

ПРОЦЮК ВОЛОДИМИР

Івано-Франківський національний технічний університет нафти і газу

<https://orcid.org/0000-0003-0055-2780>e-mail: [v.v.protsiuk@gmail.com](mailto:v.v.protsiuk@gmail.com)

## МОДЕЛІ ВИЯВЛЕННЯ АНОМАЛІЙ ДЛЯ СЕНСОРНИХ ДАНИХ ПРОЦЕСУ БУРІННЯ НАФТОГАЗОВИХ СВЕРДЛОВИН В УМОВАХ НЕВИЗНАЧЕНОСТІ

Представлено питання дослідження та впровадження алгоритмічних моделей для ефективного виявлення аномалій у сенсорних даних, що стосуються процесу буріння нафтогазових свердловин. У зв'язку з постійним стрімким розвитком сучасних технологій та підвищенням вимог до безпеки та продуктивності у виробничих процесах нафтогазової промисловості, автор статті розглядає можливості використання передових методів аналізу даних, таких як машинне навчання та штучний інтелект. В статті розглядаються основні виклики, пов'язані з аналізом сенсорних даних у контексті буріння свердловин, та обговорюються потенційні підходи до їх вирішення. Зокрема, розглядаються алгоритми класифікації, кластеризації та аналізу часових рядів з метою виявлення незвичайних змін у важливих параметрах процесу буріння, таких як тиск у свердловині, температура на різних глибинах, швидкість обертання бурової колони та інші ключові показники. Подальший розвиток та застосування подібних алгоритмічних моделей може значно покращити процес моніторингу та управління бурінням, зменшити ризики аварій та сприяти більш ефективному реагуванню на потенційні загрози для персоналу та навколишнього середовища. В контексті цифрового родовища широко застосовуються різноманітні типи датчиків для збору величезних обсягів даних під час процесу буріння нафтогазової свердловини. Ці дані передаються у різні системи, такі як системи спостереження, моніторингу та управління процесом буріння. У таких системах ці дані використовуються для прийняття рішень або прогнозування подій у реальному часі. Однак, точність цих рішень або прогнозів значно залежить від надійності даних, які надходять від датчиків. Ефективність прийняття рішень і прогнозування динаміки процесу буріння сильно ускладнюється в контексті можливої невизначеності або помилковості цих даних. Виявленню аномалій у сенсорних даних приділяється значна увага, оскільки вони потребують фільтрації перед їх використанням. Для цього застосовується ряд алгоритмів для виявлення аномалій, проте лише деякі з них здатні враховувати невизначеності, пов'язані з сенсорними даними. Такі невизначеності можуть бути пов'язані з неповнотою, невідомістю, розмитістю, неточністю та неоднозначністю. У даній статті пропонується новий алгоритм виявлення аномалій сенсорних даних процесу буріння нафтогазових свердловин на основі правил з коефіцієнтами впевненості, а також розглядається механізм побудови інтелектуальних систем на основі правил з коефіцієнтами впевненості.

**Ключові слова:** аномалії в даних, машинне навчання, цифрове родовище, невизначеність, алгоритми виявлення аномалій, правила, коефіцієнти впевненості.

PROTSIUK VOLODYMYR

Ivano-Frankivsk National Technical University of Oil and Gas

## ANOMALY DETECTION MODELS OF FOR SENSOR DATA OF OIL AND GAS WELL DRILLING PROCESS UNDER UNCERTAINTY

The article presents the issues of research and implementation of algorithmic models for effective detection of anomalies in sensor data related to the process of oil and gas well drilling. Due to the constant rapid development of modern technologies and increasing requirements for safety and productivity in the production processes of the oil and gas industry, the author of the article considers the possibilities of using advanced data analysis methods such as machine learning and artificial intelligence. The article reviews the main challenges associated with the analysis of sensor data in the context of well drilling and discusses potential approaches to their solution. In particular, the article discusses algorithms for classification, clustering and time series analysis to detect unusual changes in important parameters of the drilling process, such as well pressure, temperature at different depths, drill string rotation speed and other key indicators. Further development and application of such algorithmic models can significantly improve the process of monitoring and controlling drilling, reduce the risk of accidents, and contribute to a more effective response to potential threats to personnel and the environment. In the context of a digital field, various types of sensors are widely used to collect huge amounts of data during the oil and gas well drilling process. This data is transmitted to various systems such as surveillance, monitoring and control systems for the drilling process. In such systems, this data is used to make decisions or predict events in real time. However, the accuracy of these decisions or predictions depends heavily on the reliability of the data coming from the sensors. The effectiveness of decision-making and prediction of drilling dynamics is greatly complicated in the context of possible uncertainty or error in this data. Anomaly detection in sensor data has received considerable attention, as it requires filtering before it can be used. A number of anomaly detection algorithms have been applied to this end, but only a few are able to account for the uncertainties associated with sensor data. Such uncertainties can be related to incompleteness, unknown, blur, imprecision, and ambiguity. This article proposes a new algorithm for detecting anomalies in sensor data of the oil and gas well drilling process based on rules with confidence factors, and also considers the mechanism for building intelligent systems based on rules with confidence factors.

Keywords: data anomalies, machine learning, digital oilfield, uncertainty, anomaly detection algorithms, rules, confidence factors.

### Постановка проблеми

В процесі буріння нафтогазових свердловин використовується цілий ряд датчиків для моніторингу різних параметрів. Зібрані сенсорні дані можуть бути використані в різних експертних системах для підтримки процесів прийняття рішень або для прогнозування динаміки процесу буріння. Бездротові датчики застосовуються через їх низьке енергоспоживання, низьку вартість та стандартизацію протоколів. Зазвичай експертні системи, які працюють з такими даними, стають у нагоді, коли досліджувані події швидко змінюються і їх неможливо передбачити заздалегідь. Таким чином, технології сенсорних мереж можуть бути

