Світловодський район	2 543
Товариство з обмеженою відповідальністю «Агрофірма Дніпроагролан»	Полтавська область, Семенівський район	8 004

Залученні до процесу ресурси вимірюються в різних одиницях, земельні в площі, трудові в кількості працівників, основні виробничі фонди у грошових виразах. Кожен ресурс є необхідним у процесі. Замінна хоча б одного із них є неможливою.

Компанія працює більше ніж із 35 000 орендодавців. Власникам земель гарантують своєчасну виплату орендної плати. Більшість договорів підписанні на 10-20 років, що говорить про довіру до підприємства.

До складу «УКРПРОМІНВЕСТ-АГРО» входять «Агропродінвест» (керуюча компанія), «ПК Поділля», «ПК Зоря Поділля», «Вінницький комбінат хлібопродуктів №2», «Дніпроагролан», «Іванківці», «Мас-Агро», «Правобережне» і «Прогрес-НТ».

У період березень-жовтень 2007 завод припинив свою роботу. Для проведення реконструкції було виділено 98мільйонів гривень. Роботу виконувала ПК «Зоря Поділля» до складу якої завод входив. Під час оновлення було повністю замінено обладнання, що дало змогу збільшити потужність виробництва удвічі, до п'яти тис. тон переробленого буряка за добу.

За минулий рік структура посівних площ зображено в таблиці 9.

Таблиця 9

Структура посівних площ

Структура посівних площ	
Цукровий буряк	28 257га
кукурудза	33 446га
Соя	19 906га
Соняшник	4 214га
Ячмінь	150 га
Кормові культури	2335га

У структурі сільськогосподарських угідь найбільшу частину являє рілля. Так у 2021 97,2% сільськогосподарських угідь становила саме рілля. Це є показником високої розораності земель у компанії. З одного боку це негативне

явище, оскільки основним напрямком є рослинництво. Як показує практика передових країн то ступінь розораності в середньому має бути від 20 до 50%.

Для зберігання усіх посівів у компанії містяться елеватори у різних частинах України, місце знаходження та ємкість зображено в таблиці 10

Таблиця 10

Місце знаходження та ємкість

Компанія	Місцезнаходження	Зберігання, т
Вінницький КХП №2	Вінницька область, Вінницький район	43 340
Елеватор ПК Підпілля	Вінницька область, Крижопільський район	60 000
ПК Зоря Поділля	Вінницька область, Гайсинський район	40 500
Дніпровський Крохмалопатковий комбінат	Дніпропетровська область, Верхньодніпровський район	90 000
Агрофірма Іванківці	Кіровоградська область, Знам'янський район	25 000
АФ Дніпроагролан	Полтавська область, Семенівський район	58 000

Трудові ресурси відіграють велику роль у виробничому процесі. Саме на них покладається роль виконавця процесу та створення усіх продукції.

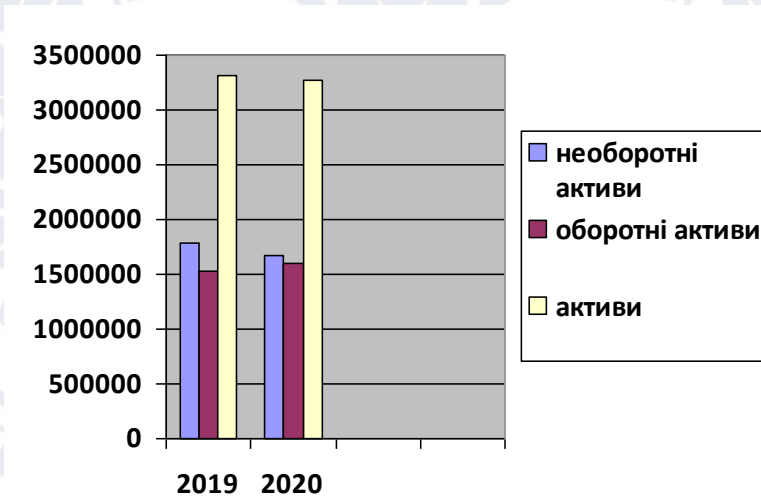
Соціальна відповідальність «УКРПРОМІНВЕСТ-АГРО» містить в собі зобов'язання своєчасно ї виплати заробітної плати та сплати податків. А також компанія бере на себе відповідальність що до власників у яких землі беруть в оренду. Вона зобов'язана робити виплати у вигляді паю, а також соціальні інвестиції, які мають напрямом на розвиток регіонів присутності, охорону навколишнього середовища та професійний розвиток а також захист персоналу.

Спостерігається зменшення господарського потенціалу, яке ми можемо побачити на зменшенні суми активів на 1,39%. Тобто майно у розпорядженні підприємства скорочується.

Факт того що активи скорочуються менш швидше ніж дохід від продажу, вказує на наявність необхідності пошуку нових резервів оптимізації даної структури активів.

Фінансовий аналіз проведено на основі звітності підприємства.

Фінансовий аналіз зображено на рис..



У таблиці 4 ми можемо побачити аналіз абсолютного приросту та відносного у відсотках.

Таблиця 11

Аналіз абсолютного приросту та відносного у відсотках

Показник	2019	2020	Абс.приріст +,-	Відн.приріст%
Особові засоби	1738764	1613295	-125469	-7,22
Необоротні активи	1791373	1671245	-120128	-6,71
Запаси	1063923	1262582	198659	18,67
Оборотні активи	1521371	1595538	74167	4,88
Активи	3312744	3266783	-45961	-1,39

Таке зменшення балансу спостерігається через скорочення необоротних активів яке складає -6,71%

Також слід зазначити що скоротилась сума джерел фінансування для залучення активів через зменшення власного капіталу та довгострокових зобов'язань. Данні зображені в таблиці5

Таблиця 12

Дані Балансу

Показник	2019	2020	Абс. Приріст, +,-	Відн. Приріст, %
Нерозподілений прибуток (непокритий збиток)	1635909	1607048	-28861	-1,76
Власний капітал	1650935	1620096	-30839	-1,87
Інші довгострокові зобов'язання	802425	731403	-71022	-8,85
Довгострокові зобов'язання	853421	732909	-120512	-14,12
Інші поточні	547633	573135	25502	4,66

зобов'язання				
Короткострокові зобов'язання	808388	913778	105390	13,04
Баланс	3312744	3266783	-45961	-1,39

Як результатом зменшення суми власного капіталу вказує на зниження рівня благополуччя власників. Скорочення сум відповідальності призводить до збільшення незалежності від інших постачальників, хоч і стримує повне розкриття потенціалу компанії. У таблиці 6 зображено показник стійкості та платоспроможності.

Таблиця 13

Показники стійкості та платоспроможності

Показник	2019	2020	Абс. Приріст, +,-	Відн. Приріст, %
Фінансова автономія	0,5	0,5	-0	-0,49
Поточна ліквідність	2,22	1,83	-0,39	-17,47

Взявши бухгалтерську звітність можемо побачити зміни у прибутках, порівнявши 2 роки між собою. Для зручного порівняння створено таблицю 7.

На ній зображено звіт про фінансовий стан.

Таблиця 14

Звіт про фінансовий стан

Назва показника	2020	2019
Нематеріальні активи	21498	17198
первісна вартість	22557	17604
накопичена амортизація	1059	406

Основні засоби	1613295	1738764
первісна вартість	2714169	2594347
знос	1100874	855583
Довгострокові біологічні активи	36208	35193
первісна вартість	36208	35193
Довгострокові фінансові інвестиції: які обліковуються за методом участі в капіталі інших підприємств	244	218

Основними напрямками підприємства є цукровиробництво. Цукор виробляється на власних цукрових заводах. Зручне положення та сприятливий клімат дають змогу здешевити собівартість товару. Також у планах збільшити посівні площі.

