

Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерної інженерії та інформаційних систем

КВАЛІФІКАЦІЙНА РОБОТА

Галузь знань _____ 12 – Інформаційні технології _____

Спеціальність _____ 126 – Інформаційні системи та технології _____

на тему «Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних»

КвРІСТ.2201144.09.09 ПЗ

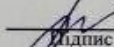
Виконав: студент 2 курсу, група ІСТм-22-1

Керівник канд.техн. наук, доцент
Науковий ступінь, вчене звання

До захисту допускаю:
Зав. кафедри КІС, д.т.н., проф.
Т.О. Говорущенко
18 12 2023 р.


Підпис

Шатровський А.О.
Ініціали, прізвище


Підпис

Гнатчук Є.Г.
Ініціали, прізвище

Хмельницький, 2023

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ ТА ІНФОРМАЦІЙНИХ СИСТЕМ

Освітній рівень МАГІСТР

Галузь знань 12 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

Спеціальність 126 ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ

Освітня програма ОСВІТНЬО-ПРОФЕСІЙНА ПРОГРАМА «ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ»

ЗАТВЕРДЖУЮ

Зав. кафедри Т.О.Говорущенко

“ 1 ” 04 2023 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

Шатровському Андрію Олександровичу

Прізвище, ім'я, по батькові студента

1. Тема проекту (роботи) Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Керівник проекту (роботи) Гнатчук Є.Г., к.т.н., доцент

Прізвище, ім'я, по батькові, науковий ступінь, вчене звання

Затверджена наказом ректора університету від 15.08.2023 р. № 30

2. Строк подання студентом проекту (роботи) на кафедру 15.12.2023 р.

3. Вихідні дані до проекту (роботи) Завдання на дипломне проектування

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити) _____

Аналіз відомих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах





Інформаційні потоки інформаційної системи керування даними спортсменів

Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

5. Перелік графічного матеріалу (із зазначенням обов'язкових креслень) _____

6. Консультанти розділів кваліфікаційної роботи магістра

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Нормоконтроль	Лисенко С.М., професор кафедри КПС		
Антиплагіат	Нічепорук А.О., доцент кафедри КПС		

7. Дата видачі завдання « 03 » 04 2023р.

КАЛЕНДАРНИЙ ПЛАН

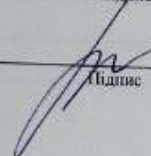
№з/п	Назва етапів (розділів) кваліфікаційної роботи магістра	Термін виконання етапів проекту (роботи)	Примітка
1	Вибір напрямку дослідження та узгодження тематики КвРМ з керівником	15.06.2023	виконано
2	Ознайомлення з предметною областю; формулювання мети та задач дослідження; визначення об'єкта та предмета дослідження	05.07.2023	виконано
3	Робота над розділом 1 – аналіз відомих моделей, методів за темою; постановка задачі	05.09.2023	виконано
4	Робота над розділом 2 – розробка моделей для вирішення поставленої задачі	05.10.2023	виконано
5	Робота над науковою статтею	05.10.2023	виконано
6	Робота над розділом 3 – розробка методів для вирішення поставленої задачі	03.11.2023	виконано
7	Робота над розділом 4 – проектування та розробка ІС для вирішення поставленої задачі, експериментальна частина	30.11.2023	виконано
8	Оформлення пояснювальної записки згідно вимог	05.12.2023	виконано
9	Попередній захист ДРМ	10.12.2023	виконано
10	Захист ДРМ на засіданні ЕК	До 19.12.2023	

Студент


Підпис

А.О. Шатровський
Ініціали, прізвище

Керівник роботи


Підпис

Є.Г. Гнатчук
Ініціали, прізвище

РЕФЕРАТ

Тема кваліфікаційної роботи магістра: Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Автор роботи: магістр гр.ІСТМ-22-1 Шатровський А.О.

Керівник роботи: к.т.н., доцент Гнатчук Є.Г.

Пояснювальна записка: 77с., 16 рис., 1 табл., 2 дод., 64 джерела.

ПЕРЕЛІК КЛЮЧОВИХ СЛІВ (6-8) ІНФОРМАЦІЙНА СИСТЕМА, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, СПОРТИВНІ ДАНІ, ЗМАГАННЯ, КЕРУВАННЯ ДАНИМИ, ПРОГНОЗУВАННЯ, КЛАСИФІКАЦІЯ.

Об'єктом дослідження є інформаційна система керування даними спортсменів в системі організації змагань.

Предмет дослідження: метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Метою кваліфікаційної роботи магістра є підвищення ефективності роботи тренера в системі організації змагань для спостереження та відбору спортсменів на змагання за допомогою інформаційної системи керування даними спортсменів на основі інтелектуального аналізу даних.

Наукова новизна отриманих результатів:

– набув подальшого розвитку метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині методів інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо;

– набула подальшого розвитку інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині додавання підсистеми керування даними спортсменів, яка

включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків.