використані для збору великих обсягів даних, особливо в контексті цифрового родовища, і в кінцевому підсумку вони подаються в системи підтримки прийняття рішень для генерації різних сценаріїв рішень і прогнозування. Однак, точне і відповідне генерування сценаріїв цими системами є ненадійним через потенційно помилкову природу сенсорних даних. Причиною цього є те, що сенсорні дані можуть містити відсутні дані, дубльовані дані або суперечливі дані через обмеження ресурсів, таких як заряд батареї, обчислювальна потужність чи пропускна здатність каналу зв'язку. Таким чином, дані, що генеруються сенсорними вузлами, стають ненадійними і неточними. Також в суворих умовах, де давачі розміщені незахищеним чином, вони можуть виходити з ладу, що може теж призводити до пошкодження даних. Наявність відсутнього значення, дублікатів або неузгодженості з даними давачів призводять до створення різних типів невизначеностей, таких як неповнота, незнання, розпливчастість, неточність і неоднозначність. Наявність невизначеності в даних з давачів, спричиненої вищезгаданими факторами, може спричинити аномалії в даних з давачів. Таким чином, дані стають ненадійними, і якщо вони не будуть відфільтровані перед подачею в експертні системи, результати, отримані такими системами, можуть виявитися неточними. Тому необхідно використовувати відповідні методи для обробки аномальних даних з можливістю обробки різних типів невизначеності в інтегрованій структурі. Для забезпечення надійності та точності сенсорних даних у будь-якій експертній системі необхідно використовувати методи виявлення аномалій. Аномалії - це патерни в даних, які відхиляються від чітко визначеної нормальної поведінки. Різні методи виявлення аномалій включають параметричні (статистичні) та непараметричні методи. У параметричних методах дані аналізуються з використанням розподілу щільності, де дані з низькою релевантністю порівнюються з розподілом, і ті, що не вписуються, вважаються аномаліями. Прикладом параметричного методу є багатовимірний метод Гауса, який використовує статистичні моделі. Непараметричні методи, з іншого боку, базуються на правилах, які генеруються на основі даних. Кожному правилу надається вагове значення залежно від частоти його використання, а аномальні дані визначаються за допомогою певних порогових значень. Прикладами непараметричних методів є асоціативний та нечіткий видобуток правил на основі асоціативної бази правил. Важливо враховувати, що ці методи не завжди включають явища невизначеності сенсорних даних, тому необхідно уважно вибирати підхід в залежності від конкретної задачі та характеру даних.

Отже, задача дослідження полягає в побудові нового алгоритму виявлення аномалій на основі правил, який здатний обробляти згадані види невизначеності в інтегрованій структурі, що дозволить точно виявляти аномальні дані та фільтрувати їх перед подачею на вхід системи для точного прогнозування подій, а також оцінки ризиків, пов'язаних з ними.

#### Аналіз останніх джерел

Дослідження з виявлення аномалій в даних [1–11] тривають вже давно, особливо в галузі статистики. Основний принцип цього методу полягає в тому, що аномальні дані слід виявляти за допомогою параметричного або гаусівського розподілу [12, 13], а також за допомогою функції щільності ймовірності [14–15]. У цьому методі остання використовується для обчислення оцінки аномальності даних. Потім використовується порогове значення для визначення аномальних даних на основі оцінки аномалії. У методі виявлення аномалій на основі Гауса припускається, що набір даних підпорядковується розподілу Гауса. Набір даних, який є одноmodalним, симетричним, асимптотичним за своєю природою, зазвичай має нормальний розподіл. Якщо набір даних не може повністю слідувати розподілу, то можна помітити неточність у виявленні аномалій. Ця неточність спричиняє невизначеність у виявленні аномалій. Тому статистичні алгоритми виявлення аномалій, такі як гаусівський розподіл, не враховують невизначеність. Крім того, всі набори даних не можуть бути змодельовані за допомогою гаусівського розподілу, якщо точки даних не згруповані навколо середнього значення набору даних. Крім того, пороговий параметр може бути важко визначити, оскільки різниця між аномальними та нормальними даними може бути дуже близькою. Більше того, якщо набір даних є асиметричним і біmodalним, то належне виявлення аномальних даних важко отримати за допомогою гаусівського розподілу. Однак природа сенсорних даних [16–17] є асиметричною або біmodalною. Тому статистичний підхід до виявлення аномалій не буде ефективним для виявлення аномалій. Натомість, підхід, заснований на знаннях, що базується на частоті точок даних у наборах даних, забезпечує краще виявлення аномальних даних. Оскільки дані з давачів є асиметричними за своєю природою, визначення частоти даних може бути використано для розробки правил. Це, в свою чергу, може сформувати базу знань і, таким чином, може бути використано для виявлення аномалій у даних давачів за допомогою різних підходів, що базуються на знаннях.

Як правило, при виявленні аномалій на основі правил або знань, детектор аномалій використовує задалегідь визначені правила для класифікації точок даних як аномалій або нормальних даних.

Правило асоціації [18–20] – це підхід, що базується на правилах для видобування даних. Метод виявлення аномалій на основі асоціативних правил заснований на виявленні частих шаблонів у наборі даних. У цьому методі кожна точка даних у наборі даних розглядається як транзакція. Таким чином, транзакції, які містять менш часті шаблони, визначаються як аномалія. Для виявлення аномальних транзакцій цей метод визначає міру, яка називається FPOF (Frequent Pattern Outlier Factor). Однак цей метод може добре працювати з точними даними, а отже, він не дуже добре підходить там, де природа даних містить нечіткість. Крім того, дані з давачів містять різні типи невизначеності, такі як незнання, неповнота, неоднозначність, розпливчастість і неточність з причин, описаних у попередньому розділі. Таким чином, за допомогою цього

методу не можна видобути відповідні правила, а отже, виявлення аномалії виходить з даних давачів.

Сенсорні дані можна розглядати як великий обсяг реальних даних, зібраних з сенсорних вузлів [21, 22]. Характеристики цих даних залежать від атрибутів даних, а також від кореляції між даними у просторі та часі. Кожен сенсорний вузол може мати один або декілька давачів. Наприклад, вузол давачів з одним датчиком, можна вважати таким, що надає одномірно атрибутовані дані. З іншого боку, сенсорний вузол, що складається з двох датчиків можна вважати багатомірними атрибутивними даними. Порівняно легше виявити відхилення з одномірних атрибутивних даних, оскільки потрібно розглядати один тип даних. Однак для виявлення аномалії в багатомірних атрибутивних даних необхідно розглядати декілька типів даних разом. Крім того, на виявлення аномалій у даних давачів також впливає спеціальна та часова кореляція із зібраними даними. Тимчасова кореляція означає, що зчитування даних з давачів в один момент часу пов'язане з попереднім моментом часу. Спеціальна кореляція, навпаки, передбачає, що існує кореляція між даними, зібраними з географічно близько розташованих давачів.

У правилах нечіткої асоціації точки даних перетворюються на нечіткі значення за допомогою функції належності. Потім генеруються правила нечіткої асоціації на основі частих точок даних або рідкісних даних. За допомогою згенерованих правил виявляються аномальні дані з даних давачів. Нечіткі множини долають проблему переоцінки або недооцінки граничних значень за допомогою функції належності. Нечітка логіка [23–24] здатна впоратися з невизначеністю, зумовленою неточністю, неоднозначністю та розмитістю, але не з іншими.

Однак метод Гауса, який є статистичним підходом, не здатний впоратися з невизначеністю через випадковість, та нечіткість. З іншого боку, підхід, що базується на правилах, наприклад, правилі асоціацій, використовує знання, які можуть бути оцінені як істинні або хибні. Таким чином, цей підхід теж не може впоратися з невизначеністю через нечіткість, незнання або неповноту. Нечітка логіка може впоратися з невизначеністю через нечіткість, але не може впоратися з незнанням і неповнотою. Вона також не здатна впоратися з невизначеністю через незнання в нечіткості. Отже, жоден зі згаданих методів не може впоратися з усіма типами невизначеності в інтегрованій структурі.