Важливо зазначити що напрямком виробничої програми є також зернові культури. Наприклад те що озимина є найкращим попередником для цукрового буряку дає можливість використовувати земельні ресурси майже цілий рік. Завдяки цьому озимина та її продаж дає 50% прибутку підприємству.

За часи використання земель для потреб сільського господарства на планеті втрачено майже 20% родючого шару, який створюється тисячі років. Щорічно руйнується 24млн т ґрунтового шару землі. Унікальні чорноземи майже зникли, тих які мали вміст гумусу 14-16%, а ті які мали гумус вмістом від 10 до 13% скоротилась в п'ятеро. Постійне використання землі призвело до того що частина родючих земель виноситься вітром та водою.

За якісь ґрунту відповідає багато чиників, серед яких вміст органічних речовин (гумус). Саме він дає можливість землі втримувати у собі вологу та поживні речовин, а також відповідає за його родючість та поліпшує структуру.

Гумус у великій кількості дає допомогу запобігти втратам які можуть виникнути через посуху, надмірну кількість опадів або навіть хвороби рослин. Через виникла потреба у впровадженню екологічно – безпечних технологій у землеробстві.

Часткову допомогу надає оранка землі. Однак під час неї усі поживні речовини та добрива загортаються у ґрунті та частково знищуються бур'яни. Частково тому що гинуть тільки ті які вже проросли, а їхнє насіння здатне проростати протягом декількох років, тому повністю знищити його не виходить.

Використовуючи цю інформацію «УПІ-АГРО» почало використовувати нові технології для обробки земель, а саме: безполицевих, поверхневий, мінімальний та нулеві технології обробки шару. Це дало змогу краще використовувати землі без надмірних витрат та збільшити продуктивність агрегатів. Позитивні результати впровадження різних технологій в обробці ґрунту дали можливість відновити родючість та на 10-15 % зменшити собівартість продукції.

Завдяки реконструкції у 2007 році на потужностях Гайсинського цукрового заводу здійснюється виробництво цукру-піску. Для можливості виготовлення такого цукру використано коштів більше ніж на 100 млн.грн. На даний момент на заводі ці технології використовуються. На цю суму було закуплено та введено в експлуатацію:

- Мийне обладнання європейського класу фірми «Putsch» та української фірми «ТМА»
- Обладнання для бурякопереробного цеху
- Установки для відділення сокодобування фірми «ТМА»
- Обладнання для фільтрування соків та сиропів європейського класу, фільтри згущувачі голандської фірми Amafilter BV та фільтрпрес Larox.

У 2012 році було прийнято рішення реконструювати вапно-випалювальну піч. Також для покращення роботи було встановлено дифузійний апарат ДС - 12, покращено роботу дефекосатурації, шляхом встановлення нового преддефектатора та сатуратора. Для швидшої фільтрації соку було добавлено по дадоткових фільтрах на кожну із сатурацій. Покращено теплову схему та встановлено фільтр сиропу тф-150.

У 2014 році було переведення котлів із газу на вугілля, що стало позитивним наслідком для зниження вартості технічного пару та вартості продукції. В тому ж році встановили 2 жомові преси «babbini»

У 2018 проведено покращення станції декальцинації соку шляхом іонообміну. В таблиці 16 показано як змінювалась продуктивність

Таблиця 16

Динаміка продуктивності

Рік	Потужність переробки, тонн на добу	Перероблено цукрового буряку, тис. тонн	Вироблено цукру, тис. тонн	Вихід цукру, %
2021	7481	844,6	133,9	15,85
2020	6914	620	80,9	13,04
2019	7300	683,9	111,8	16,3
2018	6625	940,74	120,6	12,8
2017	6485	691,78	105,52	15,25
2016	7358	738,00	117,02	15,86

В результаті проведення модернізацій на цукровому заводі, його виробнича продуктивність переробки зросла з 2,6 тис т. Буряку до 5,5 тис.т. а вихід цукру збільшився з 6,63 до 12%. Технічно-економічні показники наведено в таблиці 17.

Таблиця 17

Технічно-економічні показники виробничої продуктивності

Показники	2019 рік	2020 рік	2021 рік
Заготівля, т.	369 581	345 631	362 593
Перероблено, т.	361 177	333 521	347 526
Отримано цукру, т.	42 864,6	41 885,4	52 709,8
Середньодобова продуктивність, т.	4 006	5 164	5 000

Аналізуючи таблицю видно, що є зниження рівню середньодобової продуктивності заводу в 2021 році порівняно із 2020, це пов'язано із підвищеним вмістом цукру саме у коренях буряків.

Промислове виробництво відбувається по сезонну, а переробка зазвичай займає 4 місяці (із вересня по грудень) інколи можливе відхилення, пов'язане через погодні умови або не врожайність. Склад та структура товарів підприємства наведені в таблиці 18.

Таблиця 18

Склад та структура товарів підприємства

Види продукції	2019р.		2020р.		2021р.		В сер. За 3 роки	
	тис. Грн.	%	тис. Грн.	%	тис. Грн.	%	тис. Грн.	%
Продукція рослинництва – всього:	50234,10	77,70	52477,20	60,10	78902,00	81,40	60537,77	74,38
Зернові та зернобобові	30591,30	47,30	31452,30	36,40	73285,00	75,60	45109,53	55,42
Ріпак	8200,20	12,70	8645,10	10,00	4649,30	4,80	7164,87	8,80
Соя	1600,10	2,50	1647,70	2,00	20,80	0,02	1089,53	1,34
Інша продукція	9842,50	15,20	10732,10	12,40	946,90	1,00	7173,83	8,81
Продукція тваринництва – всього:	14379,70	22,30	32392,00	39,90	15800,90	18,60	20857,53	25,62

ВРХ	2488,10	3,90	10546,40	12,20	2904,10	4,20	5312,87	6,53
Свині	3378,30	5,10	10231,10	12,80	0,00	0,00	4536,47	5,57
Молоко	8445,10	13,20	10698,10	13,40	12750,90	14,20	10631,37	13,06
Інша продукція	67,60	0,10	916,40	1,50	145,90	0,20	376,63	0,46
Всього:	64613,80	100,00	86291,00	100,00	96904,90	100,00	81395,30	100,00

Отже, із даних наведених в таблиці видно, що 2021 рік найкращим для переробки та виробництва продукції порівняно із 2019 та 2020. Також в 2021р підприємство отримало дохід від тваринництва та рослинництва більше ніж в 2007 р на 32297тис. Грн. Дохід від продажу продукції ВРХ у 2021 зріз на 416тис. Грн. Проте підприємство закінчило продаж свинини. Також позитивним моментом є збільшення прибутку від продажу загальному так у 2021році прибуток від продажу зріз на 28668 тис. Грн. В порівнянні з тим же 2019. Зокрема дохід від зернових культур зріз на 42693,8тис.грн.

У 2022 році 10 заводів не працюють, тобто у роботі залишилось 22 цукрових заводи по всій Україні. Причиною цього стала війна та ціна на енергію. Оскільки ціни на газ піднялись виробництво цукру перестане приносити прибуток а значить буде не рентабельним. Наприклад для оплати за газ необхідно продати 400т цукру. Загалом заводи виготовляють 600-700т за добу, але враховуючи оплату за працю, за обслуговування підприємств було прийнято рішення призупинити роботу цих підприємств.