На основі проведених досліджень розроблена архітектура і компоненти інформаційної системи для керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Практична значимість отриманих результатів полягає у можливості використовувати результати досліджень для ефективного відбору спортсменів на основі аналізу його попередніх результатів, фізичного стану та готовності, а також наданні тренерам та управлінцям додаткових інструментів для прийняття обґрунтованих та докладних рішень щодо тренувальних стратегій, участі в змаганнях та медичного супроводу.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ	5
ВСТУП.....	6
1 АНАЛІЗ ВІДОМИХ МОДЕЛЕЙ, МЕТОДІВ ТА ЗАСОБІВ	
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ В СПОРТИВНИХ	
ІНФОРМАЦІЙНИХ СИСТЕМАХ.....	
1.1 Аналіз видів даних, що використовуються в системі організації змагань ...	9
1.2 Інтелектуальний аналіз даних в різних видах спорту	15
1.3 Вимоги до інтелектуального аналізу даних при аналізі даних про спортивні результати	19
1.4 Висновки. Постановка задачі	23
РОЗДІЛ 2 ІНФОРМАЦІЙНІ ПОТОКИ ІНФОРМАЦІЙНОЇ СИСТЕМИ	
КЕРУВАННЯ ДАНИМИ СПОРТСМЕНІВ	
2.1 Керування даними в спортивній індустрії	25
2.2 Інформаційні потоки в інформаційній системі керування даними спортсменів	28
2.3 Концепція інформаційної системи керування даними спортсменів	38
Зберігання даних. Опис методів зберігання даних, таких як реляційні бази даних, NoSQL рішення, а також обговорення схеми бази даних для забезпечення ефективного доступу до інформації.....	38
2.4 Висновки.....	45
3 МЕТОД СТВОРЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КЕРУВАННЯ	
ДАНИМИ СПОРТСМЕНІВ	
3.1 Формування вимог до інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних	47

3.2	Метод створення інформаційної системи керування даними спортсменів формування.....	56
3.3	Висновки.....	63
4	ІНФОРМАЦІЙНА СИСТЕМА КЕРУВАННЯ ДАНИМИ СПОРТСМЕНІВ В СИСТЕМІ ОРГАНІЗАЦІЇ ЗМАГАНЬ НА ОСНОВІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.....	65
4.1	Розробка архітектури інформаційної системи керування даними спортсменів	65
4.2	Побудова UML-діаграм інформаційної системи керування даними спортсменів	67
4.4	Висновки	78
	ВИСНОВКИ	79
	ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ	82
	ДОДАТОК А ТЕЗИ ДОПОВІДІ.....	89
	ДОДАТОК Б ПРЕЗЕНТАЦІЯ.....	94

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІС - інформаційна система

БД - база даних

ЕС - експертна система

ШНМ - штучні нейронні мережі

ВРМ - Business Process Management, управління бізнес-процесами

GDPR - General Data Protection Regulation, загальний регламент про захист даних

ВСТУП

Із швидким розвитком інформаційних технологій і спорту, аналіз спортивної інформації стає дедалі складнішим питанням. Велика кількість даних про спорт надходять з інтернету та демонструють швидку тенденцію до зростання. Дані про спорт містять багато інформації, наприклад про спортсменів, тренерів, змагання, тощо. Сьогодні можна легко отримати доступ до різних спортивних даних, а також розроблено різноманітні технології аналізу даних, які дозволяють глибше досліджувати цінність цих даних.

Дослідження спортивних великих даних може принести великі переваги популярним видам спорту, шкільним і змагальним видам спорту. Наприклад, за допомогою керування та аналізу звичайної фізичної підготовки та спортивних результатів спортсменів можна передбачити потенційних професіоналів спортсменів. Результати цих аналізів даних є сприятливою основою для тих, хто приймає рішення щодо розподілу коштів на підготовку спортсменів. Основна мотивація полягає в тому, щоб отримати знання зі спортивних великих даних, щоб надавати кращі спортивні послуги спортсменам, тренерам, особам, які приймають рішення, пов'язані зі змаганнями, і громадськості. Крім того, деякі типові сервіси великих даних, такі як продуктивність фізичних вправ, дані про стан здоров'я, статистика тренувань та аналіз, можуть ефективно допомогти тренерам і спортсменам у щоденних тренуваннях і налагодженні ігрових стратегій і відіграють незмірну роль для перемоги на змаганнях.

Аналіз спортивних великих даних спрямований на вирішення проблем у спортивній науці, спираючись на інтелектуальний аналіз даних, мережеву науку та статистичні методи. Аналіз великих даних у спорті зосереджений на відкритті цінності даних і забезпечує цінні інформаційні ресурси для підприємств і менеджерів.

Змагання різного рівня є результатом довгої виснажливої підготовки спортсменів, а отже важливим питанням для тренерів є збирання різноманітних даних про стан підготовки спортсмена, аналіз та дослідження цих даних з метою

якісного відбору спортсменів для змагань та керування цими даними. Така робота є досить часозатратною та виснажливою. Вирішити таку задачу може розроблення відповідної інформаційної системи, яка полегшить рутинну роботу тренерського штабу при аналізі даних спортсменів.

