МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ПРИРОДОКОРИСТУВАННЯ

ФАКУЛЬТЕТ МЕХАНІКИ, ЕНЕРГЕТИКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

першого (бакалаврського) рівня вищої освіти

на тему: «**Автоматизація процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу**»

Виконав: здобувач 4 курсу групи Акт-42сп
Спеціальності 151 – „Автоматизація та комп'ютерно-інтегровані технології"
Костромської І. О.

Керівник: Чаплига В.М.
Рецензент: Городецький І.М.

ДУБЛЯНИ-2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ПРИРОДОКОРИСТУВАННЯ
ФАКУЛЬТЕТ ЕНЕРГЕТИКИ, МЕХАНІКИ ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Освітній ступінь «Бакалавр» за спеціальністю –
151 – „Автоматизація та комп'ютерно-інтегровані технології"

"ЗАТВЕРДЖУЮ"
Завідувач кафедри_____
д.т.н., проф. А.М. Тригуба
"____"_____202__ р.

ЗАВДАННЯ

на кваліфікаційну роботу студенту

Костромської Ігор Олександрович

1. Тема роботи: «Автоматизація процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу».

Керівник роботи Чаплига Вячеслав Михайлович, д.т.н., професор.

Затверджені наказом по університету «27» листопада 2023 р. № 641 /к-с.

2. Строк подання студентом роботи 10.06.2024 року.

3. Початкові дані до роботи: Технічна документація об'єкту теплопостачання, завдання на автоматизацію технологічного процесу теплопостачання.

4. Зміст розрахунково-пояснювальної записки:

Вступ.

Розділ 1. Аналіз видів таблиць в підприємствах АПК та автоматизації процесу заповнення пропусків даних в них.

Розділ 2. Дослідження особливостей основних методів заповнення пропусків в таблицях даних та їх автоматизації.

Розділ 3. Розробка автоматизації технологічного процесу заповнення пропусків в таблицях даних підприємств АПК на основі інтелектуальних технологій обробки інформації.

Розділ 4. Розрахунок економічної ефективності системи автоматизації технологічного процесу.

Розділ 5. Охорона праці та безпека в надзвичайних ситуаціях.

Висновки.

Список використаної літератури.

6. Консультанти з розділів:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3, 5	*Чаплига В.М., професор кафедри інформаційних технологій*		
4	*Городецький І.М., доцент кафедри фізики, інженерної механіки та безпеки виробництва*		

7. Дата видачі завдання 04 грудня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Приміт-ка
1	*Написання першого розділу та означення головних завдань роботи*	22.01 - 21.02.24	
2	*Виконання другого розділу та формування початкових даних*	22.02 - 21.03.24	
3.	*Виконання третього розділу та узагальнення отриманих результатів роботи*	22.03 - 21.04.24	
4.	*Написання розділу: «Охорона праці»*	22.04 - 10.05.24	
5.	*Вартісна оцінка ефективності пропозицій роботи*	11.05- 31.05.24	
6.	*Завершення роботи та перевірка на плагіат*	03.06 - 10.06.24	

Студент _____ Костромської І. О.
(підпис)

Керівник роботи _____ Чаплига В. М.

АНОТАЦІЯ

УДК 681.5:635.1

Костромською І. О. Автоматизація процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу : кваліфікаційна робота. Дубляни: Львівський національний університет природокористування, 2024. _____ с.

Табл. _____; рис. _____; бібліогр. джерел _____.

Ключові слова: процес, автоматизація, заповнення пропусків, таблиці даних, підприємство, АПК.

В умовах суттєвої невизначеності в роботі сучасного агропромислового комплексу (АПК) актуальною для підприємств АПК є тема, що пов'язана з підвищенням якості та повноти попередньої обробки даних щодо усіх аспектів їх функціонування. Вирішення задачі заповнення пропусків в таблицях даних дозволяє компаніям АПК підвищити якість функціонування та підвищити їх загальну продуктивність.

Мета роботи полягає в удосконаленні методів автоматизації процесів заповнення пропусків в таблицях даних підприємств АПК на основі новітніх інформаційних технологій.

У даній роботі на основі здійсненого аналізу методів і відповідних технологічних процесів заповнення пропусків в таблицях даних підприємств АПК показано, що найбільш перспективними та ефективними є методи інтелектуальної попередньої підготовки даних для подальшої їх обробки корпоративними інформаційними системами.

Досліджено сучасні нейромережні технології заповнення пропусків в таблицях даних. Розроблено функціонал, структурні схеми, алгоритми та програмне забезпечення автоматизації процесу заповнення пропусків в таблицях даних підприємств АПК з використанням сучасних інтелектуальних інформаційних технологій.

Розроблено функціонал, структурні схеми, алгоритми та програмне забезпечення автоматизації процесу заповнення пропусків в таблицях даних підприємств АПК.

Здійснено вибір інтерфейсу автоматизації процесу заповнення пропусків в таблицях даних підприємств АПК.

Практична цінність виконаної роботи полягає у можливості використання її результатів підприємствами у різних галузях АПК для автоматизації процесу заповнення пропусків в таблицях даних, що дозволяє підвищити якість підготовки даних до їх обробки в корпоративних інформаційних системах.

Розраховано економічну ефективність автоматизації технологічних процесів заповнення пропусків в таблицях даних підприємств агропромислового комплексу. Розглянуті питання та заходи щодо охорони праці та безпеки в надзвичайних ситуаціях для приміщень з експлуатації корпоративних інформаційних систем автоматизації технологічних процесів заповнення пропусків в таблицях даних підприємств АПК.

SUMMARY

UDC 681.5:635.1

Kostromskaya I. O. Automation of the process of filling in the gaps in the data tables of agricultural enterprises : qualification work. Dubliany: Lviv National University of Environmental Management, 2024. _____ c.

Table. _____; Fig. _____; bibliography _____.

Keywords: process, automation, gap-filling, data tables, enterprise, agro-industrial complex.

In the context of significant uncertainty in the work of the modern agro-industrial complex (AIC), the topic related to improving the quality and completeness of preliminary data processing on all aspects of their functioning is relevant for AIC enterprises. Solving the problem of filling in gaps in data tables allows agricultural companies to improve the quality of their operations and increase their overall productivity.

The aim of the work is to improve the methods of automating the processes of filling in gaps in the data tables of agricultural enterprises based on the latest information technologies.

This paper, based on the analysis of methods and corresponding technological processes for filling in gaps in the data tables of agricultural enterprises, shows that the most promising and effective are the methods of intelligent preliminary preparation of data for further processing by corporate information systems.

Modern neural network technologies for filling gaps in data tables are studied. Functionality, structural schemes, algorithms and software for automating the process of filling gaps in the data tables of agro-industrial complex enterprises using modern intelligent information technologies have been developed.

Functionality, structural diagrams, algorithms and software for automating the process of filling gaps in the data tables of agro-industrial complex enterprises have been developed.

The selection of an interface for automating the process of filling gaps in the data tables of agro-industrial complex enterprises has been made.

The practical value of the performed work lies in the possibility of using its results by enterprises in various branches of the agricultural industry to automate the process of filling gaps in data tables, which allows to improve the quality of data preparation for their processing in corporate information systems.

The economic efficiency of automation of technological processes of filling gaps in data tables of enterprises of the agro-industrial complex is calculated. Considered issues and measures regarding labor protection and safety in emergency situations for the premises for the operation of corporate information systems for the automation of technological processes of filling gaps in the data tables of agro-industrial complex enterprises.

ПЕРЕЛІК СКОРОЧЕНЬ

САЗПТД - система автоматизації заповнення пропусків в таблицях даних;

Mean - Mean substitution) метод (заповнення пропусків даних в таблицях середнім значенням відповідного параметру;

MICE - (Multivariate Imputation by Chained Equations, метод заповнення пропусків даних в таблицях декількома способами з наступним об'єднанням результатів ланцюжковими рівняннями при заповненні пропущеного значення;

KNN - (K-nearest neighbor method) метод К-найбли́жчих сусі́дів, коли вибирають певну кількість сусідніх з пропущеним значенням елементів таблиці і за допомогою регресії визначають відсутнє значення;

FCM - (Fuzzy C-means) метод видалення рядків з пропусками і заповнення їх статистичною вибіркою з урахуванням нечіткої функції залежності між елементами;

SVD - (Singular value decomposition) метод сингулярного розкладу таблиці даних є доволі складним, але дозволяє побачити геометричну структуру таблиці і представити наявні дані;

bPCA - (Bayesian principal component analysis) метод баєсовського принципу вимагає великої кількості даних, але з більшою точністю ніж PCA метод визначає значення пропущених даних.

НМ – нейронні мережі;

ШНМ – штучні нейронні мережі;

ПЗ - програмне забезпечення.

ЗМІСТ

ВСТУП

Заповнення пропусків у таблицях даних підприємств агропромислового комплексу є надзвичайно актуальним завданням, яке має кілька важливих аспектів. Це пов'язано з тим, що якісні та повні дані відіграють ключову роль у прийнятті рішень, плануванні, аналітиці та оцінці ефективності аграрного бізнесу.

Так, для точного планування виробництва необхідно мати повну інформацію про попередні врожаї, погодні умови, використані добрива та засоби захисту рослин. Також, таблиці інтегрують дані з різних джерел, наприклад, від постачальників добрив, насіння та інших ресурсів; з супутників і дронів, використовуючи дані дистанційного зондування про стан полів і врожайність тощо. Повні дані дозволяють:

- оптимізувати та ефективніше управляти ресурсами, такими як вода, добрива, паливо, що сприяє зниженню витрат і підвищенню врожайності продукції:
- краще здійснювати моніторинг та контроль продуктивності полів, відстежувати зміни у врожайності та швидко виявляти проблеми.
- контролювати якість продукції на різних етапах виробництва, від посіву до збору врожаю;
- на основі повних історичних даних виявляти тенденції у виробництві продукції та робити прогнози щодо майбутніх врожаїв;
- прогнозувати ризики, що пов'язані з погодними умовами, шкідниками або хворобами, та завчасно вживати превентивних заходів;
- підвищувати якість складання реалістичних бюджетів і фінансових планів;
- приймати обґрунтовані рішення щодо інвестицій у нові технології та розширення виробництва;

- забезпечувати відповідність формам та стандартам звітування аграрних підприємств регуляторним вимогам тих країн, де провадиться господарська діяльність

- забезпечення повних даних про всі етапи виробництва для отримання сертифікатів якості або органічності продукції.

Використання сучасних методів статистичного аналізу та інтелектуальної обробки даних дозволяє забезпечити високу якість і повноту даних, що сприяє підвищенню продуктивності та зниженню ризиків у сільському господарстві.

Об'єкт дослідження – процеси автоматизації аналізу та обробки даних в науковій та виробничій діяльності.

Предмет дослідження - автоматизація процесу заповнення пропусків в таблицях даних господарської діяльності підприємств АПК.

Мета дослідження – розробка підходів та методики автоматизації процесу заповнення пропусків в таблицях даних господарської діяльності підприємств АПК.

Для досягнення визначеної мети були поставлені та рішені наступні задачі:

- Аналіз основних видів таблиць в підприємствах АПК.
- Аналіз задач заповнення пропусків в таблицях даних.
- Аналіз методів заповнення пропусків в таблицях даних та їх автоматизація.
- Класифікація та дослідження типів пропущених даних та методів їх заповнення в таблицях.
- Дослідження особливостей, алгоритмів та програмного забезпечення поширених методів заповнення пропусків в таблицях даних.
- Дослідження альтернативних традиційним методам заповнення пропусків в таблицях даних
- Розробити автоматизації процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу

- Розрахувати економічну ефективність автоматизації процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу.
- Визначити заходи з з охорони праці та безпеки в надзвичайних ситуаціях в приміщеннях з експлуатації системи автоматизації процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу.

Практичне значення отриманих результатів

Результати кваліфікаційної роботи можуть бути використані в різних сферах: як в інтелектуальній обробці наукових експериментальних даних, так і в рішенні практичних задач попередньої обробки даних для управління підприємствами різних галузей АПК в умовах використання корпоративних інформаційних ERP систем.

РОЗДІЛ 1. АНАЛІЗ ВИДІВ ТАБЛИЦЬ В ПІДПРИЄМСТВАХ АПК ТА АВТОМАТИЗАЦІЇ ПРОЦЕСУ ЗАПОВНЕННЯ ПРОПУСКІВ ДАНИХ В НИХ

1.1. Аналіз основних видів таблиць в підприємствах АПК

В управлінні підприємством агропромислового комплексу використовуються різні види таблиць даних, які допомагають ефективно організовувати, аналізувати та контролювати різноманітні аспекти його діяльності. Основні типи таких таблиць зведені в табл. 1.1.

Таблиця 1.1. Основні види таблиць даних підприємств різних галузей АПК.

Назва таблиці даних	**Зміст таблиці даних**
Головна книга	Дані всіх фінансових транзакцій підприємства
Бухгалтерські рахунки	Дані деталізації доходів, витрат, активів і зобов’язань
Податкові таблиці	Дані обліку податкових зобов'язань і платежів
Облік доходів і витрат	Дані для аналізу фінансових потоків підприємства
Облік основних засобів	Дані про фізичні активи підприємства (будівлі, обладнання)
Облік складських запасів	Дані про наявні матеріали, сировину та готову продукцію (насіння, добрива, паливо тощо).
Облік руху товарів	Вхідні та вихідні дані руху матеріалів і продукції на складі, обліку використання добрив і пестицидів
Планування виробництва	Дані для планування виробничих процесів і графіків (посівних робіт, збору врожаю тощо)
Облік виробничих процесів	Дані про виконання виробничих замовлень і операцій
	Дані обліку врожайності (з полів, теплиць тощо).
	Дані обліку роботи сільськогосподарської техніки
Облік працівників	Дані про всіх співробітників підприємства
Облік робочого часу	Дані про відпрацьований час, відпустки та лікарняні
Облік заробітної плати	Дані про нарахування та виплату заробітної плати.
Управління закупівлями	Дані про постачальників матеріалів і послуг
	Дані про замовлення матеріалів і сировини
	Дані про отримані матеріали та продукцію від постачальників

Продажі та маркетинг	Дані про клієнтів і партнерів
	Дані про замовлення від клієнтів
	Інформація про відвантажену продукцію і доставку
	Дані про маркетингові кампанії та продажі
Управління проєктами	Дані про всі проєкти, їхній статус і ключові етапи
	Дані про витрати, пов'язані з реалізацією проектів
	Дані про заплановані і фактичні терміни виконання проектів
CRM (Customer Relationship Management)	Дані контактів з контрагентами та клієнтами. Дані про комунікації з клієнтами (дзвінки, зустрічі, листування).
Моніторинг та контроль якості	Дані результатів лабораторних аналізів
	Дані обліку відповідності продукції стандартам якості
Управління ризиками	Дані обліку страхових випадків
	Дані аналізу кліматичних ризиків
	Дані аналізу ринкових ризиків

Ці таблиці забезпечують централізоване управління даними, полегшуючи доступ до інформації та підтримуючи прийняття управлінських рішень.