Таким чином, алгоритм виявлення аномалій на основі Гауса надає механізм для виявлення аномалій з багатомірних даних давачів без будь-яких попередніх знань про ці дані. Алгоритм припускає, що дані з давачів підпорядковуються нормальному або гаусівському розподілу. Однак це не так для всіх даних давачів, і в такому випадку алгоритм виявлення аномалій на основі гаусівського розподілу не може ефективно виявляти аномальні дані. Крім того, алгоритм не має механізму для виявлення та усунення невизначеності через незнання, неповноту, двозначність, нечіткість та неточність. Виявлення аномалій на основі асоціативних правил не залежить від розподілу даних. Однак, йому також не вистачає роботи з невизначеністю. Нечіткі асоціативні правила надають механізм для вирішення проблеми переоцінки або недооцінки граничних значень за допомогою функції належності. Ці методи здатні впоратися з невизначеністю, зумовленою неточністю, неоднозначністю та розмитістю, але не з іншими характеристиками, оскільки використовуються нечіткі множини.

Тому потрібен новий алгоритм для вирішення всіх типів невизначеності, які існують у даних з давачів, за допомогою інтегрованого фреймворку для виявлення аномальних даних з можливістю обробки різних типів невизначеності, таких як незнання, неповнота, неоднозначність, розпливчастість і неточність для виявлення аномальних сенсорних даних [25–27].

#### Виклад основного матеріалу

Керування буровою установкою [28] під час процесу буріння нафтогазових свердловин є складною задачею через багатомірність об'єкта, неоднорідність умов та вплив зовнішніх факторів. Бурова установка взаємодіє з навколишнім середовищем, враховуючи апріорну та поточну невизначеність. Керування включає регулювання кількох параметрів для забезпечення ефективності.

Перша група параметрів – це ті, які можна регулювати на вході у об'єкт: напрямна сила на долото, швидкість обертання долота, об'єм і характеристики промивальної рідини. Друга група параметрів – це ті, які не можна регулювати і їх значення залежить від умов буріння: тип і діаметр бурового долота, глибина свердловини, параметри бурильної колони, геометрія долота. Ці параметри визначають умови процесу. До цієї групи також відносяться зовнішні впливи, такі як опір статичного тиску, тертя колони бурильних труб, властивості гірських порід тощо. Третя група параметрів – це ті, що залежать від інших вхідних параметрів: механічна швидкість буріння, просування долота, зношення оснащення та інше. Аналіз цих параметрів цими параметрами вимагає точного контролю та аналізу у реальному часі для досягнення оптимальних результатів у бурінні.

Важливою складовою дослідження є опис практичних випадків застосування розглянутих алгоритмів у реальних умовах буріння нафтогазових свердловин. Це може включати аналіз даних з метою виявлення потенційно небезпечних умов або аварійних ситуацій, а також вдосконалення систем моніторингу та контролю процесу буріння.

Нехай  $TP_{set} = \{tp_1, tp_2, \dots, tp_l\} \in N$  є множиною з  $l$ -літералів, які будемо розглядати як режимні параметри буріння  $DB_{drill} = \{db_{tr}^1, db_{tr}^2, \dots, db_{tr}^m\}_{m+N}$  множини з  $m$ -транзакцій, кожна з яких складається множини режимних параметрів  $TP_{set}$ . Розглянемо набір елементів в формі режимних параметрів  $X$ , що є не підмножиною для  $TP_{set}$ . Відповідно довжина набору даних  $X$  визначається відповідно щільності елементів режимних параметрів в  $X$ . Тоді набір елементів довжини  $n$  будемо називати  $n$ -набором режимних

параметрів. В такому випадку транзакція  $db_{tr} \in DB_{drill}$  вважатиметься такою, що містить набір  $X$ , якщо  $db_{tr} \supseteq X$ . Підтримка набору даних  $X$  визначається як відповідна метрика:

$$supp_{set}^{item}(x) = \frac{db_{tr} \in DB_{drill} | X \subseteq db_{tr}}{db_{tr} \in DB_{drill}} \quad (1)$$

Бінарне правило асоціації створюється на основі частих наборів елементів транзакцій, що відбуваються в базі даних. Набори елементів - це колекція елементів, доступних у базі даних. Існує два основних параметри, а саме підтримка та достовірність. Підтримка може бути визначена як частота наборів елементів у всій базі даних, поділена на кількість транзакцій. Довіру можна визначити як частоту наборів елементів у правилі, поділену на частоту наборів елементів у попередній частині правила.

Метрика підтримки для асоціативного правила задаватиметься через:

$$supp_{set}^{item}(RC \Rightarrow SC) = supp_{set}^{item}(RC \cup SC) \quad (2)$$

де  $RC$  (*relevancy condition*) – умова релевантності,  $SC$  (*satisfaction condition*) – умова задоволення.

Метрика впевненості асоціативного правила задаватиметься відповідно як:

$$conf.metr(RC \Rightarrow SC) = \frac{supp_{set}^{item}(RC \cup SC)}{supp_{set}^{item}(RC)} \quad (3)$$

Тоді бінарне асоціативне правило матиме вигляд:

$$RC \Rightarrow SC \quad (4)$$

де  $RC$  і  $SC$  позначають відповідні набори даних щодо значень релевантності режимних параметрів та умов задоволення накладених обмежень.

$$POF(db_{tr}) = \frac{\sum_x [supp_{set}^{item}(x)]}{\|FPS(DB_{drill}, [supp_{set}^{item}]_{min})\|} \quad (5)$$

де  $X \subseteq db_{tr}$ ;  $X \in m(DB_{drill}, [supp_{set}^{item}]_{min})$ ;

$FPS$  (*frequently pattern set*) – набір частих патернів  $FPS(DB_{drill}, [supp_{set}^{item}]_{min})$ .

У бінарних асоціативних правилах підтримка обчислюється шляхом обчислення частоти елементів. Правило асоціації оцінюється як істинне або хибне, а отже, не дає можливості врахувати будь-які типи невизначеності. Однак, сенсорні дані містять різні типи невизначеності, і, таким чином, просте правило асоціації не підходить для виявлення аномалії в сенсорних даних. Крім того, знаходження правил зв'язку за даними давачів з кількісними атрибутами є проблематичним через поганий семантичний зміст для визначення сенсорних даних, що створює нечіткість. Нечіткі асоціативні правила створюються на основі кількісних даних, в яких кожен кількісний елемент перетворюється на нечітку множину, а нечіткі операції використовуються для знаходження нечітких асоціативних правил.

$$(x_i \text{ is } [rc_1]) \text{ AND } (x_i \text{ is } [rc_2]) \Rightarrow (y_i \text{ is } [sc_k]) \quad (6)$$

Кількісні значення з сенсорних даних представляються за допомогою лінгвістичних міток або еталонних значень нечіткими множинами в процесі видобування нечітких правил асоціації з сенсорних даних. Наприклад, значення атрибуту сенсорних даних, такого як *швидкість обертання долота*, можуть бути представлені за допомогою різних лінгвістичних міток, таких як *дуже висока*, *висока*, *середня* та *низька* у нечіткій множині. Це допомагає представити семантичний зміст сенсорних даних більш ефективно, ніж бінарне правило асоціації, надаючи значущі лінгвістичні мітки сенсорних даних. Більше того, використовуючи нечіткі функції приналежності нечітких множин, можна вирішити проблему переоцінки або недооцінки граничних значень бінарного правила асоціації, дозволивши часткову приналежність до різних нечітких множин.