Враховуючи ситуацію яка сталась в країні, було прийнято рішення про збільшення переробітку на працюючих цукрових заводах. Так використовуючи сучасні технології та різні види підходів до вирощування , стало можливим збільшення виробництва коренеплодів. Цього року середня врожайність по підприємству стала на 80 ц/га вище ніж торік. А середня цукристість більша за 18,4%. Цього року на Гайсинський завод буде відправлено понад 900 тисяч тонн буряків для переробки. Це стане найбільшим показником переробки серед вітчизняних заводів.

Станом на 15 вересня, заводи переробили 38,8 тис.тонн цукру і перероблено близько 349тис. Тонн буряків.

Пуск заводів відбувся:

1. 21 серпня – “Панда” (Селищанський цукровий завод);
2. 1 вересня – Гайсинський цукровий завод (“ПК Зоря Поділля”);
3. 3 вересня – “Червонський цукровик”;
4. 5 вересня – “Радехівський цукор” (Радехівський цукровий завод);
5. 7 вересня – “Саливонківський цукровий завод”;
6. 10 вересня – “Радехівський цукор” (Хоростківський цукровий завод);
7. 13 вересня – “Поділля” (Крижопільський цукровий завод);
8. 13 вересня – “Жданівський цукровий завод” (“Астарт-Київ”);
9. 13 вересня – “Староконстантіновсахар”.

Із негативних факторів які можуть спостерігатись у майбутньому це здороження вирощування цукрових буряків, через зменшення привабливості культури. Не дивлячись на труднощі які виникли на території країни буряком засіяно 180,4 тис га, що лише на 45тис. Га менше ніж минулий рік. За приблизними даними врожайність 2022 року знаходиться в діапазоні між 450 - 470 ц/га що є досить гарним показником врожайності.

Із нового врожаю планується виробити біля 1,1 млн.тонн цукру. Такої кількості має вистачити на внутрішній ринок, а завдяки перехідним залишкам, які накопичились то буде змога проводити експорт. На рисунку 8 видно як змінювалась динаміка виробництва цукру.



Рисунок 8

Галузь цукро-виробництва вже декілька років, постійно експортує значні обсяги цукру. Цього року також вдалося відвантажити на експорт 30 тис. тонн (станом на 5 серпня), однак сезон ще не завершено та у планах протягом місяців продовжувати експорт. Дивлячись на ситуацію у сусідніх країнах де винник ажіотаж на цукор, ЄС дозволи безмитний експорт, то сподіваємось на кращу статистику. На рисунку 9 зображено експорт цукру за 3 останні роки.



Рисунок 9 Динаміка експорту цукру

Ціни на полицях магазину від початку року дещо зросли. Дивлячись на дані Держкомстату, зростання цін укладається у загальнодержавне підвищення широкого спектру товарів:

- Загальнодержавна інфляція 17%
- Інфляція харчового кошика 23%
- Зростання цін на цукор 19%

Аналізуючи стан та дані можемо припустити що у 2022-2023 проблем із цукром не буде, однак, можливе скорочення площ для посіву буряків, причинами того є: по перше військові дії, що скорочує споживання цукру всередині країни, по друге збільшення виробничих витрат на 1 га площі, по третє перенасичення внутрішнього ринку цукру.

Такі події є не зовсім гарною новиною оскільки можуть призвести до закриття деяких цукрових заводів, рішенням цього є пошук інших стимулюючих заходів для цукробурякової галузі.

Загальний баланс цукру станом на 1 липня 2022 року зображено на рисунку 10.

Показники	2020/21 МР (фактично)	2021/22 МР (очікувано)	2021/22 МР вересень - червень
ПРОПОЗИЦІЯ			
Виробництво	1174	1450	1450
Залишки на початок періоду	255	141	141
Імпорт всього, у тому числі: (цукор білий/цукор сирець)	163 (40/123)	10	7
Пропозиція продукції, разом	1592	1601	1598
ПОПИТ			
Внутрішнє споживання*	1380	1050	930
Експорт	71	50	42
Попит на продукцію, разом (Споживання + Експорт)	1451	1100	972
ЗАЛИШКИ			
Залишки на кінець періоду (Пропозиція - Попит)	141	501	626

Рисунок 10 Загальний баланс цукру станом на 1 липня 2022 року

Отже дивлячись на таблицю можна зробити висновок що не дивлячись на скрутне становище в країні, загрози для припинення даного введення бізнесу немає.

Висновок до третього розділу

У даному розділі були розглянуті об'єкти аналізу і визначені набори даних які характеризують діяльність компаній виробництва цукру. Сформовані набори можливих слабких та сильних властивостей виробництва, виявлення яких може бути метою аналізу. Розглянули спосіб введення бізнесу, який може впливати на результати порівняльного аналізу показників компанії. Визначили фактори, які впливають на появу даного виробництва та фактори, що впливають на продуктивність праці та родючість. Визначили можливі рішення, щодо покращення на виробничому процесі, можливі загрози та способи їх усунення. Провели порівняльний аналіз динаміки виробництва по рокам, визначили характеристики становища підприємства на поточний момент.

4 Програмна реалізація аналізу виробництва

4.1 Обрання платформи та засобу реалізації аналізу

Для реалізації даного аналізу я вирішив взяти мову програмування Python та декілька бібліотек для порівняння виконання. Таким чином я обрав бібліотеки Pandas та Matplotlib.

Pandas – одне із найбільш потрібніших бібліотек в Python, для щоденної роботи із даними. Схожа до Numpy вона тримає позицію лідера в таких областях як , наука про данні , інтелектуальний аналіз, впираючись на свої багаточисельні штучно створенні методи, атрибути та функції.

Для створення датафрейму використовують таку команду

```
idx = list('abcde')
cols = ['col_' + str(i) for i in range(4)]
df = pd.DataFrame(np.arange(20).reshape(5, 4),
                  index=idx,
                  columns=cols)

df
```

Pandas дає можливість порівнювати результати через такі команди

Eq() == дорівнює

Ne() != Не дорівнює

Le() <= менше або дорівнює

Lt() < менше чим

De() >= більше дорівнює

Gt() > більше чим

За допомогою команд можна відкрити вже створений датафрейм

```
anime = pd.read_csv('anime-recommendations-database/anime.csv')
```

	anime_id	name	genre	type	episodes	rating	members
0	32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie	1	9.37	200630
1	5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	TV	64	9.26	793665
2	28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.25	114262
3	9253	Steins;Gate	Sci-Fi, Thriller	TV	24	9.17	673572
4	9969	Gintama'	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.16	151266

Для запису в формат csv використовується команда
`rating[:10].to_csv('saved_ratings.csv', index=False)`

Для перегляду початку датафрейму використовується команда `anime.head(3)`, а для кінця `rating.tail()`

Приклад виконання зображено на рисунку 11

	anime_id	name	genre	type	episodes	rating	members
0	32281	Kimi no Na wa.	Drama, Romance, School, Supernatural	Movie	1	9.37	200630
1	5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Mili...	TV	64	9.26	793665
2	28977	Gintama°	Action, Comedy, Historical, Parody, Samurai, S...	TV	51	9.25	114262

Данные из начала датафрейма

	anime_id	name	genre	type	episodes	rating	members
12293	26081	Yasuji no Pornorama: Yachimae!!	Hentai	Movie	1	5.46	142

Данные из конца датафрейма

Рисунок 11 – приклад роботи програми

Для запобігання проблем із сумісністю різних версій модулів та пакетів я вирішив писати програмний код через оболочку `jupyter notebook`.

Ця програма являє собою інтерактивний блокнот, із початку будучи веб-реалізацією і розвитком IPython, ставшу потім самостійним проектом.

