УДК: 004.9:631.544.4

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ OLAP ТА DATA MINING В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ У ТЕПЛИЧНОМУ ГОСПОДАРСТВІ

Лендєл М.І.
асистентка кафедри комп'ютерних наук
ORCID https://orcid.org/0009-0008-0042-7705
E-mail: marynalendel@gmail.com
Голуб Б.Л.
к.т.н., доцент, кафедри комп'ютерних наук
ORCID https://orcid.org/0000-0002-1256-6138
E-mail: bella.golub55@gmail.com
Національний університет біоресурсів і природокористування України
Героїв оборони 15, м. Київ, Україна

Анотація. У сучасному світі технології відіграють ключову роль у багатьох напрямках життя, включаючи сільське господарство. Вирощування овочів та фруктів у спорудах закритого ґрунту стає все більш популярним і вимагає контролю над умовами та ресурсами для збільшення врожайності, проте із зростанням обсягів даних, які збираються в системах моніторингу, стає важче ефективно аналізувати зібрану інформацію.
Дана стаття розглядає використання технологій OLAP та Data Mining для підвищення ефективності вирощування овочів та фруктів у спорудах закритого ґрунту. Метою дослідження є визначення доцільності та ефективності застосування цих технологій для аналізу великих обсягів даних, зібраних під час вирощування овочів та фруктів.
Автори дослідження аналізують, як OLAP та Data Mining можуть надати корисну інформацію для прийняття рішень щодо оптимізації процесів вирощування для підвищення врожайності.
Ключові слова: система підтримки прийняття рішень, OLAP, Data Mining, сховище даних, багатовимірний куб.

Актуальність. Сучасний світ повністю залежить від технологій, які стрімко розвиваються. Вони займають ключове місце майже у всіх напрямках нашої діяльності та допомагають у повсякденному житті майже на кожному кроці.

Під час вирощування важливим етапом є саме вивчення та аналіз всіх умов, які необхідні для нормального росту й розвитку рослини. У процесі росту рослин важливо враховувати оптимальні показники мікроклімату теплиці для підвищення ефективності використання ресурсів для вирощування врожаю. Дивлячись на згадані особливості корисним є використання програмних засобів для моніторингу, збереження та аналізу показників, які важливі для підвищення ефективності вирощування.

На основі всіх наявних даних у системі виробник може проводити аналіз всіх ключових показників, їхні зміни та вплив у розрізі часу та приймати відповідні рішення для свого підприємства. Проте створені системи з часом розширюються, а відповідно інформація у них теж, тому стає незручно та неефективно аналізувати попередньо внесені дані. У такому випадку існує необхідність у створенні системи, яка буде проводити аналіз показників на основі накопичених даних.

Саме система підтримки прийняття рішень (СППР) допомагає аналізувати накопичені дані, а також приймати рішення, які вимагають оцінки, рішучості та послідовності дій. Така інформаційна система допомагає керівництву підприємств середнього і високого рівня організації, аналізуючи великі обсяги неструктурованих даних і накопичуючи дані, у вирішенні певної проблеми та допомогти у прийнятті рішення.

На господарства СППР використовується як для одного об'єкту, так і для мережі господарств, що допоможе у кращому вивченні даних та відповідно подальшому прогнозуванні необхідних явищ, інформації про наявні та можливі факти, а також для представлення інформації користувачам у зрозумілій формі. Такі системи допомагають своєчасно прийняти керуюче рішення для покращення результатів виробництва та відповідно збільшення прибутку.

Аналіз останніх досліджень та публікацій. У процесі аналізу предметної області було проаналізовано наукові роботи, які пов'язані з вирішенням питання аналізу роботи тепличного господарства.