Розглянемо базу даних  $DB_{drill}$ , що складається з множини транзакцій  $RC_{tr}$ .

$$Rank_{DB_{drill}}(SC) = \frac{\sum_{trid=0}^{|DB_{drill}|} rank(RC_{trid}, SC)}{|DB_{drill} \cdot SC|} \quad (7)$$

де  $DB_{drill} \cdot SC$  – це підмножина транзакцій, що покривається нечітким правилом для  $SC$  у базі даних  $DB_{drill}$ .

Метрика підтримки визначається виразом:

$$supp_{DB_{drill}}(SC) = \frac{\sum_{trid=0}^{|DB_{drill}|} sup(RC_{trid}, SC)}{|DB_{drill}|} \quad (8)$$

де  $|DB_{drill}|$  – загальна кількість транзакцій.

Крім того, бінарні правила асоціації чутливі до невеликих змін значень, що є регулярним явищем у сенсорних даних. Правило асоціації також має тенденцію переоцінювати або недооцінювати граничне значення під час процесу перетворення бази даних транзакцій у бінарну базу даних шляхом розбиття значень атрибутів. Алгоритми виявлення аномалій для сенсорних даних в умовах невизначеності розглядають лише один атрибут за раз. Тому він не може врахувати всі типи невизначеності в даних з давачів, а також це призводить до генерації більшої кількості правил. Механізм виведення нечітких асоціативних правил не має жодних можливостей для запобігання невизначеності в даних давачів, таких як неповнота. Однак не виключено, що датчик не може надіслати дані через обмеженість мережевих ресурсів або зловмисну атаку,

що спричиняє невизначеність, таку як неповнота. Тому правило нечіткої асоціації не повністю підходить для виявлення аномалій. База правил впевненостей є розширенням традиційної бази правил IF-THEN. Правило впевненостей має антецедентну частину та консеквент. Атрибут антецедент приймає референтні значення, та можливі ступені вірогідності пов'язані з наслідком правила вірогідності. Вага правила, вага попереднього атрибуту та ступінь вірогідності - це параметри представлення знань, які використовуються в БЗ для відображення невизначеності. Правило є повним, якщо сума ступенів вірогідності, пов'язаних з кожним референтним значенням консеквентного атрибуту, дорівнює одиниці. Якщо сума менше одиниці, то правило вважається неповним, що може бути пов'язано з неповною інформацією або незнанням. Зв'язок між попередніми атрибутами та наступним атрибутом є нелінійним, тоді як у випадку правила IF- THEN він є лінійним.

Правило переконання можна визначити так:

$$\begin{aligned} &IF [Drill.Bitload] is High AND \\ &[Drill.BitRotationalSpeed] is Medium THEN \\ &[Drill.Mode] is \{(Rational, 0.1), (Optimal, 0.3), (Forced, 0.7)\} \end{aligned} \quad (9)$$

Загалом, сенсорні дані, які збираються з навколишнього середовища, є нелінійними за своєю природою. Таким чином призначені для представлення у короткочасній пам'яті, тоді як база правил знаходиться у довготривалій пам'яті. Загальний ступінь або комбінований ступінь відповідності до якого вхідні дані відповідають всій прецедентній частині  $k$ -го правила, можна обчислити за допомогою відповідного виразу. Слід відзначити, що кожне правило не має однакової ваги при обчисленні референтних значень атрибуту послідовності. Ця вага активації дорівнює нулю, якщо  $k$ -те правило не активоване. Коли вхідні дані для будь-якого з антецедентів ігноруються або відсутні, то ступінь правдоподібності, пов'язаний з кожним правилом в базі правил, повинен бути оновлений. Тому в процедурі оновлення впевненостей ступінь впевненостей кожного правила оновлюється. Механізм виведення використовується для генерації правил довіри на основі даних з давачів.

У процесі роботи над моделлю було виконано кілька кроків, включаючи перетворення вхідних даних, обчислення ваги активації правила, оновлення впевненостей та агрегування правил з використанням методу доказових міркувань. Завдання перетворення вхідних даних полягає в розподілі вхідних даних за сталонними значеннями атрибуту правила, який називається ступенем відповідності. Після того, як ступінь відповідності визначено, правила стають антецедентами пакетів.

$$rc_k = aggr((CF_{k1}, rc_1^k), \dots, (CF_{k\ tr_k}, rc_{tr_k}^k)) \quad (10)$$

$$rc_k = \prod_{i=1}^{tr_k} (rc_i^k)^{\overline{CF}_{k_i}} \quad (11)$$

де  $\overline{CF}_{k_i} = \frac{CF_{k_i}}{\max_{i=1, \dots, tr_k} \{CF_{k_i}\}}$  при умові що  $0 \leq \overline{CF}_{k_i} \leq 1$ ;

$tr_k$  – щільність антецедентних атрибутів у  $k$ -му правилі.

Вага активації  $weight_k$  для  $k$ -го правила буде:

$$weight_k = \frac{[RuleWeight]_k rc_k}{\sum_{i=1}^k [RuleWeight]_i rc_i} \quad (12)$$

де  $[RuleWeight]_k$  представляє вагу правила, а  $rc_k$  об'єднує ступінь релевантності для  $k$ -го правила.

$$sc_{i_k} = \overline{sc}_{i_k} \frac{\sum_{tr=1}^{tr_k} (\Delta(tr, k)) \sum_{j=1}^{tr} ([rc_{tr}]_j)}{\sum_{t=1}^{tr_k} (\Delta(tr, k))} \quad (13)$$

де

$$\Delta(tr, k) = \begin{cases} 1 & \text{якщо } tr_{def}(Rule_k(tr = 1, \dots, tr_k)) \\ 0 & \text{інші} \end{cases}$$

де  $\overline{sc}_{i_k}$  – представляє початковий ступінь довіри, в той час як  $sc_{i_k}$  є оновленим ступенем довіри для  $k$ -го правила,  $rc_{tr_j}$  – представляє ступінь довіри, що вхідне значення належить атрибуту.

$$rc_j = \frac{\theta \times [\prod_{k=1}^L (weight_k \theta_{jk} + 1 - weight_k \sum_{j=1}^N \theta_{jk}) - \prod_{k=1}^L (1 - weight_k \sum_{j=1}^N \theta_{jk})]}{1 - \theta \times [\prod_{k=1}^L (1 - weight_k)]} \quad (14)$$