Як наголошують науковці і практики, повнота даних є критично важливим аспектом якості даних, особливо в контексті діяльності підприємства в умовах функціонування корпоративних ERP систем управління підприємством. Так, С. Разневський [1] підкреслює важливість повних даних для забезпечення повноти відповідей на запит, що є ключовою вимогою для прийняття точних рішень.

Це додатково підтверджено В. Падманабханом [3], який демонструє значні переваги моделювання, які можна отримати від повної інформації, особливо в сферах лояльності клієнтів та інтенсивності перегляду. Н. Овсюк [4] підкреслює роль фінансової звітності як основного джерела інформації про діяльність підприємства, наголошуючи на необхідності повноти цих звітів для підтримки прийняття обґрунтованих рішень. Ю. Цай [5] надає теоретичну основу для оцінки повноти даних як виміру якості даних, визнаючи суб’єктивну інтерпретацію, необхідну в динамічних середовищах прийняття рішень. Представлені вище дослідження в сукупності підкреслюють критичну роль

повноти даних у підтримці прийняття точних і обґрунтованих рішень у діяльності підприємства.

Зауважимо, що таблиці можуть бути створені та оброблятись як в електронному вигляді (з використанням програмного забезпечення, наприклад, Microsoft Excel, Google Sheets, або спеціалізованих ERP систем для агробізнесу), так і в паперовому форматі залежно від потреб і можливостей підприємства.

1.2. Аналіз задач заповнення пропусків в таблицях даних

Задачі заповнення пропусків у таблицях даних виникає в багатьох ситуаціях аналізу, класифікації та прогнозування інформаційних об’єктів різного походження та природи, які функціонують в умовах невизначеності, особливо економіко-соціальної, фінансової, технологічної. Здійснений аналіз показує до них відносяться, зокрема, такі:

- аналіз даних та бізнес-аналітика (BI), коли не повні дані можуть ускладнювати виявлення закономірностей і трендів;
- наука і дослідження в АПК, наприклад, дослідження та прогнозування фінансової стійкості підприємства, медичні дослідження персоналу для встановлення діагнозу хворому, соціологічні опитування тощо;
- кредитний скоринг і управління ризиками, коли є пропуски у фінансовій інформації про клієнтів, такі як історія платежів або доходи;
- інвестиційний аналіз, коли частково відсутні дані про ринкові ціни або у фінансових звітах підприємств АПК;
- моніторинг обладнання на підприємстві, коли є пропуски в показниках сенсорів через технічні збої або нерегулярні вимірювання;
- контроль якості при неповних даних про партії продукції або результати тестів якості;
- аналіз логістичних маршрутів в АПК, коли відсутні дані про переміщення транспортних засобів або їх затримки;
- моніторинг доставки матеріалів або готової продукції з пропусками в даних про доставлені вантажі або тривалість доставок;

- аналіз поведінки клієнтів електронної комерції підприємств АПК при частковій відсутні даних про дії клієнтів на веб-сайті, такі як перегляди товарів або покупки;
- інвентаризація в підприємствах АПК при не повних даних про залишки товарів або їх поставки;
- моніторинг здоров'я або аналіз спортивних результатів персоналу АПК, коли є пропуски в даних про фізичну активність, тренувальні плани або біометричні показники;
- моніторинг споживання енергії підприємством АПК при частковій відсутності даних через збої у вимірювальних приладах або при прогнозуванні через пропуски в історичних даних про споживання енергії;
- моніторинг середовища кіберфізичних систем АПК при застосуванні Інтернету речей (IoT), коли можливі пропуски в даних з сенсорів, які вимірюють температури, вологість тощо,
- кластерний аналіз даних з різних підприємств АПК для порівняння їх діяльності та виокремлення груп з приблизно однаковими показниками;
- машинне навчання і штучний інтелект при наявності пропусків в навчальних наборах даних. що може негативно вплинути на точність моделей.

Таким чином, проблема заповнення пропусків є поширеною і може виникати в будь-якій галузі, де використовуються дані. Відсутність даних може бути обумовлена різними причинами, такими як технічні збої, людські помилки або нерегулярність вимірювань. Ефективне вирішення цієї проблеми є критичним для забезпечення якості аналізу та прийняття рішень на основі даних.

Як відомо з літературних джерел, задача заповнення пропусків у таблиці даних є некоректною задачею, тобто задачею, яка не має єдиного рішення. Існує багато способів вирішення подібних завдань, і конкретний метод буде залежати від характеристик даних таблиці та індивідуальних знань та навиків дослідника в цій царині. Тож у багатьох випадках можна лише сподіватися «розумно» заповнити прогалини в таблиці даних.

1.3. Аналіз методів заповнення пропусків в таблицях даних та їх автоматизація.

Вітчизняними та закордонними авторами було запропоновано ряд методів для заповнення пропусків у таблицях даних. Зокрема, П. Бідюк [5] пропонує багатоетапний підхід, включаючи виявлення пропусків, дослідження їх закономірностей та використання різних методів генерації даних. П. Ткаченко [6] запропонував нейромережний підхід до автоматизації процесу заповнення пропущених в таблицях даних. М. Мацута Г. Ксує-донг [7] представляє метод, заснований на грубих наборах, який є ефективним для правил прийняття рішень. Р. Камінський [8] обговорює характеристики різних методів заповнення пропусків в таблицях даних, наводячи приклади їх застосування. К. Устоорікар [9] зосереджується на заповненні пропусків у хвильових даних за допомогою генетичного програмування, яке було визнано ефективним, особливо для малої кількості пропусків даних. Ці дослідження разом підкреслюють важливість усунення розривів у таблицях даних і пропонують різноманітні підходи для цього.

Результати здійсненого нами аналізу методів заповнення пропусків в таблицях даних та їх автоматизації наведено в табл. 1. 2.

Таблиця 1.2.

Аналіз методів заповнення пропусків в таблицях даних та їх автоматизації

Назва	Сутність	Переваги	Недоліки
Mean Imputation	Заповнення середнім значенням	Легко зрозуміти та реалізувати. Швидко виконується навіть на великих наборах даних	Не враховує кореляції між змінними Систематична помилка може призвести до заниження або завищення змінності даних

Median Imputation	Заповнення медіаною	Стійкість до викидів у даних. Легко зрозуміти та реалізувати.	Не враховує кореляції між змінними. Може призвести до заниження або завищення змінності даних
Mode Imputation	Заповнення модою	Підходить для категорійних даних: Особливо корисно для обробки пропусків у категорійних змінних.	Не враховує кореляції між змінними, що може призвести до завищення частоти появи найбільш частого значення.
Constant Value Imputation	Заповнення константою, наприклад, 0 або "невідомо".	Дуже легко реалізувати. Можна обрати значення, яке має сенс в контексті даних.	Константа може не відображати дійсні значення, що призводить до втрати інформації. Може призвести до викривлення розподілу даних.
K-Nearest Neighbors Imputation, KNN Imputation	Заповнення методом K найближчих сусідів	Враховує схожість між об'єктами для заповнення пропусків. Підходить для числових і категорійних даних.	Вимогливий до обчислювальних ресурсів, особливо для великих наборів даних. Потребує налаштування параметра k (кількість сусідів).
Multivariate Imputation by Chained Equations, MICE	Мультиваріативна імпутація за ланцюговими рівняннями	Ітеративний підхід дозволяє враховувати багатовимірні взаємозв'язки між змінними. Може використовувати різні моделі для різних типів даних	Вимогливий до обчислювальних ресурсів та часу. Складність реалізації через потребу більше налаштувань і розуміння методології.
Regression Imputation	Регресійна імпутація	Враховує взаємозв'язки та використовує доступні змінні для	Вимогливий до обчислювальних ресурсів.

		передбачення пропущених значень. Використовує різні типи регресійних моделей.	Може призвести до завищених значень R^2 та недооцінки варіабельності.
Expectation-Maximization Imputation, EM Imputation	Імпутація методом очікування–максимізації	Врахування невизначеності при ймовірнісному підході для заповнення пропусків.	Вимогливий до обчислювальних ресурсів. Вимагає розуміння методів оптимізації.
Time Series Imputation	Імпутація часового ряду	Враховує послідовність даних, що важливо для часових рядів.	Просте заповнення може ігнорувати тренди і сезонні коливання в даних.
Random Forest Imputation	Імпутація за допомогою випадкових лісів (дерев прийняття рішень)	Враховує складні взаємозв'язки між змінними	Вимогливий до обчислювальних ресурсів та налаштувань моделі.
Bayesian Network Imputation	Імпутація за допомогою байєсових мереж	Використовує ймовірнісні моделі для врахування залежностей між змінними.	Вимогливий до обчислювальних ресурсів. Складність реалізації.
Cluster-Based Imputation	Імпутація за допомогою кластерного аналізу	Використовує інформацію з подібних кластерів для заповнення пропусків.	Потребує оптимального вибору кількості кластерів. Обчислювальна складність
Ensemble Methods Imputation	Імпутація за допомогою ансамблевих методів	Комбінує кілька методів для покращення результатів імпутації.	Потребує налаштування і поєднання кількох методів. Складність реалізації обчислень.
Shrinkage Imputation	Імпутація методом стислого середнього	Зменшує зміщення в оцінках шляхом стиску значень.	Потребує розуміння методів стиску. Менш ефективний для великих масивів.
Least Squares Imputation	Імпутація методом	Використовує методи найменших квадратів для	Може призвести до завищених значень

	найменших квадратів	передбачення пропущених значень.	R^2 та недооцінки варіабельності.

Аналіз показує, що кожен з методів заповнення пропусків в таблицях даних має свої переваги та недоліки. Вибір методу залежить від конкретних умов, типу даних та вимог до точності.

Автоматизація дозволяє зробити процес заповнення пропусків даних більш ефективним та якісним: прискорює процес обробки великих наборів даних; забезпечує однаковий підхід до обробки пропусків, мінімізуючи людські помилки; легко застосовується до різних масивів даних без потреби ручної обробки. Але при цьому треба враховувати наявні обчислювальні ресурси та складність реалізації методу з врахуванням у загальних алгоритмах всіх специфічних аспектів даних, а також складність налаштування і тюнінгу (точного визначення) параметрів для оптимальної автоматизованої обробки. Зазначимо, що автоматизовані методи можуть не впоратися з сильно пошкодженими або неоднорідними даними.

РОЗДІЛ 2. ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ ОСНОВНИХ МЕТОДІВ ЗАПОВНЕННЯ ПРОПУСКІВ В ТАБЛИЦЯХ ДАНИХ ТА ЇХ АВТОМАТИЗАЦІЇ

2.1. Класифікація та дослідження типів пропущених даних та методів їх заповнення в таблицях

Пропущені дані – це термін, який відноситься як до даних, відсутність яких обумовлена взагалі неможливістю їх отримати, так і до даних, які існують але їх відсутність в таблицях обумовлена зовнішніми чинниками (вимірювальними пристроями, ненавмисним або навмисним невнесенням окремих даних в таблиці).

Вчені пропонують три групи методів роботи з пропущеними даними (див. рис. 2.1).

Рис. 2.1 – Групи методів роботи з таблицями з пропусками даних

Перша група методів легко реалізується, але є не ефективною при великій кількості пропущених даних із-за сильних зміщень результатів. Друга група методів потребує задання початкових значень вагових коефіцієнтів. Тоді вибірку розбивають на підгрупи, множать їх елементи на визначені ваги, визначають величини відповідей для кожної підгрупи, порівнюють їх і розраховують вагові коефіцієнти для кожного з пропущених даних.

При аналізі пропусків даних вчені розрізняють два типи пропущених даних. Перший, коли ймовірність пропуску даних залежить від інших значень в масиві, а другий – коли не залежить. При цьому, може бути встановлено три види залежності між відсутнім значенням Xb та вимірюваним значенням Xv (див рис. 2.2).

Пропуски даних можуть мати різні причини, зокрема, включають неправильна обробка зразків, низьке співвідношення сигнал/шум, помилки вимірювання, відсутність відповіді або віддалені викиди.

Рубін в [10] визначив відсутні дані на основі трьох механізмів відсутності: повністю випадкових, коли ймовірність того, що екземпляр (випадок) має відсутнє значення для змінної, не залежить від відомого значення відсутніх даних; або відсутні дані.

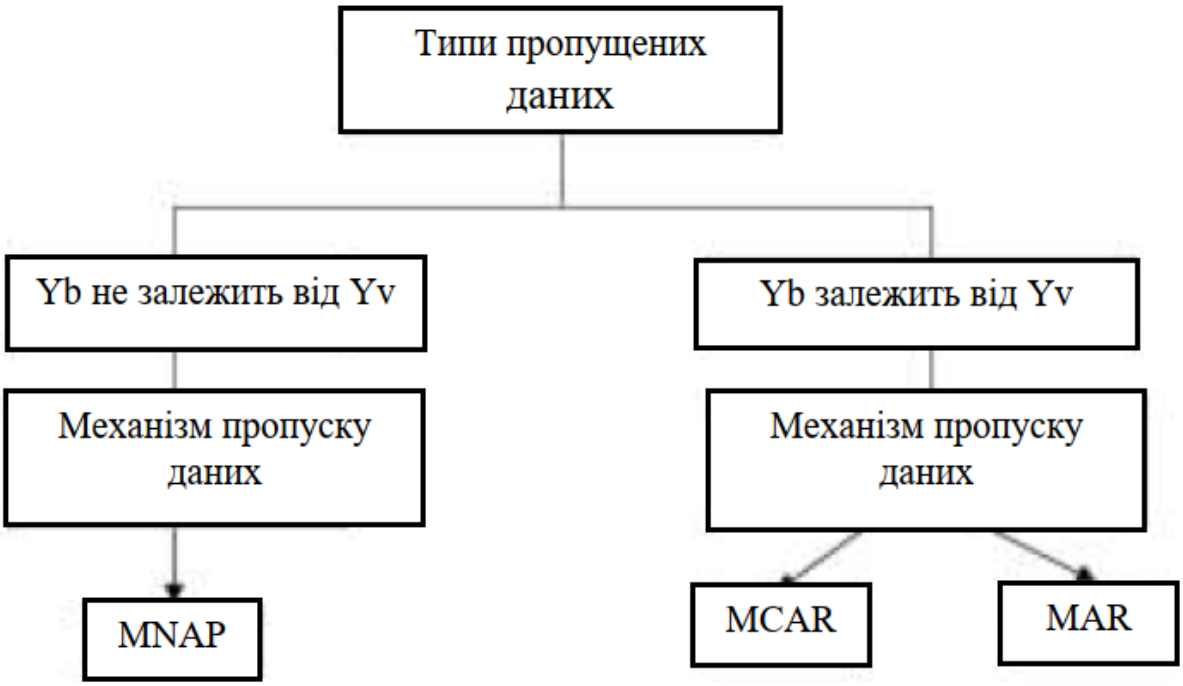


Рис. 2.2. Класифікація типів та механізмів пропуску даних.