Орієнтовну не тільки на Python а і на інші мови такі як R, Julia, Scala та багато інших

За допомогою стрічки команд достатньо перейти в місце встановлення даного блокноту та ввести команду `jupyter notebook`. Після цього система сама відкриє браузер із вашим репозиторієм. Інтерфейс блокноту зображено на рисунку 12

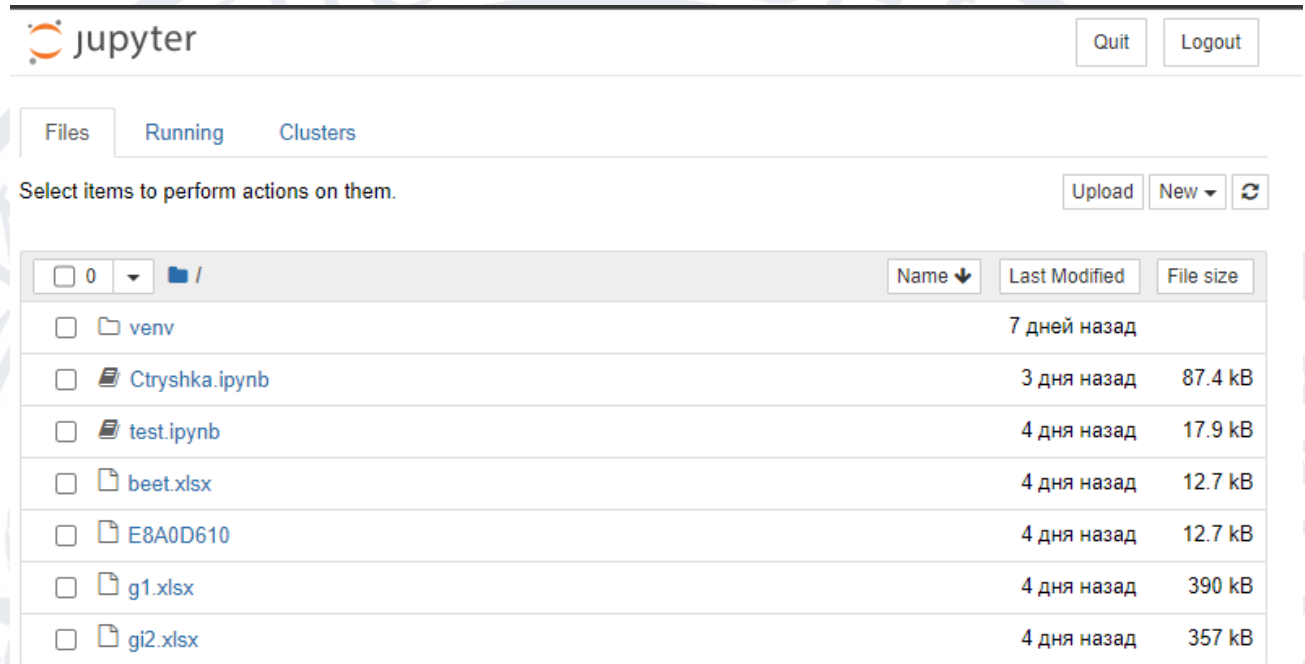


Рисунок 12 -інтерфейс блокноту

Саме місце для написання коду є дуже зручним і дає можливість перевіряти написання коду на кожній стрічці а не запускати код повністю, що є дуже добре. Адже ми можемо писати код по стрічці та відразу знати де саме не виконується та чи інша дія. Приклад зображено на рисунку 13.

```
In [4]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import openpyxl

In [5]: df = pd.read_excel('beet.xlsx')

In [6]: book = openpyxl.open("beet.xlsx", read_only=True)
sheet = book.active
wait=df['очікування'].tolist()
result=df['результат'].tolist()
print("очікування",wait)
print('-----')
print('результат',result)

очікування [36.3, 33.2, 39.0, 40.0, 40.7, 38.4, 35.8, 36.0, 37.9]
-----
результат [30.0, 33.2, 39.3, 37.0, 40.6, 38.4, 35.7, 37.0, 37.9]
```

```
In [4]: cat_par = [i for i in range (9)]
a = wait
b = result
width= 0.1
x =np.arange(len(cat_par))
print(x)

[0 1 2 3 4 5 6 7 8]

In [ ]:

In [5]: fig, ax = plt.subplots()
rects1= ax.bar(x - width/1, wait, label="wait")
rects2= ax.bar(x + width/3, result, label="result")
ax.set_xticks(x)
ax.set_xticklabels(cat_par)
plt.grid(True)
plt.show()
```

Рисунок 13 – виконання програми по рядкам

4.2 Аналіз обраної області

Аналіз був здійснений за допомогою мови програмування та бібліотек pandas, numpy, matplotlib. Pandas дає змогу працювати із аналізом даних, читаючи файли в різних типах даних він може витягнути потрібні данні та вставити їх у потрібну змінну. Matplotlib дає змогу працювати із графіками. Використовуючи інформацію по підприємству ми можемо зробити деякі висновки порівнюючи очікуємий результат та фактичним.

Головною проблемою при виробництві цукру є в'язкість патоки. Чим більша щільність тим краще буде результат.

Використовуючи данні із відкритого доступу можна порівняти різні результати за різні дні. Данні які використовувались для аналізу зображено на рисунку 14

очікування	результат
36.3	30
33.2	33.2
39	39,3
40	37
40.7	40,6
38.4	38.4
35.8	35,7
36	37
37.9	37.9

Рисунок 14- данні приставлені в табличному вигляді в форматі.xlsx.

На рисунку 15 зображений графік порівняння

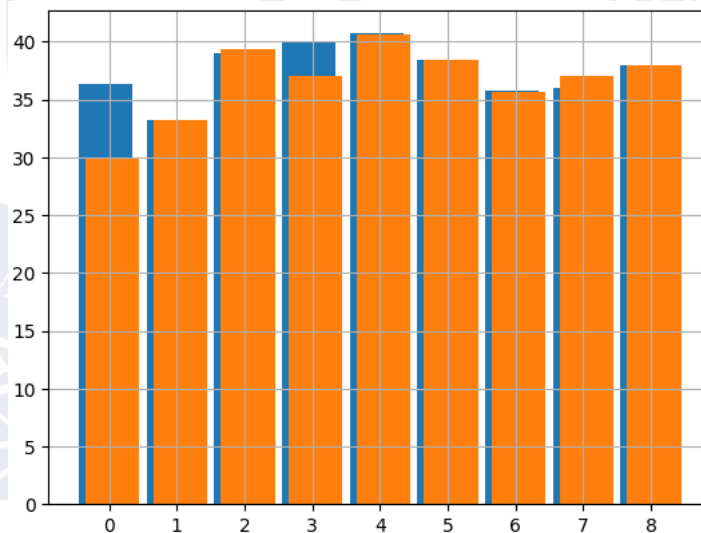


Рисунок 15- аналіз даних в графічному вигляді із рисунка 14

Отже аналізуючи рисунок ми можемо бачити що не завжди очікуємий результат співпадає із фактичним, це пов'язано із роботою підприємства в цілому. В більшості випадків під час встановлення нового обладнання продуктивність збільшується, але бувають випадки що несумісність

обладнання дає більше втрати. Використовуючи такі графіки можна очно порівняти результат.

Також на заводі прибуток отримується внаслідок продажу жому. Виробництво жому також можна взяти до аналізу адже він є головною складовою прибуткової системи.

