

Міністерство освіти і науки України
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра комп'ютерної інженерії

ДОПУЩЕНО ДО ЗАХИСТУ
Завідувач кафедри,
доктор техн. наук, професор
_____ І. М. Журавська
« __ » _____ 2024 р.

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА
Вдосконалення візуалізації дерев рішень з використанням 3D-моделей
для категоризації вхідних даних
Спеціальність 123 Комп'ютерна інженерія
123 – КМР.1 – 605.21810519

Студента: _____ І.Г.Омельченко
« __ » _____ 2024 р.

Керівник: _____ Г.П.Чуйко
« __ » _____ 2024 р.

Миколаїв 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра комп'ютерної інженерії

ЗАТВЕРДЖУЮ
Завідувач кафедри,
доктор техн. наук, професор
_____ І. М. Журавська
« __ » _____ 2024 р.

ЗАВДАННЯ

на виконання кваліфікаційної роботи

Видано студентці групи 605 факультету комп'ютерних наук

Омельченко Ігор Геннадійович

(прізвище, ім'я, по батькові студента)

1. Тема кваліфікаційної роботи: Вдосконалення візуалізації дерев рішень з використанням 3D-моделей для категоризації вхідних даних. _____

Затверджена наказом по ЧНУ від «_08_» __ листопада _____ 2023 р. No _____ 226 _____

2. Строк представлення кваліфікаційної роботи «_____» _____ 2024 р.

3. Очікуваний результат роботи та початкові дані, якщо такі потрібні

Створення дерева рішень з використанням Weka, та 3D-модель цього дерева. _____

4. Перелік питань, що підлягають розробці

Аналіз поняття теорії прийняття рішень, розглянути поняття дерево прийняття рішень, зробити опис вдосконалення візуалізації та категоризації даних, розглянути 3D-моделювання як вдосконалення візуалізації _____

5. Перелік графічних матеріалів

Зображення інтерфейсу Weka

Зображення дерева рішень

Зображення 3D-моделі дерева рішень

6. Завдання до спеціальної частини

Розглянути одне з відомих питань охорони праці для такого робочого місця (мікрокліматичні умови, електробезпека, питання безпеки роботи з обладнанням), а також розглянути питання що стосуються безпеки персоналу.

7. Консультанти:

Консультант	Кафедра (організація)	Частина роботи
викладач Л. І. Григор'єва	кафедра екології Медичного інституту ЧНУ ім. Петра Могили	спеціальна частина з охорони праці

Керівник роботи _____ доктор фіз.-мат. наук, професор , Г. П. Чуйко _ _____
(посада, прізвище, ім'я, по батькові) (підпис)

Завдання прийнято до виконання __Омельченко Ігор Геннадійович____ _
(прізвище, ім'я, по батькові студента) (підпис)

Дата видачі завдання « _____ » _____ 20 _____ р.

КАЛЕНДАРНИЙ ПЛАН

виконання кваліфікаційної роботи

Тема: Вдосконалення візуалізації дерев рішень з використанням 3D-моделей для категоризації вхідних даних

№	Найменування роботи	Початок	Закінчення	Примітки
1.	Розробка та затвердження завдання на виконання КР	01.09.2023	05.10.2023	Виконано
2	Огляд літератури за темою роботи	06.10.2023	12.10.2023	Виконано
3	Складання календарного плану КР	13.10.2023	14.10.2023	Виконано
4	Аналіз предметної області	15.10.2023	20.10.2023	Виконано
5	Розробка проектних рішень	21.10.2023	15.11.2023	Виконано
6	Моделювання дерева рішень в Weka	16.11.2023	18.12.2023	Виконано
7	Створення 3D-моделі дерева рішень	18.11.2023	13.02.2023	Виконано
8	Оформлення КР та презентації	01.11.2023	31.12.2023	Виконано
9	Попередній захист	14.02.2023	14.02.2024	Виконано
10	Рецензування	13.02.2024	14.02.2024	Виконано
11	Завершення оформлення КР та Презентації	13.02.2024	14.02.2024	Виконано
12	Відгук керівника КР	13.02.2024	16.02.2024	Виконано
13	Захист кваліфікаційної роботи	28.02.2024	28.02.2024	Виконано

Розробив студент _____ Омельченко Ігор Геннадійович _____

(прізвище, ім'я, по батькові)

(підпис)

«__» _____ 2024 р.

Керівник роботи ____ доктор фіз.-мат. наук, професор , Г. П. Чуйко _____

(посада, прізвище, ім'я, по батькові)

(підпис)

«__» _____ 2024 р.

АНОТАЦІЯ

магістерської роботи

«Вдосконалення візуалізації дерев рішень з використанням 3D-моделей
для категоризації вхідних даних»

Студента: Омельченко Ігор Геннадійович

Керівник: доктор фіз.-мат. наук, професор Г. П. Чуйко

Магістерська робота присвячена вдосконалення візуалізації дерев рішень за допомогою 3D-технологій. В даній магістерській роботі розглянуто та досліджено принципи та методи вдосконалення візуалізації дерев рішень з використанням 3D-моделей для категоризації даних. Практичне значення результати дослідження можуть бути використані у сферах аналітики, прийняття рішень та управління даними, полегшуючи роботу фахівців та покращуючи якість прийнятих рішень. Пояснювальна записка магістерської роботи складається зі вступу, чотирьох розділів, висновків та 3 додатків. У вступі визначається актуальність теми, сформульовані мета, об'єкт, предмет та завдання дослідження та розроблення, а також гіпотези, наукова новизна та практичне значення магістерської роботи. У першому розділі розглядається термін «теорія прийняття рішень». У другому розділі розглядається термін «дерево прийняття рішень» як інструмент для прийняття рішень. У третьому розділі, було розглянуті та проаналізовані методи та інструменти візуалізації та категоризації даних у сфері комп'ютерної інженерії. У четвертому розділі було проведено дослідження та отримання вихідних даних з пульсоксиметра, а також його 3D-модель в Autodesk MAYA.

У додатку А . розміщено результат Unichesk. У додатку Б та В наведено програмний код, що використовувався в проекті, а також пройдені абробації .У цілому, магістерська робота без додатків містить 77 сторінок, 20 рисунків, 5 таблиць, 24 джерел посилання.

Ключові слова: теорія прийняття рішень, дерево прийняття рішень, бінарне дерево 3D-модель, візуалізація даних

ABSTRACT

of the Master's Thesis

"Improving the visualization of decision trees using 3D models for categorization of input data"

Student: Ihor Gennadiyovych Omelchenko

Consultant: Doctor of Phys.-Math. Sciences, Professor H.P. Chuyko

The master's thesis is devoted to the improvement of visualization of decision trees using 3D technologies. In this master's thesis, the principles and methods of improving the visualization of decision trees using 3D models for data categorization are considered and researched. The practical significance of the research results can be used in the areas of analytics, decision-making and data management, facilitating the work of specialists and improving the quality of decisions made. The explanatory note of the master's thesis consists of an introduction, four chapters, conclusions and 3 appendices. The introduction determines the relevance of the topic, formulates the goal, object, subject and tasks of research and development, as well as hypotheses, scientific novelty and practical significance of the master's work. The first chapter examines the term "decision making theory". The second chapter examines the term "decision tree" as a decision-making tool. In the third chapter, the methods and tools of data visualization and categorization in the field of computer engineering were considered and analyzed. In the fourth chapter, research was conducted and the initial data from the pulse oximeter, as well as its 3D model in Autodesk MAYA, were obtained.

In Appendix A. the result of Unichack is posted. Appendices B and C provide the program code used in the project, as well as passed approvals. In general, the master's thesis without appendices contains 77 pages, 20 figures, 5 tables, 24 reference sources.

Keywords: decision theory, decision tree, binary tree 3D model, data visualization

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	9
ВСТУП	10
1 ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ В ІТ ТА ПРОМИСЛОВІЙ СФЕРІ.....	12
1.1 Базове поняття теорії прийняття рішень.....	12
1.2 Моделювання процесу прийняття рішень	16
1.3 Підходи ТПР	17
1.4 Етапи прийняття рішень	18
1.5 Прийняття рішень у реальному промисловому середовищі за допомогою підходу теорії графів	21
1.6 Теорія прийняття рішень в ІТ-галузі.....	24
Висновки до розділу 1	26
2 ДЕРЕВО ПРИЙНЯТТЯ РІШЕНЬ	27
2.1 Базове поняття дерева прийняття рішень	27
2.2 Дерева рішень у машинному навчанні та видобутку даних	30
2.2.1 Метод навчання дерева рішень.....	32
2.2.2 Типи дерев рішень	33
2.2.3 Переваги та недоліки дерев рішень.....	34
2.3 Заходи вибору атрибутів	37
2.4 Бінарні дерева	39
2.5 Розрахунок прикладу моделі бінарного дерева	41
Висновки до розділу 2	46
3 ВІЗУАЛІЗАЦІЯ ТА КАТЕГОРИЗАЦІЯ ДАННИХ.....	47
3.1 Візуалізація даних у комп'ютерній інженерії	48
3.2 Категоризація даних у комп'ютерній інженерії	50
3.3 Опис компонентів ESP 8266 та MAX 30102.....	51
3.3.1 Базова інформація ESP 8266	52
3.3.2 Базова інформація MAX 30102.....	56
3.4 Вдосконалення візуалізації за допомогою 3D-моделювання	59
Висновки до розділу 3	62

4	ДОСЛІДЖЕННЯ ТА ОТРИМАННЯ ВХІДНИХ ДАНИХ.....	63
4.1	Принцип роботи пульсоксиметра та отримання даних.....	63
4.2	Побудова дерева рішень за допомогою WEKA	67
4.3	Побудова дерева рішень з використанням Autodesk MAYA	71
	Висновки до розділу 4	73
	ВИСНОВКИ.....	74
	ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....	75
	ДОДАТОК А ДОВІДКА ПРО ПЕРЕВІРКУ У UNICHECK	78
	ДОДАТОК Б ЛІСТІНГ ПРОГРАМИ З РОЗШИРЕННЯМ JAVA	79
	ДОДАТОК В АБРОБАЦІЇ.....	81

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ТПР	–	Теорія прийняття рішень
ЗПР	–	Завдання прийняття рішень
ШНМ	–	Штучна нейронна мережа
ПЗ	–	Програмне забезпечення
ХНУ	–	Хмельницький національний університет
GTMM	–	Graph Theory and Matrix Methods
GTM	–	Graph Theoretic Method
SoC	–	System on a Chip
IDE	–	Integrated Drive Electronics
IoT	–	Internet of things
TCP/IP	–	Transmission Control Protocol/Internet Protocol

ВСТУП

Сучасний розвиток технологій та аналізу даних останнім часом відкриває широкі можливості для покращення процесу прийняття стратегічних рішень в різних галузях. Зростаюча складність та обсяги даних вимагають більш ефективних методів їхньої візуалізації та аналізу. Використання тривимірних моделей для категоризації вхідних даних у процесі візуалізації дерев може стати ключовим фактором для зрозуміння та аналізу цих великих обсягів інформації.

Ця робота націлена на удосконалення процесу візуалізації дерев рішень за допомогою тривимірних моделей, щоб зробити аналіз даних більш доступним та зрозумілим.

У цьому дослідженні буде проведено аналіз різних методів візуалізації дерев рішень та їх переваг у контексті категоризації вхідних даних. Також буде здійснено порівняльний аналіз існуючих підходів та запропоновано нові стратегії для удосконалення візуалізації за допомогою тривимірних моделей.

Результати цього дослідження можуть суттєво покращити процеси прийняття рішень у різних сферах, відкриваючи нові можливості для аналізу та оптимізації даних. Розвиток нових підходів у візуалізації допоможе розширити можливості аналізу даних та сприятиме покращенню якості та точності стратегічних рішень на основі інформації.

1. Актуальність:

З високим темпом розвитку технологій та зростання обсягу даних, важливо удосконалити методи візуалізації дерев рішень для кращого розуміння складних взаємозв'язків та категоризації великої кількості вхідних даних.

2. Мета:

Метою дослідження є розробка та вдосконалення методів візуалізації дерев рішень за допомогою 3D-моделей для покращення сприйняття та аналізу складних структур даних.

3. Об'єкт дослідження:

Об'єктом дослідження є дерева рішень, особливо їх структури та взаємозв'язки, а також можливості їх візуалізації.

4. Предмет дослідження:

Предметом дослідження є використання 3D-моделей для покращення візуалізації та розуміння дерев рішень.

5. Мета дослідження:

Головною метою є розробка ефективного та зрозумілого інструменту візуалізації, який полегшить аналіз даних, зокрема у сферах прийняття рішень.

6. Гіпотези:

Гіпотези дослідження орієнтовані на те, що використання 3D-моделей сприятиме кращому сприйняттю та розумінню структур дерев рішень.

7. Наукова новизна:

Наукова новизна полягає у застосуванні та вдосконаленні технологій 3D-моделювання для візуалізації дерев рішень в галузі обробки та аналізу даних.

8. Практичне значення:

Результати дослідження можуть бути використані у сферах аналітики, прийняття рішень та управління даними, полегшуючи роботу фахівців та покращуючи якість прийнятих рішень.

Робота пройшла **апробацію** під час XXVI Всеукраїнської науково-практичної конференції «Могилянські читання» (Миколаїв, 06–10 листопада 2022 р.). А також у «Вісник ХНУ 2023» (21 грудня.2023)[3].

Публікації. Основні положення та результати магістерської роботи опубліковані у збірнику матеріалів XXVI Всеукраїнської науково-практичної конференції «Могилянські читання–2023»[3]. А також у «Вісник ХНУ 2023»[4].

1 ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ В ІТ ТА ПРОМИСЛОВІЙ СФЕРІ

1.1 Базове поняття теорії прийняття рішень

Теорія прийняття рішень відіграє важливу роль у комп'ютерній інженерії, тому що в цій галузі часто доводиться вирішувати складні завдання та приймати рішення, пов'язані з проектуванням, розробкою та управлінням програмними та апаратними системами. До ключових аспектів теорії прийняття рішень, які застосовуються в комп'ютерній інженерії, можна віднести такі:

- аналіз вимог: перш ніж розпочати проектування чи розробку програми чи системи, необхідно провести аналіз вимог. Це включає збір інформації про цілі і завдання системи, а також обмеження і очікування користувачів. Аналіз вимог допомагає визначити, які рішення будуть найефективнішими для досягнення поставленої мети;

- моделювання та симуляція: для оцінки різних альтернативних рішень часто використовуються моделювання та симуляція. Це дозволяє інженерам створювати віртуальні моделі системи та проводити експерименти, щоб визначити, які рішення будуть найбільш ефективними та оптимальними;

- аналіз ризиків: прийняття рішень у комп'ютерній інженерії також включає оцінку ризиків. Інженери повинні аналізувати потенційні ризики, пов'язані з вибором конкретних рішень, та розробляти стратегії щодо їх управління. Це може включати заходи щодо зниження ризиків або планування резервних варіантів;

- однією з основних цілей у комп'ютерній інженерії є оптимізація продуктивності та ефективності систем. Інженери застосовують методи оптимізації для вибору найкращих рішень з урахуванням обмежень, таких як бюджет, час та інші ресурси;

– прийняття рішень за умов невизначеності: в інженерії часто існує невизначеність і невідомі чинники. Теорія прийняття рішень включає методи роботи з невизначеністю, такі як статистичний аналіз, теорія ймовірності та багато інших, щоб приймати рішення за недостатньої інформації;

– інструменти підтримки прийняття рішень: У комп'ютерній інженерії також використовуються спеціальні інструменти для підтримки прийняття рішень, такі як системи управління проектами, аналітичні інструменти та методи штучного інтелекту, включаючи машинне навчання;

– цикли зворотного зв'язку: Після прийняття рішення важливо оцінювати його результати та проводити зворотний зв'язок. Якщо рішення не відповідає очікуванням чи цілям, воно може вимагати корекції чи перегляду.

Теорія прийняття рішень у комп'ютерній інженерії допомагає інженерам та розробникам приймати обґрунтовані рішення з огляду на різноманітні аспекти проектування, розробки та управління комп'ютерними системами та програмними продуктами[5].

Рішення завжди приймається у ситуації вибору, що передбачає втрати, випадковість та певні ризики, які бажано мінімізувати. Якщо ж вибір відсутній, то й вирішувати нема чого, дій єдиним чином або взагалі не дій, як вказує альтернатива.

Сенс та мета мінімізації ризику полягає в тому, щоб застосувати ефективні заходи захисту таким чином, щоб залишковий ризик у системі став прийнятним.

Мінімізація ризику передбачає вирішення трьох питань:

- визначення тих областей, де ризик є неприпустимо великим;
- вибору найефективніших засобів захисту;
- оцінювання заходів захисту та визначення, чи прийнятний залишковий ризик у системі.

У науковому дослідженні використовуються гіпотези, які висуваються, формулюються, перевіряються, підтверджуються чи опитуються, це

природний шлях дослідження. Гіпотези можуть бути дуже різними за змістом, способами їх формулювання та методами перевірки.

Важливий клас – статистичні гіпотези, що формулюються або щодо виду закону розподілу випадкової величини, або щодо параметрів цього закону, або щодо рангової впорядкованості значень випадкової величини.

Гіпотези сформульовані щодо імовірісно-статистичних і рангових величин перевіряються і оцінюються за допомогою різноманітних статистичних прийомів і критеріїв. Результати перевірки та оцінювання статистичних гіпотез дозволяють робити якісні висновки щодо досліджуваних явищ. Наприклад, ступінь близькості емпіричного закону розподілу випадкової величини до нормального теоретичного закону або Пуассона.

Нульова та альтернативна гіпотезі. Зазвичай нульова гіпотеза полягає в тому, що висувається припущення про вид закону розподілу ймовірностей випадкової величини або параметр такого закону, або про рангову послідовність. Інша гіпотеза називається альтернативною.

Процес прийняття рішень (процес ПР) – це процес, у якого ставиться проблема (проблемна ситуація), яка знімається за кілька етапів, включаючи практичні дії з усунення проблемної ситуації (реалізація знайденого рішення).

Теорія прийняття рішень (ТПР) – прикладна міждисциплінарна наука, заснована на методах математики, штучного інтелекту, статистики, економіки, менеджменту, психології, схема якого показано на рисунку 1.1.[6]



Рисунок 1.1 – Схема ТПР

ТПР вивчає закономірності вибору людьми шляхів вирішення різноманітних завдань, а також досліджує способи пошуку найвигідніших із можливих рішень. ТПР ґрунтується на потребі людини спростувати процедуру прийняття рішень та надавати рішенням більшої надійності.

Методи оцінювання та вибору рішень поділяються на кількісні та якісні. Кількісні методи дуже різноманітні та добре опрацьовані, в тому числі в рамках таких наукових дисциплін як методи оптимізації, дослідження операцій, системний аналіз.