Отже, на сьогоднішній день актуальною задачею є розроблення методу створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Метою кваліфікаційної роботи магістра є підвищення ефективності роботи тренерського штабу при відборі спортсменів на змагання шляхом розроблення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Поставлена мета досягається розв'язанням таких основних задач:

- 1) провести огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах;
- 2) розробити метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних;
- 3) запропонувати концепцію інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Об'єктом дослідження є інформаційна система керування даними спортсменів в системі організації змагань.

Предмет дослідження: метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Метою кваліфікаційної роботи магістра є підвищення ефективності роботи тренера в системі організації змагань для спостереження та відбору спортсменів на змагання за допомогою інформаційної системи керування даними спортсменів на основі інтелектуального аналізу даних.

Наукова новизна отриманих результатів:

– набув подальшого розвитку метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині методів інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо;

– набула подальшого розвитку інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині додавання підсистеми керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків.

На основі проведених досліджень розроблена архітектура і компоненти інформаційної системи для керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Практична значимість отриманих результатів полягає у можливості використовувати результати досліджень для ефективного відбору спортсменів на основі аналізу його попередніх результатів, фізичного стану та готовності, а також наданні тренерам та управлінцям додаткових інструментів для прийняття обґрунтованих та докладних рішень щодо тренувальних стратегій, участі в змаганнях та медичного супроводу.

За темою кваліфікаційної роботи магістра опублікована одні тези доповіді авторів Швайко В.К., Кузьмін А.А., Шатровський А.О. «Гейміфікація інформаційної системи для вибору виду спорту на основі морфофункціональних показників людини» III Всеукраїнської науково – технічної конференції молодих вчених, аспірантів та студентів «комп'ютерні ігри і мультимедіа як інноваційний підхід до комунікації - 2023» 28-29 вересня 2023 р., Одеса [64].

1 АНАЛІЗ ВІДОМИХ МОДЕЛЕЙ, МЕТОДІВ ТА ЗАСОБІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ В СПОРТИВНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

1.1 Аналіз видів даних, що використовуються в системі організації змагань

Основною метою змагальних видів спорту є досягнення найвищих спортивних результатів, що в кінцевому підсумку сприяє перемозі в змаганнях. В основі даних змагальних видів спорту лежать дані про спортсмена та його поведінку. У спорті не тільки самі спортсмени мають фізичну поведінкову активність, але й поведінкова активність між спортсменами, в яких також існують просторово-часові, описані та підраховані дані про поведінку, які можна реєструвати. Тому зростання кількості даних про спортивні змагання дало поштовх для розвитку досліджень у змагальних видах спорту, а також одночасно створила основу для вивчення законів людського життя і звичок людей.

Багато дослідників використовували велику кількість відкритих онлайн-даних про змагальні види спорту для аналізу та розробки програмного забезпечення та інструментів [1 – 4].

Роботи таких авторів [5 – 7] є корисними для професійної аналітики, дозволяють ефективно приймати рішення на основі поведінки під час ігор, покращуючи результати тренувань команд та їх виступів на змаганнях. Таким чином, аналіз спортивних даних, отриманих під час змагань є необхідним і привертає значну увагу дослідників.

Однак проблема полягає в тому, що дані про спортивні змагання включають поведінку спортсменів і різні статистичні дані, тому, загалом, дані є відносно великими за обсягом, а також включають багато поведінкових патернів, невидимих неозброєним оком, що створює певні труднощі для аналізу даних. Зокрема, аналіз спортивної статистики може ефективно ідентифікувати поведінкові моделі спортсменів [2, 10] в тому числі їхній індивідуальний

внесок і ступінь активності. Однак дані про спортивні змагання містять багато вимірів, таких як простір, час та інші. Спортивні аналітики не можуть інтуїтивно сприймати дані, а покладаючись суто на цифри не може повністю представити результати аналізу даних.

Стратегічні рішення у спорті є життєво важливими для результатів. До них належать такі рішення, як, наприклад, як уникнути вибування у попередніх змаганнях з плавання без надмірної витрати енергії, що знизить результати у наступних змаганнях, або використання гравців другої лінії на ранніх стадіях хокейних турнірів, щоб уникнути травм і втоми провідних гравців.

Заняття спортом стають не виключенням для застосування інформаційних технологій. Інформаційні системи та технології допомагають спортсменам максимізувати їх продуктивність, що дозволяє їм конкурувати на вищому рівні, ніж будь-коли раніше. Схожі тенденції спостерігаються і в масових спортивних заходах [11], а також в приваблюванні людей до участі в спортивних подіях. Це породжує потребу в спеціальних системах або відповідних додатках, які могли б направляти, допомагати та підтримувати людей у відчутті задоволення від занять спортом [12]. Наприклад, багато людей у всьому світі не можуть собі дозволити професійного спортивного тренера через різні перешкоди, такі як фінансові. З іншого боку, обширні дослідження, які пов'язують інтелектуальні інструменти та методи аналізу даних із спортивною наукою, породжують нові інтелектуальні рішення, які підтримують усі етапи спортивного тренування.