Перший тип пропущених даних MCAR характеризується повною випадковістю відсутності значення Xb в множині даних таблиці.

Другий тип даних MAR характеризується тим, що значення Xb може бути відсутнім тільки для атрибуту, який не залежить від виміряних значень Xv в таблиці. Дані випадково відсутні (MAR), коли ймовірність елемента з відсутнім значенням для змінної залежить від відомого значення, але може не залежати від значення самих відсутніх даних.

Третій тип даних MNAR характерний для таблиці, де ймовірність пропущених даних Xbi залежить від вимірюваних значень Xvj. Дані є

випадковими (MNAR), якщо ймовірність того, що елемент має відсутнє значення для змінної, може залежати від значення цієї змінної.

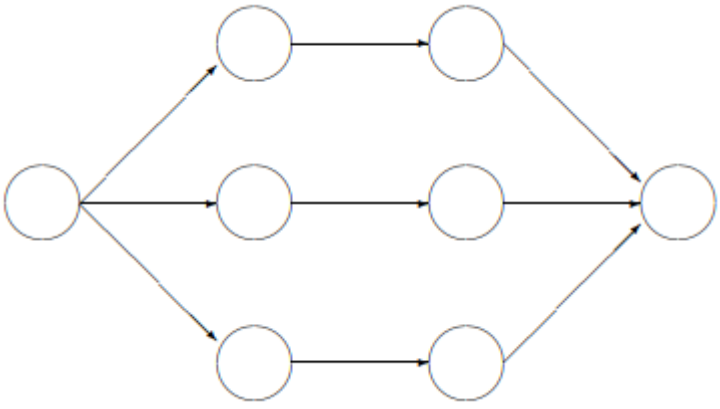
Тип та механізми пропуску даних впливають на вибір методу заповнення пропусків.

Імпутація (заповнення прогалин у даних) — це універсальний і гнучкий спосіб роботи з відсутніми даними. Це засіб прогнозного розподілу відсутніх значень, який потребує способу створення прогнозного розподілу на основі виміряних даних для заміни відсутніх значень. Методи створення повних даних шляхом заповнення пропущених значень можна розділити на:

- одиночну імпутацію;
- багатовимірну імпутацію.

Одиночний метод імпутації визначається як заповнення точно одного значення для кожного відсутнього значення.

Множинне (багатовимірне) імпутування — це метод генерації кількох змодельованих значень для кожного відсутнього Ybi, щоб правильно відобразити невизначеність, пов’язану з відсутніми даними (див.рис. 2.3).



Неповні дані - Враховані дані - Результати аналізу - Зведені результати

Рис. 2.3 – Етапи аналізу пропущених даних при множинному імпутуванні.

Багаторазове імпутування створює n > 1 повних наборів даних для заповнення пропусків в таблицях. Кожен із цих наборів обробляється стандартним ПЗ для аналізу. Результати *n* об’єднуються в кінцеву точну оцінку

та стандартну помилку за допомогою простого правила комбінування («правило Рубіна» [10]).

Перший етап аналізу починається з визначення спостережуваних і пропущених даних. При багаторазовому імпутуванні створюється кілька повних версій табличних даних шляхом заміни відсутніх значень дійсними значеннями даних. Ці близькі до реальних значення отримані з розподілів, які спеціально моделюються для кожного відсутнього набору даних.

На рисунку 2.3 показаний набір даних при n = 3, хоча на практиці n часто може бути набагато більшим. Число n = 3 нами використовується тут лише для того, щоб підкреслити, що ця техніка створює кілька версій записаних даних. Три вписані в таблиці набори даних ідентичні спостережуваному набору даних, але з різними вписаними значеннями. Величина цієї різниці відображає невизначеність щодо того, яке значення призначити.

Другим кроком є оцінка параметрів, що цікавлять, з кожного врахованого набору даних шляхом застосування аналітичних методів, які можна використовувати, якщо дані повні. Оскільки вхідні дані різні, результати також різні. Важливо визнати, що ці відмінності виникають лише через невизначеність щодо того, яке значення призначити.

Останнім кроком є об’єднання n оцінок параметрів в одну оцінку та здійснити оцінку її дисперсії. В нашому випадку, дисперсія – це комбінація нормальної дисперсії вибірки (в межах дисперсії імпутації), додаткової дисперсії, спричиненої відсутніми даними, та додаткової дисперсії, спричиненої відсутністю даних (між імпутацією) і за відповідних умов зведені оцінки є неупередженими та мають належні статистичні властивості.

Зазаначимо також, що існує два загальні методи генерації прогнозованих розподілів відсутніх значень, а саме явне та неявне моделювання. Явне моделювання ґрунтується на прогнозованому розподілі формальної статистичної моделі (наприклад, багатовимірного нормального), тому припущення є явними.

Методи включають інтерполяцію середнього значення, інтерполяцію регресії та інтерполяцію стохастичної регресії. Усереднення приносить значення

з масиву донорів замість відсутніх даних. Регресійне імпутування — це метод заміни відсутніх значень на прогнозоване регресійним значенням відсутнього елемента в одиничному спостережуваному елементі. Кінцевою формою явного моделювання є імпутація стохастичної регресії, яка замінює відсутні спостереження значеннями, передбаченими регресійною імпутацією, плюс залишки, які відображають невизначеність у прогнозованих значеннях.

Традиційно використовуються наступні методи заповнення - усереднення, обнулення, імпутації пропущених значень при математичній обробці, окремі регресійні моделі. Вважаємо, що там, де застосовуються методи регресії, слід очікувати найвищої якості результатів, Однак це можливо, лише якщо дані таблиці пов’язані один з одним і дотримано багато обмежень щодо кількості втрачених даних. Зокрема, потрібно дотримуватись певного співвідношення між кількістю рядків з пропусками та без них тощо.

У випадках, коли елементи таблиці не залежать один від одного, можна застосувати лише один із перших трьох методів, перерахованих вище. У більшості випадків якість такої процедури занадто низька, але в цьому випадку іншого виходу немає. Тому розглянемо їх особливості.

Рамлі в [11] наголошує на важливості врахування довжини розриву в даних, причому для коротких проміжків підходять методи одноразового обчислення, а для більших розривів — багаторазові обчислення.

Мохаммед в [12] порівняв різні методи імпутації відсутніх даних, зокрема, імпутації середнього значення, медіанної імпутації, k найближчих сусідів, вибіркової імпутації та кількох імпутацій за допомогою ланцюжкових рівнянь (MICE). П’ять методів імпутації порівнюються з використанням чотирьох реальних наборів даних. Дев'ять різних відсотків відсутності вводяться в набори даних абсолютно випадковим чином. Для оцінки ефективності методів використовується статистичний показник, середньоквадратична помилка (RMSE). Результати показують, що численні імпутації за допомогою ланцюжкових рівнянь (MICE) перевершують інші методи імпутації. Середнє значення та k найближчого сусіда (KNN) показали кращі результати порівняно з

методами вибірки та медіанного імпутації. Ефективність п’яти методів імпутації не залежить від набору даних і відсотка пропусків. Таким чином, множинні умовні оцінки перевершують середню умовну оцінку, яка, у свою чергу, ефективніша, ніж середня умовна оцінка.

2.2. Особливості, алгоритми та програмне забезпечення поширених методів заповнення пропусків в таблицях даних

Класифікація алгоритмів заповнення пропусків даних показана на рис. 2.4.

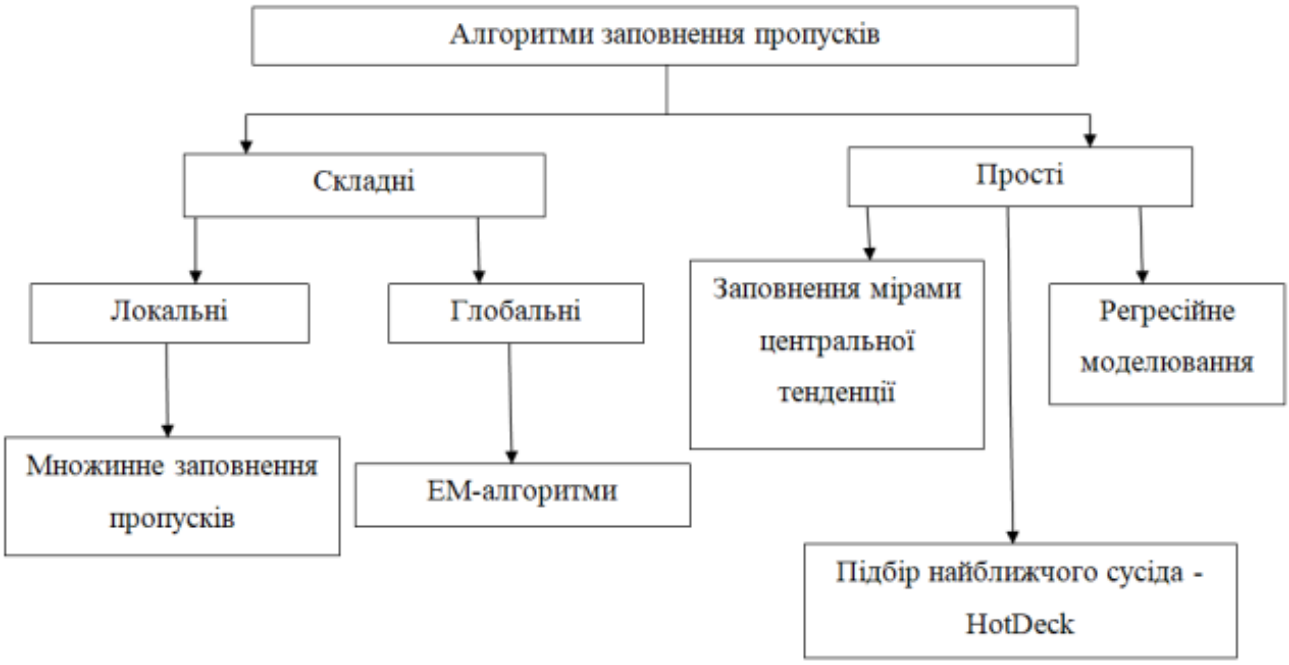


Рис. 2.4 - Класифікація алгоритмів заповнення даних

Нехай маємо таблицю об’єктів-ознак Y^{n*m}, де n – кількість стрічок даних (о’єктів), m – кількість стовпців (ознак) даних. В таблиці частина значень відсутня. Приклад таблиці даних наведено на рис. 2.1.

Найпростіший спосіб вирішити проблему обробки відсутніх значень у таблиці об’єктів — це видалити об’єкти (тобто рядки в таблиці ознак), які мають прогалини в даних. Перевагою цього методу є його простота і неможливість спотворити дані шляхом заміни пробілів. Цей метод використовується лише тоді, коли невелика частина об’єктів вибірки має відсутні значення. Недоліком цього методу є втрата інформації під час виключення об’єктів.

Таблиця 2.1.

Приклад матриці (Об’єкт – Атрибут) з пропусками даних.

Objects	**Attributes**					
	Height	**Length**	**Width**	. . .	**Weight**	**Volume**
Object Y_1	h_1	d_1	l_1	⋯	m_1	v_1
Object Y_2	h_2	∅	l_2	⋯	m_2	v_2
. . .	⋯	⋯	⋯	⋯	⋯	⋯
Object Y_{n-1}	h_{n-1}	d_{n-1}	∅	⋯	m_{n-1}	∅
Object Y_n	h_n	∅	l_n	⋯	m_n	∅

Для зменшення вибіркової оцінки при відборі об'єктів без пропусків використовується метод, який присвоює кожному об'єкту певну вагу перед обробкою даних. Якщо невелика кількість ознак відсутня, альтернативою є видалення таких ознак (стовпців матриці ознак) із вибірки. Алгоритм видалення об’єктів або ознак, що мають пропуски даних представлено на рис. 2.5.

Одним із найпростіших способів заповнити прогалини є заповнення відсутніх даних вибірковим середнім для кожної характеристики. Вибіркове середнє – це математичне очікування емпіричного розподілу. Емпіричний розподіл — це розподіл ймовірностей, визначений за допомогою вибірки для оцінки справжнього розподілу F.

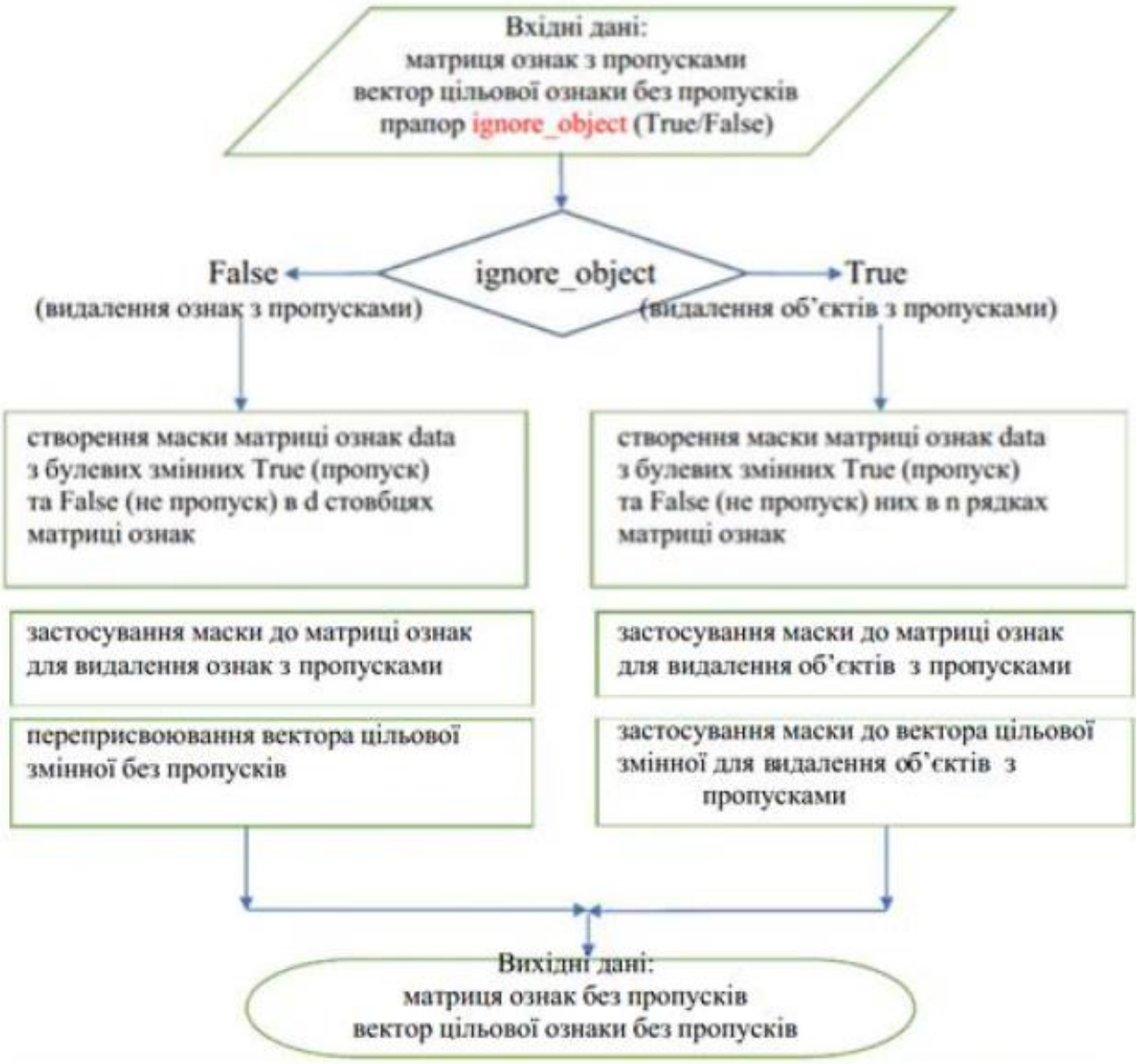


Рис. 2.5 – Алгоритм видалення об'єктів (тобто рядків в таблиці ознак) або ознак (стовпців), які мають пропуски в даних.