До кількісних належать такі методи:

- класичні методи математичного програмування та обчислювальної математики (сфера їх застосування часто обмежена умовами диференційності та унімодальності цільових функцій);
- перерахувальні методи, що ґрунтуються на перебірному пошуку в дискретному просторі рішень, що часто призводять до проблеми комбінаторного вибуху («прокляття розмірності»);
- методи, що використовують елемент випадковості (імовірнісні, ігрові, еволюційні), що довели свою ефективність при вирішенні важких завдань, але не гарантують отримання глобально оптимального рішення.

Разом з кількісними методами у ТПР застосовуються методи, які дозволяють отримувати та аналізувати якісну інформацію, наприклад, методи змістового (вербального) аналізу ситуацій, багатокритеріального аналізу, експертного оцінювання тощо.

Теоретично прийняття рішень виділяють декілька підходів, які називають моделями прийняття рішень, до яких відносять такі види моделей: нормативна (класична), дескриптивна (описова), модель Карнегі, модель інкрементального процесу прийняття рішень, модель сміттевого кошика та ін.

1.2 Моделювання процесу прийняття рішень

Поняття моделі в сучасній науці стало звичним і необхідність з'ясування змісту поняття перестала усвідомлюватися. Насправді поняття моделей, процедур, схем і методів, які часто поєднуються і перестають відрізняти одне одного. Можливості моделювання переваг багаторазово перекривають такі у людини і часто можливості моделі виявляються багатшими за реальність.

Говорити про модель прийняття рішення слід лише у зв'язку з конкретним завданням рішення (ЗПР), яке має рішення. Це означає, що обраний клас базових структурних переваг, у межах якого забезпечується пошук кращого рішення.

Різні моделі рішень однієї й тієї ж ЗПР відрізняються одними принципами, закладеними в тому підставі. Вважаємо, що розглядається кілька вихідних структур переваг (відносин), заданих у матричній формі, наприклад, матриця парних порівнянь. На цій множині досліджується певна ЗПР і кажуть, що на множині вихідних структур задана модель рішення поставленої ЗПР.

До моделей прийняття рішень перед'являються досить жорсткі вимоги: коректність, адекватність, повнота, універсальність та ін.

Коректність у математиці зменшення кількості рішень, єдиним рішенням та його стійкістю.

Адекватність – відповідність оригіналу, тобто правильність відображення в моделі принципів, що моделюються, і особливостей процесу прийняття рішення. Істотними є відмінності між нормативними (прескриптивними) та дескриптивними підходами.

У першому панують апіорні припущення про те, якими повинні бути загальні принципи, що формулюються як аксіоми, якими повинні задовольняти моделі рішень, що розробляються.

У другому – особливості моделей, що розробляються, описуються не аксіоматично, а атрибутивно, за допомогою системи влади, кожна з яких змістовно інтерпретується ЛПР і представляється його розумовою і т. д. чи іншою мірою бажання.

Повнота для моделей діє в тому, що основні принципи, які лежать на основі ухвалення рішення, повинні відображатися не тільки точно, але й достатньо.

Універсальність моделі збільшенням можливості її застосування до широкого класу вихідних структурних переваг.

1.3 Підходи ТПР

Нормативну (класичну) модель прийняття рішень, запропоновану Г.А. Саймоном, характеризується можливістю технічного процесу розробки стратегій, який допомагає виявити найбільш ефективні шляхи досягнення поставленої мети. Ця модель включає функціональні рівняння, що відображають зв'язки між незалежними та залежними змінними. Незалежні змінні в цих моделях є параметрами дій, тоді як залежні змінні представлені очікуваними результатами, отриманими в результаті впливу незалежних змінних.

Дескриптивні (описові) моделі, засновані на емпіричних спостереженнях, містять обмежену кількість елементів та пояснюють економічні взаємозв'язки у спрощеній формі, відображаючи реальні процеси прийняття рішень у важких ситуаціях. Ці моделі використовуються для опису непрограмованих рішень та ситуацій невизначеності, коли менеджерам складно ухвалити економічно обгрунтовані рішення.

Модель Карнегі (Політична модель прийняття рішень), розроблена Г.А. Саймоном, Дж. Марчем та Р. Кайертом, показує, що в організаціях менеджери часто приймають рішення у вигляді коаліцій – неформальних альянсів між кількома менеджерами, що мають за мету вирішення проблем та встановлення пріоритетів.

Модель інкрементального процесу ухвалення рішень, запропонована Г. Мінцбергом, зосереджується на структурній послідовності дій, що здійснюються протягом усього процесу прийняття рішень.

Організація здійснює послідовні «дрібні» вибори, що можуть призводити до повторення циклу прийняття рішень у разі переривань процесу.

Модель «сміттевого кошика», розроблена М. Коеном, Дж. Марчем та Дж. Олсеном, пояснює прийняття рішень в умовах крайньої невизначеності, де рішення виникають внаслідок незалежних потоків подій всередині організації.

Теорія локальних збільшень Ч. Ліндблома вказує на два підходи до прийняття управлінських рішень – синоптичний та стратегію локальних прирощень, де управлінські рішення приймаються послідовно, з невеликими змінами, і їх результати визначаються умовами конкретної ситуації.

Модель прийняття рішень В. Врума враховує можливі стилі прийняття рішень керівниками, такі як автократичний, консультативний та груповий, залежно від ситуації та умов.

1.4 Етапи прийняття рішень

Етап 1. Цільове виявлення. Процес формулювання проблем -складне завдання. Причинами цього є складність, багатовимірність та багатозв'язковість проблем організаційного управління, неструктурований характер багатьох з них, коли потрібна додаткова інформація про них, труднощі вимірювання змінних, відсутність апріорних відомостей про зв'язки між ними.

Етап 1 дає відповіді на питання: яку проблему та в яких умовах потрібно вирішувати; коли її слід вирішувати; якими силами та засобами вона вирішуватиметься.

Етап 2. Формування цілей. Для визначення бажаного стану усунення проблемної ситуації необхідно сформулювати безліч цілей системи. Чим точніше буде сформульовано цілі системи, тим легше вибрати засоби їх досягнення. На цьому етапі визначається, що треба зробити для зняття проблеми. Методологічною основою етапу формування цілей є системний аналіз із використанням експертних методів.

Етап 3. Вироблення рішень. На цьому етапі виробляються альтернативні варіанти рішень, здійснюється пошук різних шляхів досягнення поставленої мети. Без альтернативних варіантів рішень відпадає і завдання вибору, більше, безліч вихідних альтернатив має бути досить повним, характеризуватись великим ступенем впевненості у наявності оптимальної альтернативи, для знаходження якої і вирішується завдання вибору. Формами генерування альтернатив є мозковий штурм, розробка сценаріїв та ділові ігри.

Етап 4. Вибір рішень. Порівняння та вибір альтернативних рішень можливий, якщо ввести вимірник ступеня досягнення поставленої мети. Таким вимірником є критерій. Таким чином, будується система критеріїв, що однозначно характеризує відповідні цілі ЛПР. Сформовані критерії надалі замінюють цілі. Критерієм корисності альтернативного рішення може бути будь-яка її ознака, виміряна на якісному чи кількісному рівні. Для опису мети часто вводяться кілька критеріїв те щоб вони повно характеризували мета. Критерії вибору рішень визначаються методами експертного аналізу та математичної статистики.

Етап 5. Оцінка рішень. При прийнятті рішень ЛПР та експерти, описуючи ситуацію, цілі, обмеження, варіанти рішень, будують модель оцінки альтернативних рішень через систему переваг ЛПР для вибору найкращого рішення. Через вплив довкілля оцінка альтернативних рішень містить елементи невизначеності, що з наявністю у зовнішньому середовищі кількох можливих ситуацій, і кожна їх випадково стає дійсністю. Ця невизначеність враховується з допомогою теорії ймовірностей.

Окремий випадок фізичної невизначеності – це ситуація, коли у зовнішньому середовищі присутні сили, що протидіють ЛПР, тобто. має місце конфлікт, тоді застосовують методи теорії ігор. За відсутності невизначеності (у разі визначеності) завдання ПР вирішують методами оптимізації.

Етап 6. Ухвалення рішення. Тут необхідно вибрати рішення для його подальшої реалізації за певним алгоритмом, що вибирає одне-єдине рішення, найкраще за деяким критерієм або принципом оптимальності.

При прийнятті рішень за багатьма критеріями (векторних завдань оптимізації) виникають труднощі визначення кращого з погляду ЛПР компромісного рішення з безлічі допустимих локальних критеріїв. Якщо потрібно визначити єдине найкраще рішення, безліч допустимих рішень зводиться до безлічі Парето і в ньому відбувається пошук рішення на основі деякої схеми.

Етап 7. Реалізація рішення. План реалізації обраного рішення має дати відповіді на питання, хто і що має робити, якими засобами та у які терміни. Конкретизація рішення щодо виконавців може здійснюватися у вигляді завдання про призначення виконавців виконання комплексу робіт, за термінами і об'єктах робіт – методами мережевого планування і управління. Конкретизація ресурсного забезпечення виконується розв'язанням задачі розподілу ресурсів методами математичного програмування.[8]

– Класифікація задач ПР

- За характером виграшів: антагоністичні (ігри з нульовою сумою – виграш партії дорівнює нулю), неантагоністичні (ігри з ненульовою сумою – виграш партії не дорівнює нулю)
- За кількістю ходів: однокрокові та багатокрокові. Багатокрокові діляться на позиційні, стохастичні, диференціальні. Позиційні ігри – кілька гравців роблять кілька послідовних ходів. Якщо у грі виробляються ходи, що призводять до вибору певних позицій, причому є певна можливість повернення на попередню позицію, така гра називається Стохастичною. Якщо у багатокроковій грі допускається робити ходи безперервно і дії гравців описуються диференціальними рівняннями, така гра називається диференціальною
- Залежно від стану інформації: ігри з повною (кожному гравцеві відомо які вибори зроблено гравцями раніше) та неповною інформацією (Не все відомо).

- По виду функцій виграшу: матричні (кінцева гра двох гравців з нульовою сумою, в якій виграш одного гравця дорівнює програшу іншого), біматричні (кінцева гра двох гравців з ненульовою сумою, в якій виграш кожного гравця зосереджені в матриці ігор даного гравця), безперервні (функція виграшу кожного гравця є безперервною залежно від стратегій).

1.5 Прийняття рішень у реальному промисловому середовищі за допомогою підходу теорії графів

Підхід, який називається «теорією графів і матричним підходом», добре використовують у чисельних дослідженнях з метою прийняття рішень, коли ситуація стає заплутаною або коли існує дуже сильна відносна важливість одного параметра над іншим. У таких випадках згадана теорія графів та матричний метод забезпечує дуже відповідні та плідні рішення для прийняття залишкового ефективного рішення.

Подальші удосконалення та покращення результатів також можуть бути виявлені шляхом використання комбінованої практики результатів теорії графів разом із деякими логіками та практиками, вдохненими штучним інтелектом, такими як нечітка логіка, штучна нейронна мережа тощо. Значення та застосовність зазначеного методу також доведено величезні галузі науки, техніки та досліджень.

Основний огляд теорії графів

Прийняття рішень тепер розроблено та практикується з використанням кількох нещодавно з'явилися методологій, щоб зробити традиційні практики більш ефективними та вражаючими [9]. Поведінка людей була ефективно і обмежено імітована за допомогою моделей штучного інтелекту.

Штучні нейронні мережі, еволюційні обчислення, нечітка логіка, моделі імовірнісного аналізу, інтелектуальні агенти тощо є деякими основними інструментами, які зазвичай визначають базову систему штучного інтелекту.

Досліджування застосування штучного інтелекту (ШІ), практикуючи здорове прийняття рішень у практичних галузевих проблемах. Існує так багато методів штучного інтелекту, які досліджувалися у минулих дослідженнях з метою прийняття рішень [10].

Такими методологіями є, зокрема, аналіз тенденцій даних, координація доставки даних, надання прогнозів, розробка узгодженості даних, кількісна оцінка невизначеності тощо.

Добре відомий метод прийняття рішень, який називається теорія графів, є системним і логічним підходом, який також виявився корисним для аналізу та моделювання широкого спектру застосувань в інженерії та багатьох інших галузях. Ця методологія, зокрема, базується на передовій теорії графів, і тому її застосування дуже добре відомо. Демонстрація за допомогою моделі орграфа «або» виявилася корисною для моделювання та аналізу різноманітних систем і проблем у багатьох галузях науки і техніки [11].

Підхід, заснований на матричному методі, є корисним для швидкого аналізу моделей графів/орграфів для отримання системної функції та індексу для досягнення цілей.

Теорія графів є предметом комбінаторної математики і багато чого черпає з теорії матриць. Матричне представлення графа формує проблему використання комп'ютерів для різноманітних складних операцій. Теорія графів та матричні методи (GTMM) складаються з подання орграфа, матричного подання та математичного подання, тобто постійної функції. Орграф – це візуальне подання змінних та їх взаємозалежностей. Матриця перетворює орграф у математичну форму, а перманентна функція є математичним уявленням, яке допомагає визначити числовий індекс [12].

Підхід теорії графів є систематичним підходом для перетворення якісних факторів у кількісні значення, а математичне моделювання дає перевагу запропонованій техніці перед звичайними методами, такими як діаграми причинно-наслідкових зв'язків, блок-схеми тощо.

Результати цих методів тепер можна розмістити у деяких інструментах та кроках на основі штучного інтелекту для створення більш плідних і розумних рішень.

Більш конкретно, підхід на основі нечіткої логіки може бути реалізований далі, щоб використовувати людські думки щодо проблеми, що розглядається.

Генерація логіки базуватиметься на чистому людському досвіді, який допоможе реалізувати підхід теорії графів на основі штучного інтелекту більш розумним та практичним. Штучний інтелект, який зазвичай описують як дисципліну розробки машин, що виконують дії, що передбачають людську кмітливість, часто спантеличується робототехнікою, особливо гуманоїдною робототехнікою, оскільки вона є суміжною з «людським» інтелектом. Теорія графів має широкий спектр застосувань у техніці та багатьох інших областях. Основні кроки до виконання вищезазначеної методології такі.

Підбір атрибутів

На цьому головному етапі визначаються різні характеристики, що впливають на досліджуваний результат, і завершується план експерименту, який задовольняє вимоги до операції.

Таблиця 1.1 – Кількісна оцінка факторів, що впливають на оброблюваність титану [12].

С.№	Якісна міра факторів, що впливають на оброблюваність титану	Призначене значення коефіцієнтів оброблюваності (T_i)
1	Винятково низький	0
2	Надзвичайно низький	1
3	Дуже низький	2
4	Нижче середнього	3
5	Середній	4
6	Вище середнього	5
7	Помірний	6
8	Високий	7
9	Дуже високо	8
10	Надзвичайно високий	9
11	Винятково висока	10

Значення ознак (T_i) і відносної важливості (u_{ij}) отримані з використанням таблиць 1. і 2 .

Таблиця 2. – Відносна важливість атрибутів оброблюваності (u_{ij}) [12].

С.№	Опис категорії	Взаємодія атрибутів	
		u_{ij}	$u_{ij} = 10 - u_{ij}$
1	Два атрибути однаково важливі	5	5
2	Один атрибут трохи важливіший за інший	6	4
3	Один атрибут важливіший за інший	7	3
4	Один атрибут набагато важливіший за інший	8	2
5	Одна ознака надзвичайно важливіша за іншу	9	1
6	Одна ознака надзвичайно важливіша за іншу	10	0

Додатково важливо враховувати вибір атрибутів для моделювання графових структур, таких як ваги ребер, мітки вершин, чи інші параметри, що відображають ключові аспекти системи. Це допомагає впроваджувати аналітичні інструменти теорії графів у процесі дослідження та розробки оптимальних стратегій для визначених завдань.

1.6 Теорія прийняття рішень в ІТ-галузі

Теорія прийняття рішень грає величезну роль ІТ-сфері. У цій галузі прийняття рішень часто пов'язане з вибором оптимальних стратегій розробки програмного забезпечення, управління проектами, оптимізацією процесів та ресурсів.

Ось які аспекти теорії прийняття рішень важливі в ІТ:

- Оптимізація процесів розробки: Розробка ПЗ вимагає ухвалення безлічі рішень щодо архітектури, вибору технологій, управління часом та ресурсами. Теорія прийняття рішень допомагає визначити оптимальні підходи та інструменти для досягнення цілей проекту.

- **Управління проектами:** Прийняття рішень в ІТ-проектах включає оцінку ризиків, планування ресурсів, розподіл завдань та контроль виконання. Теорія прийняття рішень надає методи для аналізу даних та прогнозування, що допомагає приймати обґрунтовані рішення під час управління проектами.
- **Аналіз даних та машинне навчання:** У сфері аналізу даних та машинного навчання, прийняття рішень ґрунтується на статистичних методах та алгоритмах машинного навчання. Це включає прийняття рішень на основі даних, класифікацію, кластеризацію, прогнозування і багато іншого.

Додатково до вищевказаного, можна відзначити, що теорія прийняття рішень в ІТ є ключовою складовою для досягнення успіху в цій галузі. Вона допомагає інженерам програмного забезпечення та менеджерам проектів ефективно керувати складними завданнями та приймати обґрунтовані рішення на кожному етапі розробки та впровадження програм. Аналіз даних та машинне навчання в ІТ також неможливі без використання методів та принципів теорії прийняття рішень.

Заключно, можна сказати, що в контексті швидкозмінюваного технологічного середовища, розуміння та використання принципів теорії прийняття рішень стає критично важливим для успішної реалізації проектів, вдосконалення процесів та досягнення стратегічних цілей в області інформаційних технологій.

Висновки до розділу 1

В останні десятиліття спостерігається значне розвиток технологій, зокрема в галузі аналізу та обробки даних. Цей прогрес призвів до збільшення обсягу та складності інформації, що вимагає більш ефективних методів візуалізації для раціонального аналізу та прийняття стратегічних рішень.

Сучасний розвиток технологій та аналізу даних надає широкі можливості для покращення процесу прийняття стратегічних рішень у різних галузях. Зростаюча складність та обсяги даних вимагають більш ефективних методів їхньої візуалізації та аналізу. Використання тривимірних моделей для категоризації вхідних даних у процесі візуалізації дерев може стати ключовим фактором для розуміння та аналізу цих великих обсягів інформації.

Цей розділ націлений на удосконалення процесу візуалізації дерев рішень за допомогою тривимірних моделей, щоб зробити аналіз даних більш доступним та зрозумілим. Головна мета полягає в розробці та використанні методів візуалізації, які нададуть можливість створювати більш динамічні та інтерактивні моделі дерев рішень, що сприятиме кращому представленню та розумінню даних.