де

$$\theta = \left[ \sum_{j=1}^N \prod_{k=1}^L \left( weight_k rc_{jk} + 1 - weight_k \sum_{j=1}^N rc_{jk} \right) - (N - 1) \times \prod_{k=1}^L \left( 1 - weight_k \sum_{j=1}^N rc_{jk} \right) \right]^{-1}$$

Кінцеві значення можуть бути перетворені в чіткі значення за допомогою оцінки корисності, пов'язаної з кожним референтним значенням, для отримання кінцевого результату. Отже, підсумовуючи, ступені вірогідності еталонних значень виразів мають дорівнювати одиниці, якщо наявні всі дані давачів для прецедентної частини правил, які враховують невизначеність через неповноту. Вираз враховує невизначеність, пов'язану з незнанням або відсутністю значень від давачів, шляхом оновлення ступеня вірогідності кожного з правил під час процедури оновлення вірогідності. Крім того, невизначеність, пов'язана з нечіткістю, неточністю та неоднозначністю, враховується у виразі під час процесу агрегування правил. Дані з давачів, в свою чергу, будуть виступати в якості початкових правил впевненостей для представлення бази знань експертної системи. Отже, необхідно дослідити відповідні методи виведення. Однак, у світлі процедур виведення на основі правил впевненостей охоплюють початкові правила впевненостей. Задача перетворення вхідних даних може бути виконана шляхом розробки бази даних вхідних транзакцій у вигляді перетворення бази даних транзакцій впевненостей у базу даних транзакцій ступенів довіри. Оскільки в цьому випадку використовуються як бінарні, так і нечіткі правила, необхідно розробити процедуру для обчислення підтримки бази даних транзакцій впевненостей.

Однак, типові процедури виведення, які складаються з перетворення вхідних даних, обчислення ваги активації правил, оновлення впевненостей та агрегації правил бази правил впевненостей, не можуть бути безпосередньо застосовані для виявлення правил впевненостей на основі даних з давачів. Причиною цього є те, що у випадку сенсорних даних немає необхідності мати початкову базу правил, *оскільки метою інтелектуального аналізу сенсорних даних є виявлення наборів правил впевненостей розрахунок ступенів довіри до бази даних транзакцій впевненостей, що може бути досягнуто шляхом розробки матриці впевненостей та розрахунку відстані Хеммінга.* Це дозволить обчислити достовірність кожної транзакції бази даних впевненостей. Зрештою, для даних з давачів можна знайти правила асоціації впевненостей, які виступатимуть в якості початкової бази правил впевненостей для експертної системи. Можна продемонструвати, що за допомогою правил асоціації впевненостей можна усунути аномалії значень достовірності з сенсорних даних.

На першому кроці дані з давачів, серед яких ми хочемо знайти аномалію, в кожній точці даних присвоюється ідентифікатор. Відтепер кожна транзакція заноситься до бази даних транзакцій. Потім база даних транзакцій перетворюється в базу даних транзакцій впевненостей за допомогою вхідного перетворення. На наступному кроці, який називається обчислення підтримки, розраховується підтримка даних з давачів. Згодом створюється матриця впевненостей. Потім обчислюється відстань Хеммінга, щоб знайти відмінності між транзакціями. Обчислюється довіра до кожної операції. Використовуючи матрицю впевненостей, а також на основі значень довіри, визначаються правила асоціації на основі правил впевненостей, які не містять аномальних даних з давачів і, таким чином, можуть бути використані як початкові правила впевненостей в базі правил. Це демонструє новий спосіб вилучення правил впевненостей з сенсорних даних.

Таблиця 1

Приклад бази даних транзакцій

Номер транзакції	Глибина	Навантаження на долото	Кількість обертів долота	Швидкість буріння
$t_1$	3305.0	26217.863999999900	1.314720300000	0.004088054000
$t_2$	3310.0	83492.293000000000	1.328673900000	0.005159383000
$t_3$	3315.0	97087.882000000000	1.420116300000	0.005971469300
$t_4$	3320.0	54793.206000000000	1.593930700000	0.005419108900
$t_5$	3325.0	50301.579000000000	1.653262000000	0.005435248300
$t_6$	3330.0	56835.194000000000	1.657450000000	0.004738274000
$t_7$	3335.0	37638.028000000000	2.139238300000	0.008811767200
$t_8$	3340.0	37638.028000000000	2.139238300000	0.008811767200
$t_9$	3345.0	44564.583000000000	1.974806400000	0.008456208300
$t_{10}$	3350.0	60928.602999999900	1.735741900000	0.007391549200

Таблиця 2

Орієнтовні значення навантаження на долото

Навантаження на долото	Низький рівень	Середній рівень	Високий рівень	Надвисокий рівень
Орієнтовне значення	10 000	40 000	70 000	100 000

Таблиця 3

## Орієнтовні значення кількості обертів долота

Кількість обертів долота	Низький рівень	Середній рівень	Високий рівень	Надвисокий рівень
Орієнтовне значення	1.2	1.7	2.2	2.7

Крім того, для підтримки видобування сенсорних даних необхідно розробити новий спосіб для обчислення підтримки та достовірності, як обговорювалося під час попереднього формування бінарних та нечітких правил асоціації. Однак ці методи обчислення підтримки та довіри не можуть враховувати різні типи невизначеності. Отже, необхідно розробити нові методи, які б враховували різні типи невизначеності при обчисленні підтримки та достовірності. Таким чином, це дослідження демонструє нові методи розрахунку підтримки та довірчої ймовірності шляхом включення різних типів невизначеності, як буде продемонстровано нижче.

Вхідна база даних транзакцій буде містити всі дані з давачів. Щоб ідентифікувати дані з давачів, кожному рядку даних присвоюється унікальний ідентифікаційний номер, який називається ідентифікатором транзакції. Однак, дані з давачів мають кількісну природу. Тому ці дані містять невизначеність, таку як неточність, розмитість. Більше того, ці дані також є семантично бідними. Тому, щоб усунути вищезазначені невизначеності та вирішити проблему поганого семантичного наповнення вводяться еталонні значення та лінгвістичні мітки. Дані з бази вхідних транзакцій будуть використані для отримання бази транзакцій впевненостей, яка міститиме дані сенсорів з референтними значеннями. В якості лінгвістичних міток та референтних значень на вхід беруться транзакційні бази даних, а потім формується сенсифікована база даних.

$$NormValue(x_i + 1) = \frac{|x_i - rc|}{|x_i - x_{i+1}|} \quad (15)$$

$$NormValue(x_i) = 1 - NormValue(x_i + 1) \quad (16)$$

де  $x_i < rc < x_{i+1}$

$$\theta_{jk} \geq 0, \sum_{j=1}^N \theta \leq 1 \quad (17)$$

Отримані дані перетворюються в еталонні значення за допомогою функції корисності. Це полегшує обчислювальну процедуру розрахунку підтримки. Цей крок дозволяє врахувати невизначеність, пов'язану з неоднозначністю, нечіткістю та неточністю сенсорних даних, шляхом розподілу ступеня вірогідності на еталонні значення. Перетворення вхідної бази даних транзакцій у базу даних транзакцій впевненостей нагадує вхідне перетворення механізму виведення бази правил впевненостей. Згодом еталонні значення використовуються для обчислення опорних значень сенсорних даних, що є важливим кроком для видобутку аномалій.