В цьому методі спостереження *Xi , i= [1, n]* є незалежними і однаково розподіленими випадковими величинами з функцією розподілу *F(x)*, і нехай *Xn* є флуктуаційним рядом *Xn*. Емпіричний розподіл називається дискретним розподілом, у якому кожному значенню *Xi* присвоюється ймовірність *1/n*. Емпірична функція розподілу називається ступінчастою функцією зі стрибками, кратними *1/n*. Емпіричне (вибіркове) середнє — це неупереджена та надійна оцінка математичного очікування *F*-розподілу.

Алгоритм методу Mean - заповнення пропусків даних в таблицях середнім значенням відповідного параметру представлено на рис. 2.6, а принцип роботи програми показано на рис. 2.7.

Рис. 2.6 - Алгоритм методу (Mean substitution, Mean) заповнення пропусків даних в таблицях середнім значенням відповідного параметру.

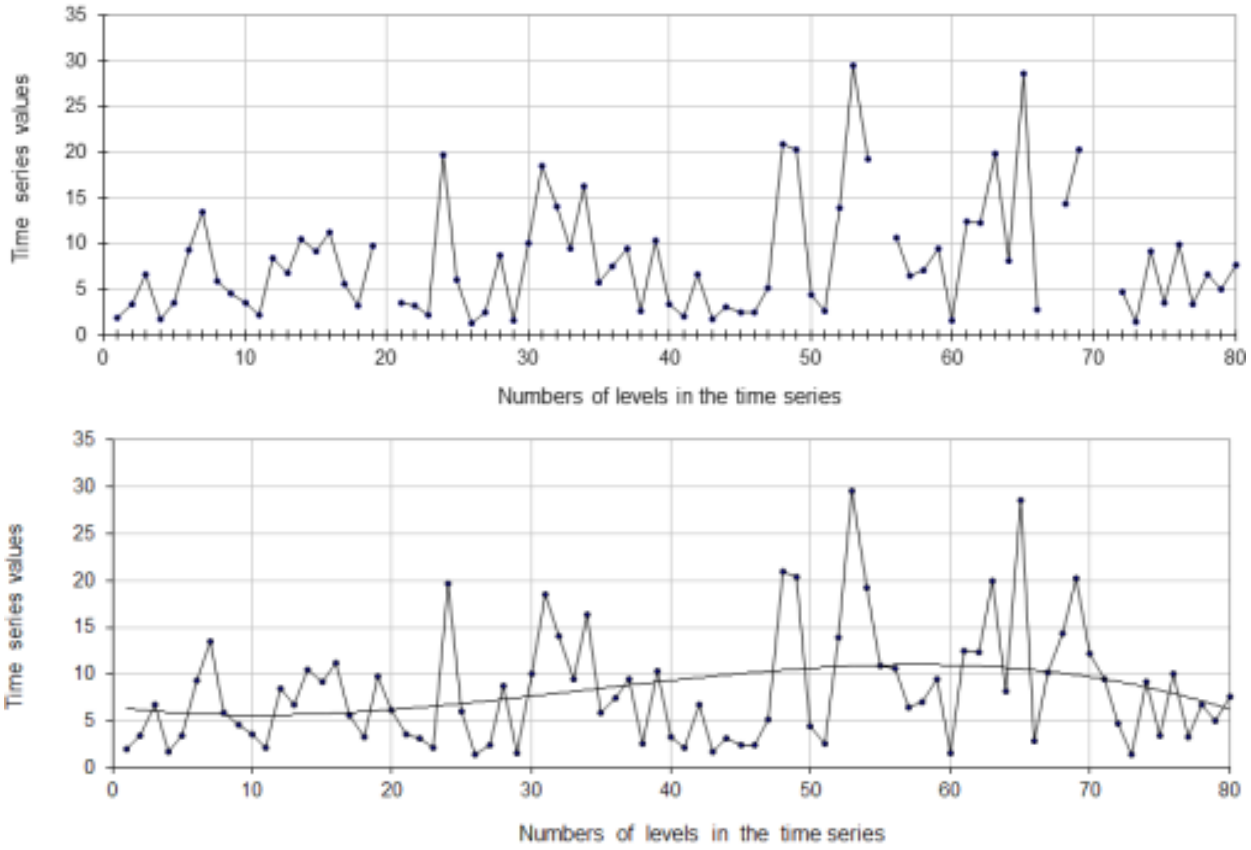


Рис. 2.7 - Результати заповнення пропусків за методом (Mean substitution, Mean) середнім значенням відповідного параметру.

Для реалізації SVD-методу (Singular value decomposition) написана програма в MATLAB

Синтаксис

s = svd(X)

[U,S,V] = svd(X)

[U,S,V] = svd(X,0)

Опис

Команда svd обчислює розкладання таблиці за сингулярним значенням.

s = svd(X) returns a vector of singular values.

[U,S,V] = svd(X) produces a diagonal matrix S of the same dimension as X, with nonnegative diagonal elements in decreasing order, and unitary matrices U and V so that X = U*S*V'.

[U,S,V] = svd(X,0) produces the "economy size" decomposition.

If X is m-by-n with m > n, then svd computes only the first n columns of U and S is n-by-n.

Для таблиці, наприклад:

X =

1 2

3 4

5 6

7 8

the statement [U,S,V] = svd(X)

produces

U =

-0.1525 -0.8226 -0.3945 -0.3800

-0.3499 -0.4214 0.2428 0.8007

-0.5474 -0.0201 0.6979 -0.4614

-0.7448 0.3812 -0.5462 0.0407

S =

14.2691 0

0 0.6268

0 0

0 0

V =

-0.6414 0.7672

-0.7672 -0.6414

The economy size decomposition generated by

[U,S,V] = svd(X,0)

produces

U =

-0.1525 -0.8226

-0.3499 -0.4214

-0.5474 -0.0201

-0.7448 0.3812

S =

14.2691 0

0 0.6268

V =

-0.6414 0.7672

-0.7672 -0.6414

Алгоритм SVD використовує підпрограми LAPACK для обчислення сингулярного розкладання.

2.3. Дослідження альтернативних традиційним методам заповнення пропусків в таблицях даних

Мадху [13] представляє алгоритм нормалізованого середнього для числових наборів даних, який перевершує інші методи. Шмітт [14] визначає метод bayesian principal component analysis (bPCA) та FCM як багатообіцяючі альтернативи традиційним методам імпутації.

Для дослідження виберемо найбільш популярні за оцінкою Google методи (див. рис. 2.8),

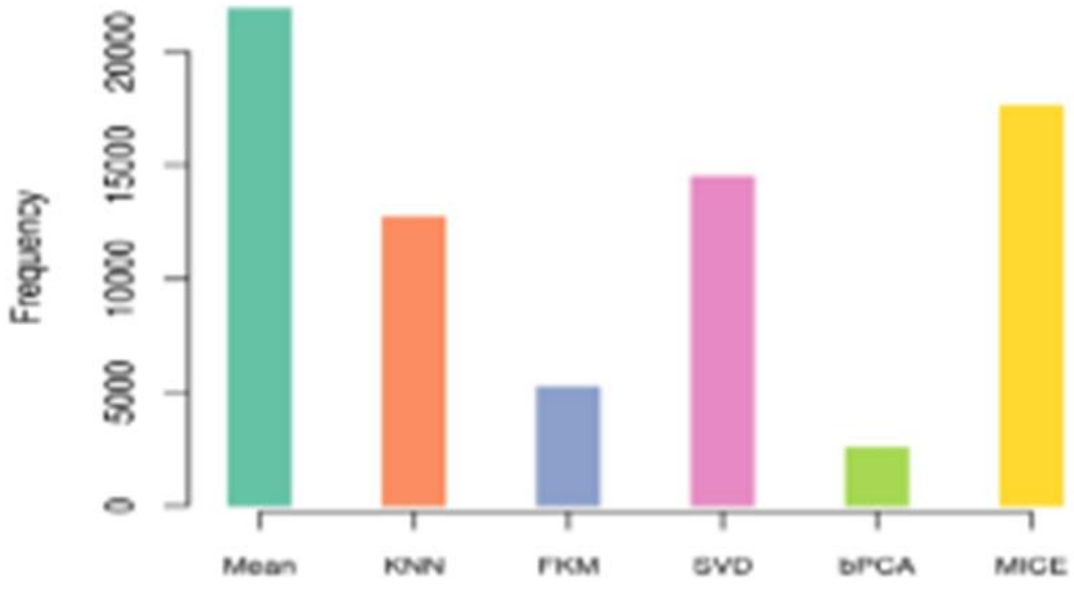


Рис. 2.8 – Частота публікацій в Google щодо різних методів заповнення пропусків в таблицях даних.

Виберемо шість найбільш популярних методів заповнення пропусків в таблицях даних, а саме:

- Метод (Mean substitution, Mean) заповнення пропусків даних в таблицях середнім значенням відповідного параметру;
- Метод (Multivariate Imputation by Chained Equations, MICE) - заповнення пропусків даних в таблицях декількома способами з наступним об'єднанням результатів ланцюжковими рівняннями при заповненні пропущеного значення;
- Метод К-найбли́жчих сусі́дів (K-nearest neighbor method, KNN), коли вибирають певну кількість сусідніх з пропущеним значенням елементів таблиці і за допомогою регресії визначають відсутнє значення;
- Метод (Fuzzy C-means, FCM) видалення рядків з пропусками і заповнення їх статистичною вибіркою з урахуванням нечіткої функції залежності між елементами;
- Метод (Singular value decomposition, SVD) сингулярного розкладу таблиці даних є доволі складним, але дозволяє побачити геометричну структуру таблиці і представити наявні дані;
- Метод (Bayesian principal component analysis, bPCA) баєсовського принципу вимагає великої кількості даних, але з більшою точністю ніж PCA метод визначає значення пропущених даних.

Принцип, за яким буде здійснено аналіз показано на рис. 2.9, де:

1. Формування таблиці даних з відкритих джерел без пропущених значень розміром 80 об'єктів на 65 атрибутів;
2. Введення 5%, 10%, 15%,…. 45% пропусків значень у кожну з таблиць даних;
3. Застосування вибраних методів імпутації до кожної з отриманих таблиць;
4. Здійснення розрахунків при конкретній кількості відсотків пропущених даних для кожного методу;

5. Визначення при конкретній кількості відсотків пропущених даних продуктивності кожного методу (1000 циклів обчислень);

6. Розрахунок для кожного методу усереднених результатів продуктивності заповнення пропущених даних.

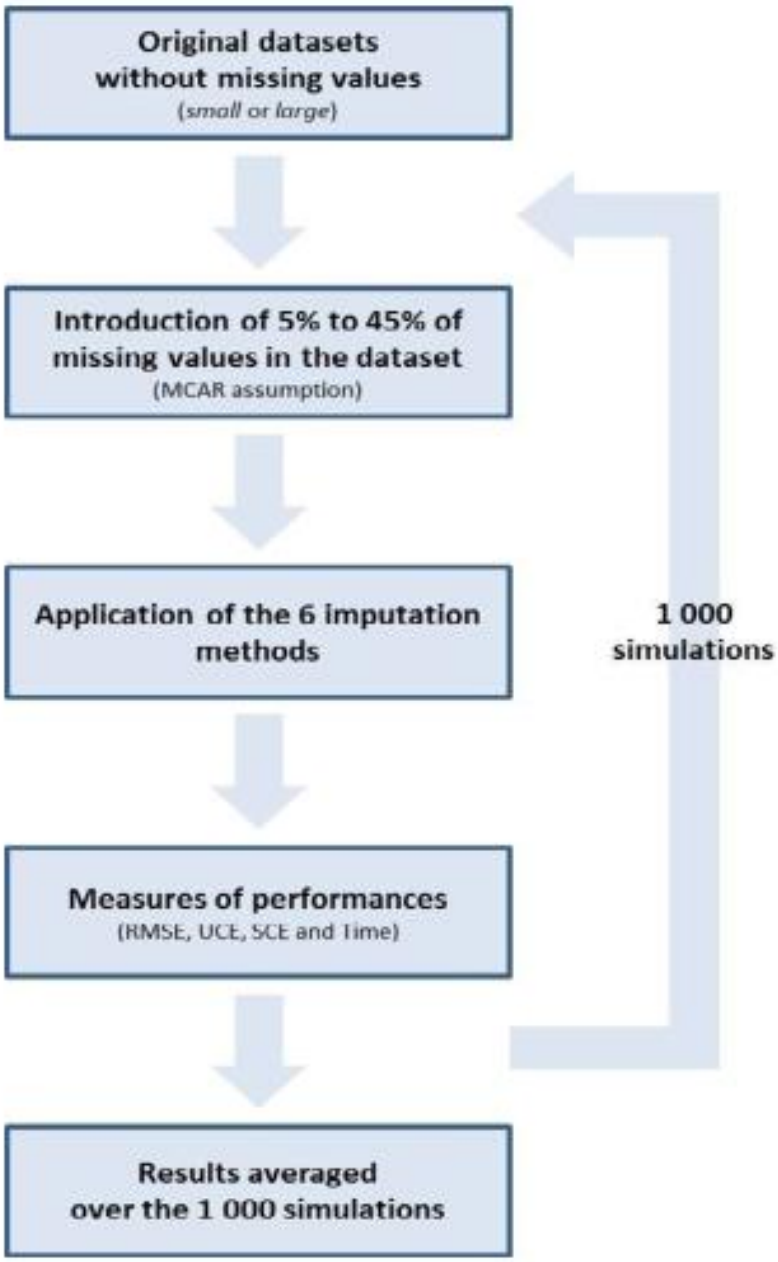


Рис. 2.9. Прийняті принципи дослідження вибраних методів заповнення пропусків в таблицях даних.