У цьому дослідженні буде проведено аналіз різних методів візуалізації дерев рішень та їх переваг у контексті категоризації вхідних даних. Також буде здійснено порівняльний аналіз існуючих підходів та запропоновано нові стратегії для удосконалення візуалізації за допомогою тривимірних моделей.

Результати цього дослідження можуть суттєво покращити процеси прийняття рішень у різних сферах, відкриваючи нові можливості для аналізу та оптимізації даних. Розвиток нових підходів у візуалізації допоможе розширити можливості аналізу даних та сприятиме покращенню якості та точності стратегічних рішень на основі інформації.

2 ДЕРЕВО ПРИЙНЯТТЯ РІШЕНЬ

Дерево рішень є одним із найпотужніших інструментів алгоритмів навчання під наглядом, який використовується як для завдань класифікації, так і для регресії. Він створює структуру дерева, подібну до блок-схеми, де кожний внутрішній вузол позначає перевірку атрибуту, шкірна гілка представляє результат тесту, а кожний кінець (кінцевий вузол) містить мітку класу. Він створюється шляхом рекурсивного розбиття навчальних даних на підмножині на основі значень атрибутів, доки не буде виконано критерій зупинки, такий як максимальна глибина дерева або мінімальне кількість зразків, необхідних для розбиття вузла.

Під час навчання алгоритм дерева рішень вибирає найкращий атрибут для поділу даних на основі таких показників, як ентропія чи домішка Джині, яка вимірює рівень домішок чи випадковості у підмножинах. Мета полягає у тому, щоб знайти атрибут, який максимізує приріст інформації або зменшення домішок після поділу.

2.1 Базове поняття дерева прийняття рішень

Дерево рішень — це структура дерева, подібна до блок-схеми, де кожен внутрішній вузол позначає функцію, гілки позначають правила, а листові вузли позначають результат алгоритму. Це універсальний контрольований алгоритм машинного навчання, який використовується як для завдань класифікації, так і для регресії. Це один із найпотужніших алгоритмів. Він також використовується в Random Forest для навчання на різних підмножинах навчальних даних, що робить випадковий ліс одним із найпотужніших алгоритмів машинного навчання. Нижче наведено деякі поширені термінології, які використовуються у деревах рішень:

- Кореневий узел: це найвищий у дереві, який представляє повний набір даних. Це відправна точка процесу прийняття рішень.
- Рішення/внутрішній узел: узел, який символізує вибір щодо вхідної функції. Відгалуження внутрішніх вузлів з'єднує їх із листовими вузлами або іншими внутрішніми вузлами.
- Листовий/кінцевий узел: узел без дочірніх вузлів, який вказує на мітку класу чи числове значення.
- Поділ: процес поділу вузла на два або більше підвузлів за допомогою критерію поділу та вибраної функції.
- Гілка/піддерево: Підрозділ дерева рішень починається у внутрішньому вузлі та закінчується кінцевими вузлами.
- Батьківський узел: узел, який ділиться на один чи кілька дочірніх вузлів.
- Дочірній вузол: вузли, що виникають, коли батьківський узел розділено.
- Домішка вимірювання однорідності цільової змінної у підмножині даних. Це стосується ступеня випадковості чи невизначеності у наборі прикладів. Індекс Джіні та ентропія є двома часто використовуваними вимірюваннями домішок у деревах рішень для завдання класифікації
- Дисперсія : дисперсія вимірює, наскільки прогнозовані та цільові змінні відрізняються у різних вибірках набору даних. Він використовується для завдань регресії в деревах рішень. Середня квадратична помилка, середня абсолютна помилка, `friedman_mse` або відхилення половини Пуассона використовуються для вимірювання дисперсії для завдань регресії у дереві рішень.

- Інформаційний приріст: інформаційний приріст є мірою зменшення домішок, досягнутого шляхом поділу набору даних на певну функцію в дереві рішень. Критерій розбиття визначається ознакою, яка забезпечує найбільший приріст інформації. Він використовується для визначення найбільш інформативної ознаки для розбиття на кожному вузлі дерева з метою створення чистих підмножин.
- Обрізка : процес видалення гілок з дерева, які не надають жодної додаткової інформації або призводять до надмірного облаштування.

Дерево рішень є формою схеми, де відображаються можливі наслідки прийняття послідовності взаємопов'язаних рішень. Такий метод дозволяє як окремим особам, так і організаціям оцінити різні варіанти дій з огляду на витрати, переваги та ймовірності. Древа рішень використовуються як і неформальних обговореннях, і у створенні алгоритмів для математичного прогнозування оптимальних дій. Приклад такого дерева з на рисунку 2.1

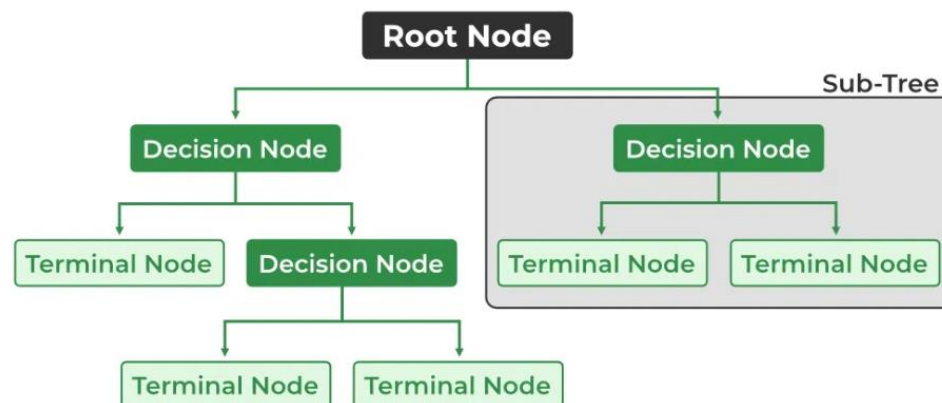


Рисунок 2.1 – Приклад дерева прийняття рішень

Зазвичай, дерево рішень починається з одного вузла, від якого відходять різні варіанти. Кожен із них у свою чергу веде до додаткових вузлів, створюючи додаткові варіанти розвитку. Таким чином, структура всього дерева набуває форми дерева, що гілкується.

Вузли в дереві рішень є одним з трьох: ймовірність, рішення або кінцевий результат. Вузол ймовірності позначається кругом і забезпечується мітками, що вказують на ймовірність настання певного результату. Вузол рішення представлений квадратом і позначає прийняте рішення. Завершальний вузол свідчить про кінцевий результат, якого призводить обраний варіант рішення.

2.2 Древа рішень у машинному навчанні та видобутку даних

Навчання дерева рішень є керованим методом навчання, що широко застосовується в статистиці, інтелектуальному аналізі даних і машинному навчанні. У рамках цього формалізму дерево рішень класифікації або регресії використовується як прогностична модель для висновків на основі набору спостережень.

Древа рішень широко застосовуються під час створення автоматизованих моделей прогнозування у сфері машинного навчання, видобутку даних, і статистики. Метод, відомий як "навчання через дерево рішень", дозволяє передбачити цінність об'єкта, враховуючи спостереження, що є про нього.

У таких деревах вузли швидше представляють дані, ніж рішення, і вони відомі як "древа класифікації". Кожна гілка містить набір атрибутів або правил класифікації, прив'язаних до позначки класу, розташованої в кінці гілки.

Ці правила, також відомі як правила прийняття рішень, формуються за умови "якщо". Наприклад, "якщо задоволені умови 1, 2 і 3, це призведе до результату X з ймовірністю Y".

Кожне нове спостереження дозволяє моделі точніше передбачати, яке значення заданого набору відповідає об'єкту. Ця інформація може бути вхідними даними для складніших моделей прийняття рішень.

Древа рішень можуть бути дискретними, що передбачають конкретні класи, так і регресійними, що передбачають безперервні значення.

Іноді підвищення точності застосовують кілька дерев рішень одночасно, наприклад, з допомогою методів беггинга, випадкового лісу, бустингу чи ротаційного лісу.

Ідеальне дерево рішень має передавати максимум інформації за мінімальної кількості питань або рівнів. Існує кілька алгоритмів для створення оптимізованих дерев рішень, таких як CART, ASSISTANT, CLS та ID3/4/5. Дерева рішень також можуть будуватися на основі набору асоціативних правил та цільової змінної.

Мета кожного методу – визначити найбільш ефективний спосіб поділу інформації за рівнями. У цьому допомагають такі інструменти, як коефіцієнт Джіні, приріст інформації та зниження дисперсії.

Використання дерев рішень у машинному навчанні надає кілька переваг, таких як зменшення витрат зі збільшенням даних, застосування різних типів даних, здатність моделювати безліч результатів і можливість пояснення результатів. Тим не менш, у дерев рішень є і недоліки, включаючи їх обмеження оператором "I" та складність у разі невизначеності та пов'язаних результатів.

Деревоподібні моделі, де цільова змінна набуває дискретного набору значень, називаються деревами класифікації. У цих структурах листя представляють мітки класів, а гілки відображають з'єднання об'єктів, що призводять до цих міток. Дерева рішень, де цільова змінна може набувати безперервних значень, називаються деревами регресії. У більш загальному контексті концепція дерева регресії може бути розширена будь-які об'єкти з попарними відмінностями, наприклад, категоріальні послідовності.

Дерева рішень вважаються одними з найпопулярніших алгоритмів машинного навчання завдяки своїй зрозумілості та простоті. У процесі аналізу рішень дерево рішень може бути візуальним і явним уявленням рішень та їх прийняття. При інтелектуальному аналізі даних дерево рішень визначає дані, і отримане дерево класифікації може бути вхідними даними для прийняття рішень.

2.2.1 Метод навчання дерева рішень

Навчання дереву рішень є поширеним методом, широко застосовується в області інтелектуального аналізу даних. Його мета полягає в побудові моделі, здатної прогнозувати значення цільової змінної на основі множини вхідних змінних.

Дерево рішень є зручним засобом для класифікації прикладів. Припускаючи, що всі вхідні об'єкти мають кінцеві дискретні області, і існує єдина цільова ознака, відома як "класифікація", дерево рішень може бути використане для класифікації елементів в область класів. Кожен внутрішній вузол дерева позначений вхідною ознакою, а дуги, що виходять із вузла, позначаються можливими значеннями цільового об'єкта. Листя дерева може бути позначені класами чи розподілом ймовірностей по класам.

Процес побудови дерева включає поділ вихідного набору даних, що становить кореневий вузол, на підмножини, які стають дочірніми елементами. Поділ ґрунтується на правилах, що залежать від ознак класифікації. Цей процес повторюється рекурсивно кожному похідного підмножини у випадках, коли додавання нового поділу приносить цінність прогнозам. Процес, відомий як "рекурсивний поділ", завершується, коли підмножина має однакові значення цільової змінної або подальші поділу не приносять користі.

У контексті інтелектуального аналізу даних, дерева рішень є комбінацією математичних та обчислювальних методів, які допомагають описувати, категоризувати та узагальнювати заданий набір даних.

Дані надходять у записі виду:

$$(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$$

Залежна змінна, що позначається як Y , є цільову величину, яку ми прагнемо зрозуміти, класифікувати або узагальнити. Вектор x включає різні особливості (x_1, x_2, x_3) , які служать вхідними даними для даної задачі.

Ці особливості є важливими змінними, що використовуються для аналізу та побудови моделі, спрямованих на пошук зв'язку або залежності із залежною змінною Y . У контексті аналізу даних, успішне використання цих особливостей може допомогти у виявленні закономірностей та дозволити більш точно прогнозувати значення чи класифікації для залежної змінної.

2.2.2 Типи дерев рішень

Дерева рішень, які застосовуються в рамках інтелектуального аналізу даних, поділяються на два основні типи:

Дерево класифікації:

- Прогнозує дискретні класи, яких належать дані.
- Внутрішні вузли позначені вхідними ознаками, а дуги становлять можливі значення цільового класу.
- Ефективно застосовується до завдань, де метою є категоризація даних.

Дерево регресії:

- Прогнозує дійсні числа, такі як ціна будинку чи тривалість перебування пацієнта у лікарні.
- Використовує самі основи, як і дерево класифікації, але спрямоване передбачення числових значень.
- Застосовується для завдань, пов'язаних із кількісною оцінкою.

Термін «аналіз дерева класифікації та регресії» (CART) охоплює обидві ці процедури. Однак, незважаючи на схожість, існують відмінності у процедурі поділу для дерев класифікації та регресії.

Існують також ансамблеві методи, що використовують кілька дерев рішень:

- Посилені дерева: Побудова ансамблю з акцентом на коригуванні раніше неправильно змодельованих даних. Прикладом є метод AdaBoost, що застосовується для регресії та класифікації.

- Бутстреп-агрегація: Побудова кількох дерев шляхом повторної вибірки навчальних даних із заміною. Прогнози поєднуються для консенсусного результату.

Відомі алгоритми дерев рішень включають ID3, C4.5 і CART. Древа умовного висновку використовують статистичні тести як критерії поділу, запобігаючи перенавченню.

Додатково існує поняття списків рішень - односпрямованих дерев, легших для розуміння, але менш виразних.

Важливі алгоритми включають ID3, C4.5, CART, CHAID та MARS, кожен з яких призначений для оптимального розбиття інформації на різних рівнях.

Невід'ємні аспекти використання дерев рішень у машинному навчанні включають зниження витрат зі збільшенням даних, застосовність до категоріальних та числових даних, підхід до завдань з кількома результатами, зрозумілість результатів та можливість чисельного тестування. Тим не менш, дерева рішень можуть стикатися з обмеженнями при роботі з категоріальними даними та можуть ускладнюватися за наявності невизначеності та пов'язаних результатів.

2.2.3 Переваги та недоліки дерев рішень

Древа рішень, що використовуються в інтелектуальному аналізі даних, надають низку переваг:

1. Простота розуміння та інтерпретації:
 - Моделі дерева рішень легко сприймаються людьми навіть без спеціальних знань.
 - Можливість графічного відображення полегшує інтерпретацію для нефахівців.
2. Обробка числових та категоріальних даних:
 - Древа рішень можуть працювати з різними типами даних, включаючи як числові, і категоріальні змінні.

- На відміну від деяких методів дерева не обмежені певним типом змінних, що робить їх більш універсальними.
3. Мінімальна підготовка даних:
- Немає потреби у складній передобробці даних, оскільки дерева можуть обробляти якісні предиктори, уникаючи створення фіктивних змінних.
4. Модель "білої скриньки":
- Дерева рішень є " білі ящики " , де пояснення рішень легко розуміється через булеву логіку.
 - На відміну від "чорних ящиків", таких як штучні нейронні мережі, результати легко інтерпретувати.
5. Можливість статистичної перевірки:
- Модель дерева рішень може бути піддана статистичним тестам, що забезпечує контроль та оцінку надійності моделі.
6. Непараметричний підхід:
- Дерева рішень не роблять припущень про розподіл даних або залишки прогнозу, що робить їх гнучкими та універсальними.
7. Ефективність із великими даними:
- Дерева рішень успішно справляються з аналізом великих обсягів даних, вимагаючи стандартних обчислювальних ресурсів.
8. Точність та гнучкість моделювання:
- Висока точність та адаптивність роблять дерева рішень ефективними для медичних досліджень та точного моделювання прийняття рішень.
9. Моделювання людських рішень:
- Дерева рішень краще відбивають процес прийняття рішень людиною, що корисно при моделюванні людських рішень та поведінки.

10. Стійкість до колінеарності:

- Стійкість до колінеарності, особливо в порівнянні з іншими методами, робить їх надійнішими.

11. Вбудований вибір функцій:

- Дерева рішень автоматично визначають значення атрибутів, що дозволяє видаляти непотрібні функції при наступних запусках.

12. Апроксимація будь-якої логічної функції:

- Дерева рішень можуть апроксимувати будь-яку логічну функцію, включаючи складні, такі як XOR.

Дерева рішень мають певні недоліки і викликають певні виклики, які важливо враховувати:

1. Нестійкість до змін у даних:

- Навіть невеликі зміни у навчальних даних можуть призвести до істотних змін у структурі дерева і, отже, у прогнозах моделі.

2. NP-повнота та евристики:

- Завдання навчання оптимального дерева рішень є NP-повним, що робить його складним для вирішення у певних сценаріях.
- Практичні алгоритми ґрунтуються на евристичних, таких як жадібний алгоритм, що може не гарантувати глобальної оптимальності.

3. Перенавчання:

- Навчання дерев може призводити до створення надмірно складних моделей, які невдало узагальнюють дані навчання, відоме як перенавчання.
- Заходи, такі як обрізка (за винятком деяких алгоритмів), можуть використовуватися для запобігання цьому явищу.

4. Складність дерева:

- Середня глибина дерева, яка визначається кількістю вузлів або тестів до класифікації, не завжди мінімальна або невелика за різних умов поділу.

5. Упередженість категоріальних змінних:

- Для даних із категоріальними змінними різного рівня складність дерева може бути нерівномірною на користь атрибутів із великою кількістю рівнів.
- Цей дисбаланс може бути пом'якшений, вибираючи атрибути з найбільшим коефіцієнтом приросту інформації серед тих, приріст яких перевищує середній рівень.

6. Альтернативні методи для упередженого вибору змінних:

- Підходи, такі як умовний висновок або адаптивний вибір виключених ознак, пропонуються як альтернативи для вирішення проблем упередженого вибору предикторів.
- Важливо врахувати ці обмеження та вживати відповідних заходів для оптимізації та забезпечення ефективності моделей на основі дерев рішень.

2.3 Заходи вибору атрибутів

Побудова дерева рішень: дерево можна «вивчити», розділивши вихідний набір на підмножини на основі міри вибору атрибутів. Міра вибору атрибутів (ASM) — це критерій, який використовується в алгоритмах дерева рішень для оцінки корисності різних атрибутів для поділу набору даних. Мета ASM полягає в тому, щоб визначити атрибут, який створить найбільш однорідні підмножини даних після поділу, тим самим максимізуючи приріст інформації. Цей процес повторюється для кожної похідної підмножини рекурсивним способом, що називається рекурсивним розбиттям. Рекурсія завершується, коли вся підмножина у вузлі має однакове значення цільової змінної або коли розбиття більше не додає значення прогнозів.

Для побудови класифікатора дерева рішень не потрібні знання предметної області чи налаштування параметрів, і тому він підходить для дослідницького виявлення знань. Деревя рішень можуть обробляти дані великого розміру.