Автори Ченг Ванг, Лілі Венг, Пінг Донг, Ксяо Кяо у своїй роботі «Дослідження системи підтримки прийняття рішень (DSS) для теплиць на основі видобутку даних» розглянули систему підтримки прийняття рішень для теплиці, побудованої зі сховища даних та технології видобутку дат. У системі було створено сховище даних з метою диверсифікації пам'яті, використання онлайн аналітичної обробки та видобутку даних збагачує базу знань новою інформацією про сільське господарство. Дана робота показує структуру системи підтримки прийняття рішень загалом: використання технології аналітичної обробки OLAP, яка допомагає аналітично обробляти дані, які зберігаються у базі знань. Також пояснює структуру модуля видобутку даних, а саме побудову окремих модулів, кожен з яких відповідає за моделювання та прогнозування, пошуку схожості, побудови асоціативних правил тощо [7].

В іншій роботі «Експертна система управління тепличним виробництвом» автори Йонгуанг Ху, Джиганг Ванг та Пінгпінг Лі розповіли про важливість використання теплиці та правильного управління всіма внутрішніми процесами і це підштовхує до розробки експертних систем.

У роботі описано та зображено функціональну структуру системи, яка складається з декількох підсистем. Важливу частину займає база знань, яка містить словник правил. Також було описано систему збору даних у режимі реального часу.

У результаті аналізу було описано короткі функціональні характеристики системи:

• моніторинг мікроклімату теплиці у режимі реального часу;

• оновлення інформаційної бази довкілля для експертної системи;

• виведення та відображення інформації з текстом, графічною таблицею або роздруківкою;

• тривожний сигнал у разі екстремальних ситуацій або зниження врожайності.

Крім програмного у роботі було розглянуто апаратне забезпечення та впровадження системи та її результати роботи, де показано приклад діагностики захворювання на основі введених ознак, а також аналізу оптимальних умов мікроклімату та поливу [8].

Мета дослідження. Метою дослідження є визначення доцільності застосування технологій OLAP і Data Mining для підвищення ефективності вирощування овочів та фруктів у спорудах закритого ґрунту.

Матеріали і методи дослідження. У процесі дослідження були використані такі методи:

1. Технології OLAP для накопичення даних у більш зручній структурі: багатомірний куб, який ефективний для проведення аналізу у розрізі різних параметрів, формування звітної інформації тощо.
2. Технології Data Mining для знаходження зв'язків між наповненими даними, які відображають результати роботи тепличного господарства.

Результати дослідження та їх обговорення. Для робочої станції оператора теплиці було розроблено підсистему моніторингу, яка працює з даними, які зберігаються в оперативній базі даних, а також апаратне забезпечення з використанням датчиків температури й вологості, які підключені до мікроконтролера Arduino. Для вимірювання показників використано датчик температури і вологості повітря DHT11. Підсистему моніторингу реалізовано з використанням мови програмування Python, де забезпечується

передача даних про отримані показники температури й вологості до оперативної бази даних. На рис. 1 зображена схема оперативної бази даних, де зберігаються показники мікроклімату у теплицях, врожайність, оптимальні показники тощо.