Для порівняння методів сформулюємо критерії їх оцінки. В якості критеріїв використаємо наступні.

1. Стандартне відхилення (середньоквадратична помилка – RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i^{obs} - X_i^{imputed}\right)^2}{n}}$$

2. Величина неконтрольованої помилки класифікації (UCE), яка дорівнює відсотку неправильно визначених значень пропусків в таблиці.

3. Величина помилки контрольованої класифікації (SCE) при використанні лінійного дискримінантного аналізу.

4. Час (хвилин) здійснення розрахунків кожного з методів

Результати порівняльного аналізу наведені на рис. 2.10, де представлено значення вибраних критеріїв в залежності від кількості відсотків відсутніх даних. Методи заповнення пропусків в таблицях даних

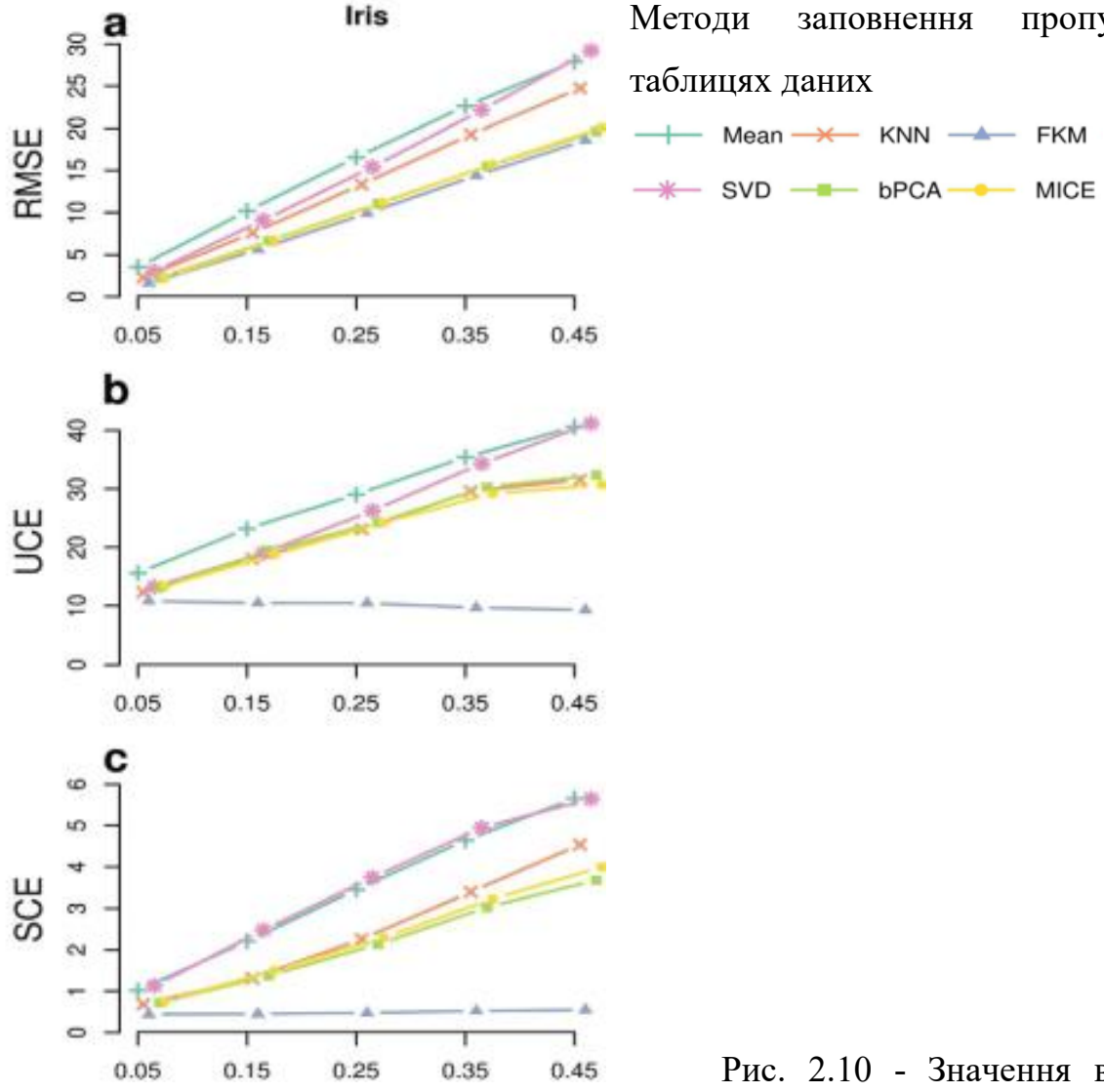


Рис. 2.10 - Значення вибраних критеріїв (RMSE, UCE, SCE) в залежності від кількості відсотків відсутніх даних.

Очікувано частка правильного заповнення даних зменшилася із зростанням частка відсутніх значень зросла в усіх наборах даних.

Беручи до уваги відтворюваність чотирьох наборів даних відповідно до критеріїв RMSE, UCE та SCE, середнє значення було методом, який є менш ефективним при застосуванні до менших наборів даних, де відмінності з іншими методами більш виражені.

Поведінка SVD і MICE була непослідовною в різних наборах даних.

Насправді, хоча MICE показав хороші результати на невеликих наборах даних, це був другий найгірший метод (після середнього) на великих наборах даних.

Навпаки, протилежна ситуація спостерігається для SVD, де SVD добре працює на великих наборах даних, але погано, коли застосовується до малих наборів даних.

KNN розмістився між найкращими та найгіршими методами.

bPCA і FCM-методи належать до кращих для різних наборів даних і показників ефективності. Зокрема, FCM-метод кращий при застосуванні до невеликих наборів даних на основі критеріїв UCE та SCE.

Порівняння часу виконання процесів заповнення пропусків в таблицях даних різними методами в залежності від обсягу даних наведено на рис. 2.11.

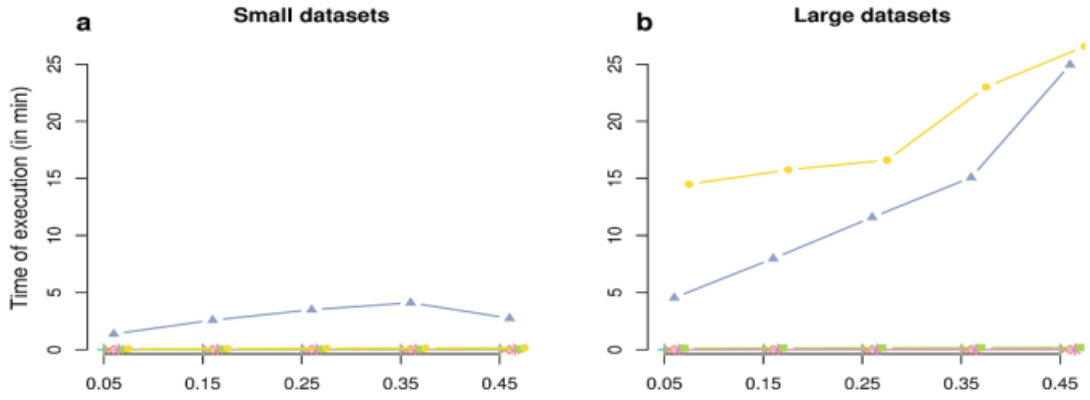


Рис. 2.11 – Час виконання процесів заповнення пропусків в таблицях даних різними методами в залежності від обсягу даних.

Середнє значення, KNN, SVD і bPCA є дуже швидкими з тривалістю обчислень від 0,5 до 10 секунд після відсутнього значення.

FCM є повільнішим, але все одно показує хороший час виконання, за винятком випадків, коли застосовується з великим набором даних для 45% відсутніх значень (близько 25 хв), в діапазоні від 1 хв до 15 хв залежно від розміру даних та частоти відсутніх значень.

Час виконання MICE залежить від розміру набору даних, особливо від довжини змінних. Дуже швидко (приблизно 5-10 секунд) для невеликих наборів даних і близько 30 хвилин на повній швидкості для найбільших наборів даних.

РОЗДІЛ 3. РОЗРОБКА АВТОМАТИЗАЦІЇ ПРОЦЕСУ ЗАПОВНЕННЯ ПРОПУСКІВ В ТАБЛИЦЯХ ДАНИХ ПІДПРИЄМСТВ АГРОПРОМИСЛОВОГО КОМПЛЕКСУ

3.1. Обґрунтування застосування нейромережних технологій до автоматизації заповнення пропусків в таблицях даних

Технології нейронних мереж успішно використовуються для автоматизації заповнення пробілів у таблицях даних. А. Десілетс (Desilets) розробив модель для придушення клітинок у табличних даних, тоді як Вей () запропонував загальну структуру для імпутації відсутніх даних. В Сміейа (Śmieja) представив механізм обробки відсутніх даних шляхом заміни відповіді нейрона на його очікуване значення, а В. Краснопольський застосував техніку нейронної мережі для заповнення прогалин у супутникових вимірюваннях, зокрема у спостереженнях за кольором океану. Ф. Війера (A.Vieira) використовував ШНМ для заповнення пропусків у хвильових даних в таблицях, досягнувши трохи кращої продуктивності, ніж спектральна хвильова модель. Так само Н. Ахр (N. Ahr) успішно відновив пропуски даних в записах геомагнітних обсерваторій за допомогою нейронної мережі. Р. Бхаттачарія використовував штучні нейронні мережі (ШНМ) для заповнення пропусків в часових рядах хвильових даних із достатньою точністю. Ці дослідження в цілому підкреслюють потенціал нейронних мереж у автоматизації заповнення пропусків у даних, пропонуючи покращену продуктивність та ефективність обчислень. Ці дослідження в цілому демонструють ефективність нейронних мереж в автоматизації заповнення пропусків у таблицях даних.

Зазначимо, що нейромережні технології (або штучні нейронні мережі, ШНМ) відкривають нові можливості для автоматизації процесу заповнення пропусків у таблицях даних. Використання ШНМ для цієї задачі має ряд переваг та може значно покращити якість та ефективність обробки даних. Нижче наведемо обґрунтування застосування цих технологій.

1. Підвищення точності заповнення даних, висока здатність до машинного

навчання та адаптивність до складних залежностей. Нейронні мережі здатні виявляти складні нелінійні залежності між даними, що дозволяє більш точно передбачати відсутні значення. Глибокі нейронні мережі (глибоке навчання) можуть використовувати багатошарові структури для більш точного моделювання складних взаємозв'язків у великих наборах даних.

2. Автоматизація та ефективність паралельної обробки великих обсягів даних, що дозволяє швидко заповнювати пропуски у великих таблицях. ШНМ можуть автоматично навчатися на нових даних, що забезпечує постійне вдосконалення моделей і адаптацію до змін у даних.

3. Гнучкість і універсальність за рахунок різноманітності архітектур нейронних мереж, зокрема, такі як рекурентні нейронні мережі (RNN) для тимчасових даних, згорткові нейронні мережі (CNN) для просторових даних, та автоенкодери для заповнення пропусків шляхом навчання представлень даних. Також використовуються спеціалізовані моделі. Наприклад, варіаційні автоенкодери (VAE) та генеративно-змагальні мережі (GAN), які можуть забезпечити більш точне і реалістичне заповнення пропусків.

4. Підтримка різних типів даних, зокрема, обробка числових, категорійних і текстових даних. Зауважимо, що нейронні мережі можуть ефективно працювати з різними типами даних одночасно, включаючи числові, категорійні, текстові та зображення, а завдяки механізмам типу attention, ШНМ можуть враховувати контекстні залежності в даних, що покращує якість заповнення пропусків.

5. Виявлення і усунення аномалій в таблицях даних за рахунок навчання виявляти аномалії у даних, що дозволяє автоматично ідентифікувати і коригувати нетипові значення або помилки. При цьому моделі можуть також інтегрувати механізми перевірки і виправлення для забезпечення цілісності даних під час заповнення їх пропусків.

6. Переваги у порівнянні з традиційними методами, особливо, над простими методами заповнення пропусків середніми і медіанними значеннями, коли нейронні мережі забезпечують більш точні і контекстно обґрунтовані

результати. Також традиційні регресійні моделі можуть бути недостатньо гнучкими для складних наборів даних, тоді як нейронні мережі краще справляються з нелінійними залежностями і великою кількістю змінних.

Таким чином, можемо стверджувати, що застосування нейромережних технологій для автоматизації заповнення пропусків у таблицях даних має значні переваги в порівнянні з традиційними методами. Висока точність, масштабованість, гнучкість, здатність працювати з різними типами даних та виявляти аномалії роблять нейронні мережі потужним інструментом для забезпечення якості та цілісності даних. Впровадження цих технологій дозволяє значно покращити процес обробки і аналізу даних, забезпечуючи більш надійні і достовірні результати.

3.2. Розробка системи заповнення пропусків даних на основі нейромережної структури автоасоціативного типу.

Автоасоціативні нейронні мережі (автоенкодери) є ефективним інструментом для заповнення пропусків у даних. Вони навчаються стискати та відновлювати дані, що дозволяє їм ефективно відновлювати втрачену інформацію. Нижче наведено етапи розробки системи заповнення пропусків даних за допомогою автоенкодера.

1 етап - підготовка даних (їх отримання та очищення).

Завдання цього етапу зібрати відповідний набір даних в таблицю, який буде використовуватися для навчання і тестування моделі, та виконати попередню обробку даних, включаючи їх нормалізацію або стандартизацію, для покращення якості навчання моделі. Для навчання водяться випадкові пропуски у даних, щоб навчити модель відновлювати втрачені значення. При цьому дані поділяються на тренувальний і тестовий набори для оцінки ефективності моделі.

2 етап - створення моделі автоенкодера, який містить операції визначення архітектури мережі автоенкодера, включаючи кількість шарів, кількість нейронів у кожному шарі, активаційні функції тощо. При цьому стаорюються дві основні частини автоенкодера - кодувальник (encoder) і декодувальник (decoder).

Нижче розроблена архітектура автоенкодера на Python з використанням Keras:

```
import numpy as np
import pandas as pd
from keras.layers import Input, Dense
from keras.models import Model
from sklearn.preprocessing import MinMaxScaler
# Приклад набору даних
data = pd.DataFrame({
    'A': [1.0, 2.0, np.nan, 4.0, 5.0],
    'B': [5.0, np.nan, 3.0, 2.0, 1.0]
})
# Заповнення пропусків середніми значеннями для навчання
data.fillna(data.mean(), inplace=True)
# Масштабування даних
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(data)
# Архітектура автоенкодера
input_dim = data_scaled.shape[1]
encoding_dim = 2  # Розмір латентного простору
input_layer = Input(shape=(input_dim,))
encoded = Dense(encoding_dim, activation='relu')(input_layer)
decoded = Dense(input_dim, activation='sigmoid')(encoded)
autoencoder = Model(input_layer, decoded)
# Кодувальник
encoder = Model(input_layer, encoded)
# Декодувальник
encoded_input = Input(shape=(encoding_dim,))
decoder_layer = autoencoder.layers[-1]
```

```
decoder = Model(encoded_input, decoder_layer(encoded_input))
autoencoder.compile(optimizer='adam', loss='mean_squared_error')
# Навчання моделі
autoencoder.fit(data_scaled, data_scaled, epochs=50, batch_size=2, shuffle=True, validation_split=0.2)
```

3 етап - заповнення пропусків даних, який починається з введення пропусків у тестовий набір даних, щоб перевірити ефективність моделі. Далі використовується навчений автоенкодер для відновлення пропусків у тестовому наборі даних.