Ентропія:

Ентропія є мірою ступеня випадковості чи невизначеності в наборі даних. У разі класифікацій він вимірює випадковість на основі розподілу міток класів у наборі даних. Ентропію для підмножини вихідного набору даних, що має K кількість класів для i -го вузла, можна визначити як:

$$H_i = - \sum_{k \in K}^n p(i, k) \log_2 p(i, k)$$

Де:

- S — набір даних.
- k — окремий клас із K класів
- $p(k)$ — частка точок даних, що належать до класу k , до загального кількості точок даних у вибірці S набору даних.

Важливі моменти, пов'язані з ентропією:

- Ентропія дорівнює 0, коли набір даних повністю однорідний, тобто кожен примірник належить до одного класу. Це найнижча ентропія, яка вказує на відсутність невизначеності у вибірці набору даних.
- коли набір даних порівну розподілено між кількома класами, ентропія має максимальне значення. Таким чином, ентропія найвища, коли розподіл міток класів рівномірний, що вказує на максимальну невизначеність у вибірці набору даних.
- Ентропія використовується для оцінки якості поділу. Метою ентропії є вибір атрибуту, який мінімізує ентропію результуючих підмножин шляхом поділу набору даних на більш однорідні підмножини щодо міток класу.
- Найвищий атрибут приросту інформації вибирається як критерій розщеплення (тобто зменшення ентропії після розщеплення за цим атрибутом), і процес повторюється рекурсивно для побудови дерева рішень.

2.4 Бінарні дерева

Бінарні дерева – це особливий тип дерев, де кожен вузол може мати не більше двох нащадків. У бінарних деревах шкірний вузол має лівого та правого сина. При відображенні цих дерев можна використовувати нахили сполучних ліній для розрізнення між лівим і правим синами. Однак, з погляду структури, упорядковані дерева завжди мають єдиного першого сина (лівий нащадок), тому з точки зору упорядкованого дерева, всі бінарні дерева є ідентичними. Таке співвідношення підкреслює, що бінарні дерева є специфічним видом дерев, а не окремим випадком упорядкованих дерев.

Бінарні дерева, як абстрактні математичні об'єкти, мають ряд цікавих властивостей, які можуть стати в нагоді при аналізі різних алгоритмів, тому зупинимося докладніше на цьому моменті. На будь-якому рівні n бінарне дерево може містити від 1 до 2^n вузлів. Число вузлів, що припадає на рівень, є показником густини дерева. На рис. 2.2 дерево А містить 8 вузлів при висоті 3, тоді як дерево В містить 5 вузлів при висоті 4.

Останній випадок є особливою формою, яка називається виродженим (degenerate) деревом, у якого є єдиний лист (e) і кожен внутрішній вузол має лише одного сина. Вироджене дерево вважатиметься аналогом лінійного зв'язкового списку, його висота дорівнює кількості вузлів без одиниці. Це дуже можливе значення для висоти бінарного дерева. У більшості алгоритмів, що використовують бінарні дерева, вироджене дерево є найгіршим випадком в оцінці продуктивності.

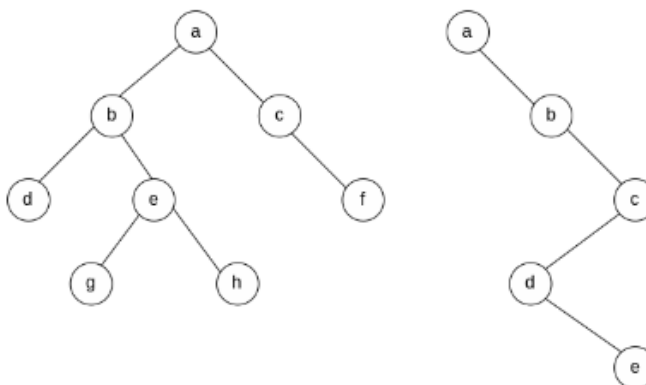


Рисунок 2.2 – Приклади бінарних дерев різної щільності

Натомість, дерева з великою щільністю дуже важливі як структури даних, оскільки вони містять пропорційно більше елементів поблизу кореня, тобто. з короткими шляхами від кореня. Найвищий рівень щільності мають повні бінарні дерева, які мають 2^k вузли на кожному рівні k . На рис. 2.3 зображено повне бінарне дерево висоти два. Звернімо увагу на кілька ключових моментів.

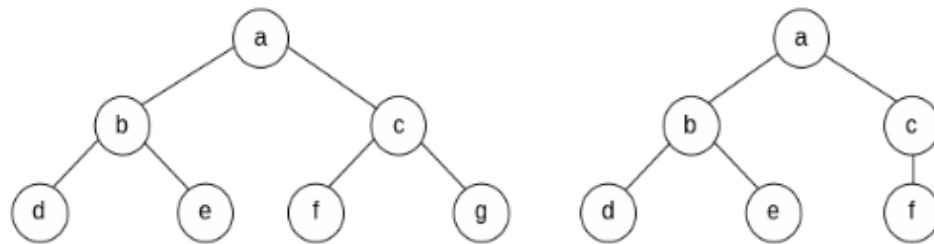


Рисунок 2.3 – Повне, та майже повне бінарні дерева

На нульовому рівні міститься 2^0 вузлів, на першому – 2^1 , на другому – 2^2 тощо. Також потрібно зауважити, що на перших $k-1$ рівнях кількість вузлів становить суму арифметичної прогресії, де кожен наступний рівень додає по одному вузлу до загального числа. Це є важливим аспектом при аналізі та характеристиці структури бінарних дерев.

$$1 + 2 + 4 + \dots + 2^{k-1} = 2^k - 1$$

На рівні k кількість вузлів становить 2^k , тобто рівно на один більше. Це призводить до висновку, що у повному бінарному дереві кількість внутрішніх вузлів на одиницю менша за кількість листя. Знаючи кількість листя, легко визначити висоту h повного бінарного дерева:

$$h = \log_2 n \text{ де } n - \text{кількість листя, або:}$$

$$h = \log_2(N + 1) - 1 \text{ де } N - \text{кількість вузлів повного бінарного дерева.}$$

Для повного бінарного дерева зазначені формули надають точне значення, яке відповідає мінімальній висоті при заданій кількості вузлів. Якщо кількість вузлів унеможливорює побудову повного бінарного дерева, для створення бінарного дерева з мінімально можливою висотою необхідно заповнити всі рівні, крім останнього, максимально можливим числом вузлів.

Якщо вузли, що залишилися, розмістити на останньому рівні зліва направо, бінарне дерево, що вийшло, буде називатися майже повним. На рис. 2.4 зображено таке майже повне бінарне дерево.

Повні і майже повні бінарні дерева мають ще одну цікаву властивість: якщо вузли нумерувати, починаючи з одиниці, зверху вниз і зліва направо, то лівому синові завжди відповідатиме код, вдвічі більший, ніж код його батька, а правому синові — код, на одиницю більшу, ніж код лівого сина (рис. 3.8). Номер кореня завжди дорівнює 1, його лівий нащадок отримує номер 2, а правий – номер 3. Лівий нащадок вузла 2 отримає номер 4, а правий – 5, лівий нащадок вузла 3 отримає номер 6, правий – 7 тощо. Така схема нумерації використовується для представлення дерев з використанням масивів, і ми повернемося до цієї теми пізніше.

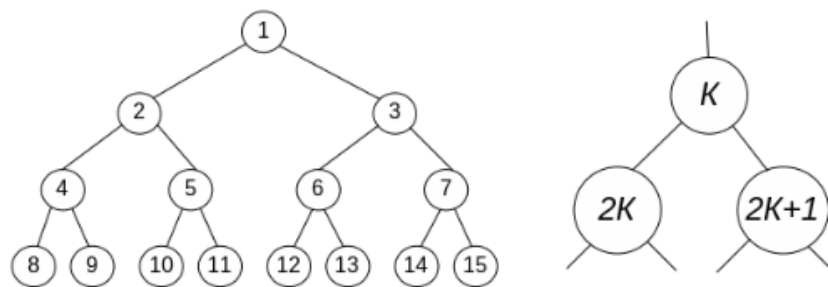


Рисунок 2.4 - Нумерація вузлів повного або майже повного бінарного дерева

Такої нумераційної схемою можна користуватися для пронумерованих вузлів бінарних дерев, які не є майже повними, оскільки в цьому випадку гарантується унікальність кожного номера, навіть якщо в процесі роботи до дерева додаються нові листки. Використовуючи такий метод нумерації, можна реалізувати структуру, що ґрунтується на дереві, з використанням масиву.

2.5 Розрахунок прикладу моделі бінарного дерева

Базова модель передбачає, що ціна акції або зросте до \$115 або впаде до \$85 протягом одного періоду. Насправді, по-перше, вартість активу зростає і падає дрібними кроками, вибудовуючи згодом цілу динаміку руху.

По-друге, чинники u і d , змінюють ціну акції від спотового значення, використовуються лише ілюстрації. Насправді розробники опціонних моделей прагнуть вивести найбільш підходящі значення цих чинників, обчислюючи волатильність ціни акції. З логічних міркувань випливає, що за інших рівних більш висока волатильність базового активу призведе до більшого відхилення ціни від поточного рівня i , отже, до вищої вартості опціону.

Щоб показати, як можна вирішити питання з волатильністю, збудуємо 3-х крокове бінарне дерево для знаходження справедливої вартості європейського опціону на акцію, яка не сплачує дивіденди.

Спотова ціна акції $S = \$300$

Ціна виконання опціону $X = \$250$

Безризикова процентна ставка $r = 10\%$ річних

Час до експірації $t = 0,25$ років

Волатильність $\sigma = 40\%$

Щоб побудувати дерево, що починається на рисунку 2.5, та співвідносить волатильність базового активу з факторами u і d , ми будемо використовувати формули, запропоновані американськими академіками Коксом, Россом і Рубінштейном:

$$u = e^{\sigma\sqrt{t/n}}$$

$$d = \frac{1}{u}$$

e – математична константа (приблизно дорівнює 2,71828)

σ – волатильність, стандартне відхилення денних доходностей ціни базового активу, виражена в річному еквіваленті (у нашому випадку 0,4)

t – час до погашення опціону в роках (0,25 років або 3 місяці)

n – кількість кроків бінарного дерева (три кроки)



Рисунок 2.5 - Трикрокова модель бінарного дерева

Виконаємо розрахунки:

$$u = e^{0.4\sqrt{0.25/3}} = 1.1224$$

$$d = \frac{1}{1.1224} = 0.8909$$

Загальний час дії опціону ділиться на три рівні періоду. На цьому рисунку показано, як значення u та d використовуються для побудови бінарного дерева. Наприклад, перший рух ціни активу вгору має значення \$337, отриманий шляхом множення \$300 на величину u . Значення \$267 розраховується шляхом множення \$300 величину d . Якщо ціна акції досягає \$267 у період 1, вона може або збільшитися до \$300 у період 2 (шляхом множення \$267 на u), або продовжити зниження до \$238. Далі відбувається створення дерева, що відображає вартість кол-опціону із забігом \$250, ґрунтуючись на динаміці зміни ціни акцій. Почати найлегше з кінцевої точки - у момент експірації вартість колл-опціону визначається виключно його внутрішньою вартістю, як показано на Рис. 4. Наприклад, у верхньому правому куті дерева ціни базового активу на Рис. 3, вартість акції складає \$424. У цьому випадку кол з виконавчою ціною \$250 оцінюватиметься в \$174.

На рисунку. 2.6 буква «C» є вартість опціону в момент 0. C_u – вартість опціону при зростанні ціни акції на один рівень; C_{uu} – вартість опціону у разі зростання акції на два рівні. C_d – вартість опціону при падінні ціни акції на один рівень; C_{dd} – вартість опціону при падінні на два рівні; C_{ud} – вартість опціону при зростанні ціни акції на один рівень та подальшому падінні на один рівень.

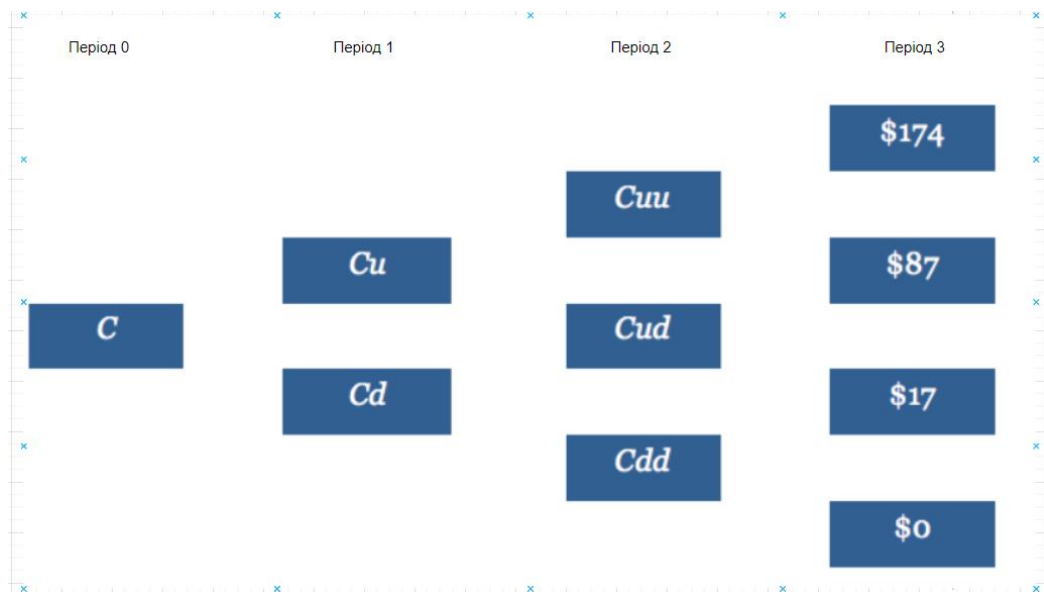


Рисунок 2.6 – Трикрокове дерево для кол опціону

Оцінка вартості кол-опціону в кожному періоді може бути досягнута шляхом розбиття трикрокового дерева на три окремі дерева з одним кроком, що полегшить обчислення. Спочатку необхідно обчислити ймовірності p і $(1 - p)$, використовуючи чинники зростання u і падіння d .

$$p = \frac{e^{0.1 \times 0.25/3} - 0.8909}{1.1224 - 0.8909} = 0.5074$$

$$1 - p = 0.4926$$

C_{uu} являє собою дисконтовану суму, отриману таким чином: по-перше, множиться вартість кол-опціону при зростанні акції до \$424 (тобто \$174) на ймовірність досягнення акцією рівня \$424 за період; по-друге, вартість кол-опціону при зростанні акції до \$337 (тобто \$87) множиться на ймовірність зростання акції до рівня \$337 за той же період. Отримані результати підсумовуються і потім дисконтуються.

Розрахунок C_{uu} :

$$C_{uu} = [(\$174 \times 0.5074) + (\$87 \times 0.4926)] \times e^{0.1 \times 0.25/3} = \$130$$

Завершена трикрокова бінарна діаграма представлена на рисунку 2.7. Поточна оцінна вартість кол-опціону із закінченням терміну через три місяці приблизно становить \$61.

У разі оцінки вартості американського пут-опціону критичним є регулярна перевірка на кожному кроці, щоб гарантувати, що ціна не опускається нижче за внутрішню вартість опціону. Така необхідність обумовлена можливістю вигідного дострокового виконання американського пут-опціону у деяких випадках.



Рисунок 2.7 – Трикрокове дерево для кол опціону після розрахунку

У контексті фінансового аналізу та машинного навчання, дерево рішень відіграє значну роль, надаючи універсальний та потужний алгоритм для завдань класифікації та регресії. Цей метод дозволяє візуалізувати та інтерпретувати процес прийняття рішень, що є важливим аспектом аналізу даних. Дерево рішень також широко застосовується у складі випадкового лісу, підвищуючи його ефективність та точність. У сукупності ці інструменти забезпечують надійні засоби для аналізу фінансових ринків та прийняття обґрунтованих інвестиційних рішень.

Висновки до розділу 2

Дерево рішень представляє собою важливий алгоритм машинного навчання, здатний вирішувати завдання класифікації та регресії. Його структура, схожа на блок-схему, дозволяє візуалізувати та інтерпретувати прийняті рішення. Цей універсальний контрольований алгоритм став основою для випадкового лісу, який об'єднує кілька дерев рішень. Випадковий ліс ефективно використовується для навчання на різних підмножинах даних, що дозволяє зменшити ризик перенавчання та забезпечити високу точність прогнозування. Цей інтегрований підхід робить дерево рішень та випадковий ліс важливими інструментами для аналізу даних і передбачення в різноманітних областях, забезпечуючи високий рівень надійності та результативності. А також подібні алгоритми виявляють особливу ефективність у розв'язанні завдань, пов'язаних із складними структурами даних, де важлива якість аналізу та точність передбачень. Зокрема, їх можливість робити адаптивний аналіз на різних підмножинах даних робить їх надзвичайно придатними для сучасних завдань аналітики, включаючи фінансовий аналіз, медичні дослідження та інші області, де точність та надійність важливі для прийняття важливих рішень. Такий підхід дозволяє застосовувати їх як в ізольованому вигляді, так і в складі більш складних систем обробки інформації для досягнення високого рівня ефективності в аналізі та передбаченні динаміки даних.

3 ВІЗУАЛІЗАЦІЯ ТА КАТЕГОРИЗАЦІЯ ДАНИХ

Терміни "візуалізація даних" та "інфографіка" часто розглядаються як взаємозамінні, але фахівці в галузі подання інформації підкреслюють їх відмінності. Згідно з однією точкою зору, візуалізація даних сприймається як складова інфографіки. У цій концепції інфографіка є поєднанням візуалізації даних, ілюстрацій, малюнків і тексту, об'єднаних для передачі комплексного повідомлення.

Інфографіка прагне створювати сильне візуальне враження, використовуючи елементи дизайну, колірну гаму та композицію для привернення уваги до ключових аспектів представленої інформації. У той час як візуалізація даних зосереджується на чіткому та точному відображенні даних, орієнтованому на їхній аналіз. Ця форма подання даних націлена на точність передачі інформації та зручність інтерпретації, забезпечуючи об'єктивність та доступність для широкої аудиторії.

Інфографіка і візуалізація даних, незважаючи на свою природу, що перекривається, мають відмінності в цілях і підходах до подання інформації, особливо в контексті категоризації даних. Інфографіка виділяє властивості мистецтва та дизайну, приділяючи увагу емоційному впливу та креативному візуальному сприйняттю, що в контексті категоризації даних передбачає вибір специфічних елементів для передачі певних категорій інформації, забезпечуючи високу естетичну якість та унікальність візуального досвіду.

З іншого боку, візуалізація даних фокусується на чіткому та точному поданні даних для їх аналізу та інтерпретації. У контексті категоризації даних візуалізація може використовуватися для угруповання, виділення закономірностей чи трендів усередині конкретних категорій, забезпечуючи об'єктивність та доступність візуалізованої інформації.