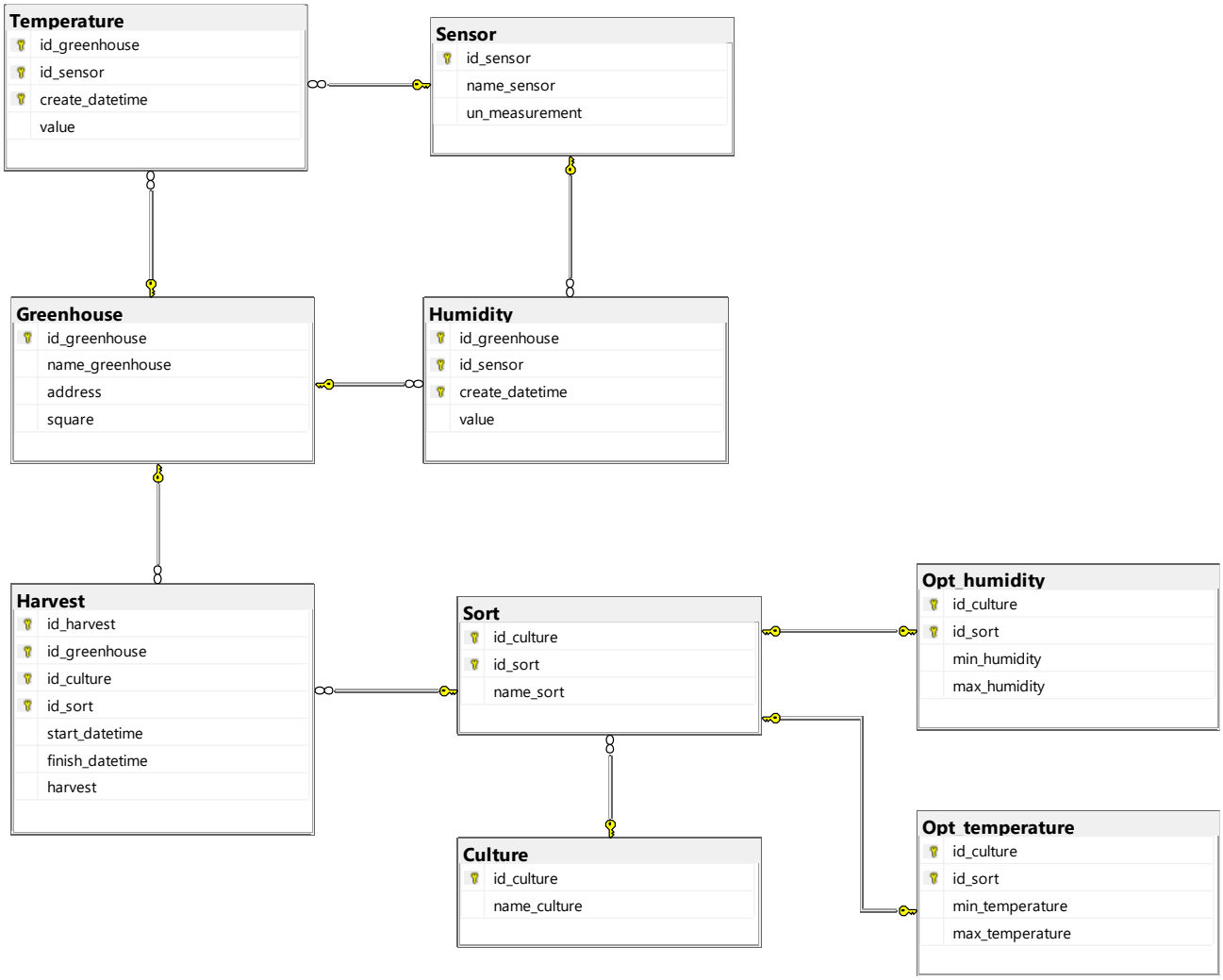

Рисунок 1 - Схема оперативної бази даних

На основі отриманого оперативного джерела даних для проведення аналізу використовується сховище даних (СД), яке дозволить проводити аналіз у різних розрізах. Це забезпечується наявністю декількох вимірів, які являють собою сукупність довідкової інформації про вимірювану подію. У контексті сховищ даних подіями являються факти, які описують результати певного бізнес-процесу. Структура сховища даних зображена на рис. 2.