Аналіз результатів, як засіб створення та аналізу достовірних записів виступів спортсменів за допомогою систематичних спостережень, набув важливого значення в останнє десятиліття. Цьому сприяв розвиток інформаційних технологій та цифрової фотографії. Однією з таких передових технологій є інтелектуальний аналіз даних, який є галуззю комп'ютерних наук і штучного інтелекту. Методи інтелектуального аналізу даних можуть бути використані для аналізу спортивних даних, зокрема, даних про спортивні результати в елітних видах спорту, мається на увазі дані про спортивні

результати елітних видів спорту, як їх зазвичай називають, тобто дані, отримані на міжнародних змаганнях, таких як чемпіонати світу або Олімпійські ігри.

Технології штучного інтелекту та комп'ютерного зору стають модними для аналізу відео у спортивній сфері. Основна мета спортивних змагань – досягти кращих спортивних результатів і, зрештою, допомогти виграти змагання через заохочення вищих стандартів досягнень, розширення можливостей фізичної активності, формування дисципліни, навчання, як добре програвати та справлятися з розчаруванням, а також розвивати товариську і командну роботу [11]. В такому випадку, даними вважаються є спортсмен і його поведінка в центрі змагання [12].

Модель на основі CNN або згорткової нейронної мережі широко використовується для ефективного вирішення складних завдань машинного перекладу, обробки сигналів і комп'ютерного зору [12]. Однак комп'ютерний зір наразі коливається від математичних методів до підходів машинного навчання через його ефективність у вилученні складових функцій без участі людини [11]. Ця галузь досліджень є найперспективнішою для інновацій та розвитку. Наприклад, прогнозний аналіз за допомогою штучного інтелекту можна застосувати для покращення здоров'я та фізичної форми [12]. Так звані носимі програми можуть надавати інформацію про розрив і розтягнення гравця, зменшуючи пошкодження спортсменів. Крім того, штучний інтелект може виявити тенденції в ігрових тактиках, планах і слабких місцях гравців та тренувань [13].

Дані є значною частиною спортивної індустрії для тренерів, виконавців, керівництва, працівників спортивної медицини та вболівальників [14]. Аналітика даних може допомогти командам перемогти в іграх, і ці дані можуть допомогти підвищити ефективність спортсменів, зменшити кількість травм і надихнути вболівальників приєднуватися до ігор [15]. Крім того, великі дані допомагають розробляти покращені спортивні стратегії. Незалежно від того, чи це індивідуальний чи командний вид спорту, стратегічне управління є важливим для будь-якого виду спорту. Ці методи залежать від професійних спортсменів і

команд, які змагаються проти своїх суперників. Наприклад, сучасний коучинг використовує великі масиви даних для розробки успішної тактики для індивідуальних і командних спортсменів [16]. Люди справді підуть за надійними лідерами, а відомі тренери мають перевагу над молодшими, оскільки вони мають доступ до значно більшого банку даних інформації.

З точки зору тренера, найбільш очевидним застосуванням даних є збір цифр і статистики. Дані спортивної індустрії є життєво важливими для гравців, тренерів і керівництва, а також для спеціалістів зі спортивної медицини та глядачів. У той час як аналітика даних може допомогти команді виграти ігри та підвищити ефективність гравців, ті самі цифри також можуть мотивувати глядачів відвідувати ігри. Наука про дані дозволяє тренерам професійних команд будувати гіперперсоналізовані матчі спортсменів та інші плани зокрема для кожного матчу, який грає команда. Стратегії команди залишаються несподіваними, але ефективними [17]. Аналіз результатів у спорті допомагає тренерам і гравцям досягати своїх цілей, визначаючи дії, які можуть спрямовувати прийняття рішень, максимізувати ефективність і допомагати їм на шляху до досконалості. Вони часто включають тактичну оцінку, аналіз рухів, відео та статистичні бази даних, моделі та відображення даних тренера та гравця. Завдяки очевидним удосконаленням технологій змінилися вимоги до збирання, зберігання та навчання даних щодо представлення даних, тому аналітикам тепер потрібно набагато більше знань у багатьох пристроях відстеження та програмному забезпеченні. Відеозапис гри може допомогти усунути такі упередження та дати більш неупереджену перспективу того, що відбувалося на полі. Щоб гравці та тренери могли зрозуміти, що вийшло добре, а що не так, аналітики ефективності збирають дані про всі події, що відбуваються на полі. Такі види спорту, як бейсбол, футбол, футбол, баскетбол, і такі види спорту, як фентезі, підвищили ефективність гравців і прогнозували майбутні результати відповідно до великих даних [18]. Незалежно від того, чи йдеться про історичні дані, ведення життєво важливих результатів, алгоритмічні прогнози чи чітку статистику гравців, великі дані є невід'ємним елементом спортивної індустрії [19]. Великі дані дозволяють