Таким чином, під час роботи з категоризацією даних вибір між інфографікою та візуалізацією даних залежить від унікальних вимог мети представлення інформації та взаємодії з аудиторією.

3.1 Візуалізація даних у комп'ютерній інженерії

Візуалізація даних – це метод представлення інформації у формі графічних елементів, що сприяє більш ефективному дослідженню даних. У комп'ютерній інженерії візуалізація даних, схема якої показана на рис 3.1, широко використовується для аналізу великих обсягів інформації та представлення результатів досліджень. Вона відіграє ключову роль у різних галузях, включаючи наукові та статистичні дослідження, прогнозування, інтелектуальний аналіз даних та бізнес-аналіз. Також важливу роль візуалізація даних відіграє у педагогічному дизайні для навчання та тестування, а також у створенні зведень новин та аналітичних оглядів.

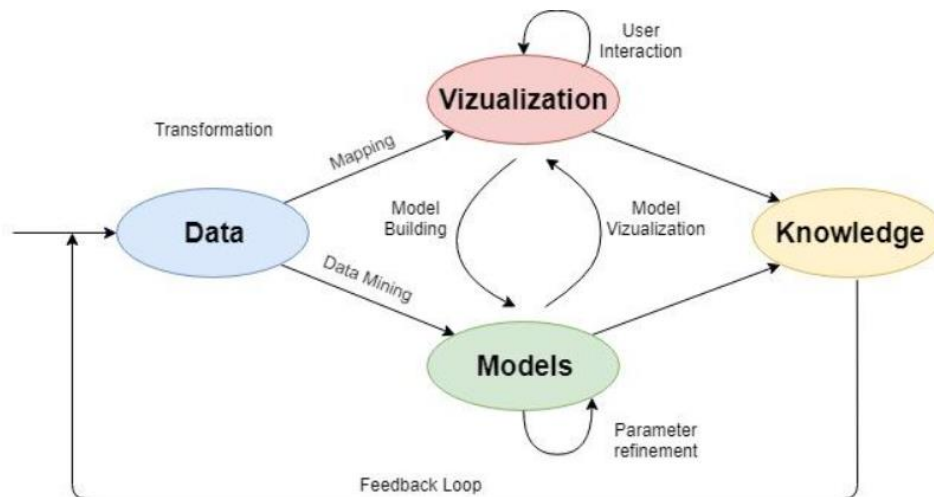


Рис. 3.1 – Схема візуалізації даних

Хоча терміни "візуалізація даних" та "інфографіка" часто розглядаються як синоніми, фахівці в галузі подання інформації проводять між ними різницю. В одному підході візуалізація даних розглядається як частина інфографіки, що є поєднанням візуалізації даних, ілюстрацій, малюнків і тексту для передачі комплексного повідомлення. Інший підхід розмежовує ці поняття за методом створення, естетичними якостями та обсягом даних. Відповідно до цього підходу, інфографіка відноситься до подання даних, створених вручну, специфічним для інформації, що подається, високоестетичним і не містить великого обсягу даних.

З іншого боку, візуалізація даних створюється алгоритмічним шляхом, що легко відтворюється для різних вибірок і не містить зайвих декоративних елементів, що дозволяє відображати великі обсяги даних.

У контексті комп'ютерної інженерії візуалізація даних використовується для полегшення аналізу та інтерпретації великих обсягів інформації. Вона допомагає візуалізувати вихідні дані, результати обробки та проміжні результати, що покращує сприйняття інформації та забезпечує розуміння закономірностей та трендів у даних. Візуалізація даних також є важливою частиною систем інтелектуального аналізу даних, забезпечуючи стислість, відносність, концентрацію та контекст, масштабованість та орієнтацію на "праву півкулю" для користувачів.

Візуалізація даних у комп'ютерній інженерії може також використовуватися для дослідження продуктивності програмного забезпечення, виявлення вузьких місць в алгоритмах, аналізу ефективності алгоритмів оптимізації та оцінки якості різних компонентів системи. Завдяки візуалізації даних інженери можуть швидше виявляти проблеми та покращувати продуктивність, а також легше пояснювати та демонструвати результати своєї роботи колегам та замовникам.

Таким чином, візуалізація даних є невід'ємною частиною процесу розробки програмного забезпечення та інженерних рішень, забезпечуючи ефективніший аналіз даних, покращення продуктивності та зручність взаємодії з інформацією.

Якісна візуалізація даних відіграє вирішальну роль в аналізі інформації та прийнятті обґрунтованих рішень. Подання даних у графічному форматі дозволяє нам швидко виявляти взаємозв'язки та тренди, які могли б бути непоміченими під час роботи з сирими даними. Крім того, візуалізація робить інформацію більш доступною та зрозумілою, що знижує ризик неправильної інтерпретації. Це особливо важливо, враховуючи, що більшості людей простіше та приємніше сприймати інформацію через графічні елементи, ніж через текстові документи чи таблиці.

Ефективна візуалізація не тільки містить у собі дані, а й акцентує увагу на ключових аспектах інформації, привертаючи увагу та утримуючи інтерес аудиторії. Це засноване на принципі того, що візуальні образи можуть бути більш незабутніми та легше засвоюваними, ніж текстова інформація. Таким чином, графічне подання даних не тільки допомагає нам швидше і краще зрозуміти інформацію, але й робить цей процес більш приємним та захоплюючим.

3.2 Категоризація даних у комп'ютерній інженерії

У галузі штучного інтелекту та аналізу даних, категоризацією розуміється процес віднесення спостережень або об'єктів до певних груп, які називаються категоріями. На відміну від класифікації, де об'єкти належать до заздалегідь визначених класів, завдання категоризації має кілька особливостей:

1. Категорія описує лише загальні характеристики об'єктів, не задаючи точних кордонів чи суворих правил віднесення.
2. Підхід до категоризації менш жорстко визначений, ніж класифікації, і часто вимагає суб'єктивної оцінки.
3. Кордони між категоріями часто нечіткі, що визначаються скоріше порівнянням з іншими категоріями, ніж формальними правилами.

Завдання категоризації є складнішим завданням порівняно з класифікацією, оскільки категорії не зумовлені, а формуються у процесі. У контексті навчання з учителем процес формування категорій здійснюється шляхом проб і помилок на основі навчальних прикладів з відомими відповідями, що надаються експертом.

Система-категоризатор грає ключову роль процесі формування узагальнених ознак з урахуванням наявних даних. При збільшенні числа прикладів значимі ознаки посилюються, а випадкові значення згладжуються, що призводить до поступового уточнення меж між категоріями.

У контексті комп'ютерної інженерії завдання категоризації даних має важливе значення, особливо в галузях машинного навчання та обробки великих даних. Категоризація даних часто використовується для організації та структуризації інформації, що дозволяє системам та алгоритмам комп'ютерного навчання краще розуміти та аналізувати дані.

Комп'ютерні інженери розробляють та реалізують системи категоризації, які здатні автоматично обробляти та класифікувати дані відповідно до певних критеріїв. У цьому процесі використовуються різні методи машинного навчання, включаючи алгоритми кластеризації та класифікації, щоб визначити оптимальні категорії для набору даних.

Більше того, комп'ютерні інженери розробляють та оптимізують алгоритми категоризації для забезпечення високої точності та ефективності обробки даних. Це включає роботу над методами визначення оптимальної кількості категорій, поліпшення алгоритмів виділення ознак і оптимізацію обчислювальних процесів для обробки великих обсягів даних.

Таким чином, категоризація даних відіграє важливу роль у комп'ютерній інженерії, допомагаючи системам та алгоритмам обробляти, аналізувати та використовувати інформацію більш ефективно.

3.3 Опис компонентів ESP 8266 та MAX 30102

В данному пункті магістерської роботи буде розглянуто процес отримання даних за допомогою інтегрованих модулів ESP8266 та MAX30102. ESP8266 є популярним мікроконтролером з вбудованим модулем Wi-Fi, здатним забезпечити бездротовий зв'язок, тоді як MAX30102 є інтегрованим датчиком серцевого ритму та кисню в крові, призначеним для точного вимірювання фізіологічних показників. У цьому розділі будуть розглянуті технічні аспекти з'єднання та конфігурації цих компонентів для забезпечення ефективного збору даних.

Важливими аспектами будуть визначення протоколу взаємодії між ESP8266 та MAX30102, розробка програмного забезпечення для зчитування даних з датчика та їх передачі через мережу Wi-Fi, а також вивчення питань щодо забезпечення безпеки та надійності збору та передачі інформації. Виконання цього етапу дослідження є важливим кроком у розвитку системи моніторингу фізіологічних параметрів та її подальшого застосування в медичних та фітнес-технологіях.

3.3.1 Базова інформація ESP 8266

ESP8266 – це мікроконтролер, розроблений компанією Espressif Systems, який відрізняється високою інтеграцією та вбудованим Wi-Fi інтерфейсом.

Таблиця 3.1 – Порівняльна характеристика МК ESP

Мікроконтролер	Частота процесора	Wi-Fi	Bluetooth	GPIO	Пам'ять (Flash/RAM)	Особливості
ESP8266	80/160 МГц	802.11b/г/н	Hi	До 17	512 Кб / 80 Кб	Вбудований Wi-Fi, низьке енергоспоживання
ESP32	160/240 МГц	802.11b/г/н	4,2/5,0	До 34	4 Мб / 520 Кб	Вбудований Wi-Fi та Bluetooth, два ядра CPU
ESP32-C2	240 МГц	802.11b/г/н	Hi	До 43	4 Мб / 320 Кб	Вбудований Wi-Fi, низьке енергоспоживання
ESP32-C3	160 МГц	802.11b/г/н	5.0	До 22	2 Мб / 384 Кб	Вбудований Wi-Fi та Bluetooth, низьке енергоспоживання

На відміну від інших мікроконтролерів ESP8266, схема якої показана на рис.3.2, не має вбудованої флеш-пам'яті всередині системи на кристалі (SoC).

Цей SoC (система на кристалі) інтегрований зі стеком протоколів TCP/IP, який може забезпечити доступ мікроконтролера до будь-якого типу Wi-Fi. У 2014 році мікроконтролер привернув увагу багатьох розробників своєю неймовірно низькою ціною та можливістю інтеграції Wi-Fi у проекти з мінімальними витратами. Він швидко став популярним інструментом для створення Інтернету речей (IoT) та інших проектів, які потребують бездротового з'єднання.

Він може використовуватися як підлеглий або автономний додаток. Натомість програма користувача виконується із зовнішньої флеш-пам'яті, підключеної через інтерфейс SPI.

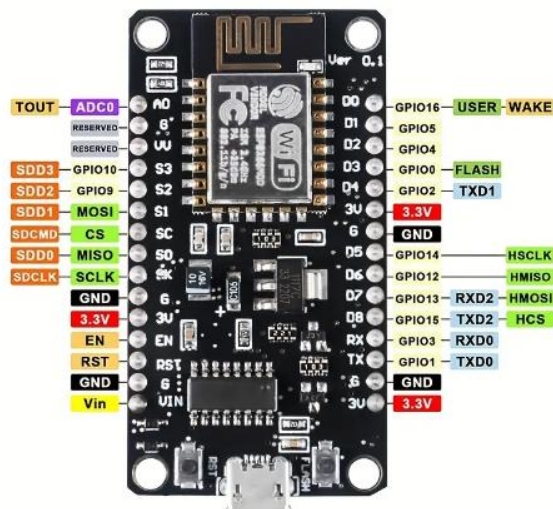


Рис. 3.2 – Схема ESP8266

Модуль Wi-Fi ESP8266 є мікрочіп SOC, який в основному використовується для розробки кінцевих додатків Інтернету речей (IoT). Його називають автономним бездротовим трансівером, доступним за дуже низькою ціною, що привернув увагу багатьох розробників у 2014 році. Цей модуль надає рішення, що відповідають вимогам галузей Інтернету речей, таких як вартість, потужність, продуктивність та дизайн.

Якщо ESP8266 Wi-Fi працює як підлеглий пристрій хоста мікроконтролера, його можна використовувати як адаптер Wi-Fi для будь-якого типу мікроконтролера, який використовує UART або SPI. Коли модуль використовується як самостійний додаток, він забезпечує функції мікроконтролера і мережі Wi-Fi.

Модуль Wi-Fi ESP8266 тісно інтегрований з радіочастотним балуном, модулями живлення, радіочастотним передавачем та приймачем, аналоговим передавачем та приймачем, підсилювачами, фільтрами, цифровою основною смугою, модулями живлення, зовнішніми схемами та іншими необхідними компонентами. Цей модуль є мікросхемою, яка може легко взаємодіяти з мікроконтролером через послідовний порт.

Інші модулі Espressif Systems ESP8266 доступні в різних варіантах, включаючи такі, як ESP-01, ESP-12F, NodeMCU, Wemos D1 Mini, Adafruit HUZZAH ESP8266, SparkFun ESP8266 Thing, Olimex MOD-WIFI-ESP8266-DEV та ESP-32. Кожен з цих модулів має свої унікальні характеристики та можливості, що робить їх ідеальними для різних застосувань в Інтернеті речей (IoT), розробки прототипів та інших проектів з підтримкою Wi-Fi зв'язку. Інші модулі Espressif Systems ESP8266 доступні в різних варіантах, включаючи ті, що в табл.3.2:

Таблиця 3.2- Порівняння модулів ESP8266

Model	Antennas	Dimensions	GPIO and ADC pins	Flash	LEDs	Shielded
ESP-ROOM-02	PCB trace	18×20	18	2 MB	No	Yes
ESP-WROOM-02D	PCB trace	18×20	18	2 MB	No	Yes
ESP-WROOM-02U	U.FL socket	18×20	18	2 MB	No	Yes
ESP-ROOM-S2	PCB trace	16×23	20	2 MB	No	Yes

Процесор WiFi модуля ґрунтується на Tensilica Xtensa Diamond Standard 106 micro та легко працює на частоті 80 МГц. Існують різні типи модулів ESP, розроблених сторонніми виробниками, такі як ESP8266-01, ESP8266-02, ESP8266-03 та ESP8266-04, кожен з яких має свої характеристики та контактні інтерфейси.

Модуль Wi-Fi ESP8266 поставляється із завантажувальним ПЗУ об'ємом 64 КБ, ОЗУ даних об'ємом 80 КБ і ОЗУ інструкцій об'ємом 32 КБ. Він підтримує мережу Wi-Fi 802.11 b/g/n на частоті 2,4 ГГц та функції інтерфейсу I2C, SPI, I2C з DMA та 10-бітного АЦП.

Набір АТ-команд потрібний мікроконтролеру для зв'язку з модулем Wi-Fi ESP8266. Це дозволяє використовувати функції Arduino Wi-Fi та завантажувати різне програмне забезпечення для розробки власної програми у пам'яті та процесорі модуля.

Модуль ESP8266 широко використовується в робототехніці та додатках Інтернету речей завдяки своїй низькій вартості та компактним розмірам.

Всі ці технологічні новинки відіграють ключову роль у розвитку сучасних систем Інтернету речей та смарт-пристроїв, забезпечуючи стабільне та надійне з'єднання у бездротових мережах та розширюючи можливості впровадження технологій у повсякденне життя. Мікроконтролер ESP8266 оснащений 80 МГц 32-бітним процесором Tensilica Xtensa L106, який забезпечує високу продуктивність та ефективність.

Існує можливість негарантованого розгону до 160 МГц, що дає змогу досягти ще більшої швидкості обробки даних.

Основною особливістю ESP8266 є наявність підтримки Wi-Fi стандартів IEEE 802.11 b/g/n, що робить його ідеальним для проектів, що потребують бездротового підключення до Інтернету. Він підтримує протоколи безпеки WEP та WPA/WPA2, забезпечуючи надійний захист даних.

Мікроконтролер має 14 портів вводу-виводу, з яких 11 можуть бути використані для підключення зовнішніх пристроїв. Він також забезпечує інтерфейси SPI, I²S та UART, а також 10-бітний аналого-цифровий перетворювач (АЦП), що розширює його можливості для підключення різних датчиків та пристроїв.

Однією з особливостей мікроконтролера є її здатність працювати з низьким енергоспоживанням. Він може переходити до різних режимів зниженого енергоспоживання, таких як Modem sleep, Light sleep та Deep sleep, що дозволяє збільшити термін служби пристроїв на його основі.

Незважаючи на відсутність на кристалі мікроконтролера енергонезалежної пам'яті, він підтримує виконання програм із зовнішньої SPI флеш-пам'яті, що дозволяє розширити доступний простір для програм та даних до 16 МБ. Для роботи з периферією мікроконтролера надається набір бібліотек, через API яких програміст може отримати доступ до різних інтерфейсів та функцій.

Детальні електричні параметри, схеми включення та інші технічні характеристики можна знайти в документації, включаючи 0A-ESP8266EX__Datasheet і 0B-ESP8266__System_Description з Espressif SDK. Також важливо відзначити, що джерело програми ESP8266, що виконується, задається станом певних портів GPIO0, GPIO2 і GPIO15, що дозволяє вибирати між виконанням коду з UART або із зовнішньої ПЗУ в залежності від потреб проекту.

3.3.2 Базова інформація MAX 30102

MAX30102 - інноваційний пристрій з низьким енергоспоживанням, призначений для вимірювання пульсу та рівня кисню в крові. Працюючи через інтерфейс I2C.

Таблиця 3.3 – Порівняльна характеристика датчиків MAX

Датчик	Додаток	Інтерфейс	Діапазон вимірювання	Bit	Примітки
MAX30102	Пульс, SpO2	I2C	Пульс: 30-250 уд/хв SpO2: 70-100%	18	Медичні додатки
MAX31855	Температура	SPI	-270 до + 1800 °C	0.25°C	Висока точність
MAX44009	Освітленість	I2C	0.045 лк до 188000 лк	16	Низьке споживання енергії
MAX30105	Пульс, SpO2	I2C	Пульс: 30-250 уд/хв SpO2: 70-100%	18	Алгортми фільтрації шумів
MAX31865	Температура	SPI	Різні типи термопар	Різні	Промислове застосування
MAX30101	Пульс, SpO2	I2C	Пульс: 30-250 уд/хв SpO2: 70-100%	18	Схож з MAX30102

На відміну від інших, цей біометричний датчик забезпечує просте підключення та стабільну роботу без зайвих складнощів. Він придатний для широкого кола користувачів: від студентів та ентузіастів до досвідчених інженерів та професійних розробників. Крім того, MAX30102 ідеально підходить для створення ігор та мобільних програм, де необхідний моніторинг пульсу в реальному часі.

Цей датчик має високу точність вимірювань та широкі функціональні можливості, що робить його ідеальним вибором для медичних, спортивних та фітнес-додатків, а також для пристроїв моніторингу здоров'я.