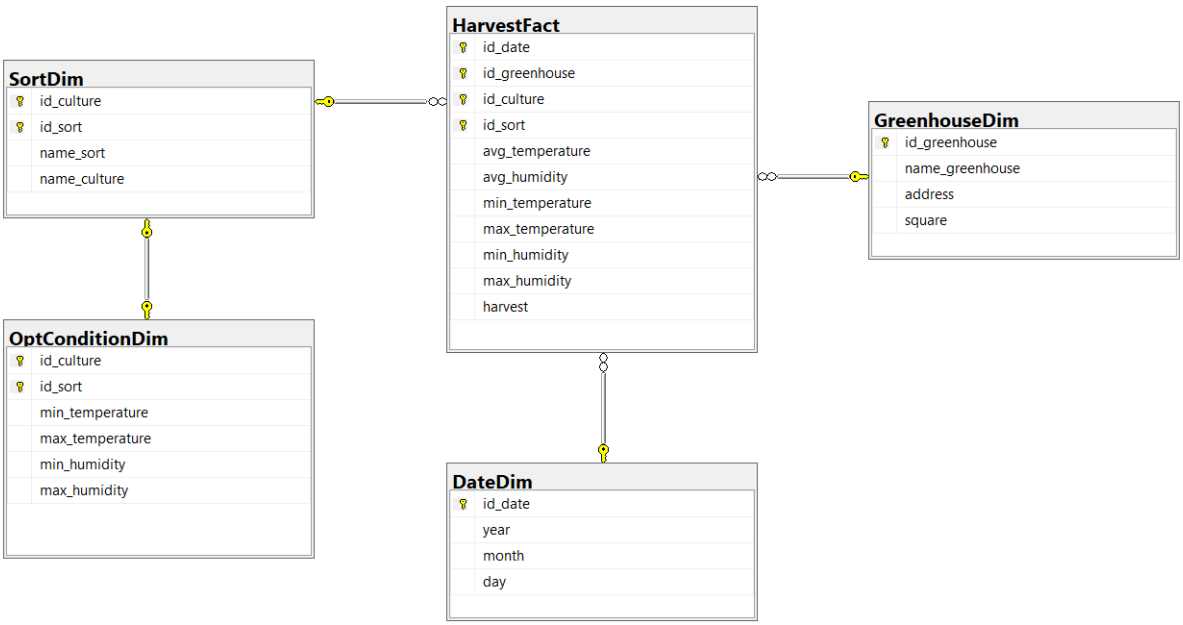

Рисунок 2 - Вітрина тепличного господарства

Розроблене сховище даних було використано для розгортання багатовимірного кубу, використовуючи службу SQL Server Analysis Services (SSAS). Саме для аналізу даних було реалізовано наповнення сховища даних, яке відбувається на основі даних з оперативної бази даних, зображеної на рис. 1. Процес передачі даних був реалізований за допомогою служби SQL Server Integration Services.

У процесі аналізу отриманих даних у СД було реалізовано розрахунок ключового показника ефективності (KPI), який був реалізований з використанням мови запитів MDX для доступу до багатовимірних структур даних (рис. 3).

Отобразить структуру	Значение	Цель	Состояние
KPI_Sort1_Opt_Avg_Humidity	66.00%	67.5 %	
KPI_Sort1_Opt_Avg_Temp	21.11°C	22.0 °C	
KPI_Sort4_Opt_Avg_Humidity	75.22%	77.5 %	
KPI_Sort4_Opt_Avg_Temp	19.89°C	22.5 °C	

Рисунок 3 - Результати розрахунку ключового показника ефективності

Для проведення інтелектуального аналізу даних було використано SQL Server Data Tools-Business Intelligence(SSDT-BI), який містить набір технологій для бізнес аналізу: створення моделей даних Analysis Services (AS), пакетів Integration Services (IS) та звітів Reporting Services (RS).

Використовуючи розгорнутий куб було побудовано графік зміни врожайності з використанням алгоритму часових рядів, який зображено на рис. 4.