командам і корпораціям бути в курсі продуктивності, виконувати прогнози та бути великими даними, що дозволяє командам і організаціям залишатися актуальними щодо продуктивності, прогнозувати та бути рішучими у світі спорту [20]. Крім того, усі сторони, задіяні в галузі, включно з аналітиками, експертами та прихильниками, постійно змінюють дані, щоб оновлювати послідовність дій або робити прогнози [21]. Аналітика великих даних і штучний інтелект зробили революцію в спортивному секторі, уточнюючи статистичну інформацію та керуючи кількісними та якісними даними в зрозумілій і стабільній вміст [22]. Першим кроком до розуміння даних є їх візуалізація. Використовуючи візуалізацію даних, користувачі можуть створювати історії, організовуючи дані в більш просту для сприйняття форму, виділяючи шаблони та викиди. На додаток до усунення шуму в даних і відображення відповідної інформації, чудові візуалізації створюють історію. Щоб створити ефективну візуалізацію даних, потрібен ретельний баланс між дизайном і функціями. У результаті аналітики даних використовують широкий спектр інструментів, таких як графіки та діаграми, а також карти, серед іншого, для перекладу та відображення даних і з'єднань даних. Щоб зробити дані зрозумілими, зазвичай потрібен правильний підхід і його налаштування.

Що стосується вищезазначеної проблеми, із зростаючими вимогами до аналізу даних стали широко використовуватися користувацькі інтерфейси, засновані на візуалізації та візуальній аналітиці (Lei et al. 2015 рік). Останніми роками дослідники в області візуальної аналітики запропонували багато корисних методів і інструментів, які допомагають аналітикам і тренерам знаходити моделі поведінки, вирішуючи певні труднощі, що виникають під час процесу аналізу.

На рисунку 1.1 представлена таксономія методів, що використовуються при аналізі даних в спортивній галузі. Слід зазначити, що не всі вони є однаково популярними для використання в спортивній галузі.

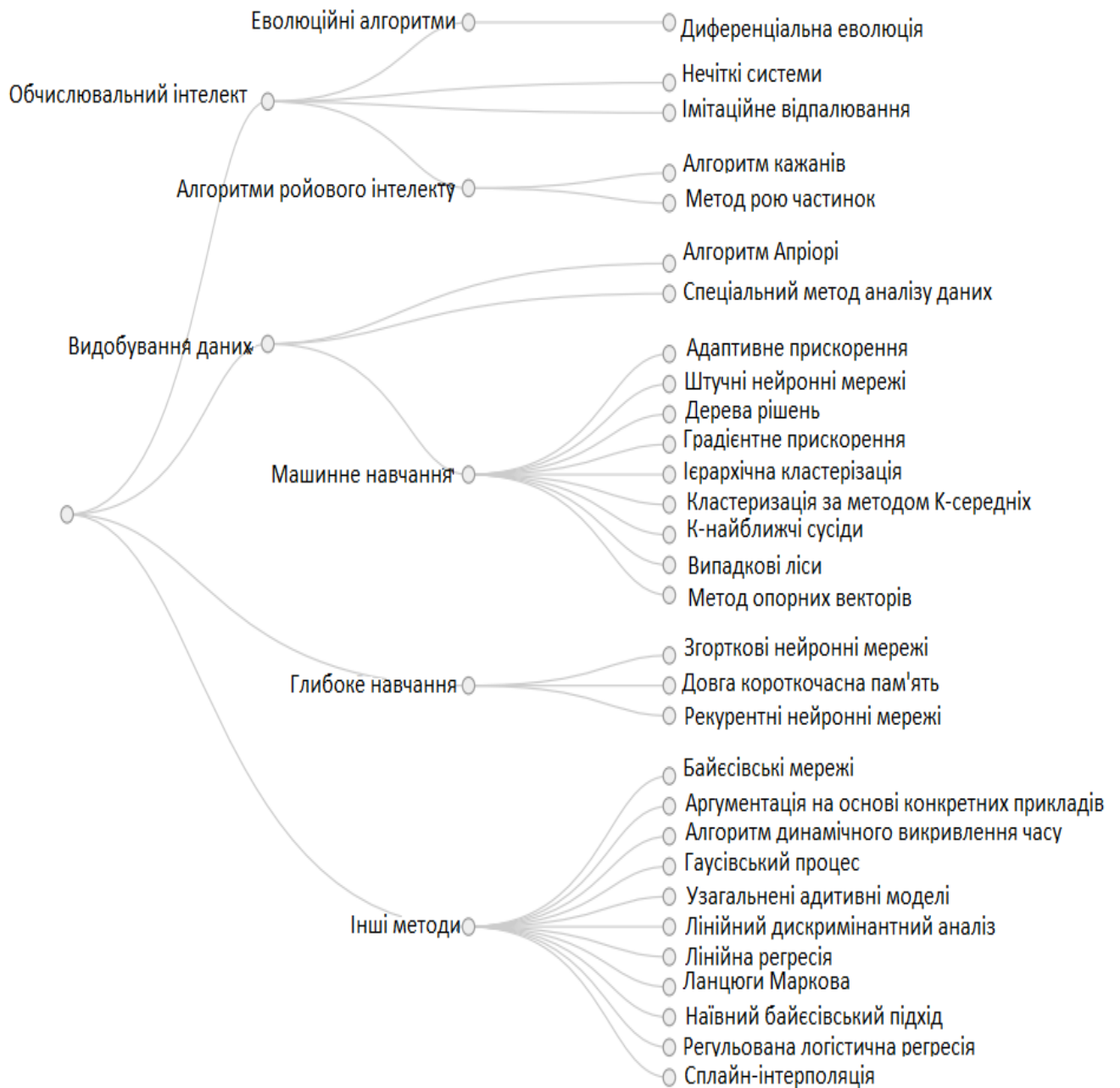


Рисунок 1.1 – Таксономія методів, що використовуються при аналізі даних в спортивній галузі [13]