На модулі розташовані два світлодіоди – червоний та інфрачервоний, як показано на рис. 3.3, а також чутливий фотодетектор. Це забезпечує одночасне освітлення шкіри та вимірювання кількості світла, відбитого назад на датчик. Завдяки цій технології, заснованій на аналізі сигнатури світла, датчик здатний точно вимірювати рівень кисню в крові та частоту серцевих скорочень.

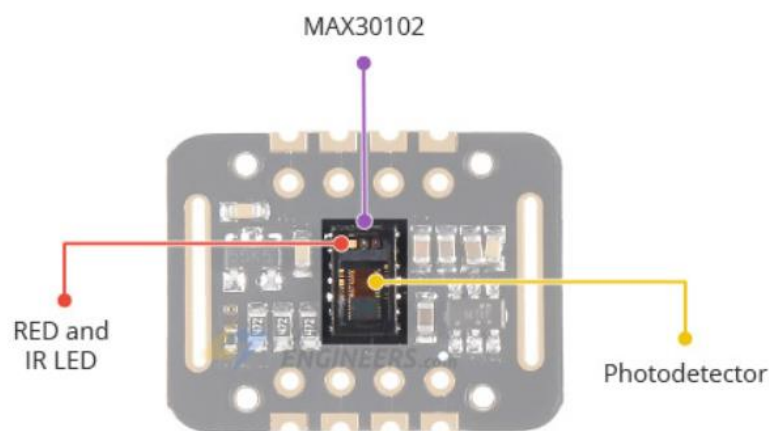


Рисунок 3.3 – Розташування світлодіодів на MAX30102

Модуль включає сучасний пульсоксиметр і датчик серцевого ритму MAX30102 від Analog Devices, що є поліпшеною версією попередньої моделі MAX30100. Він поєднує два світлодіоди, фотодетектор, оптимізовану оптику та аналогову обробку сигналів для точного визначення пульсоксиметрії (SpO₂) та частоти серцевих скорочень (HR).

Мікросхема MAX30102 вимагає дві окремі напруги живлення: 1,8 В для самої мікросхеми та 3,3 В для роботи червоного та ІЧ-світлодіодів, як показано на рис. 3.4. Для цього модуль комплектується регуляторами 3,3 В та 1,8 В.

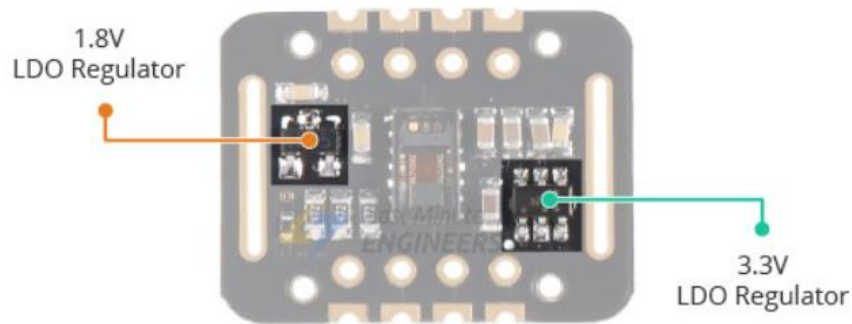


Рисунок 3.4 – Схема розташування напруг живлення MAX30102

На звороті друкованої плати передбачена перемичка, яка використовується для вибору логічного рівня між 3,3 В і 1,8 В. За замовчуванням встановлено логічний рівень 3,3 В, сумісний з логічними рівнями Arduino, але ви також можете вибрати логічний рівень 1,8 В залежності від ваших потреб.

Це забезпечує можливість підключення модуля до будь-якого мікроконтролера з входом/виходом на рівні 5, 3,3 і навіть 1,8 Ст.

Однією з ключових особливостей MAX30102 є низьке енергоспоживання: під час вимірювання споживання становить менше 600 мкА. Крім того, MAX30102 може бути переведений в режим очікування, де його енергоспоживання знижується до 0,7 мкА. Це низьке енергоспоживання дозволяє використовувати його в пристроях з батарейним живленням, таких як мобільні телефони, пристрої, що носяться, і розумний годинник.

MAX30102 має вбудований датчик температури, призначений для компенсації впливу навколишнього середовища на вимірювання та калібрування пристрою.

Цей датчик має високу точність і здатний вимірювати "температуру кристала" в широкому діапазоні -40°C до $+85^{\circ}\text{C}$ з точністю $\pm 1^{\circ}\text{C}$. Його застосування дозволяє коригувати вимірювання та забезпечувати більш надійну роботу пристрою у різних умовах навколишнього середовища.

3.4 Вдосконалення візуалізації за допомогою 3D-моделювання

У світі тривимірне моделювання грає ключову роль різних галузях, і комп'ютерна інженерія перестав бути винятком. Завдяки безперервному розвитку технологій та програмного забезпечення, тривимірне моделювання стає все більш актуальним та затребуваним у галузі комп'ютерної інженерії.

Техніки тривимірного моделювання дозволяють інженерам візуалізувати та проектувати складні системи та конструкції, створюючи цифрові прототипи, аналізуючи їхню поведінку та оптимізуючи дизайн до того, як фізичний прототип буде створено. Це дозволяє скоротити час та витрати на розробку, підвищити якість продукції та покращити загальну ефективність проекту.

Таким чином, тривимірне моделювання відіграє важливу роль у різних аспектах комп'ютерної інженерії, включаючи проектування механізмів, електроніки, архітектури, створення віртуальних середовищ та багато інших. Воно допомагає інженерам у створенні інноваційних рішень, підвищенні продуктивності та зниженні ризиків при розробці нових технологій та продуктів.

3D-моделювання – це процес створення об'ємних зображень у спеціалізованих комп'ютерних програмах. Результатом 3D-моделювання є 3D-модель, яка дозволяє уявити майбутній об'єкт у обсязі та детально описати всі його характеристики. Програми для 3D-моделювання забезпечують можливість створювати реалістичні зображення людей, тварин, предметів, поверхонь та міських ландшафтів, що дає змогу відстежити всі нюанси проекту та усунути помилки на ранніх етапах.

Методи, що використовуються у 3D-моделюванні:

- Параметричне моделювання, де об'єкти створюються на основі параметрів та правил, що дозволяє легко вносити зміни та редагувати моделі.
- Воксельне моделювання, яке використовує вокселі (тривимірні пікселі) для створення об'ємних форм.

- Полігональне моделювання, що ґрунтується на використанні багатокутників для опису поверхні об'єктів.
- Сплайнове моделювання (NURBS), де для створення кривих та поверхонь використовуються неунормовані раціональні B-сплайни.
- Скульптуринг, який дозволяє моделювати об'єкти, додаючи та видаляючи матеріал на кшталт скульптури.

Існує безліч програм для 3D-моделювання як платних, так і безкоштовних, кожна з яких призначена для виконання певних завдань у конкретних областях.

Autodesk 3Ds Max: Це один з найвідоміших і найпотужніших інструментів для створення 3D-моделей, анімації та візуалізації. Він широко використовується в різних галузях, включаючи ігрову індустрію, кіно, архітектуру та дизайн. 3Ds Max пропонує багатий набір інструментів для створення складних об'єктів, анімації персонажів, моделювання довкілля та багато іншого. Він також інтегрується з іншими продуктами Autodesk, що полегшує співпрацю над проектами.

Autodesk Maya: Це програмне забезпечення, яке надає широкі можливості для моделювання, анімації, симуляції та рендерингу. Maya є одним з найпопулярніших інструментів в індустрії анімації та спецефектів у кіно та телебаченні. Вона пропонує просунуті інструменти для створення складних персонажів, динамічних сцен та реалістичних ефектів.

AutoCAD: Це програмне забезпечення для створення 2D- та 3D-моделей, що широко використовується в інженерних та архітектурних проектах. AutoCAD пропонує потужні інструменти для проектування будівель, машин та інших об'єктів. Він дозволяє створювати точні креслення, моделювати об'єкти з різними параметрами та виконувати складні розрахунки.

Cinema 4D: Ця програма для створення високоякісної 3D-графіки та анімації. Воно надає інтуїтивно зрозумілий інтерфейс та безліч інструментів для моделювання, текстурування, анімації та рендерингу. Cinema 4D широко використовується в рекламній індустрії, телевізійному та кінематографічному виробництві, а також у створенні анімації та спецефектів.

Rhinoceros (Rhino): Це програма для тривимірного моделювання, яка особливо популярна у дизайні та промисловому дизайні. Rhino пропонує широкий набір інструментів для створення складних геометричних форм, поверхонь і текстур. Він також підтримує широкий спектр файлів для спільної роботи з іншими програмами.

Blender 3D: Це безкоштовне та відкрите програмне забезпечення для створення 3D-графіки, анімації та ігор. Blender надає потужні інструменти для моделювання, анімації, текстурування та рендерингу. Він має активну спільноту користувачів і регулярно оновлюється з додаванням нових функцій та покращень.

SketchUp: Це інструмент для створення 3D-моделей, який широко використовується в архітектурному проектуванні та будівництві. SketchUp має простий та інтуїтивно зрозумілий інтерфейс, що робить його доступним навіть для новачків. Він дозволяє швидко створювати прості та складні 3D-моделі будівель, меблів, ландшафтів та багато іншого.

Програми для 3D-моделювання надають широкий набір інструментів та функцій, які дозволяють користувачам створювати реалістичні та деталізовані моделі різних об'єктів. З їх допомогою можна моделювати людей, тварин, рослини, техніку, архітектурні споруди, ландшафти та багато іншого.

Висновки до розділу 3

Удосконалення візуалізації за допомогою Autodesk Maya є важливим кроком для розвитку сучасних технологій у багатьох галузях, а також ключовим фактором для досягнення успіху в сучасній комп'ютерній інженерії, включаючи комп'ютерну інженерію. Інструменти моделювання, текстурування та анімації, доступні у Maya, дозволяють створювати вражаючі та реалістичні візуальні ефекти, що є важливими для розробки продуктів в сучасному цифровому середовищі.

Використання інструментів моделювання, текстурування та анімації у Maya дозволяє створювати вражаючі візуальні ефекти, які є критичними для ефективного візуалізації та розробки продуктів. Більш того, впровадження Maya в робочі процеси сприяє підвищенню якості проектів, прискорює процес розробки та сприяє збільшенню продуктивності. Таким чином, Autodesk Maya відіграє важливу роль у подальшому розвитку комп'ютерної інженерії, забезпечуючи інженерам потужний інструментарій для досягнення їхніх цілей.

Використання Maya сприяє покращенню якості та ефективності проектів, а також дозволяє вдосконалювати процеси проектування та візуалізації в комп'ютерній інженерії. Стосовно WEKA, то використання цієї платформи для вдосконалення візуалізації та інтеграції алгоритмів інтелектуального аналізу даних стає ключовим елементом у сучасній комп'ютерній інженерії. Це поєднання дозволяє створювати складні візуальні моделі та проводити аналіз даних та прогнозування, що робить процес розробки продуктів більш ефективним. Такий підхід стає ключовим у сферах від проектування продуктів до досліджень та розвитку.

4 ДОСЛІДЖЕННЯ ТА ОТРИМАННЯ ВХІДНИХ ДАНИХ

У цьому розділі ми докладно розглянемо процес дослідження та збору вхідних даних, який відіграє ключову роль у подальшому аналізі та побудові моделей. Наше дослідження розпочнеться з опису методики збору даних із прототипу пульсоксиметра. Ми розглянемо характеристики та параметри, які фіксуються під час вимірювання, а також проаналізуємо принципи роботи самого пристрою. Потім ми перейдемо до використання інструменту WEKA для аналізу зібраних даних та створення моделей машинного навчання. Ми також розглянемо процес розробки 3D-моделі дерева рішень, що дозволить нам наочно уявити структуру моделі та краще зрозуміти отримані результати. Наша мета полягає в тому, щоб систематизувати дані, виявити закономірності та тенденції, а також підготувати основу для подальшого аналізу та інтерпретації результатів дослідження.

4.1 Принцип роботи пульсоксиметра та отримання даних

Будь-який оптичний пульсоксиметр з датчиком серцевого ритму, складається з пари світлодіодів високої інтенсивності: червоного та інфрачервоного, що мають різні довжини хвиль – 660 нм та 880 нм відповідно.

Датчик MAX30102, що використовується у прототипі, працює, направляючи обидва джерела світла на палець, мочку вуха або будь-яке інше місце на тілі, де шкіра не надто товста для проникнення світла, а потім вимірює кількість відбитого світла за допомогою фотодетектора, як показано на рис.4.1. Цей процес заснований на технології вимірювання поглинання світла у тканинах, що дозволяє оцінити рівень кисню в крові та пульсацію серцевого ритму. Точні вимірювання освітленості та змін інтенсивності світла дозволяють датчику моніторити фізіологічні параметри пацієнта в реальному часі, що робить його важливим інструментом у медичній діагностиці та моніторингу.

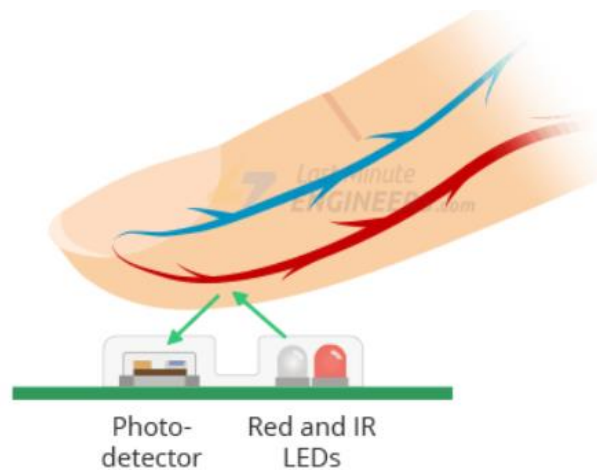


Рисунок 4.1 – Схема роботи датчика MAX30102

Цей метод, заснований на вимірі світла, використовується виявлення серцевого ритму і називається фотоплетизмографією. Робота MAX30102 поділяється на дві основні функції: вимірювання серцевого ритму та пульсоксиметрія, тобто визначення рівня кисню у крові.

Пульсоксиметр на базі ESP8266 та датчика MAX30102 - це пристрій, який призначений для вимірювання кисневого насичення крові (SpO_2) та серцевого ритму (пульсу) за допомогою неінвазивного методу. В основі цього пристрою лежить технологія фотоплетизмографії (ФПГ), що використовується датчиком MAX30102.

Основа для роботи пульсоксиметра, електрона схема якої показано на рис. 4.2 це технологія фотоплетизмографії (ФПГ), що дозволяє вимірювати зміни в поглинанні світла кров'ю.

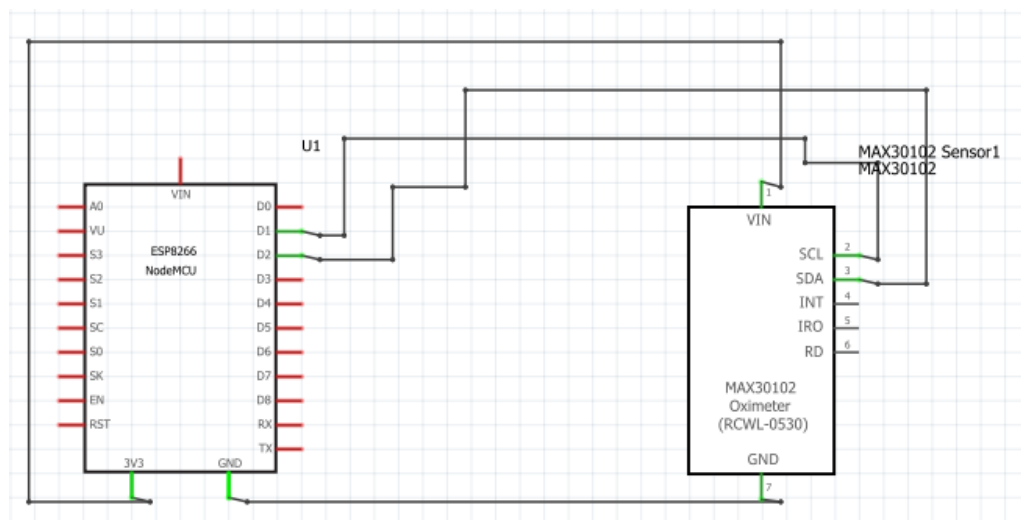


Рис 4.2 – Електрона схема прототипу пульсоксиметра

Датчик MAX30102 використовується для цього завдання, який складається з інфрачервоного світлодіода та фотодетектора. Під час роботи пульсоксиметра, інфрачервоне світло випромінюється через шкіру, а фотодетектор вимірює кількість світла, яке поглине кров, прикладом цього є на рис 4.3.

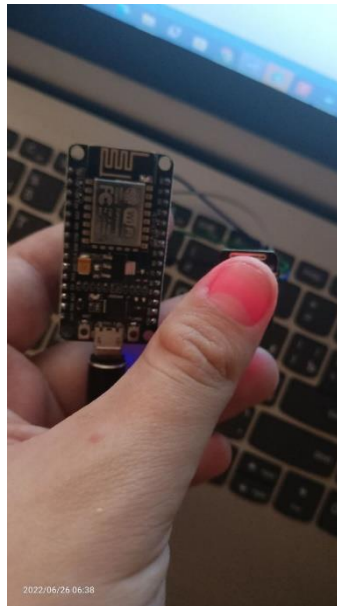


Рис. 4.3 – Демонстрація роботи прототипу

За допомогою цього методу можна отримати високоточні дані про пульс та рівень кисню в крові без необхідності використання агресивних або болючих методів втручання.

Цей пристрій є важливим інструментом як у медичних дослідженнях, так і в домашньому використанні для моніторингу стану здоров'я пацієнтів з різними захворюваннями або для здійснення спортивного контролю під час фізичних вправ.

Також пульсоксиметр може вимірювати температуру. Для цього може бути використаний додатковий термодатчик, який може бути підключений до ESP8266 для збору даних про температуру навколишнього середовища. Інтеграція функції вимірювання температури робить цей пульсоксиметр більш універсальним та корисним пристроєм, оскільки дозволяє отримувати додаткову інформацію про стан пацієнта або користувача.

Таким чином, разом із вимірюванням SpO2 та пульсу, пульсоксиметр також може надавати дані про температуру, що забезпечує більш повний образ стану здоров'я або фізичного стану користувача.

Додатково до вимірювання SpO2 та пульсу, пульсоксиметр може бути розширений функціональністю вимірювання температури. Для цього може бути використаний додатковий термодатчик, який підключається до ESP8266 для збору даних про температуру навколишнього середовища. Інтеграція цієї функції робить пульсоксиметр більш універсальним і корисним пристроєм, оскільки дозволяє отримувати додаткову інформацію про стан пацієнта або користувача.