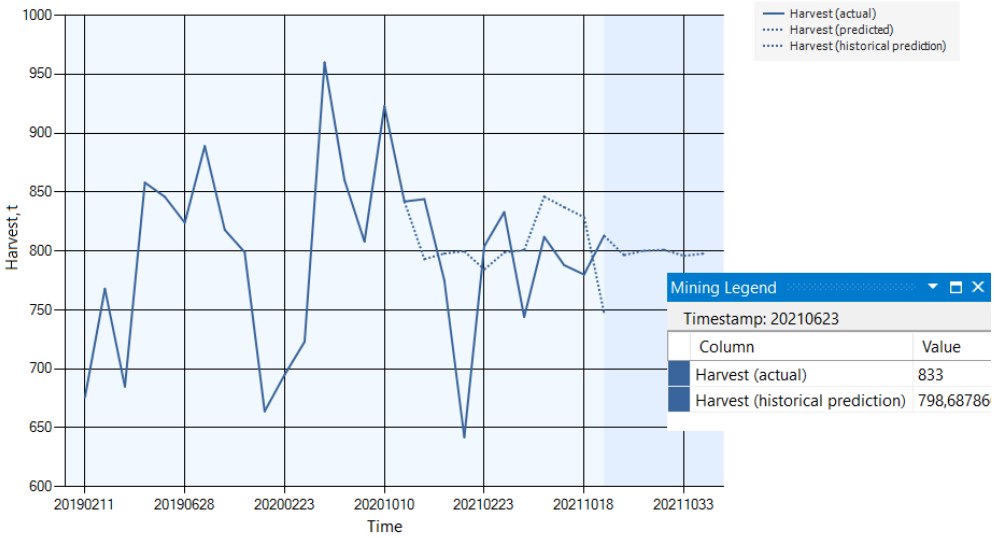

Рисунок 4 - Сформований графік з використання алгоритму часових рядів

У процесі аналізу за допомогою часових рядів було проаналізовано зміну врожайності. Часові ряди дозволяють прогнозувати як буде далі розвиватися ситуація. Тому на основі отриманого графіку можна побачити, що починаючи з початку 2021 року сумарна врожайність у теплицях зменшилась. Наприклад взимку 2021 врожайність досягла найнижчого значення, прогнозоване значення відрізняється від фактичного.

Також на основі розгорнутої структури було побудовано мережу залежностей назви культури та сорту від заданих діапазонів температури та вологості (рис. 5).