1.2 Інтелектуальний аналіз даних в різних видах спорту

В інформатиці перетворення необроблених даних на інформацію, інформації на знання, а знань на мудрість утворює відому ієрархію обробки даних, також відому як ієрархія мудрості [14]. Інтелектуальний аналіз даних в основному стосується першої частини ієрархії мудрості - перетворення даних в інформацію. Це методологія вирішення проблем, яка знаходить логічний або математичний опис шаблонів і закономірностей у наборі даних [25]. За допомогою методів інтелектуального аналізу даних можна витягти корисну і раніше невідому інформацію з архівних або потокових даних. Видобута інформація може бути у формі прогнозування подій, знаходження подій, що відбуваються одночасно, та послідовності їх настання, а також поділ даних на групи схожих об'єктів.

Найбільш показовими для дослідження є елітні види спорту, наприклад види спорту, які входять в програми Олімпійських ігор. Оскільки саме в таких видах спорту широкого розповсюдження набуло використання інформаційних систем та технологій для досягнення більш високих спортивних результатів, підвищення спортивної майстерності спортсменів, плануванні та організації змагань різного рівня та спортивного скаутингу.

В елітних видах спорту методи інтелектуального аналізу даних зазвичай використовуються для моделювання взаємозв'язків між показниками ефективності та атрибутами, а також для вилучення моделей поведінки спортсменів з попередніх змагань [26]. Вони можуть бути використані в процесах прийняття рішень для підтримки стратегічного планування та відбору спортсменів на змагання.

Найпоширенішими технологіями інтелектуального аналізу даних в елітних видах спорту є кластеризація, класифікація, моделювання взаємозв'язків, регресійний аналіз, та видобуток правил. Розглянемо їх більш детально.

Кластеризація – це одна з форм неконтрольованого навчання, яка полягає у визначенні способу організації даних та узагальненні/поясненні ключових

особливостей даних [22]. Результатом кластерного аналізу є формування низки груп. Члени кожної групи схожі між собою за деякими критеріями (тобто атрибутами подібності) і найбільше відрізняються від членів інших груп за тими самими критеріями.

Класифікація – це форма керованого навчання, метою якого є вивчення характеру відображення від заданого входу до заданого виходу [24]. Системі повторно надаються факти про різні випадки разом з (відомими) очікуваними результатами. Потім вона коригує вагові коефіцієнти функції відображення, яка може давати результати, подібні до очікуваних. Після того, як функція відображення стає стабільною (тобто, навчання завершено), її можна використовувати для прогнозування групової приналежності для невидимих випадків даних (тобто, нових індивідуальних випадків). На відміну від кластеризації, де попередні знання про групи/приналежність даних відсутні або обмежені, при класифікації клас тренувальних даних відомий, а клас невидимих даних можна передбачити після завершення тренування.

Моделювання взаємозв'язків. Метою моделювання зв'язків у складному середовищі є пошук функції або моделі, яка найкраще описує взаємозв'язок між предиктором і залежними атрибутами. Основна мета в цьому завданні – знайти модель, яка відповідає даним з найменшою похибкою.

Регресійний аналіз – це один з методів, який використовується для підбору лінії або полінома до даних. Ідея цього аналізу полягає в тому, щоб пов'язати змінну відгуку з вектором предикторних змінних [25]. Знайдений зв'язок можна використовувати для прогнозування. Більш досконалими методами моделювання зв'язків є нейронні мережі, які можна використовувати для пошуку взаємозв'язків між атрибутами.

Нейронна мережа отримала свою назву через схожість з нервовою системою мозку. Вона складається з безлічі самоналагоджуваних обробних елементів, які співпрацюють у щільно взаємопов'язаній мережі. Кожен процесор генерує один вихідний сигнал, який передається іншим процесорам. Вихідний сигнал обчислювального елемента залежить від вхідних сигналів: кожен вхід

має ваговий коефіцієнт, який визначає ступінь впливу вхідного сигналу на вихідний. Сила вагових коефіцієнтів регулюється оброблювальним елементом автономно в процесі обробки даних [27].

Видобуток правил. Пошук правил, які відображають взаємозв'язки між серіями подій (наприклад, пресинг у першій половині футбольного матчу і перемога у грі) або станів (наприклад, емоційний спад або фізична слабкість) має дві основні форми в інтелектуальному аналізі даних, а саме: видобуток асоціативних правил і послідовний видобуток шаблонів. У той час як видобуток асоціативних правил стосується лише пошуку подій або станів, що повторюються, послідовний видобуток шаблонів зосереджений на послідовності подій, що повторюються з високою частотою у впорядкованому за часовими мітками наборі подій.