Вхідні данні зображено на рисунку 16

час	стружка		
23:50:01.0000000	215950	215809	217180
23:45:00.0000000	216937	215809	217180
23:40:02.0000000	217109	215809	217180
23:35:00.0000000	217020	215809	217180
23:30:02.0000000	217099	215809	217180
23:25:00.0000000	216293	215809	217180
23:20:01.0000000	215991	215809	217180
23:15:00.0000000	216662	215809	217180
23:10:02.0000000	216726	215809	217180
23:05:00.0000000	216303	215809	217180
23:00:03.0000000	216596	215809	217180
22:55:00.0000000	215918	215809	217180
22:50:02.0000000	216614	215809	217180
22:45:00.0000000	215891	215809	217180
22:40:02.0000000	216708	215809	217180
22:35:00.0000000	216574	215809	217180
22:30:02.0000000	216431	215809	217180
22:25:00.0000000	217066	215809	217180
22:20:01.0000000	215942	215809	217180
22:15:00.0000000	216815	215809	217180
22:10:02.0000000	216419	215809	217180
22:05:01.0000000	216930	215809	217180
22:00:03.0000000	216462	215809	217180
21:55:01.0000000	216785	215809	217180
21:50:01.0000000	216662	215809	217180
21:45:00.0000000	216185	215809	217180
21:40:02.0000000	216314	215809	217180
21:35:00.0000000	216139	215809	217180

Рисунок 16. Вхідні дані аналізу

Головною перевагою програм для інтелектуального аналізу є те що дані можна змінювати в реальному часі та дивитись як це змінить статистику та картинку у побудованому графіці, також читання даних стає в рази простіше, оскільки ми звертаємось тільки за тими даними які дійсно нам потрібні, а все інше пропускаючи.

За допомогою бібліотеки `pandas` програма звертається до необхідної нам комірки та дає можливість проводити різні операції над ними. Взявши дані ми отримуємо на виході графік із даними про вироблення жому протягом 100 перших днів.

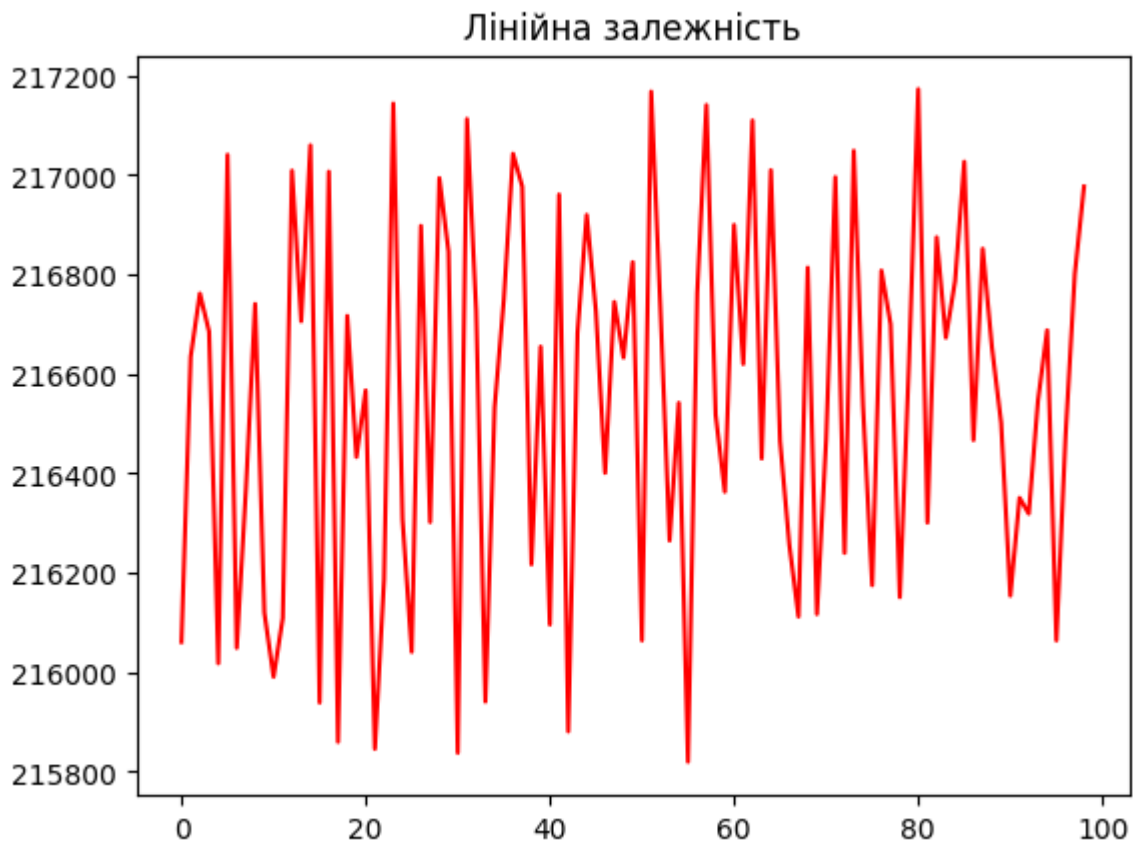


Рисунок 17 – Вихідні данні після аналізу даних із рисунку 16

Головною продукцією заводу є саме цукор. На його видобуток та переробку впливає багато факторів. Однак для порівняння потрібно взяти загальну кількість переробленого та роки що порівняти їх між собою. Отож взявши данні та використавши програму `python` із бібліотекою `pandas` ми можемо побачити скільки цукру в якому році було вироблено. Данні містяться в файлі `excel` зображено на рисунку 18.

рік	Потужність	Перероблено цукрового буряку	Вироблено цукру	Вихід цукру
2021	7481	644,6	133,9	15,85
2020	6914	620	80,9	13,04
2019	7300	683,9	111,8	16,3
2018	6625	940,74	120,6	12,8
2017	6485	691,78	105,52	15,25
2016	7358	738	117,02	15,86
2015	7612	692	104,05	15,04
2014	6952	720,3	106,31	14,76
2013	6885	633,4	82,35	13
2012	6295	688	99,02	14,39
2011	5590	570,2	80,57	14,13
2010	5336	402,6	46,61	11,58
2009	4653	347,5	52,71	15,17
2008	3374	333,5	41,9	12,56

Рисунок 18 – загальні данні по виробництву за роки

Для зручності порівняння використано графік побудований у програмі. Графік зображено на рисунку 19.