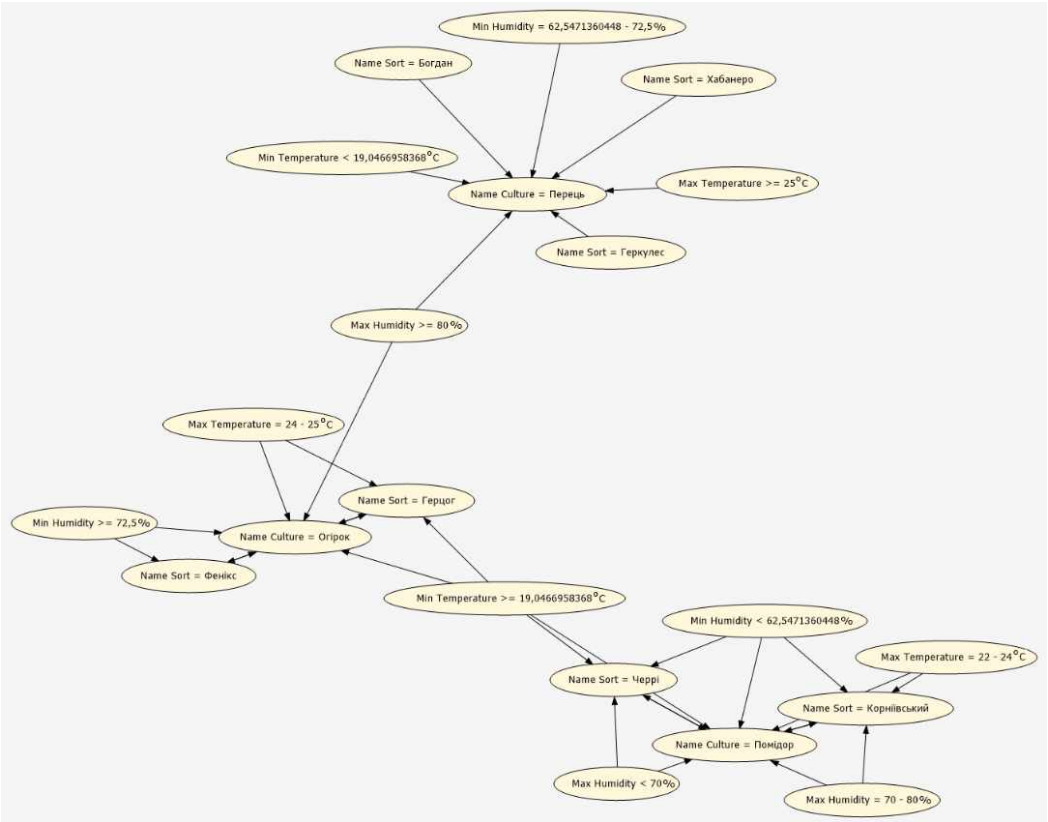

Рисунок 5 - Мережа залежностей на основі реалізованого алгоритму асоціативних правил

З наведеної схеми асоціативних правил можна визначити найбільш сприятливі умови для вирощування як окремих сортів та культур, так і про можливість комбінування сортів різних культур для вирощування в єдиному просторі. Виокремимо значущі особливості:

• сорти однієї культури мають доволі схожі вимоги до середовища, що робить процес вирощування більш зрозумілим, адже дає підстави для їх вирощування в єдиному просторі;

• перці та огірки мають однакові вимоги до максимальних та мінімальних показників вологості та температури, що дає можливість комбінувати їх сорти під час вирощування в єдиному просторі;

• щодо помідорів, то їх мінімальне значення температури співпадає із показником для огірків, але через більшу чутливість сортів цієї культури, найкращим рішенням буде вирощування їх в окремому просторі.

Розгорнувши структуру аналізу алгоритмом кластеризації було отримано модель, наведену на рис. 6.