База даних транзакцій впевненостей може бути визначена як сукупність еталонних значень сенсорних даних. Дані показано в таблицях 4 та 5 відповідно. У стовпчику 1 таблиць 4 і 5 вказано ідентифікатор транзакції. Назва атрибуту та значення датчика наведені у стовпчику 2 таблиць 4 та 5.

Таблиця 4

## Приклад бази даних транзакцій на основі впевненостей експерта для навантаження на долото

Номер транзакції	Атрибут	Низький рівень	Середній рівень	Високий рівень	Надвисокий рівень
t1	atr1 → 26217	0.459404533	0.540595467	0	0
t2	atr1 → 83492	0	0	0.5502569	0.4497431
t3	atr1 → 97087	0	0	0.097071	0.902929
t4	atr1 → 54793	0	0.506893	0.493107	0
t5	atr1 → 50301	0	0.656614	0.343386	0
t6	atr1 → 56835	0	0.438827	0.561173	0
t7	atr1 → 37638	0.078732	0.921268	0	0
t8	atr1 → 37638	0.078732	0.921268	0	0
t9	atr1 → 44564	0	0.847847	0.152153	0
t10	atr1 → 60928	0	0.30238	0.69762	0

Таблиця 5

**Приклад бази даних транзакцій на основі впевненостей експерта для кількості обертів долота**

Номер транзакції	Атрибут	Низький рівень	Середній рівень	Високий рівень	Надвисокий рівень
t1	atr2 → 1.3147203	0.770559	0.229441	0	0
t2	atr2 → 1.3286739	0.742652	0.257348	0	0
t3	atr2 → 1.4201163	0.559767	0.440233	0	0
t4	atr2 → 1.5939307	0.212139	0.787861	0	0
t5	atr2 → 1.6532620	0.093476	0.906524	0	0
t6	atr2 → 1.6574500	0.0851	0.9149	0	0
t7	atr2 → 2.1392383	0	0.121523	0.878477	0
t8	atr2 → 2.1392383	0	0.121523	0.878477	0
t9	atr2 → 1.9748064	0	0.450387	0.549613	0
t10	atr2 → 1.7357419	0.071484	0.928516	0	0

Розрахунок підтримки визначається як функція даних давачів і референтних значень щодо бази даних транзакцій впевненостей. Дані з давачів та референтні значення беруться як вхідні дані для розрахунку підтримки, а частота даних з давачів по відношенню до бази даних транзакцій впевненостей надається послідовно. Алгоритми виявлення аномалій на основі бінарних та нечітких асоціативних правил використовують функцію підтримки для знаходження ймовірності набору елементів у бази даних. У випадку **DrillWiseIfPro** еталонні значення сенсорних даних також включаються в розрахунок підтримки.

$$supp_{set}^{item}(X_i) = \frac{\sum_{i,j}^n X_i * X_i.value_j}{|ID.Number|} \tag{18}$$

Таблиця 6

**Обчислення значення підтримки для навантаження на долото та кількості обертів долота**

Навантаження на долото		Кількість обертів долота	
Значення	Підтримка	Значення	Підтримка
26217.863999999900	2621.786	1.314720300000	0.131472
83492.293000000000	8349.229	1.328673900000	0.132867
97087.882000000000	9708.788	1.420116300000	0.142012
54793.206000000000	5479.321	1.593930700000	0.159393
50301.579000000000	5030.158	1.653262000000	0.165326
56835.194000000000	5683.519	1.657450000000	0.165745
37638.028000000000	3763.803	2.139238300000	0.213924
37638.028000000000	3763.803	2.139238300000	0.213924
44564.583000000000	4456.458	1.974806400000	0.197481
60928.602999999900	6092.86	1.735741900000	0.173574

Таким чином, підтримка може враховувати такі невизначеності, як неповнота, незнання, розпливчастість, неточність і неоднозначність. Крім того, правило нечіткої асоціації враховує лише одне з референтних значень як показано в таблицях 4 і 5, і цей атрибут пов'язаний зі ступенем вірогідності. Включення цього явища до правил асоціації впевненостей дає змогу вирішувати проблеми незнання та неповноти даних з давачів. Розрахунок підтримки допомагає врахувати невизначеність сенсорних даних, таких як неточність, неоднозначність і розпливчастість, оскільки шуканий вираз використовує еталонні значення.

Матрицю довіри можна визначити як комбінацію ступенів довіри до еталонних значень та допоміжних значень сенсорних даних. База даних транзакцій довіри та опорні значення сенсорних даних використовуються як вхідні дані для цієї процедури. Результатом цієї процедури є значення комірок матриці довіри, отримані шляхом множення відповідних ступенів довіри та опорних значень атрибутів у сенсорних даних, який використовується як вхідні дані для обчислення довіри. Оскільки значення сенсорних даних, які є кількісними за своєю природою, розподіляються між різними еталонними значеннями для подолання семантичної розрідженості сенсорних даних. Ступені достовірності, прикріплені до еталонних значень, що відповідають сенсорним даним, враховують невизначеність через неоднозначність, неточність і розпливчастість. Однак це не може усунути невизначеність через неповноту, а отже, ступені довіри, пов'язані з еталонними значеннями, необхідно помножити на відповідні опорні значення даних давачів, щоб усунути



невизначеність через неповноту. Таким чином, за допомогою такого виразу можна сформувати матрицю впевненостей.

$$CF\_Matrix\_Item_{i,j} = \text{supp}(rc_{k_i}) \times CF.DB_{drill_{k_j,x}} \tag{19}$$

*Впевненість* - це оцінка ступеня достовірності виявленого зв'язку між антецедентом і консеквентом правила. Активація правила, механізму виведення на основі правил впевненостей дуже схожа на обчислення впевненості. Однак, комбінований ступінь відповідності, для активації правила обчислюється за допомогою мультиплікативної функції агрегування. Оскільки вона не підходить для сенсорних даних через свою природу, її можна замінити популярною мірою схожості, яка називається відстанню Хаммінга. Крім того, відстань Хаммінга підходить для роботи з сенсорними даними, оскільки вона призначена для роботи з кількісними даними, що є характерною особливістю сенсорних даних. Таким чином, достовірність правила асоціації на основі правил впевненостей можна визначити як функцію відстані Хаммінга транзакцій та загальної суми відстаней Хаммінга всіх транзакцій матриці впевненостей.