Наведемо нижче приклад відновлення пропусків даних

```
# Введення пропусків у тестові дані
data_with_missing = pd.DataFrame({
   'A': [1.0, 2.0, np.nan, 4.0, 5.0],
   'B': [5.0, np.nan, 3.0, 2.0, 1.0]
})
data_with_missing_scaled = scaler.transform(data_with_missing.fillna(0))
# Відновлення даних
reconstructed_data = autoencoder.predict(data_with_missing_scaled)
reconstructed_data = scaler.inverse_transform(reconstructed_data)
# Заповнення пропусків відновленими значеннями
data_filled = data_with_missing.copy()
data_filled[np.isnan(data_filled)] = reconstructed_data[np.isnan(data_with_missing)]
print(data_filled)
```

4 етап - оцінка ефективності моделі за визначеними метриками. Зазвичай, для оцінки ефективності заповнення пропусків використовуються такі метрики, як середня абсолютна похибка (MAE) або середньоквадратична похибка (MSE).

Ефективність автоенкодера видно у порівнянні з базовими моделями, такими як заповнення середнім або медіанним значенням.

Нижче наведемо приклад такої оцінки ефективності (python).

```
from sklearn.metrics import mean_absolute_error
# Оригінальні дані без пропусків для порівняння
original_data = pd.DataFrame({
   'A': [1.0, 2.0, 3.0, 4.0, 5.0],
   'B': [5.0, 4.0, 3.0, 2.0, 1.0]
})
# Обчислення метрики MAE
mae = mean_absolute_error(original_data, data_filled)
print(f'MAE: {mae}')
```

5 етап - розгортання і використання системи заповнення пропусків даних складається з її інтеграції у існуючі бізнес-процеси та інформаційні системи компанії та налаштування автоматизованого процесу заповнення пропусків на основі нових даних, що надходять у систему. В процесі експлуатації системи необхідно постійно здійснювати моніторинг ефективності системи та оновлювати модель. Також, за необхідності, здійснюється розширення моделі для обробки нових типів даних або збільшення обсягів даних.

Таким чином, розроблена система заповнення пропусків даних на основі нейромережної структури автоасоціативного типу (автоенкодера) дозволяє ефективно відновлювати втрачену інформацію, підвищуючи якість і цілісність даних. Впровадження такої системи включає підготовку даних, створення і навчання моделі, оцінку ефективності та інтеграцію у бізнес-процеси. Використання автоенкодерів забезпечує більш точні результати порівняно з традиційними методами і дозволяє автоматизувати процес заповнення пропусків у таблицях даних.

3.3. Адаптація нейроподібних мереж машини геометричних перетворень для заповнення пропусків в таблицях даних підприємств АПК

Як показано вище, результати заповнення прогалин у таблиці даних найвищої якості отримують при застосуванні методів регресії, але це можливо, лише якщо дані таблиці корельовані та відповідають багатьом критеріям. Обмеження на кількість відсутніх даних. Зокрема, потрібно дотримуватися певного співвідношення між кількістю рядків з пропусками і без них тощо.

У випадках, коли елементи таблиці не залежать один від одного, можна застосувати лише один із перших трьох методів, перерахованих вище. У більшості випадків якість цієї методики занадто низька, але в цьому випадку іншого виходу немає. Значно розширені параметри доступні для варіантів таблиці, де дотримуються певні залежності між елементами, зокрема:

- снують взаємозалежності між елементами рядків кожного окремого стовпця; подібні залежності мають місце для часових послідовностей, або при наяності окремого стовпця (стовпців), що є маркером кожного рядка таблиці;
- існують взаємозалежності між елементами по стовпцях для кожного окремого рядка; даний варіант носить більш універсальний характер, так як з нього випливає також варіант залежності і між окремими рядками.

Для розв’язання задачі адаптуємо нейроподібну структуру Машини Геометричних Перетворень (МГП) [15], яка забезпечує універсальний підхід, придатний для різноманітних варіантів таблиць даних, з наявністю різнотипних взаємозв’язків між їх елементами. В основу методу покладено просторове тлумачення таблиці взаємозалежних даних як тіла, що є геометричним місцем точок - рядків таблиці в координатах, кожній з яких відповідає один стовпець таблиці. Тобто, кожен рядок таблиці розглядається як точка в просторі реалізацій, вимірність якого визначається числом стовпців таблиці, а кожен стовпець представляє одну з координат тіла. Метод заповнення пропусків на основі МГП базується на принципі побудови тіла, що може бути частково, або повністю представлене лише проекціями точок на певні координатні гіперплощини і знаходження всіх координат точок за їх проекціями, виходячи з

належності кожної точки побудованому тілу. Даний метод переважає по якості існуючі, в першу чергу, через те, що забезпечує врахування всієї повноти інформації, представленої заданими компонентами елементів таблиці a , отже, дозволяє отримати більш точні представлення, якщо це загалом можливе для заданої сукупності і структури даних.

Розглянемо основи функціонування системи заповнення пропусків на основі нейроподібної структури МГП автоасоціативного типу (рис. 1).

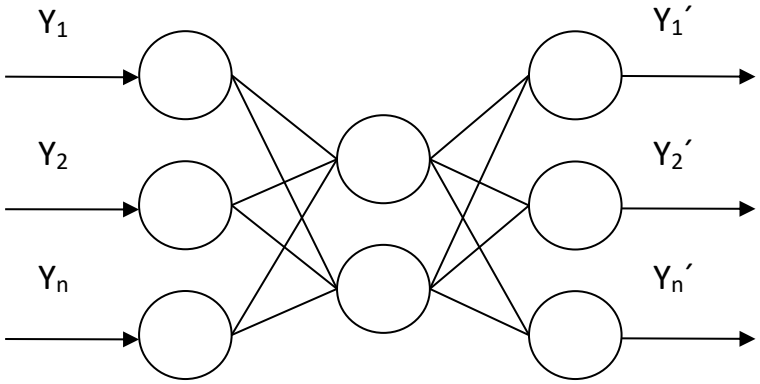


Рис. 3.1 - Нейроподібна структура МГП автоасоціативного типу

Початково всі пропущені елементи кожного з стовпців заміняємо середньоарифметичними величинами по наявних елементах кожного стовпця. Особливістю автоасоціативного режиму навчання нейромережі є ідентичність вхідних і вихідних компонентів тренувальних векторів, однак застосування подібної структури для заповнення пропусків в даних можливе лише для зменшеного (відповідно до числа входів) числа нейронних елементів прихованого шару – варіант вузького горла. Лише в цьому випадку нейронними елементами прихованого шару формується тіло інформаційного об’єкта, яке на виході мережі представляється дещо спотвореними компонентами первинних векторів. Похибки такого представлення визначаються точністю апроксимації тіла нейромережею і можуть бути в багатьох випадках достатньо малими. У випадку відсутності корельованості між елементами таблиці, коли довільний її елемент не містить жодної інформації про всі інші елементи, а двохвимірна густина імовірності $p(y_i,y_j)$ рівна добутку одновимірних функцій густини

$$p(y_i, y_j) = p(y_i) * p(y_j),$$

тіло інформаційного об'єкта є аморфним і не має певної форми. Подібний випадок даних такого типу легко виявляється за допомогою цієї ж нейроподібної структури МГП. Це матиме місце, коли для числа нейронних елементів в прихованому шарі, що лише на одиницю менше за число входів (виходів), похибка трансформації вхідних векторів на вихід є надто великою. В такій ситуації наступні дії не виконуються і замість пропущених елементів таблиці використовуються раніше встановлені усереднені значення по її стовпцях, так як подальше застосування нейромережі для заповнення пропусків втрачає зміст через великі похибки апроксимації.

Інформаційна технологія заповнення пропусків в даних на основі застосування автоасоціативної нейромережі МГП передбачає виконання наступних кроків перетворення інформації:

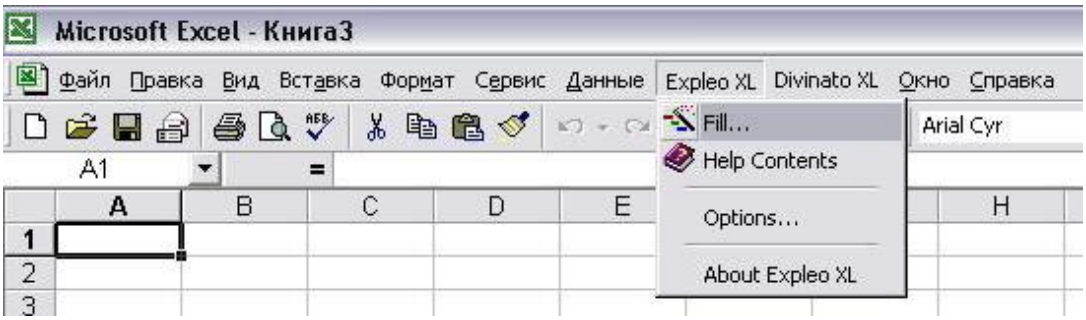
1. навчання нейромережі на основі тренувальних векторів - рядків даних, в яких пропуски замінені середніми значеннями;
2. трансляція тренувальних векторів-рядків через навчену нейромережу на вихід;
3. заміни пропущених початково компонентів векторів значеннями, отриманими на виходах нейромережі ;
4. вихід з циклу за умови досягнення заданого порогу змін між двома послідовними перетвореннями
5. перенавчання нейромережі;
6. перехід на п.2.

Даний підхід базується на унікальній властивості нейромереж МГП швидко перенавчатися. Він є значно точнішим за інші методи заповнення пропусків, універсальний, так як придатний і для проведення попереднього аналізу даних, і для здійснення замін в таблицях різної структури і наповнення, не вимагає спеціальних знань від користувача.

Програмна реалізація адаптованої нейроподібної мережі для заповнення пропусків в таблицях даних. Адаптовано програмний продукт Expleo XL , що реалізує нелінійний метод, який не ставить додаткових вимог до даних, які обробляються, і тому може застосовуватись практично у всіх випадках. Допустимим є також випадок, коли кожен рядок даних містить пропуски в різних стовпцях.

Наведемо основні принципи роботи з адаптованою нейроподібною мережею для заповнення пропусків в таблицях даних підприємств АПК Expleo XL.

Expleo XL легко інтегрується в меню MS Excel.



Для заповнення пропусків в даних слід здійснити наступні кроки:

- Вибрати пункт меню Fill. З’явиться вікно настройки параметрів:

- В таблиці з даними Excel вибрати відповідний діапазон комірок для заповнення пропусків в ньому:

- Натиснути кнопку Fill і пропущені значення будуть заповнені та помічені кольором:

Microsoft Excel - pima-indians-diabetes.xls

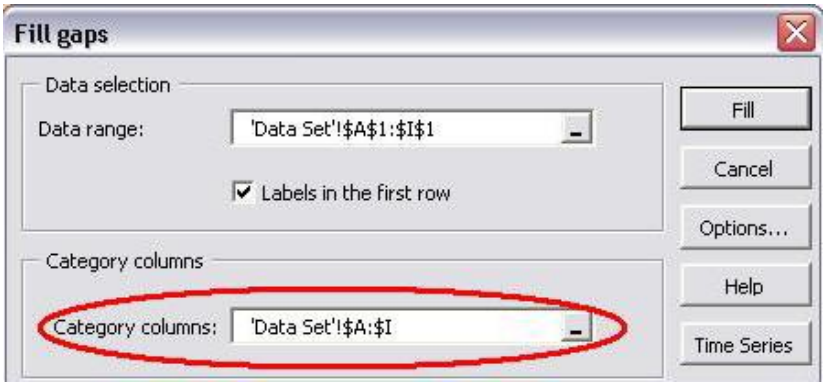
Файл Правка Вид Вставка Формат Сервис Данные Expleo XL Divinato XL Окно Справка

Arial

A1 = Incomplete data set - Diabetes

	A	B	C	D	E	F	G
4	Inputs						
5	6	148	72	35	260	33,6	0,627
6	1	85	66	29	94	26,6	0,351
7	8	183	64	37	299	23,3	0,672
8	1	89	66	23	94	28,1	0,167
9	0	137	40	35	168	43,1	2,288
10	5	116	74	28	152	25,6	0,201
11	3	78	50	32	88	31	0,248
12	10	115	74	29	166	35,3	0,134
13	2	197	70	45	543	30,5	0,158
14	8	125	96	31	178	0	0,232
15	4	110	92	28	159	37,6	0,191
16	10	168	74	38	323	38	0,537
17	10	139	80	30	187	27,1	1,441
18	1	189	60	23	846	30,1	0,398

Expleo XL автоматично розрізняє дані, що представлені текстом (класи) і дані, що представлені цілими числами та числами з плаваючою комою. Якщо, в деяких колонках містяться класи представлені числами, потрібно вручну вказати тип цих колонок:



Крім звичайної інформацій в таблицях даний метод може заповнювати пропуски також в часових послідовностях. Режим часових послідовностей активується натисненням кнопки Time Series.

Слідом з’являється ще один параметр настройки – Period, який визначає періодичність даних. Періодичність залежить від характеру даних: напр., коли таблиця містить дані місячних звітів, то напевно слід попробувати застосувати період 6, що охоплює 0,5 року, або12 - для одного року.

Зазначимо, що адаптований вище підхід є універсальним, достатньо простим і забезпечує, загалом, помітно вищу якість заповнення пропусків в даних .

Зі збільшенням об’єму наявних даних підвищується точність заповнення (стосується також збільшення кількості колонок в таблиці з даними).

Зауважимо, що повністю пусті колонки не будуть заповнені, а повністю пусті рядки можуть бути заповнені лише для часових послідовностей. При цьому кількість розташованих підряд пустих комірок не може перевищувати довжину періоду.

РОЗДІЛ 4. РОЗРАХУНОК ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ АВТОМАТИЗАЦІЇ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ ЗАПОВНЕННЯ ПРОПУСКІВ В ТАБЛИЦЯХ ДАНИХ

4.1. Аналіз витрат на розробку та розгортання автоматизації процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу

Для характеристики прототипу програмного забезпечення автоматизації процесу заповнення пропусків в таблицях даних підприємств агропромислового комплексу використовуємо параметри X1-X4.

На основі даних, що представлені у літературі, визначаємо їх мінімальні, середні отримуванні та максимально допустимі значення (див. табл. 4.1).

Таблиця 4.1.