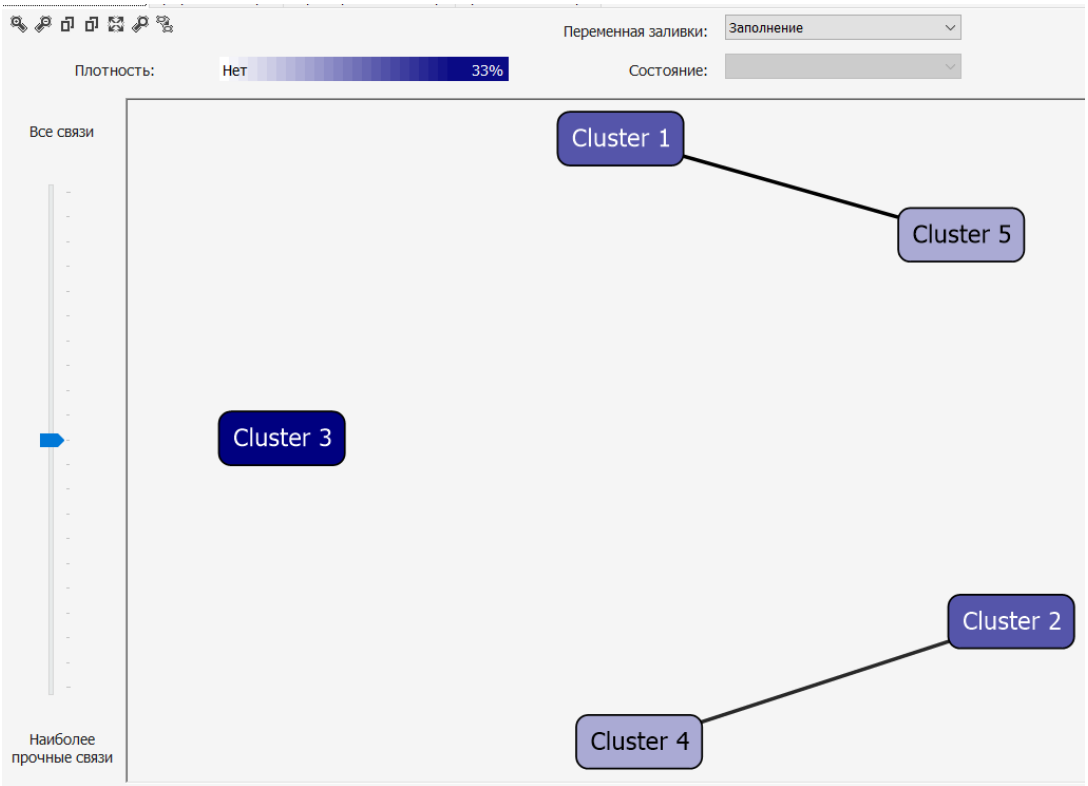

Рисунок 6 - Кластеризація наявних фактів у сховищі даних

За допомогою методу кластеризації були отримані зв'язки між різними кластерами, схожі за ознаками, культурою і умовами вирощування. У результаті розгорнутої структури можна зробити висновок, що дані є однорідними та впорядкованими. Особливо це помітно на схемі мережі зв'язків кластери майже не пов'язані один з одним.

Висновки і перспективи. Під час розробки системи підтримки прийняття рішень було представлено застосування технологій OLAP і Data Mining на прикладі тепличного господарства. Обидві технології доповнюють одна одну і в той час як DM знаходить закономірності на основі відомих знань, то OLAP аналізує дані в режимі реального часу та може стверджувати або спростовувати гіпотези, наведені у процесі інтелектуального аналізу. Отримані знання корисні у процесі формування керуючих рішень тепличного господарства, що дозволить спрямувати стратегію щодо управління окремими бізнес-процесами таким чином, щоб підвищити врожайність у теплицях та відповідно прибутковості господарства загалом. Проте застосування не обмежується лише двома параметрами, так як систему можна розширювати застосовуючи інші показники, які допоможуть розширити пошук залежностей. До прикладу це може бути освітленість, рівень CO_2 тощо. Також предметною областю може виступати не тільки тепличне господарство, дана система також може поширюватись і на інші сфери.

Список використаних джерел

1. Research of Decision Support System (DSS) for Greenhouse Based on Data Mining / Wang, C., Wang, L., Dong, P., & Qiao, X. – CCTA, 2007. URL: https://link.springer.com/content/pdf/10.1007%2F978-0-387-77253-0_40.pdf
2. Expert System for Greenhouse Production Management / Yongguang Hu, Jizhang Wang and Pingping Li. DOI: 10.5772/7080, 2010. – Назва з екрану. URL: https://www.intechopen.com/chapters/6706
3. Data Mining Concepts [Електронний ресурс] – Режим доступу до сайту: https://docs.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions
4. Data Mining Techniques [Електронний ресурс] – Режим доступу до сайту: https://www.tutorialspoint.com/what-are-the-basic-concepts-of-data-mining
5. Overview of Online Analytical Processing (OLAP) [Електронний ресурс] – Режим доступу до сайту: https://support.microsoft.com/en-us/office/overview-of-online-analytical-processing-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6
6. Data Source Views in Multidimensional Models [Електронний ресурс] – Режим доступу до сайту: https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/data-source-views-in-multidimensional-models?view=asallproducts-allversions
7. Key Performance Indicators (KPIs) in Multidimensional Models [Електронний ресурс] – Режим доступу до сайту: https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/key-performance-indicators-kpis-in-multidimensional-models?view=asallproducts-allversions
8. Monitoring The Greenhouse Environment [Електронний ресурс] – Режим доступу до сайту: https://www.greenhousegrower.com/production/plant-culture/special-series/monitoring-the-greenhouse-environment/