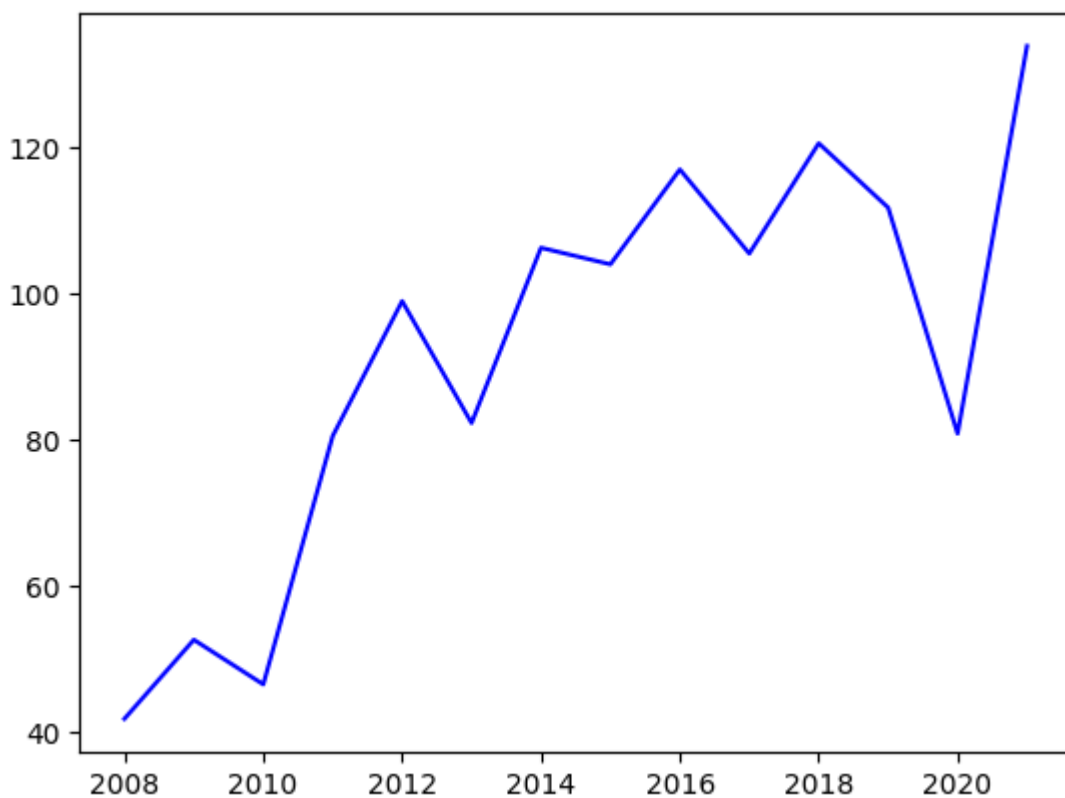


Рисунок 19 – аналіз виробництва цукру.

Дивлячись на графік можна зробити такі висновки що кращим роком для переробки був 2021, цього року вироблено більше 120тис.тон.

Для загального аналізу необхідно також порівняти скільки буряків було перероблено. Переробка цукрових буряків зображено на рисунку 20

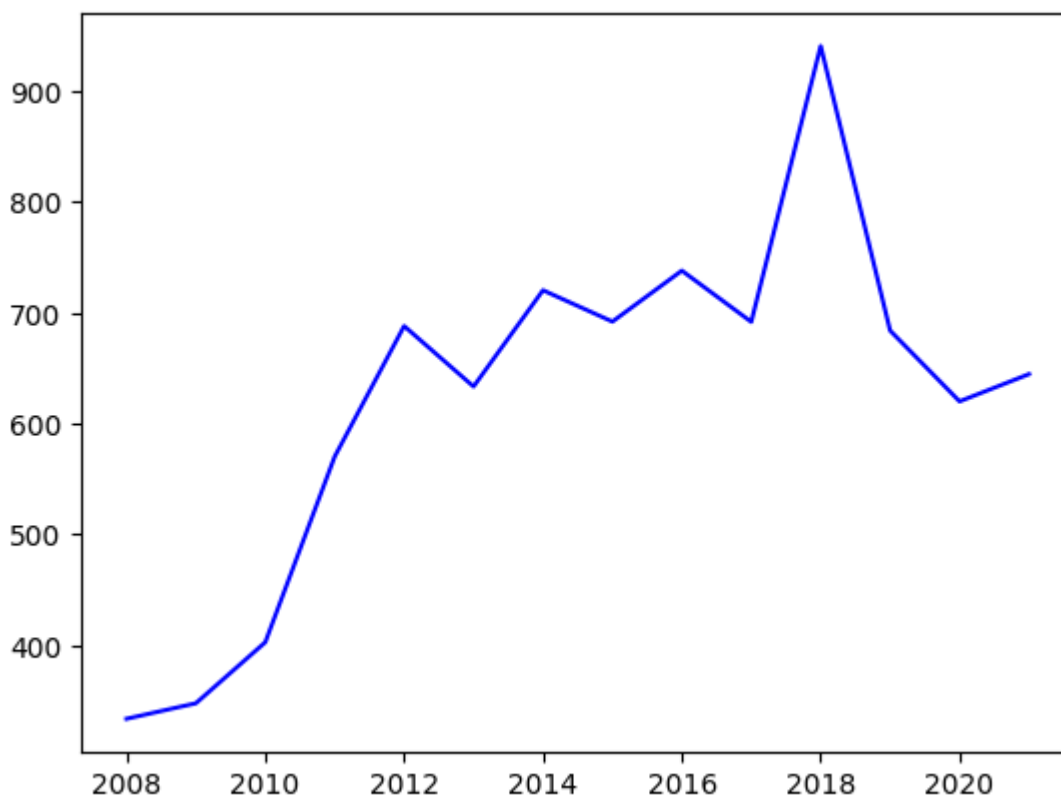


Рисунок 20 – Графік перероблення цукрового буряку

Порівнюючи графік ми бачимо що найбільше буряку було перероблено у 2018 році, однак найбільша маса цукру становила у 2021. Це пояснюється тим що втрати буряки були великими, адже буряк що замерз та подрібнився шляхом на цукор не підходить. Найважливішим показником є цукористість буряку. Для аналізу продемонструю графік цукристості на рисунку 21.

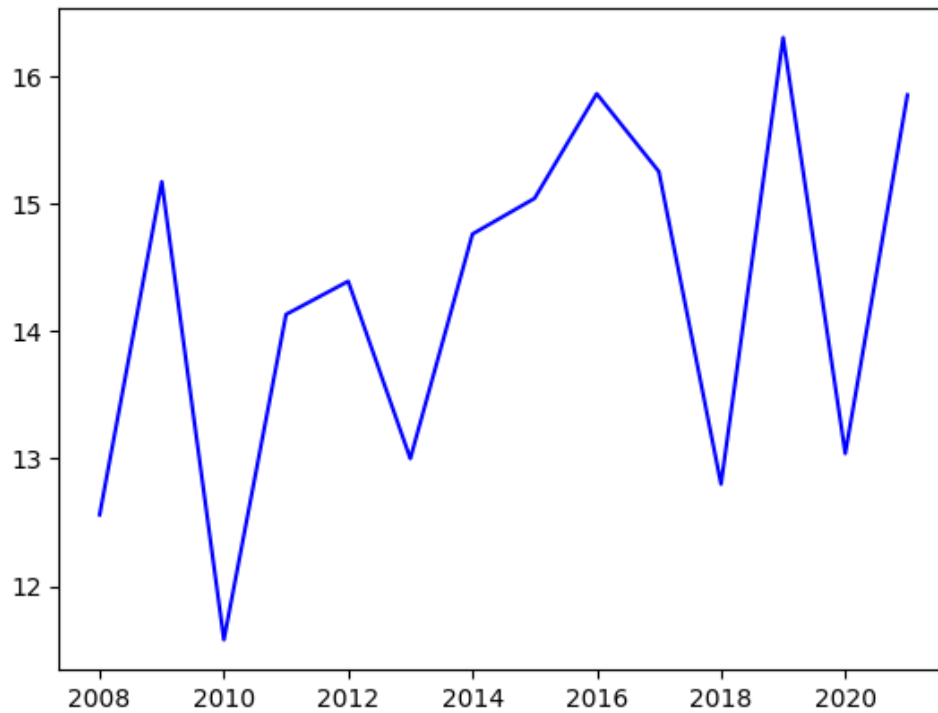


Рисунок 21 – вихід цукру

Отож із даного графіку можна зробити висновок такий, у 2018 році була найбільша врожайність, однак низька цукористість. Швидше за все це пов'язано із неправильним вживанням добрив, оскільки не всі добрива дають змогу зберегти цукористість при великій вазі буряка.

Також одним із найважливіших чинників є потужність промисловості. Вона характеризує силові можливості компанії тобто переробка тис.тон. за добу данні зображено на рисунку 22.