Комплекс параметрів, що характеризують прототипу програмного забезпечення автоматизації процесу заповнення пропусків в таблицях даних підприємств АПК

Найменування параметру	Позначення параметру	Значення параметру		
		Мінімальне	Середнє	Максимальне
Час розробки, людина*год	X1	352	528	704
Час роботи алгоритму, мс	X2	100	550	1000
Рекомендована швидкість запису на диск, МБ/с	X3	0	32	64
Об'єм пам'яті для збереження даних, МБ	X4	0	16	32

Області змін значень параметрів X1-X4 для прототипу програмного забезпечення автоматизації процесу заповнення пропусків в таблицях даних (див.рис.4.1 – 4.3).

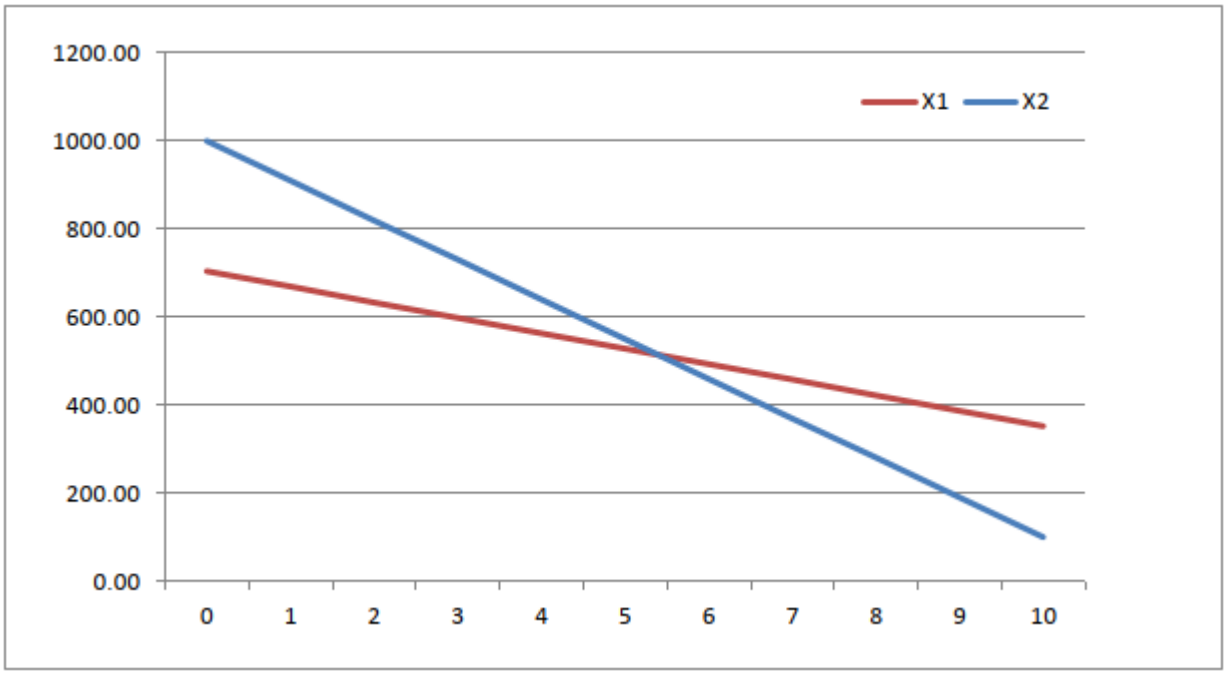


Рис. 4.1 – Області значень параметрів X1, X2.

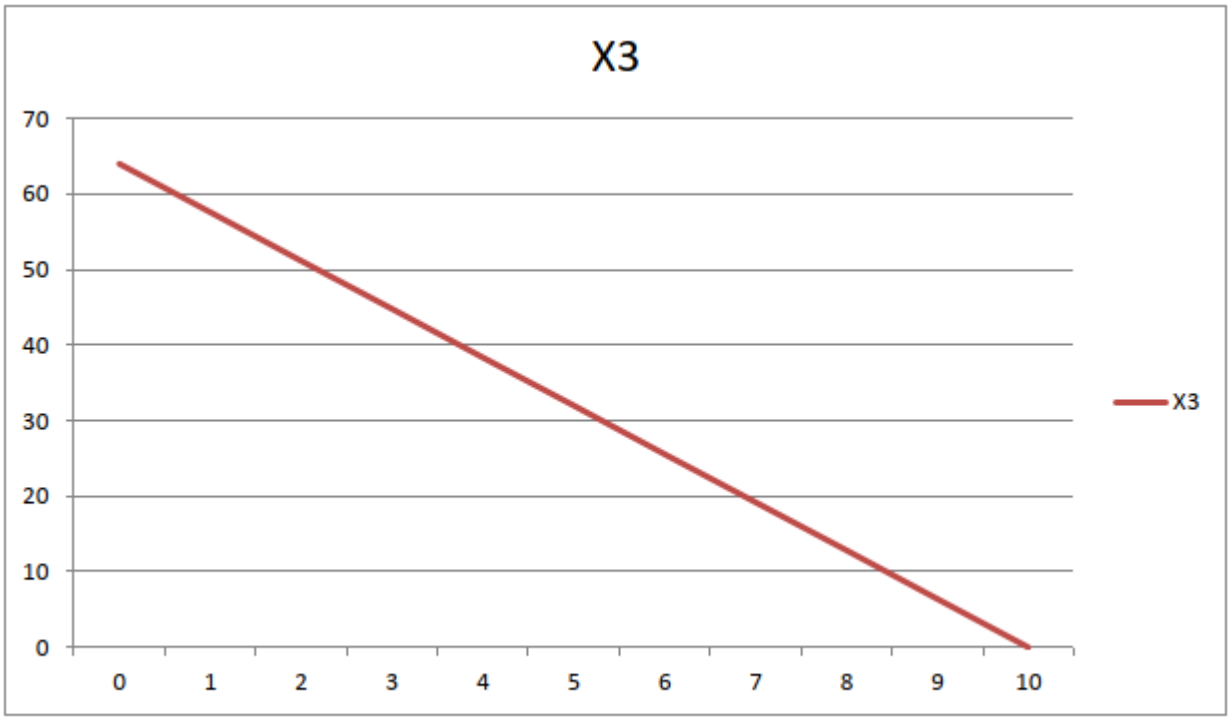


Рисунок. 4.2 - Значення параметра X3

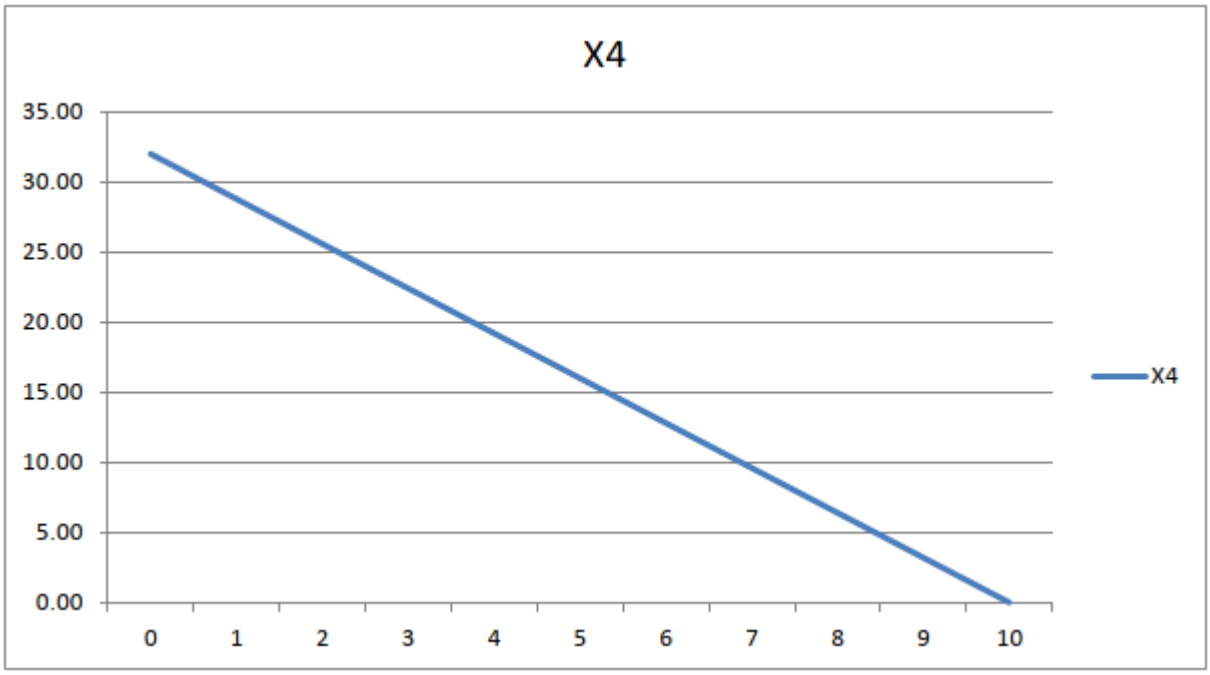


Рис. 4.3 – Область значень параметру X4.

Для оцінки трудомісткості розробки спочатку проведемо розрахунок трудомісткості. Усі варіанти мають наступні основні завдання:

1) Видалення зайвих атрибутів.

2) Вивід результатів на екран.

Також кожний з варіантів проєктування системи автоматизації має додаткові завдання, які є реалізаціями розгалужених варіантів розробки.

Далі наведено варіанти додаткових завдань (два завдання, які мають номери 1 в реалізаціях)

Завантаження даних з пам'яті.

Штучне створення даних.

Збереження отриманих результатів.

Ваги параметрів оцінюються за допомогою методів попарного порівняння. Ранги варіюються від 1 до 4. Результати наведені в табл. 4.2-4.3

Таблиця. 4.2 - Результат оцінки параметрів

Параметр	Ранг параметру по оцінці експерта							Сума рангів, R_i	Відхилення Δ_i	Квадрат відхилення, $(\Delta_i)^2$
	1	2	3	4	5	6	7			
X1	3	2	1	1	2	1	2	12	-5.5	30.25
X2	1	1	2	2	3	2	1	12	-5.5	30.25
X3	4	4	4	4	4	4	4	28	10.5	110.25
X4	2	3	3	3	1	3	3	18	0.5	0.25
Разом	10	10	10	10	10	10	10	70	0	171

За 1 приймається найбільший ранг, за 4 – найменший.

Таблиця 4.3 - Попарне порівняння параметрів

Параметри	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 та X2	<	<	>	>	>	>	<	>	1.5
X1 та X3	>	>	>	>	>	>	>	>	1.5
X1 та X4	<	>	>	>	<	>	>	>	1.5
X2 та X3	>	>	>	>	>	>	>	>	1.5
X2 та X4	>	>	>	>	<	>	>	>	1.5
X3 та X4	<	<	<	<	<	<	<	<	0.5

Спираючись на норми розрахункового часу визначимо трудомісткість. Вона складає для першого завдання $T_р$=43 людино-днів. Поправочний коефіцієнт, що враховує новизну ПП складає $К_н$=1 (нормативно-довідкова інформація). Оскільки під час виконання даного завдання використовуються новостворенні модулі, врахуємо це за допомогою коефіцієнта $К_{ст}$ = 0,6. Коефіцієнти $К_м$ і $К_{ст.м}$, які враховують відповідно програмування на мові низького рівня та розробку стандартного програмного забезпечення, для всіх завдань дорівнюють 1.

Загальна трудомісткість обчислюється як

$$Т_О = Т_Р * К_Н * К_М * К_{СТ} * К_{СТ.М},$$

де $T_Р$ - трудомісткість розробки ПП;

$К_Н$ - поправочний коефіцієнт;

$К_М$ - коефіцієнт рівня мови програмування;

$К_{СТ}$ - коефіцієнт використання стандартних модулів і прикладних програм.

Далі розрахуємо витрати на оплату однієї машино-години. Враховуючи, що вона обслуговує одного спеціаліста з окладом 13500 грн та одного з окладом 10000 грн з коефіцієнтом зайнятості 0,6, то для двох машин отримаємо

Сг= 12*13500*0,6+12* 10000*0,6 = 169200 грн

Враховуючи додаткову заробітну плату

СЗП = 169200 *(1+0,4)= 270720

Відрахування на соціальне страхування 22%

СВІД = 270720 * 0,22 = 59558,4

Розрахуємо амортизаційні підрахунки (амортизація 25%, вартість ЕОМ 25000 грн)

СА= КТМ*КА*ЦПР= 1,15*0,25*25000 = 7187,5 грн

Розрахуємо витрати на ремонт та профілактику:

СР = КТМ * ЦПР * КР = 1,15 * 25000 * 0,05 = 1437,5грн

Розрахуємо ефективний годинний фонд часу ПК за рік

ТЕФ= (365-142-16)* 8 *0,8 = 1324,8 год

Розрахуємо витрати на електроенергію

СЕЛ = 1324,8 * 0,6 * 0,6 * 1,75 * 3 = 2191,33 грн

Накладні витрати рівні:

СН = 25000 * 0,67 = 16750 грн.

Отже експлуатаційні витрати(грн):

СЕКС = 270720 +59558,4 +7187,5 + 1437,5 + 2191,33 +16750

= 3357844,73

Тоді собівартість однієї машино-години ЕОМ дорівнюватиме:

3357844,73

СМ-Г = 1324>8 = 270,11 грн/год

4.2. Розрахунок терміну окупності автоматизації процесу заповнення пропусків в таблицях даних

Розрахунок терміну окупності автоматизації процесу заповнення пропусків в таблицях даних може бути зроблений за допомогою наступних кроків:

Визначення витрат на автоматизацію:

Одноразові витрати (розробка або закупівля програмного забезпечення, встановлення, навчання персоналу).

Поточні витрати (підтримка, оновлення, ліцензії).

Визначення економії від автоматизації:

Час: кількість часу, яку заощаджують співробітники завдяки автоматизації процесу (години на день або місяць).

Грошова вартість часу: вартість роботи співробітників, яка була б витрачена на заповнення пропусків вручну.

Якість даних: зниження витрат на виправлення помилок у даних.

Розрахунок економії:

Визначте вартість години роботи співробітника.

Помножте кількість заощаджених годин на вартість години роботи, щоб отримати економію в грошах.

Розрахунок терміну окупності (Payback Period):

Поділіть загальні витрати на автоматизацію на річну економію.