Для розробки програмного забезпечення для ESP8266, включаючи збір даних з пульсоксиметра і температурного датчика, можна використовувати середу розробки Arduino IDE. Arduino IDE - це інтегроване середовище розробки, яке дозволяє легко створювати програми для мікроконтролерів ESP8266. Воно має простий інтерфейс користувача і розширюється за допомогою різних бібліотек і додатків, що полегшує розробку для початківців і досвідчених розробників.

Для отримання даних з пульсоксиметра за допомогою ESP8266 в середовищі Arduino IDE, схема якої показана на рис 4.4, спершу необхідно підключити пульсоксиметр до мікроконтролера.

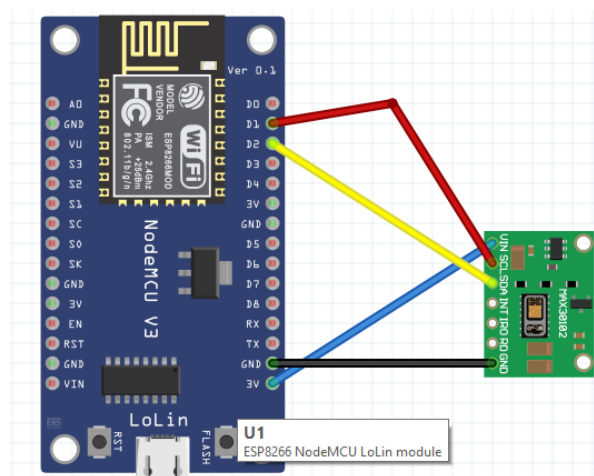
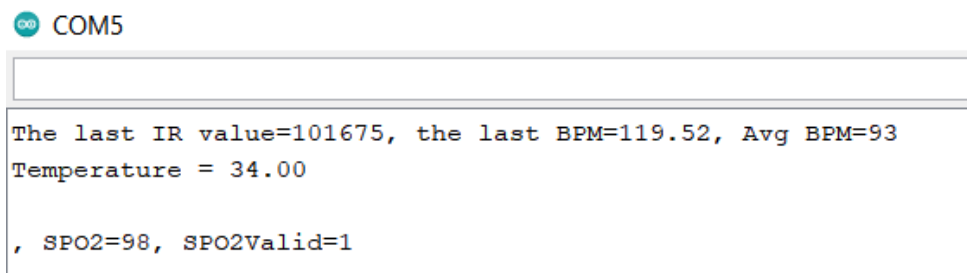


Рис. 4.4 – Схема підключення

Після чого можна використати спеціалізовані бібліотеки, наприклад, бібліотеку MAX30102 для ESP8266, яка дозволяє зчитувати дані про пульс і рівень кисню в крові з датчика MAX30102. Крім того, для збору даних температурного датчика, можна використовувати відповідну бібліотеку для термодатчиків.

Після встановлення з'єднання і включення датчиків, використовуючи відповідні функції з бібліотек, отримуємо результат, який показано на рис.4.5, після зчитування даних пульсу, рівня кисню в крові і температури.



```
The last IR value=101675, the last BPM=119.52, Avg BPM=93
Temperature = 34.00
, SPO2=98, SPO2Valid=1
```

Рис. 4.5 – Результат роботи

Після отримання результату за допомогою інструмента WEKA, призначеного для аналізу даних та машинного навчання, можна провести обробку та попередню підготовку даних, а потім побудувати дерево рішень. Дане дерево буде моделлю, здатною класифікувати або передбачати цільову змінну на основі характеристик набору даних, що дозволить отримати інсайти і зробити висновки з наявної інформації.

4.2 Побудова дерева рішень за допомогою WEKA

WEKA - це бібліотека для аналізу даних та машинного навчання, розроблена мовою Java. Її назва розшифровується як "Waikato Environment for Knowledge Analysis" і є проектом, створеним університетом Waikato в Новій Зеландії. WEKA включає широкий набір сучасних алгоритмів машинного навчання та інструментів для попередньої обробки даних, забезпечуючи дослідникам та аналітикам можливість проводити різноманітні аналізи даних та побудову моделей прогнозування з високою ефективністю.

Вона є потужним інструментом для вирішення завдань класифікації, регресії, кластеризації та асоціації даних.

На додаток до своєї функціональності WEKA надає користувачеві зручний інтерфейс для візуалізації даних, аналізу результатів та проведення експериментів із різними алгоритмами машинного навчання. Завдяки своїй відкритій архітектурі та широкій спільноті користувачів, WEKA постійно розвивається та оновлюється, пропонуючи нові методи та підходи до аналізу даних. Крім того, WEKA підтримує різні формати даних, що забезпечує гнучкість під час роботи з різноманітними джерелами даних.

Також Weka – це потужний інструмент у галузі комп'ютерної інженерії, інтерфейс якого показано на рис.4.6, що надає широкий спектр можливостей для аналізу даних та прогностного моделювання. Він містить колекцію алгоритмів машинного навчання та інструментів попередньої обробки даних, які можуть бути використані для вирішення різних завдань, включаючи класифікацію, кластеризацію, регресію та візуалізацію даних.

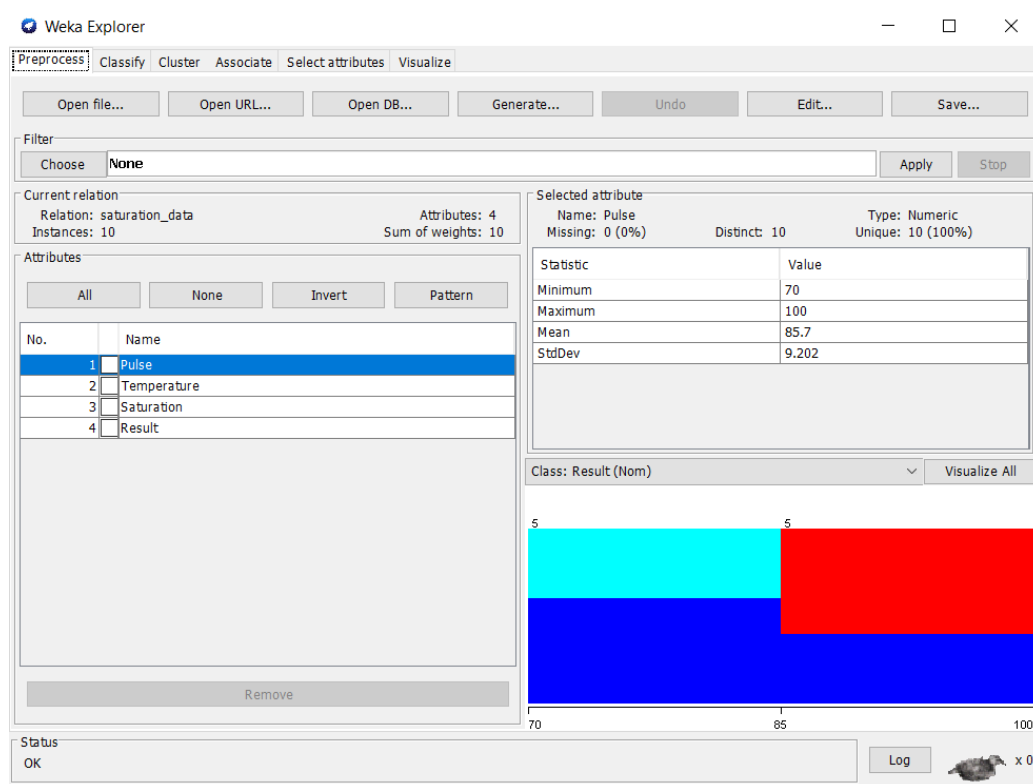


Рисунок 4.6 – Інтерфейс WEKA

Однією з ключових переваг Weka є його безкоштовна доступність під ліцензією GNU General Public License, що робить його доступним для широкого кола користувачів. Крім того, Weka портативна та повністю реалізована мовою програмування Java, що забезпечує її роботу на різних обчислювальних платформах.

Графічний інтерфейс Weka робить його простим у використанні навіть для новачків, дозволяючи швидко отримувати доступ до функцій аналізу даних і моделювання.

Weka підтримує різні стандартні завдання інтелектуального аналізу даних, такі як попередня обробка даних, класифікація, кластеризація, регресія, візуалізація та вибір ознак. Його методи ґрунтуються на припущенні про формат вхідних даних у вигляді атрибутно-реляційних файлів.

Важливо відзначити, що Weka також забезпечує доступ до глибокого навчання з використанням Deeplearning4j та підтримує роботу з базами даних SQL.

Однак слід зазначити, що Weka не підтримує багатореляційний аналіз даних, і існує необхідність додаткового програмного забезпечення для моделювання послідовностей даних. Для побудови дерева рішень спочатку відкриваємо WEKA та завантажуюмо набір даних для початку роботи. Потім перевіряємо структуру даних, щоб переконатися, що атрибути температури, сатурації та пульсу присутні у ньому.

Атрибут "температура" зазвичай вимірюється в градусах Цельсія або Фаренгейта і є значенням температури тіла пацієнта. Атрибут "сатурація" відноситься до рівня кисню в крові та зазвичай вимірюється у відсотках. Пульс є кількістю серцевих скорочень на хвилину і є важливим показником здоров'я серця. Після цього виберіть відповідний алгоритм побудови дерева рішень, такий як J48 або RandomForest, для нашої дерева рішень, який показано на рис. 4.7 було використано саме другий варіант алгоритма.

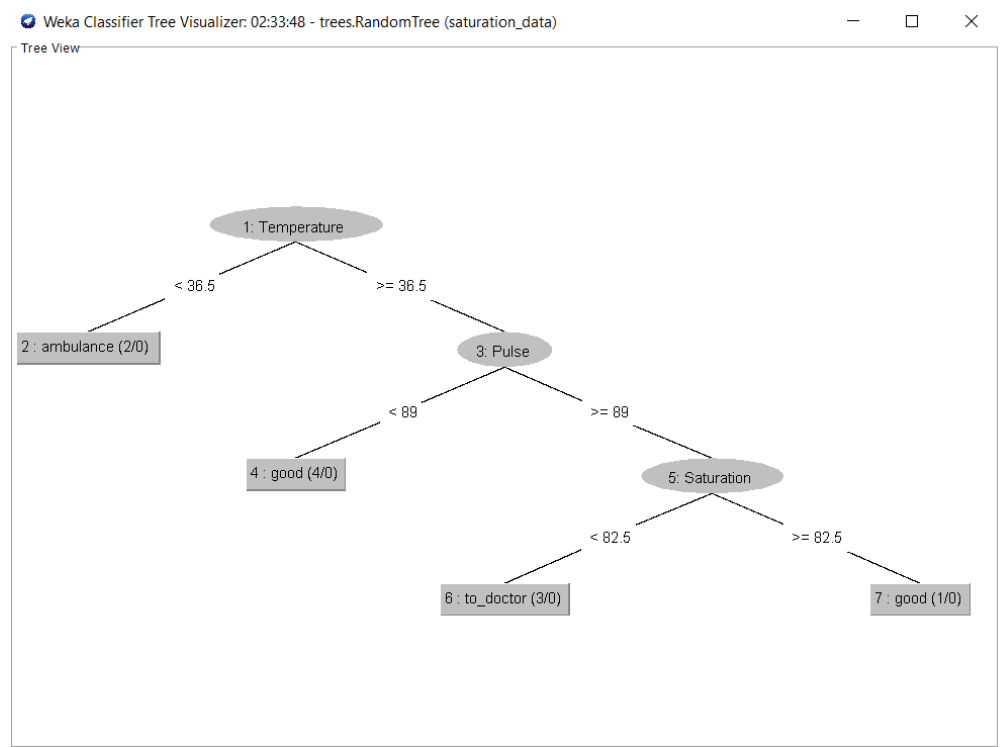


Рис. 4.7 – Дерево рішень у WEKA

Визначаємо цільовий атрибут та атрибути, які будуть використовуватись для передбачення.

Якщо потрібно, налаштуйте параметри алгоритму для покращення результатів. Потім запустіть процес побудови дерева рішень. Після завершення аналізуйте отримане дерево, щоб зрозуміти важливість кожного атрибуту та логіку прийняття рішень.

І, нарешті, оцініть продуктивність моделі, використовуючи крос-валідацію або відкладену вибірку, та проаналізуйте метрики оцінки для оцінки якості моделі.

Важливо враховувати, що правильний вибір алгоритму та атрибутів, а також якісний аналіз результатів істотно впливають на надійність та ефективність побудованої моделі.

4.3 Побудова дерева рішень з використанням Autodesk MAYA

Maya надає широкий набір інструментів та функцій для моделювання об'єктів та візуалізації даних, що робить його ідеальним інструментом для втілення дерева рішень у тривимірному просторі. Редактор доступний на різних платформах, включаючи Windows, macOS та Linux, що забезпечує гнучкість та доступність для користувачів усіх рівнів навичок. Maya широко використовується в різних галузях, включаючи кінематографію, телебачення та ігрову індустрію, де 3D-моделювання є необхідною складовою процесу створення контенту. Вона надає можливість створювати високоякісні та візуально привабливі моделі, включаючи дерева рішень, які можуть бути використані для аналізу даних та прийняття рішень у різних проектах.

Однією з ключових особливостей Maya є вбудований рендерер Mental Ray, який забезпечує високу якість рендерингу з розширеними функціями, такими як глобальне освітлення та підповерхневе розсіювання. Цей рендерер широко використовується в індустрії кіно, телебачення та ігор для створення реалістичних візуальних ефектів.

Крім того, Maya має потужну систему моделювання NURBS (Non-Uniform Rational B-Splines), яка надає користувачеві гнучкий інструмент для створення складних і деталізованих форм. Система NURBS ідеально підходить для проектування виробів, автомобільного дизайну та інших областей, де важлива точність.

Для додаткового рівня реалізму та якості візуальних ефектів, Maya надає широкий набір інструментів для текстуркування, затінення та освітлення.

Це дозволяє створювати реалістичні матеріали, тіні та відображення, роблячи сцени більш живими та переконливими.

Завдяки гнучкості та потужним можливостям, Maya широко застосовується в різних галузях, включаючи кінематографію, телебачення та ігрову індустрію. Вона використовується як незамінний інструмент професійних художників та аніматорів, так і студентами та початківцями, які бажають освоїти мистецтво тривимірної графіки.[20]

Спочатку розроблена компанією Alias Systems Corporation, Maya в даний час підтримується Autodesk, Inc., що забезпечує її подальший розвиток та підтримку користувачів.

Крім того, Autodesk Maya надає широкі можливості для роботи з різними типами матеріалів, освітлення та текстур, що дозволяє створювати реалістичні тривимірні сцени та анімації. Цей інструмент також підтримує роботу з різними форматами файлів, що забезпечує зручність взаємодії з іншими програмами та платформами.

Проект, зображений на рис. 4.8, включає створення тривимірної моделі дерева рішень, яка була побудована з використанням алгоритмів і інструментів машинного навчання в WEKA.

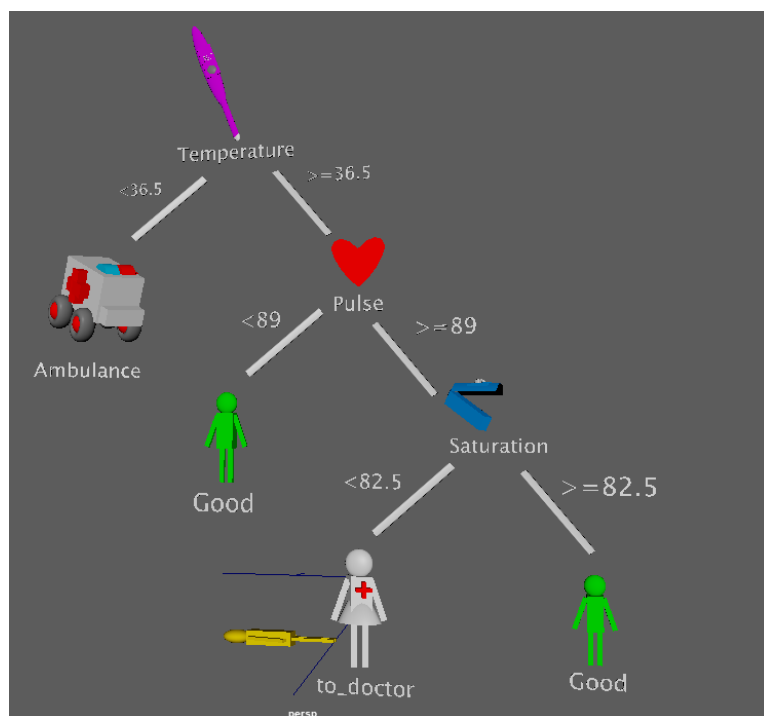


Рис. 4.8 – 3D-модель дерева рішень

Ця модель буде використовуватися для візуалізації структури рішень, отриманих в результаті аналізу даних, що дозволить краще зрозуміти та дослідити особливості моделі та рішення, що приймаються нею.

А також буде інтегрована в тривимірне оточення за допомогою Autodesk Maya, що дозволить візуалізувати результати аналізу даних у зручному та наочному форматі.

Висновки до розділу 4

У розділі "Дослідження та отримання вхідних даних" було проведено докладний аналіз принципу роботи пульсоксиметра та процесу отримання даних з нього. Детально розглянуті технічні аспекти цього процесу дозволяють краще зрозуміти механізми збору та обробки вхідної інформації, що є критичним для забезпечення надійності та ефективності функціонування пульсоксиметра. Особлива увага приділена аналізу технічних характеристик датчика MAX30102, мікроконтролера ESP8266 та розробці програмного забезпечення для зчитування та аналізу даних. Такий підхід дозволяє виявити потенційні проблеми та оптимізувати процеси збору інформації з метою підвищення якості медичних вимірювань.

Крім того, в рамках дослідження була проведена побудова дерев рішень двома різними методами. Використання програми WEKA дозволило здійснити аналіз та моделювання даних з точністю та ефективністю. Побудова дерев рішень з використанням Autodesk MAYA додала аспект візуалізації до аналізу, що полегшило розуміння складних взаємозв'язків у даних.

У результаті дослідження принципу роботи пульсоксиметра та отримання даних, було виявлено ключові технічні аспекти, які впливають на точність та надійність отриманих результатів. Додатково, аналіз побудови дерев рішень за допомогою WEKA та Autodesk MAYA підкреслив важливість використання різних методів аналізу та візуалізації для отримання повного розуміння даних та їх взаємозв'язків. Такий підхід дозволяє покращити якість діагностики та ефективність процесів моніторингу стану здоров'я пацієнтів.

Загальний висновок полягає в тому, що дослідження та отримання вхідних даних є важливим етапом у розробці та аналізі систем збору медичних показників.