Ці два типи правил/патернів (правило асоціації та послідовний видобуток патернів), коли вони використовуються у спортивній сфері, потенційно можуть виявити низку умов, рухів, рішень, позицій або подій загалом, а також їх послідовність у часі, які можуть призвести до певних позицій, балів або результатів.

У таблиці 1.1 наведено приклади використання методів інтелектуального аналізу даних для аналізу спортивних досягнень в елітних видах спорту.

З точки зору аналізу даних, кожен метод інтелектуального аналізу даних (наприклад, кластеризація та класифікація) може бути реалізований за допомогою різних технік (наприклад, k-середніх, дерев рішень, байєсівських мереж, машин опорних векторів та регресійного аналізу). Кожен метод можна охарактеризувати за трьома основними характеристиками:

1. Інтерпретованість: наскільки легко результати, отримані за допомогою певного методу, можуть бути інтерпретовані експертами з аналізу даних і зрозумілі фахівцям (спортивним), які не є експертами в галузі аналізу даних.

2. Точність. Точність і надійність результатів, отриманих за допомогою цього методу.

3. Гнучкість. Ступінь, до якого певний метод може бути використаний для аналізу певних проблем з різними параметрами та/або різними даними.

Таблиця 1.1 – Приклади використання методів інтелектуального аналізу даних для аналізу спортивних досягнень в елітних видах спорту

Метод	Алгоритм	Вид спорту
Кластеризація	k-середніх	гольф, велоспорт на треку
	середній зв'язок (ієрархічний)	плавання
	алгоритм Варда	гольф
	змішаний	десятиборство
	карти, що самоорганізуються	баскетбол
Класифікація	наївний Байєс	велоспорт на треку
	лінійний дискримінантний аналіз	регбі, веслування, стрибки в довжину
Моделювання відносин	лінійна регресія	плавання
	лінійна та поліноміальна регресія	плавання
	нейронні мережі	плавання, аеробіка, футбол
Видобуток правил	Правила асоціації	настільний теніс, баскетбол

Кожна з вищезгаданих цілей аналізу спортивних результатів вимагає використання відповідного специфічного методу аналізу даних, тоді як кожна технологія інтелектуального аналізу даних для цього методу має свої особливості з точки зору інтерпретованості, точності та гнучкості. З іншого боку, кожна вимога до аналізу спортивних результатів вимагає різних рівнів від кожної технології.

1.3 Вимоги до інтелектуального аналізу даних при аналізі даних про спортивні результати

За результатами проведеного аналізу розглянемо вимоги до інтелектуального аналізу даних при аналізі даних про спортивні результати в елітних видах спорту.

При аналізі даних про спортивні результати є три основні атрибути, які цікавлять спортивних науковців та професіоналів, а саме: рейтинг, час та бали (РЧБ). Ці три показники відображають результативність і використовуються для оцінювання спортсменів у більшості спортивних змагань, а також відбору спортсменів на змагання.

Існуючий розрив між прикладним спортом і інтелектуальним аналізом даних можна подолати за допомогою мапування, як показано на рисунку 1а, від спортивної сфери до сфери інтелектуального аналізу даних. Таке відображення дозволяє вибрати відповідну техніку аналізу даних для конкретного спортивного питання. У той час як спортивна область включає основні правила, положення, тактику, стратегії, результати, умови і здібності, пов'язані з конкретними видами спорту, область інтелектуального аналізу даних включає репрезентативні показники ефективності, а саме показники РЧБ.

Методи попередньої обробки та аналізу даних, які можуть бути використані в просторі інтелектуального аналізу даних, можуть лише інтерпретувати наявні дані у вигляді показників ефективності.

Глибше розуміння знань (спортивної) предметної області, а також кращого розуміння доступних і відповідних інструментів інтелектуального аналізу даних допомагає мінімізувати проблеми в аналізі спортивних результатів, які можуть виникнути через недостатню точність, обґрунтований підхід або достовірність аналізу.

Попередня обробка даних, пов'язаних зі спортом, у сфері інтелектуального аналізу даних передбачає підготовку та сортування даних для аналізу. Залежно

від поточного та необхідного формату даних і дослідницької проблеми, попередня обробка може складатися з одного або декількох наступних завдань:

1. Фільтрація. Фільтрація записів даних за певними категоріями змагань (наприклад, розподіл результатів веслування на швидкі, середні та повільні дистанції).

2. Перетворення формату. Перетворення даних у формат, який може бути інтерпретований спеціальним програмним забезпеченням для аналізу даних та інструментами, які використовуються для проведення фактичного аналізу даних (наприклад, перетворення часу чч:мм:сс у збірні секунди).

3. Видобування. Пошук нових даних, які не є доступними у явному вигляді, на основі зібраних даних (наприклад, вилучення абсолютних і кумулятивних рейтингів човнів відносно різних секторів веслувальних змагань на 2000 м на основі часу човнів).