Формули

Витрати на автоматизацію: Савтоматизації=Содноразові+СпоточніC_{\text{автоматизації}} = C_{\text{одноразові}} + C_{\text{поточні}}Савтоматизації=Содноразові+Споточні

Економія часу: Ечас=Тзаощаджені години×СгодиниE_{\text{час}} = T_{\text{заощаджені години}} \times C_{\text{години}}Ечас=Тзаощаджені години×Сгодини

Термін окупності: Токупності=СавтоматизаціїЕчас+ЕякістьT_{\text{окупності}} = \frac{C_{\text{автоматизації}}}{E_{\text{час}} + E_{\text{якість}}}Токупності=Ечас+ЕякістьСавтоматизації

Приклад розрахунку

Витрати на автоматизацію:

Одноразові витрати: $10,000

Поточні витрати: $1,000 на рік

Савтоматизації=10,000+1,000=11,000 C= 10,000 + 1,000 = 11,000

Савтоматизації=10,000+1,000=11,000

Економія часу:

Заощаджені години: 100 годин на місяць

Вартість години роботи: $30

Ечас=100×30×12=36,000 на рікE_{\text{час}} = 100 \times 30 \times 12 = 36,000 \text{ на рік}Ечас=100×30×12=36,000 на рік

Термін окупності:

Токупності=11,00036,000≈0.31 роківT_{\text{окупності}} = \frac{11,000}{36,000} \approx 0.31 \text{ років}Токупності=36,00011,000≈0.31 років

Тобто, термін окупності автоматизації складе приблизно 0.31 року, або близько 4 місяців.

РОЗДІЛ 5. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

5.1. Нормативно-правові документи України з охорони праці та безпеки в надзвичайних ситуаціях.

Національна система нормативно-правових актів України з охорони праці та безпеки в надзвичайних ситуаціях включає в себе закони, постанови, накази та інші акти, які регулюють права та обов'язки працівників і роботодавців щодо забезпечення безпечних умов праці і захисту від надзвичайних ситуацій. Ось декілька основних нормативно-правових актів у цій галузі:

Закон України "Про охорону праці" (від 14 грудня 1992 року № 2694-XII) - цей закон встановлює загальні принципи та вимоги щодо охорони праці в Україні.

Закон України "Про надзвичайні ситуації та станом надзвичайної ситуації" (від 21 грудня 1992 року № 2693-XII) - цей закон регулює організацію та управління діяльністю в галузі захисту населення та територій від надзвичайних ситуацій.

Закон України "Про цивільний захист" (від 5 лютого 1993 року № 3206-XII) - цей закон визначає порядок організації цивільного захисту та заходи щодо захисту населення від надзвичайних ситуацій.

Закон України "Про працю" (від 10 грудня 1971 року № 322-VIII) - цей закон встановлює основні права та обов'язки працівників і роботодавців, включаючи вимоги до охорони праці та безпеки на робочому місці.

Постанова Кабінету Міністрів України "Про затвердження Порядку розслідування нещасних випадків на виробництві та професійних захворювань" (від 23 жовтня 1996 року № 1248) - ця постанова визначає процедуру розслідування нещасних випадків на виробництві та професійних захворювань.

Постанова Кабінету Міністрів України "Про затвердження Положення про організацію та проведення заходів з охорони праці" (від 10 грудня 2003 року №

1913) - ця постанова встановлює загальні вимоги до організації та проведення заходів з охорони праці в підприємствах та організаціях.

Накази Державної служби України з надзвичайних ситуацій (ДСНС) та інших відповідних органів, які регулюють конкретні аспекти безпеки та охорони праці в різних сферах діяльності.

Це лише загальні приклади нормативно-правових актів, які стосуються охорони праці та безпеки в надзвичайних ситуаціях в Україні. При вирішенні конкретних питань, пов'язаних з цими питаннями, важливо враховувати чинне законодавство та консультуватися з фахівцями з охорони праці та безпеки.

5.2. Розрахунок заземлення в приміщеннях, де на серверах встановлена Інформаційна система «Інтерактивний тезаурус для дистанційного навчання»

Блискавкозахист — це комплекс захисних пристроїв, призначених для забезпечення безпеки людей, збереження будинків і споруджень, електронного устаткування і матеріалів від можливих вибухів, руйнувань і пожеж, що виникають від удару блискавки, а в будинках сільськогосподарських підприємств — також для забезпечення безпеки тварин і птахів.

Відповідно до курсу України на гармонізацію національної нормативної бази з міжнародною, прийнято чотири державні стандарти, а саме:

- ДСТУ EN 62305-1:2012 «Захист від блискавки. Частина 1. Загальні принципи» (EN 62305-1:2011, IDT);
- ДСТУ IEC 62305-2:2012 «Захист від блискавки. Частина 2. Управління ризиками» (IEC 62305-2:2010, IDT;)
- ДСТУ EN 62305-3:2021 «Захист від блискавки. Частина 3. Фізичні руйнування споруд та небезпека для життя людей» (EN 62305-3:2021, IDT, далі — ДСТУ EN 62305-3:2021);

- ДСТУ EN 62305-4:2012 «Захист від блискавки. Частина 4. Електричні та електронні системи, розташовані в будинках і спорудах» (EN 62305-4:2011, IDT).

Основним елементом блискавкозахисту є правильно спроєктоване заземлення. При виносній системі заземлення заземлювачі розташовуються на деякій відстані від заземленого обладнання. Тому заземлене обладнання знаходиться поза полем розтікання струму і людина, торкаючись його, опиниться під повною напругою відносно землі. Виносне заземлення захищає тільки за рахунок малого опору ґрунту.

При використанні заземлюючого пристрою одночасно для електроустановок напруга вище 1000 В мережі з ізольованою нейтраллю і для електроустановок до 1000 В з глухозаземленою нейтраллю, опір заземлюючого пристрою має бути не більше 4 Ом при лінійній напрузі 380 В.

Контур штучного заземлення овочесховища має форму прямокутника. Заземлювач передбачається виконати з сталевих електродів завдовжки 3,5 метри. Верхні кінці вертикальних електродів з'єднуються за допомогою горизонтального електроду - сталевої смуги розміром 50х4 мм, укладеної в землю на глибину 0, 7 м.

Початкові дані для розрахунку штучних заземлювачів зведені в табл. 5.1.

Таблиця 5.1 - Початкові дані для розрахунку захисного заземлення

Вид заземлення	**Виносне**
Довжина вертикального електроду *l*, м	3
Діаметр вертикального електроду, м	0,016
Глибина заставляння заземлювачів у грунт *h*, м	0,5
Питомий опір грунту *p*, Ом*м	50
Кліматична зона	II
Розміри горизонтального електроду *b* х *c*, мм	40 х 4
Опір заземлюючого пристрою *R3.n.*, Ом	4

Розрахунок заземлюючого пристрою робитимемо згідно ДСТУ.

Визначаємо значення електричного опору розтіканню струму в землю від поодинокого заземлювача:

$$R_з = \frac{\rho \cdot K_c}{2 \cdot \pi \cdot l}\left(\ln\frac{2 \cdot l}{d} + 0{,}5 \cdot \ln\frac{4t + l}{4t - l}\right),$$

де ρ – питомий опір грунту, Ом · м;

K_c – коефіцієнт сезонності, що враховує промерзання і просихання грунту, в нашому випадку рівний 2;

l – довжина вертикального електроду, м;

d – діаметр вертикального електроду, м;

t – відстань від поверхні грунту до середини вертикального електроду, м.

$$t = h + 0{,}5 \cdot l,$$

де h – глибина заставляння заземлювача в грунт, м

$$t = 0{,}5 + 0{,}5 \cdot 3 = 2\text{м};$$

$$R_з = \frac{50 \cdot 2}{2 \cdot 3{,}14 \cdot 3}\left(\ln\frac{2 \cdot 3}{0{,}016} + 0{,}5 \cdot \ln\frac{4 \cdot 2 + 3}{4 \cdot 2 - 3}\right) = 33{,}6\,\text{Ом}.$$

Розраховуємо число заземлювачів без урахування взаємних перешкод, що робляться заземлювачі один одному, так званим явищем взаємного екранування:

$$n' = \frac{R_{з.}}{R_{З.П}};$$

$$n' = \frac{33{,}6}{4} = 8{,}4 \approx 8\,\text{шт}.$$

Розраховуємо число вертикальних електродів з врахуванням екранування.

$$n = \frac{n'}{\eta_з}$$

де $\eta_з$ – коефіцієнт екранування.

Коефіцієнт екранування приймаємо, за умови, що відстань між вертикальними електродами а = l = 3 м (рис. 5.1).

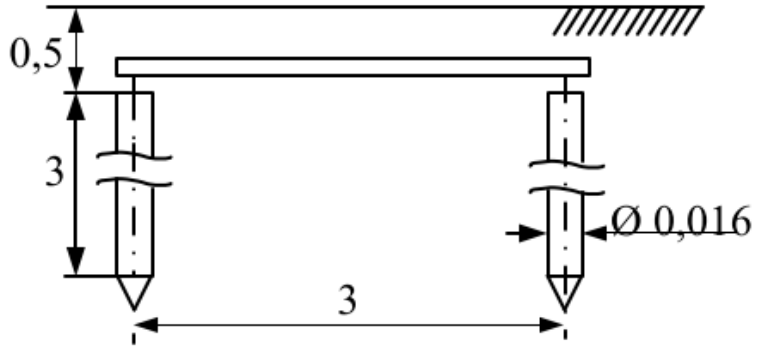


Рис. 5.1 - Схема розташування електродів.

$$n = \frac{n'}{0,49} = \frac{8}{0,58} = 13,8 \approx 14 \text{ шт.}$$

Визначаємо довжину сполучної смуги :

$$l_{П} = 1,05 \cdot n \cdot a;$$

$$l_{П} = 1,05 \cdot 14 \cdot 3 = 44,1 \text{ м.}$$

Розраховуємо повне значення опору заземлюючого пристрою :

$$R_{зп} = \frac{R_з \cdot R_п}{R_з \cdot \eta_п + R_п \cdot \eta_з \cdot n},$$

де $\eta_з$ – коефіцієнт екранування смуги, [24];

$$R_{зп} = \frac{33,6 \cdot 4,4}{33,6 \cdot 0,46 + 4,4 \cdot 0,58 \cdot 14} = 2,9 \text{ Ом.}$$

Опір $R_{3п}$ = 2,9 Ом менше ніж допустимий опір 4 Ом. Таким чином, розрахована система заземлення забезпечить захист при винесенні заземлювачів.

ВИСНОВКИ

Сучасний агропромисловий комплекс (АПК) працює в умовах суттєвої невизначеності. Тому актуальною для підприємств АПК залишається тема даної кваліфікаційної роботи, пов'язаної з підвищенням якості та повноти попередньої обробки даних щодо усіх аспектів їх функціонування.

У даній роботі на основі здійсненого аналізу методів і відповідних технологічних процесів заповнення пропусків в таблицях даних підприємств АПК показано, що найбільш перспективними та ефективними є методи інтелектуальної попередньої підготовки даних для подальшої їх обробки корпоративними інформаційними системами.

При цьому, здійснено:

- аналіз методів заповнення пропусків в таблицях даних та їх автоматизація;
- класифікація та дослідження типів пропущених даних та методів їх заповнення в таблицях підприємств АПК.
- дослідження особливостей, алгоритмів та програмного забезпечення поширених методів заповнення пропусків в таблицях даних, а також шести альтернативних традиційним методам заповнення пропусків в таблицях даних, зокрема: метод (Fuzzy C-means, FCM) видалення рядків з пропусками і заповнення їх статистичною вибіркою з урахуванням нечіткої функції залежності між елементами; метод (Singular value decomposition, SVD) сингулярного розкладу таблиці даних, який є доволі складним, але дозволяє побачити геометричну структуру таблиці і представити наявні дані; метод (Bayesian principal component analysis, bPCA) баєсовського принципу, що вимагає великої кількості даних, але з більшою точністю ніж PCA-метод визначає значення пропущених даних. Показано, що bPCA і FCM-методи належать до кращих для різних наборів даних і показників ефективності. Зокрема, FCM-метод кращий при застосуванні до невеликих наборів даних.

Досліджено сучасні нейромережні технології заповнення пропусків в таблицях даних.

Розроблено функціонал, структурні схеми, алгоритми та програмне забезпечення автоматизації процесу заповнення пропусків в таблицях даних підприємств АПК на основі нейромережної структури автоасоціативного типу.

Здійснено адаптацію нейроподібних мереж машини геометричних перетворень для заповнення пропусків в таблицях даних підприємств АПК

Розрахована економічна ефективність автоматизації технологічних процесів заповнення пропусків в таблицях даних підприємств АПК на основі аналізу капітальних витрат та розрахунку терміну її окупності.

Практична цінність виконаної роботи полягає у можливості використання її результатів підприємствами у різних галузях АПК для підвищенню якості попередньої обробки інформації їх функціонування.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Razniewski S. Completeness of queries over incomplete databases. 205. URL: https://doi.org/10.14778/3402707.3402715

2. Padmanabhan B. An Empirical Analysis of the Value of Complete Information for eCRM Models. 2016. URL: https://doi.org/10.2307/25148730

3. Ovsiuk N. Financial statements as the main source of information on the enterprise activity. 2021. URL: https://doi.org/10.37634/efp.2021.11(1).5

4. Cai Y. Evaluating Completeness of an Information Product : Proceedings of Americas Conference on Information Systems (AMCIS). 2013. URL: https://aisel.aisnet.org/amcis/?utm_source=aisel.aisnet.org%2Famcis2003%2F294&utm_medium=PDF&utm_campaign=PDFCoverPages

5. Bidyuk P. An Approach to Identifying and Filling Data Gaps in Machine Learning Procedures. 2021. URL: https://link.springer.com/chapter/10.1007/978-3-030-82014-5_11 t

6. Ткаченко П. Р. Нейромережний метод заповнення пропусків у таблицях фінансових показників. *Соц.-екон. пробл. сучас. періоду України*. 2018. С. 568-575.

7. Xue-dong G. A Method for Filling up the Missed Data in Information Table. 2019. URL: https://doi.org/10.58729/1941-6687.1287

8. Kaminskyi R. Recovery Gaps in Experimental Data. 2018. URL: https://ceur-ws.org/Vol-2136/10000108.pdf

9. Ustoorikar K. Filling up gaps in wave data with genetic programming. URL: https://doi.org/10.1016/J.MARSTRUC.2007.12.001

10. Ramli M. Roles of Imputation Methods for Filling the Missing Values: A Review. *Advances in Environmental Biology*. Vol. 7(12) Special Issue 2013. P.3861-3869

11. M. B. Mohammed. Comparison of five imputation methods in handling missing data in a continuous frequency table. 2021. URL: https://doi.org/10.1063/5.0053286

12. Madhu G. A Normalized Mean Algorithm for Imputation of Missing Data Values in Medical Databases. 2020. URL: http://dx.doi.org/10.1007/978-981-15-3172-9_72

13. Schmitt P. A Comparison of Six Methods for Missing Data Imputation. 2015. URL: http://dx.doi.org/10.472/2155-6180.1000224

14. Polishchuk U., Tkachenko P. Features of the auto-associative neurolike structures of the geometrical transformation machine (GTM). URL: https://ieeexplore.ieee.org/document/5069708

15. R. Tkachenko, P. Tkachenko,O. Tkachenko, J.Schmitz. Geometrical transformation machine // Комп'ютерні науки та інформаційні технології : Матеріали міжнародної конференції CSIT"2017. – Львів, 2017.- с. 52-54.