Таблиця 7

**Матриця впевненостей експерта для навантаження на долото та кількості обертів долота**

<i>t1</i>	<i>t2</i>	<i>t3</i>	<i>t4</i>	<i>t5</i>	<i>t6</i>	<i>t7</i>	<i>t8</i>	<i>t9</i>	<i>t10</i>
1204.461	0	0	0	0	0	296.3332	296.3332	0	0
1417.326	0	0	2777.43	3302.872	2494.081	3467.47	3467.47	3778.396	1842.358
0	4594.221	942.4379	2701.891	1727.286	3189.438	0	0	678.0625	4250.502
0	3755.008	8766.35	0	0	0	0	0	0	0
0.101307	0.098674	0.079493	0.033813	0.015454	0.014105	0	0	0	0.012408
0.030165	0.034193	0.062518	0.12558	0.149872	0.15164	0.025997	0.025997	0.088943	0.161166
0	0	0	0	0	0	0.187927	0.187927	0.108538	0
0	0	0	0	0	0	0	0	0	0

$$rc_{t_j,k} = \sum_{j=1}^n hd(tr_j) \tag{20}$$

$$\eta_k = \sum_{i=1}^n rc_{t_i,k} \tag{21}$$

$$CF(tr_i) = \frac{\sum_{k=1}^n rc_{t_i,k}}{\sum_{k=1}^n \eta_k} \tag{22}$$

Процедура виявлення правил асоціації складається на основі еталонних значень з бази даних транзакцій впевненостей та значень довіри, виявлених у попередній процедурі. Традиційні правила ступенів впевненостей в наступних частинах правил даних пропонують нову процедуру виявлення через недоступність ступеня впевненостей аномалій. Система здатна враховувати різні типи невизначеності, такі як неповнота, незнання, розпливчастість, неточність і неоднозначність, які є загальними характеристиками сенсорних даних. Запропоновано нову процедуру розрахунку підтримки, яка враховує невизначеності, пов'язані з неповнотою. Крім того, запропоновано вдосконалений до сенсорних даних метод обчислення впевненості за допомогою відстані Хаммінга замість використання функції мультиплікативного агрегування. Крім того, запропоновано робастне правило асоціації впевненостей шляхом вбудовування ступенів впевненостей з еталонними значеннями в передню частину правил, які будуть використовуватися як початкова база правил для експертної системи.

Таблиця 8

**Кінцеві результати обчислення**

Номер транзакції	Впевненість	Навантаження на долото	Кількість обертів долота
<i>t1</i>	0,131465	26217.863999999900	1.314720300000
<i>t2</i>	0,132865	83492.293000000000	1.328673900000
<i>t3</i>	0,14201	97087.882000000000	1.420116300000
<i>t4</i>	0,159388	54793.206000000000	1.593930700000
<i>t5</i>	0,165321	50301.579000000000	1.653262000000
<i>t6</i>	0,16574	56835.194000000000	1.657450000000
<i>t7</i>	0,213912	37638.028000000000	2.139238300000
<i>t8</i>	0,213912	37638.028000000000	2.139238300000
<i>t9</i>	0,197472	44564.583000000000	1.974806400000
<i>t10</i>	0,173569	60928.602999999900	1.735741900000

Веб-експертна система на основі правил впевненостей експертів буріння щодо значень його режимних параметрів [28] здатна обробляти дані з давачів, а також виконувати прогнозування значень режимних параметрів.

Це означає, що виявлення аномалій за даними давачів за допомогою DRILLWISEIFPRO є кращим за інші методи завдяки врахуванню різних типів невизначеності, таких як неповнота, незнання, нечіткість, неточність та неоднозначність.

Метод виявлення гаусових аномалій припускає, що дані вибірки підпорядковуються нормальному розподілу. Тому він погано працює для даних про режимні параметри буріння НГС, оскільки більшість з таких даних є дані з давачів, що не підпорядковуються нормальному розподілу. Крім того, такий метод не враховує жодного типу невизначеності, що також впливає на ефективність виявлення аномалій. Навпаки, правило асоціації не залежить від розподілу даних. Тому воно працює краще, ніж гаусові алгоритми для обох наборів даних (з давачів і регламентних). Однак, через те, що бінарне правило асоціації не враховує жодного типу невизначеності, воно не працює краще, ніж нечіткі випадки. Нечітке правило асоціації обробляє невизначеності, пов'язані з неточністю, неоднозначністю та розмитістю, що допомагає йому працювати краще, ніж бінарне правило асоціації. Таким чином, враховуючи невизначеності, пов'язані з неточністю, неоднозначністю та розпливчастістю, нечітке правило асоціації працює краще, ніж бінарне правило асоціації в досліджуваній предметній області.

$$[rc_1 \rightarrow sc_1]: \begin{cases} [Drill.Bitload]\{(Низький, 0.46), (Середній, 0.54), \\ (Високий, 0), (Надвисокий, 0)\} \\ [Drill.BitRotationalSpeed]\{(Низький, 0.77), (Середній, 0.23), \\ (Високий, 0), (Надвисокий, 0)\} \end{cases} \quad (23)$$

$$[rc_2 \rightarrow sc_2]: \begin{cases} [Drill.Bitload]\{(Низький, 0), (Середній, 0), \\ (Високий, 0.55), (Надвисокий, 0.45)\} \\ \Rightarrow [Drill.BitRotationalSpeed]\{(Низький, 0.74), (Середній, 0.256), \\ (Високий, 0), (Надвисокий, 0)\} \end{cases} \quad (24)$$

де  $[Drill.Bitload]$  – навантаження на долото,  $[Drill.BitRotationalSpeed]$  – кількість обертів долота.

Нарешті, система враховує всі типи невизначеності в інтегрованій структурі, що призводить до кращої продуктивності, ніж у гаусівських, бінарних та нечітких правил асоціації. З наведеного вище можна побачити, що виявлення аномалій з сенсорних даних за допомогою **DrillWiselfPro** працює краще, ніж інші методи, завдяки врахуванню різних типів невизначеності, таких як неповнота, незнання, розпливчастість, неточність і неоднозначність. Більше того, пропонується система не залежить від навчального набору даних для виявлення аномалій, як це відбувається в класичних алгоритмах машинного навчання.