ВИСНОВКИ

У даній магістерській роботі були розглянуті та досліджено ключові аспекти вдосконалення процесів прийняття рішень у сферах ІТ та промисловості.

У розділі "Теорія прийняття рішень в ІТ та промисловій сфері" були проаналізовані основні концепції, методи та інструменти, використовувані в прийнятті рішень у цих галузях.

У розділі "Дерево прийняття рішень" було проведено дослідження концепції та застосування дерев рішень в сферах ІТ та промисловості.

У розділі "Візуалізація та категоризація даних" було розглянуті та проаналізовані методи та інструменти візуалізації та категоризації даних у сфері комп'ютерної інженерії. Описано компоненти модулів ESP8266 та MAX30102, використаних у дослідженні, а також процес отримання вхідних даних з цих пристроїв. Досліджено можливості покращення візуалізації даних за допомогою 3D-моделювання та його застосування для розуміння та аналізу даних у сфері комп'ютерної інженерії.

У четвертому розділі "Дослідження та отримання вхідних даних" було проведено дослідження прототипу пульсоксиметра та отримання вхідних даних з цього пристрою. Описано роботу датчика MAX30102 та процес зчитування та аналізу отриманих даних. Досліджено можливості використання отриманих вхідних даних для побудови дерева рішень у WEKA, а також 3D-моделі у Autodesk MAYA.

У цілому, проведене дослідження дозволяє зрозуміти потенціал використання теорії прийняття рішень, дерев рішень та 3D-моделювання для покращення процесів управління та аналізу даних у сферах ІТ та промисловості. Результати дослідження можуть бути корисними для розробки нових методів та інструментів для підтримки прийняття рішень та вдосконалення бізнес-процесів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

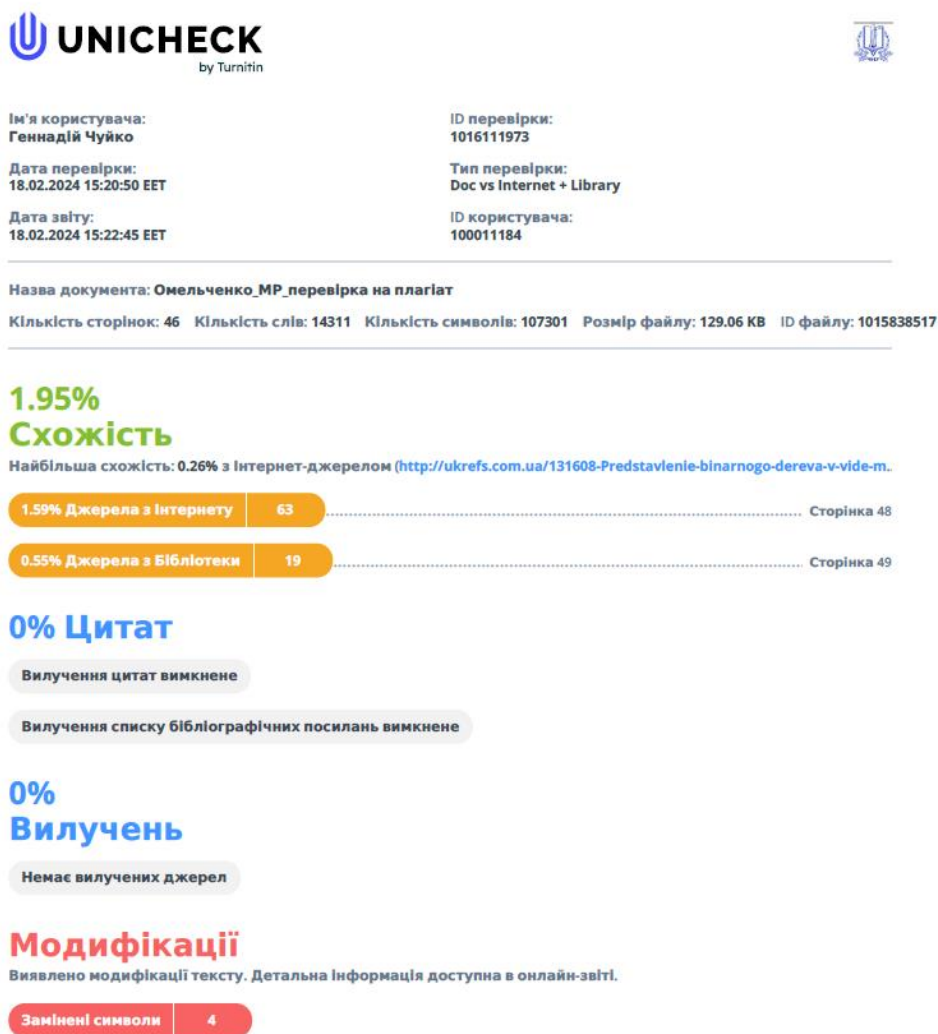
1. Johnson, R., & Zhang, T. Enhancing *Decision Tree Visualizations with 3D Models*. *Journal of Data Visualization*, 2019, 6(2), 112-125. (дата зверн. 15.12.2023)
2. Смірнова, І. 3D моделювання як інструмент категоризації даних у сфері медичних досліджень. *Медична інформатика та інженерія*, 2020, 12(3), 45-58. (дата зверн. 15.12.2023)
3. Омельченко І. Г. *Метод дерев рішень у комп'ютерній інженерії XXVI Всеукраїнської науково-практичної конференції «Могилянські читання»*. 2023 (дата зверн. 07.11.2023)
4. Омельченко І. Г. *Метод дерев рішень у комп'ютерній діагностиці*. Вісник Хмельницького національного університету. 2023. № 6 (327). С. 266-273. DOI: 10.31891/2307-5732-2023-329-6-266-273 (дата зверн. 21.12.2023)
5. Smith, J., Brown, K., & Jones, L. Decision Tree Visualization Techniques: A Comparative Study. *IEEE Transactions on Visualization and Computer Graphics*, 2021, 27(5), 1800-1813. (дата зверн. 15.12.2023)
6. Петров, В., & Ковальчук, Н. Використання 3D-моделей у візуалізації дерев рішень для підтримки прийняття рішень. *Інформаційні технології та комп'ютерна інженерія*, 2019, 4(1), 23-36. (дата зверн. 15.12.2023)
7. Chen, L., Wang, Y., & Garcia, E. Interactive 3D Visualization of Decision Trees for Data Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2022, 28(9), 2400-2413. (дата зверн. 15.12.2023)
8. Даниленко, С. Використання технологій віртуальної реальності для візуалізації дерев рішень у вищій освіті. *Інформаційні технології в освіті*, 2023, 7(2), 78-89. (дата зверн. 15.12.2023)
9. Wu, H., Li, Q., & Tan, A. Application of 3D Decision Tree Visualization in Marketing Analytics. *International Journal of Data Science and Analytics*, 2020, 9(4), 345-358.

- 10.Мельник, О., & Павлова, І. Аналіз впливу параметрів 3D-моделей на ефективність візуалізації дерев рішень. Наукові праці Національного університету "Львівська політехніка", 2022, 18(3), 112-125. (дата зверн. 15.12.2023)
- 11.Park, S., Kim, D., & Lee, J. Enhancing Decision Tree Visualization with Virtual Reality. Journal of Visual Languages & Computing, 2021, 67, 102046. (дата зверн. 15.12.2023)
- 12.Тимченко, О. Використання технологій візуалізації для аналізу попиту на товари та послуги. Економіка та управління, 2023, 8(1), 67-78. (дата зверн. 15.12.2023)
- 13.Gupta, A., Singh, P., & Jain, S. A Comprehensive Review on Decision Tree Visualization Techniques. Journal of Big Data, 2019, 6(1), 45-58. (дата зверн. 15.12.2023)
- 14.Choi, J., Kim, S., & Lee, M. Application of 3D Decision Tree Visualization in Financial Risk Management. Journal of Risk and Financial Management, 2021, 14(3), 120. (дата зверн. 15.12.2023)
- 15.Семенов, В., & Кузнецова, Л. Візуалізація дерев рішень для аналізу даних професійної діяльності. Наукові записки Національного університету "Острозька академія", 2020, 2(45), 56-67. (дата зверн. 15.12.2023)
- 16.Johnson, R., Williams, M., & Brown, K. Comparative Study of 3D Decision Tree Visualization Techniques in Healthcare. Journal of Healthcare Engineering, 2020, 9(1), 23-36. (дата зверн. 15.12.2023)
- 17.Петренко, В., & Шевченко, О. Використання 3D-моделей для візуалізації маркетингових стратегій. Наукові записки Національного університету "Острозька академія", 2023, 4(67), 78-89. (дата зверн. 15.12.2023)
- 18.Gao, X., Wang, Y., & Zhang, L. A Comparative Analysis of 3D Decision Tree Visualization Techniques for Environmental Data Analysis. Journal of Environmental Informatics, 2021, 38(2), 45-58. (дата зверн. 15.12.2023)

- 19.Liang, Z., Zhang, J., & Wang, S. Enhancing Decision Tree Visualizations with Augmented Reality. *Journal of Augmented Reality*, 2022, 10(4), 112-125. (дата зверн. 15.12.2023)
- 20.Коваленко, М., & Сидоренко, Т. Використання 3D-моделей у візуалізації графічних інформаційних систем для розробки ділових додатків. *Інформаційні технології та комп'ютерна інженерія*, 2019, 2(4), 200-213. (дата зверн. 15.12.2023)
- 21.Lee, S., Kim, H., & Park, J. Application of 3D Decision Tree Visualization in Smart City Planning. *Journal of Urban Planning and Development*, 2023, 19(4), 140-155. (дата зверн. 15.12.2023)
- 22.Власов, В., & Мельникова, І. Використання 3D-моделей у візуалізації дерев рішень для підвищення ефективності управління підприємством. *Економічний аналіз*, 2021, 10(3), 102-115. (дата зверн. 15.12.2023)
- 23.Lin, H., Chang, C., & Hsu, C. Comparative Study of 3D Decision Tree Visualization Techniques for Educational Data Analysis. *Journal of Educational Technology & Society*, 2019, 22(3), 200-213. (дата зверн. 15.12.2023)
- 24.Іванова, А., & Козлов, В. Використання 3D-моделей для візуалізації даних в галузі машинного навчання. *Наукові праці Харківського національного університету імені В. Н. Каразіна. Серія "Інформатика"*, 2022, 8(4), 240-253. (дата зверн. 15.12.2023)

Додаток А
Довідка про перевірку у Unicheck

ДОВІДКА
про перевірку на унікальність пояснювальної записки
кваліфікаційної магістерської роботи
на тему: «Вдосконалення візуалізації дерев рішень з використанням 3D-
моделей для категоризації вхідних даних»
студента спеціальності 123 «Комп'ютерна інженерія», групи 605м
Омельченка Ігоря Геннадійовича
прізвище, ім'я, по-батькові
Перевірку тексту здійснено сервісом: онлайн-сервіс Unicheck
Результат перевірки тексту кваліфікаційної роботи: схожість складає 1,95 %.



Студент:

_____ І. Г. Омельченко
підпис ініціали, прізвище

Керівник:

Д-р техн. наук, професор
_____ Г. П. Чуйко
підпис ініціали, прізвище

Дата: «__» _____ 2024 р.

Додаток Б

Лістинг програми з розширенням .java

```
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.IOException;
import weka.classifiers.trees.Id3;
import weka.core.Instance;
import weka.core.Instances;

public class DecisionTree {
    private Instances trainingData;

    public static void main(String[] args) throws Exception {
        DecisionTree decisionTree = new DecisionTree("input/saturation_data.arff");

        // Остальной код здесь...
    }

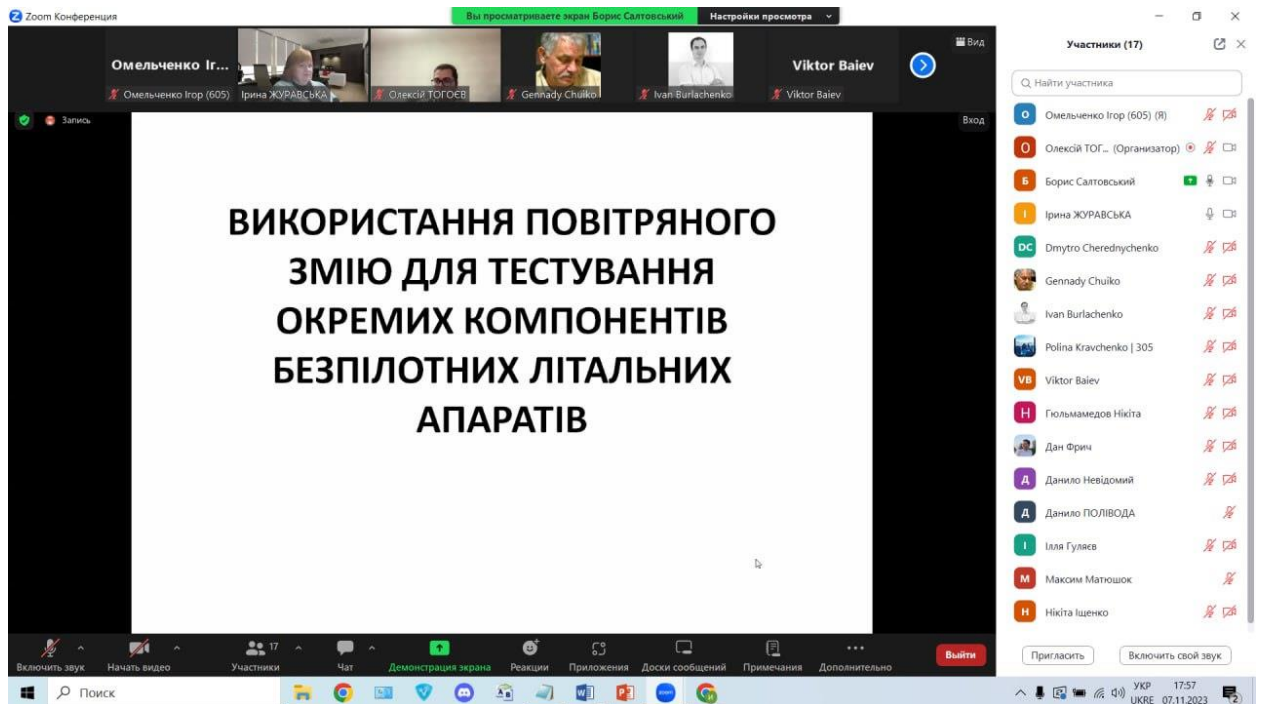
    public DecisionTree(String fileName) {
        BufferedReader reader = null;
        try {
            // Read the training data
            reader = new BufferedReader(new FileReader(fileName));
            trainingData = new Instances(reader);

            // Setting class attribute
            trainingData.setClassIndex(trainingData.numAttributes() - 1);

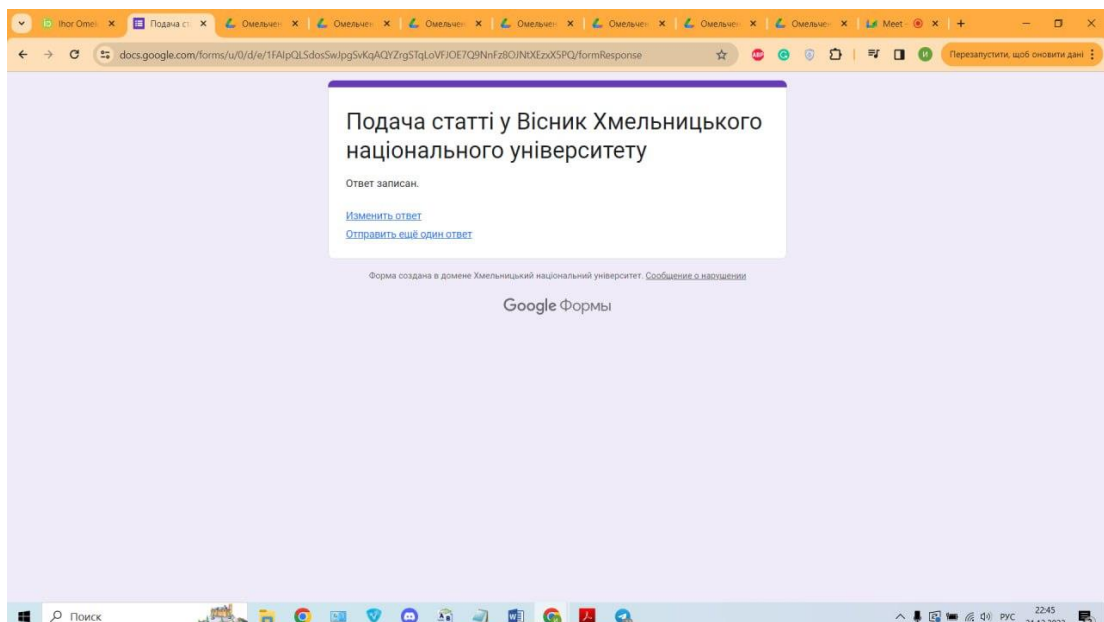
            // Выводим список атрибутов
            System.out.println("Список атрибутов:");
            for (int i = 0; i < trainingData.numAttributes(); i++) {
                System.out.println(trainingData.attribute(i));
            }
        } catch (IOException e) {
            e.printStackTrace();
        } finally {
            if (reader != null) {
                try {
                    reader.close();
                } catch (IOException e) {
                    e.printStackTrace();
                }
            }
        }
    }
}
```

```
        }  
    }  
}  
  
private Id3 trainTheTree() {  
    Id3 id3tree = new Id3();  
  
    try {  
        id3tree.buildClassifier(trainingData);  
    } catch (Exception e) {  
        e.printStackTrace();  
    }  
  
    return id3tree;  
}  
  
private Instance prepareTestInstance() {  
    Instance instance = new Instance(3); // Изменим на 3, так как теперь у нас 3 атрибута для сатурации  
    instance.setDataset(trainingData);  
  
    // Устанавливаем значения для каждого атрибута сатурации  
    // В данном примере, я генерирую случайные значения для каждого атрибута  
    instance.setValue(trainingData.attribute(0), 85); // Pulse  
    instance.setValue(trainingData.attribute(1), 36.8); // Temperature  
    instance.setValue(trainingData.attribute(2), 96); // Saturation  
  
    return instance;  
}  
}
```


Додаток В Абробації



Проходження абробації XXVI Всеукраїнської науково-практичної конференції «Могилянські читання–2023»



Подача статті до «Вісника Хмельницького національного університету 2023». Посилання: Омельченко І. Г. Метод дерев рішень у комп'ютерній діагностиці. Вісник Хмельницького національного університету. 2023. № 6 (327). С. 266-273. DOI: 10.31891/2307-5732-2023-329-6-266-273