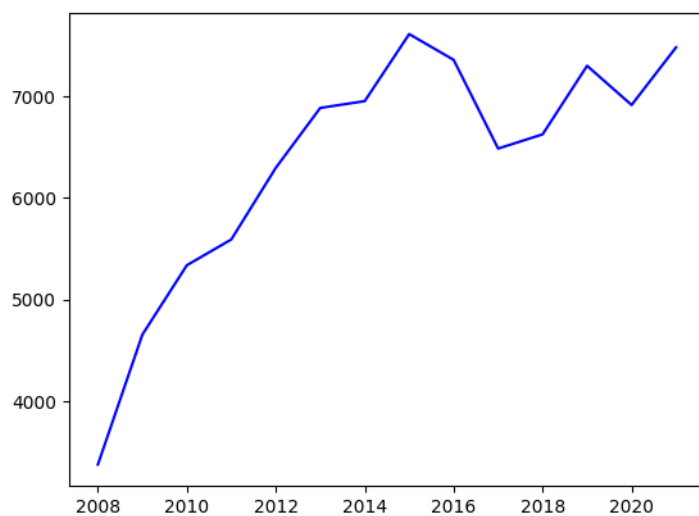
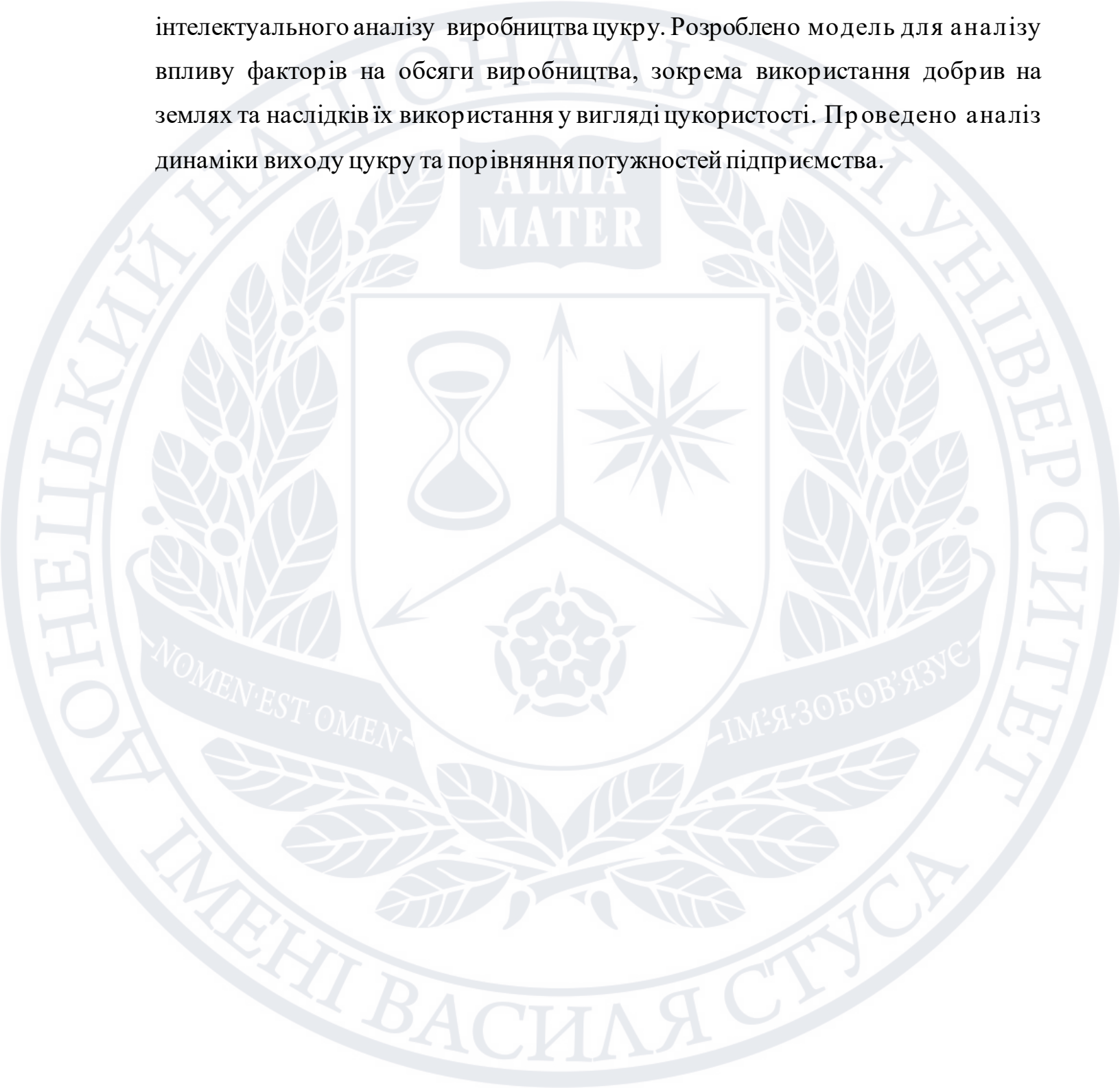


Рисунок 22- потужність даного підприємства

Висновок до четвертого розділу

У даному розділі продемонстрована програмна реалізація елементів інтелектуального аналізу виробництва цукру. Розроблено модель для аналізу впливу факторів на обсяги виробництва, зокрема використання добрив на землях та наслідків їх використання у вигляді цукористості. Проведено аналіз динаміки виходу цукру та порівняння потужностей підприємства.



Висновок до роботи

Потенціал Data Mining дає "зелене світло" для розширення меж застосування технології. Щодо перспектив Data Mining можливі такі напрями розвитку:

- виокремлення типів предметних областей із відповідними їм евристичними, формалізація яких полегшить вирішення відповідних завдань Data Mining, які стосуються цих областей;
- створення формальних мов та логічних засобів, за допомогою яких буде формалізовано міркування та автоматизація яких стане інструментом вирішення завдань Data Mining у конкретних предметних галузях;
- створення методів Data Mining, здатних як витягувати з даних закономірності, а й формувати деякі теорії, що спираються на емпіричні дані;
- подолання суттєвого відставання можливостей інструментальних засобів Data Mining від теоретичних досягнень у цій галузі.

Інтелектуальний аналіз даних є одним із найбільш актуальних та затребуваних напрямків прикладної математики. Сучасні процеси бізнесу та виробництва породжують величезні масиви даних, і людям стає все важче інтерпретувати та реагувати на велику кількість даних, які динамічно змінюються у часі виконання, не кажучи вже про попередження критичних ситуацій. «Інтелектуальний аналіз даних» отримати максимум корисних знань з багатовимірних, різномірних, неповних, неточних, суперечливих, непрямих даних. Допомагає зробити це ефективно, якщо обсяг даних вимірюється гігабайтами чи навіть терабайтами. Допомагає будувати алгоритми, здатні навчатися прийняття рішень у різних професійних галузях.

Кошти «Інтелектуального аналізу даних» оберігають людей від інформаційного навантаження, переробляючи оперативні дані на корисну інформацію те щоб потрібні дії було прийнято у потрібні часи.