### Висновки

В представленому дослідженні запропоновано новий алгоритм виявлення аномалій для *сенсорних даних процесу буріння НГС*, що дає можливість обробляти різні види невизначеності, такі як неповнота, незнання, нечіткість, неточність, неоднозначність, які є загальними характеристиками сенсорних даних процесу буріння. Через природу сенсорних даних не можна використовувати традиційний механізм виведення за правилами з коефіцієнтами впевненостей. Тому пропонується новий механізм виведення, який складається з бази даних вхідних транзакцій, виду перетворення в базу даних транзакцій щодо впевненостей користувача, обчислення рівня підтримки користувача, матриці впевненостей користувача, обчислення ступенів довіри та виявлення правил асоціації. Запропоновано нову процедуру розрахунку підтримки користувача, яка враховує невизначеності, пов'язані з неповнотою даних. Крім того, запропоновано вдосконалений метод обчислення ймовірності ступенів довіри до даних з давачів, з використанням відстані Хаммінга замість мультиплікативної функції агрегування. Оскільки відстань Хаммінга більше підходить для сенсорних даних, ніж мультиплікативна функція агрегування, та оскільки сенсорні дані є в більшій мірі кількісними за своєю природою. Крім того, запропоновано робастне правило асоціації ступенів довіри шляхом вбудовування ступенів довіри з еталонними значеннями в попередню частину правила, яке буде використовуватися як початкова база правил для експертної системи. Оскільки в традиційному правилі з коефіцієнтами впевненостей відсутні ступені довіри по відношенню до референтних значень попередніх частин правил. Результати були порівняні з іншими методами виявлення аномалій. Було продемонстровано, що отримано кращі результати, ніж ті що дають інші методи для заданих наборів даних. Причиною цього є те, що Гаусс, будучи статистичним підходом, не може впоратися з невизначеністю через неповноту, незнання, розпливчастість, неточність і неоднозначність, тоді як бінарне правило асоціації використовує стверджувальні знання, які можуть бути оцінені як істинні або хибні відповідно. Нечіткі асоціативні правила можуть впоратися з невизначеністю через нечіткість, двозначність і неточність, але не можуть впоратися з незнанням і неповнотою. Однак система може впоратися з усіма типами невизначеності в інтегрованій структурі. Більше того, як правило Гауса, так і бінарне правило асоціації не мають кращого представлення семантичного змісту, а отже, невизначеність, пов'язану з лінгвістичними мітками, не може бути вирішена за

допомогою цих методів. Крім того, отримані криві показують, що отриманий результат є кращим за вищезгадані методи, оскільки запропонована методика обробляє всі типи невизначеності, як було зазначено вище. Запропонований алгоритм виявлення аномалій демонструє спосіб вилучення початкової бази правил на основі даних з давачів. Система забезпечує кращий результат визначення стану режимних параметрів для даних без аномалій, ніж з аномальними даними. Крім того, система допомагає виконувати більш надійне і точне прогнозування, використовуючи дані, отримані з давачів, розміщених у свердловині, шляхом видалення аномалій. Таким чином, можна стверджувати, що нова методика покращить підхід до виявлення аномалій для інших сфер застосування, таких як спостереження, моніторинг навколишнього середовища та задач керування в умовах невизначеності. Цей новий алгоритм виявлення аномалій також дозволить покращити прогнозування в експертних системах загалом, оскільки аномальні дані можна буде видалити більш ефективно.

## Література

1. Foorthuis, Ralph. "On the nature and types of anomalies: a review of deviations in data." *International journal of data science and analytics* 12.4 (2021): 297-331.
2. Foorthuis, Ralph. "A typology of data anomalies." *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Cham: Springer International Publishing, 2018.
3. Agrawal, Shikha, and Jitendra Agrawal. "Survey on anomaly detection using data mining techniques." *Procedia Computer Science* 60 (2015): 708-713.
4. Bakumenko, Alexander, and Ahmed Elragal. "Detecting anomalies in financial data using machine learning algorithms." *Systems* 10.5 (2022): 130.
5. Xu, Dong, et al. "An improved data anomaly detection method based on isolation forest." *2017 10th international symposium on computational intelligence and design (ISCID)*. Vol. 2. IEEE, 2017.
6. Taha, Ayman, and Ali S. Hadi. "Anomaly detection methods for categorical data: A review." *ACM Computing Surveys (CSUR)* 52.2 (2019): 1-35.
7. Shukla, Amit K., et al. "Explaining data regularities and anomalies." *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2020.
8. Zhang, Aoqian, et al. "Time series data cleaning: From anomaly detection to anomaly repairing." *Proceedings of the VLDB Endowment* 10.10 (2017): 1046-1057.
9. Astrid, Marcella, et al. "Learning not to reconstruct anomalies." *arXiv preprint arXiv:2110.09742* (2021).
10. Goldstein, Markus, and Seiichi Uchida. "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data." *PloS one* 11.4 (2016): e0152173.
11. Togbe, Maurras Ulbricht, et al. "Anomalies detection using isolation in concept-drifting data streams." *Computers* 10.1 (2021): 13.
12. Rippel, Oliver, et al. "Gaussian anomaly detection by modeling the distribution of normal data in pretrained deep features." *IEEE Transactions on Instrumentation and Measurement* 70 (2021): 1-13.
13. Wu, Di, et al. "LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT." *IEEE Transactions on Industrial Informatics* 16.8 (2019): 5244-5253.
14. Cha, Sung-Hyuk. "Comprehensive survey on distance/similarity measures between probability density functions." *City* 1.2 (2007)
15. Ren, Mifeng, Qichun Zhang, and Jianhua Zhang. "An introductory survey of probability density function control." *Systems Science & Control Engineering* 7.1 (2019): 158-170.
16. Krishnamurthi, Rajalakshmi, et al. "An overview of IoT sensor data processing, fusion, and analysis techniques." *Sensors* 20.21 (2020): 6076.
17. Patni, Harshal, Cory Henson, and Amit Sheth. "Linked sensor data." *2010 International Symposium on Collaborative Technologies and Systems*. IEEE, 2010.
18. Kotsiantis, Sotiris, and Dimitris Kanellopoulos. "Association rules mining: A recent overview." *GESTS International Transactions on Computer Science and Engineering* 32.1 (2006): 71-82.
19. Agrawal, Rakesh, et al. "Fast discovery of association rules." *Advances in knowledge discovery and data mining* 12.1 (1996): 307-328.
20. Srikant, Ramakrishnan, and Rakesh Agrawal. "Mining generalized association rules." (1995).
21. Tubaishat, Malik, and Sanjay Madria. "Sensor networks: an overview." *IEEE potentials* 22.2 (2003): 20-23.
22. Glaser, Steven D. "Some real-world applications of wireless sensor nodes." *Smart Structures and Materials 2004: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Vol. 5391. SPIE, 2004.
23. Stubberud, Stephen C., and Kathleen A. Kramer. "Data association for multiple sensor types using fuzzy logic." *IEEE Transactions on Instrumentation and Measurement* 55.6 (2006): 2292-2303.
24. Maksimović, Mirjana, Vladimir Vujović, and Vladimir Milošević. "Fuzzy logic and wireless sensor networks—a survey." *Journal of Intelligent & Fuzzy Systems* 27.2 (2014): 877-890.
25. Gaddam, Anuroop, et al. "Detecting sensor faults, anomalies and outliers in the internet of things: A

---

survey on the challenges and solutions." *Electronics* 9.3 (2020): 511.

26. Hill, David J., and Barbara S. Minsker. "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach." *Environmental Modelling & Software* 25.9 (2010): 1014-1022.

27. Hayes, Michael A., and Miriam AM Capretz. "Contextual anomaly detection in big sensor data." *2014 IEEE International Congress on Big Data*. IEEE, 2014.

28. Protsiuk V. (2024). The construction of intelligent rule-based system with deep-learning functionality for oil and gas wells drilling. *Herald of Khmelnytskyi National University. Technical Sciences*, 331(1), 173-179.