Прикладні розробки ведуться за такими напрямками: - прогнозування в економічних системах; автоматизація маркетингових досліджень та аналіз клієнтських середовищ для виробничих, торгових, телекомунікаційних та Інтернет-компаній; автоматизація прийняття кредитних рішень та оцінка кредитних ризиків; моніторинг фінансових ринків; автоматичні торговельні системи.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. <https://www.upi-agro.com.ua/ua/Home/MapDetails/5>
2. Економічна оцінка виробництва цукрових буряків у 2019 році” (Журнал „Цукрові буряки” №4, 2019р.).
3. https://zvitnist.com/34009446_tovarystvo_z_obmezhenu_vdpovdalnstu_prodovalcha_
4. <https://landlord.ua/news/shche-dva-zavody-prystupyly-do-pererobky-tsukrovikh-buriakiv/>
5. <http://ukrsugar.com/uk/post/pusk-cukrovih-zavodiv-20212022-mr-onovleno>
6. Зюбровська Т.А. Вдосконалення оперативного аналізу в рослинництві при ринкових відносинах / Т.А. Зюбровська, Л.М. Будняк. // Збірник наукових праць Подільської державної аграрно-технічної академії - Кам'янець-Подільський, 2003.-Вип.10.-с.133-135.
7. Калінін Ф.Л. Біологічно активні речовини в рослинництві. К.: Наукова думка
8. Основні чинники, що впливають на ціни внутрішнього ринку цукру” (Журнал Пропозиція” №6, 2010 р.);
9. Стиснюк О.І. Оптимізація структури посівних площ і матеріальних витрат на виробництво продукції рослинництва // Економіка
10. <https://vkursi.pro/card/tov-pk-zoria-podillia-34009446>
11. https://thedigital.gov.ua/lms_ai/intelektualnij-analiz-danih
12. <http://kiis.knu.ua/intelektualnij-analiz-danih/>
13. <http://www.tsatu.edu.ua/kn/course/intelektualnyj-analiz-danyh/>
14. <https://github.com/>
15. <https://agravery.com/uk/posts/show/v-ukraini-z-01-veresna-2022-roku-planuut-zdijsnuvati-virobnictvo-23-cukrovi-zavodi>

16. <https://www.ukrinform.ua/rubric-economy/3558837-cukrovij-golod-nam-ne-zagrozue-ale-nizkih-cin-ne-bude.html>
17. <https://www.ukr.net/news/details/economics/92444112.html>
18. <https://www.upi-agro.com.ua/>
19. Alotaibi N.M., Abdullah M., Mosli H. Agent-based Big Data Mining // International Journal of Advanced Trends in Computer Science and Engineering, 2019. Vol. 8. N.1.1. P. 245-252.
20. Anushree R. Survey on Big Data Mining Algorithms. // International Journal for Research in Applied Science and Engineering Technology, 2019. N7. P. 1363-1367.
21. Bathla G., Aggarwal H., Rani R. Evolution of Data Analytics Techniques: From Data Mining to Big Data Mining // International Journal of Advanced Science and Technology, 2019. T 28. N19. P. 779-795.
22. Flask is a microframework for Python based on Werkzeug [Електронний ресурс] — Режим доступу: <http://flask.pocoo.org/>
23. Keras: The Python Deep Learning library [Електронний ресурс] — Режим доступу: <https://colab.research.google.com>
24. 2. Introducing Colaboratory [Електронний ресурс] — Режим доступу: <https://colab.research.google.com>
25. Hornik K., Stinchcombe M., White H. Multilayer Feedforward Networks are Universal Approximators / Neural Networks, 1989, vol. 2, no. 5. — P. 359-366.
26. Aitkenhead-Peterson et al. 2012 [Електронний ресурс] — Режим доступу: https://www.academia.edu/1537285/Aitkenhead-Peterson_et_al._2012
27. С. Николенко, А. Кадурин, Е. Архангельская; Глубокое обучение погружение в мир нейронных сетей. ББК 32.973.236

28. Is the Braess Paradox related to Dropout in Neural Nets? [Електронний ресурс] — Режим доступу: <https://hackernoon.com/is-the-braess-paradox-related-to-dropout-in-neural-nets-270ecb97cde>
29. Types of Neural Network Activation Functions: How to Choose. [Електронний ресурс] — Режим доступу: <https://missinglink.ai/guides/neuralnetwork-concepts/7-types-neural-network-activation-functions-right/>
30. Robert M. Haralick, Linda G. Shapiro; Image segmentation techniques [Електронний ресурс] — Режим доступу: <https://www.sciencedirect.com/science/article/pii/S0734189X85901537>
31. Dilpreet Kaur, Yadwinder Kaur; Various Image Segmentation. Techniques: A Review// IJCSMC, Vol. 3, Issue. 5, May 2014, p.g.809 – 814. ISSN 2320–088X
32. Обзор алгоритмов сегментации. [Електронний ресурс] — Режим доступу: <https://habr.com/ru/company/intel/blog/266347/>
33. Su Hnin Hlaing, Aung Soe Khaing; Weed and crop segmentation and classification using area thresholding //International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308
34. Sebastian Haug, Andreas Michaels; Plant classification system for crop /weed discrimination without segmentation// Published in IEEE Winter Conference 2014. DOI:10.1109/WACV.2014.6835733
35. Philipp Lottes, Cyrill Stachniss; Semi-Supervised Online Visual Crop and Weed Classification in Precision Farming Exploiting Plant Arrangement. [Електронний ресурс] — Режим доступу: http://flourishproject.eu/fileadmin/user_upload/publications/lottes17iros.pdf
36. Yao Wang; Yisong Chen; Peng Lu; Heng Wang Sobel Heuristic Kernel for Aerial Semantic Segmentation// 2018 25th IEEE International Conference on Image Processing (ICIP), Electronic ISSN: 2381-8549

37. The 2016 Sugar Beets Dataset Recorded at Campus Klein Alten dorf in Bonn, Germany [Електронний ресурс] — Режим доступу: <http://www.ipb.unibonn.de/data/sugabeets2016/>
38. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989. 51
39. Le Hou; Dimitris Samaras; Tahsin M. Kurc; Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification //2016 IEEE Conference on Computer Vision and Pattern Recognition, ISSN: 1063-6919
40. Fully Convolutional Networks for Semantic Segmentation [Електронний ресурс] — Режим доступу: https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf
41. U-Net: Convolutional Networks for Biomedical Image Segmentation [Електронний ресурс] — Режим доступу: <https://lmb.informatik.unifreiburg.de/people/ronneber/u-net/>
42. Improve Deep Learning Models performance & deep network tuning [Електронний ресурс] — Режим доступу: https://medium.com/@jonathan_hui/improve-deep-learning-models-performancenetwork-tuning-part-6-29bf90df6d2d
43. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Deep Residual Learning for Image Recognition [Електронний ресурс] — Режим доступу: <https://arxiv.org/abs/1512.03385>

Додаток А

Лістинг коду

Імпорт бібліотек

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import openpyxl

df = pd.read_excel('beet.xlsx') (вказання файлу який необхідно відкрити)

book = openpyxl.open("beet.xlsx", read_only=True)
sheet = book.active
wait = df['очікування'].tolist() (обираєм стовпець який потрібно для аналізу)
result = df['результат'].tolist()
print("очікування", wait)
print('-----')
print('результат', result)

очікування [36.3, 33.2, 39.0, 40.0, 40.7, 38.4, 35.8, 36.0, 37.9]
-----
результат [30.0, 33.2, 39.3, 37.0, 40.6, 38.4, 35.7, 37.0, 37.9]

cat_par = [i for i in range(9)]
a = wait
b = result
width = 0.1
x = np.arange(len(cat_par))

fig, ax = plt.subplots() (виводимо обробку даних у вигляді графіку)
rects1 = ax.bar(x - width/1, wait, label="wait")
rects2 = ax.bar(x + width/3, result, label="result")
ax.set_xticks(x)
ax.set_xticklabels(cat_par)
plt.grid(True)
plt.show()
```