4. Структурне перетворення. Перетворення частин даних у формат, який дозволяє проводити більш точний аналіз даних (наприклад, узагальнення підсумкових результатів від 1 до кількості учасників у категоріях призерів (1-3 місця) та не призерів (більше 3 місць)).

5. Описова конвертація. Перетворення певних частин даних у формат, який краще описує природу конкретного виду спорту/проблеми (наприклад, перетворення абсолютного часу у відносний/диференціальний час, який показує різницю в часі між лідером та іншими спортсменами).

Якщо перші три типи попередньої обробки даних є простими, то різниця між останніми двома (структурним та описовим перетворенням) може бути більш делікатною та менш відчутною.

При структурному перетворенні семантика даних не змінюється (наприклад, рейтинги до і після узагальнення мають однакову семантику), тоді як при описовому перетворенні значення нових перетворених даних не збігається з їх початковою формою (наприклад, час проти різниці в часі або позначені міткою часу і впорядковані позиції проти необроблених позицій).

На завдання попередньої обробки даних, такі як фільтрація, перетворення формату та вилучення, не впливає специфіка конкретного виду спорту. Але структурне перетворення зазвичай залежить від:

- специфіки методу аналізу даних, коли деякі методи аналізу даних можуть ефективніше працювати з номінальними значеннями порівняно з числовими типами даних (наприклад, системи класифікації);
- обсягу наявних даних, коли часто невеликі обсяги даних є несуттєвими, обсягу наявних даних, коли часто невеликі обсяги даних, що стосуються конкретних параметрів, можуть вимагати узагальнення значень до більш грубих величин (наприклад, узагальнені місця фінішу, такі як категорії, визначені для олімпіади з велоспорту на треку, які охоплюють більшу кількість записів даних).

Перетворення описових даних тісно пов'язане з розумінням виду спорту та його особливостей. Відповідна методика попередньої обробки даних може бути обрана або зумовлена комбінацією особливостей виду спорту, таких як кількість подій, кількість спортсменів, тривалість або критерії перемоги. У випадках, коли потрібно врахувати кілька особливостей виду спорту, може бути доцільним поєднання завдань попередньої обробки. Більшість однокомпонентних видів спорту, наприклад велосипедні гонки на час та марафонський біг зазвичай не вимагають і не впливають на дескриптивне перетворення при попередній обробці даних.

Основною причиною цього є те, що для цих видів спорту предикторні змінні (тобто показники РЧБ) чітко відображають характер виду спорту. Отже, ці змінні можуть бути використані для моделювання базової структури для прогнозування змін залежної змінної (наприклад, фінального заліку). Наприклад, плавання на 200 м вільним стилем не потребує спеціальної попередньої обробки часу проходження кіл для прогнозування фінішного часу.

Однак у багатокомпонентних видах спорту може виникнути потреба у перетворенні конкретних даних в інші форми.

Ми вважаємо, що це залежить від того, чи проводяться окремі змагання послідовно (наприклад, триатлон), чи незалежно (наприклад, олімпіада з велоспорту на треку). РЧБ у незалежних багатокомпонентних видах спорту зазвичай надає достатньо інформації для моделювання та прогнозування залежних змінних (наприклад, загальний час та загальний підсумковий залік). Наприклад, аналіз на основі машинного навчання (кластеризація і класифікація) необхідних здібностей для перемоги у змаганнях з велоолімпіади на треку, проведений Ofoghi та ін., не включає ніякої подальшої попередньої обробки рейтингів гонщиків у кожному окремому змаганні, окрім узагальнення підсумкового заліку за трьома категоріями. Цими категоріями є: володарі медалей, володарі немедалей, які посіли місця від 4 до 10, і володарі немедалей, які посіли місця вище 10. Така попередня обробка в основному пов'язана з невеликою кількістю історичних даних, доступних для цього конкретного виду спорту.

У послідовних багатокомпонентних видах спорту, навпаки, сирі показники РЧБ можуть вводити в оману при аналізі результатів. Триатлон є прикладом послідовних змагань з кількох видів спорту, де сирі рейтинги або сирій час спортсменів у кожному окремому компоненті можуть не дати багато інформації про те, яка дисципліна відіграє найважливішу роль у вирішенні питання про те, чи зможе триатлоніст перемогти у змаганнях.

Це пов'язано з безпосередньою послідовністю змагань: абсолютний час або рейтинг триатлоністів у кожному компоненті не має вирішального значення. Важливою є різниця в часі відставання від лідера в кожному компоненті.

Друге місце з великою різницею в часі в будь-якому компоненті триатлону означає досить малі шанси на перемогу в змаганнях порівняно з четвертим або п'ятим місцем в будь-якій дисципліні з набагато меншою різницею в часі. Тому попередня обробка часу і перетворення його в диференціальний час стає вирішальним фактором для інтелектуального аналізу даних у триатлоні.

Існує тісний взаємозв'язок між виявленням патернів продуктивності та аналізом попиту. Аналіз попиту також може бути більш ефективно проведений за допомогою методів кластеризації та класифікації. Взаємозв'язок між цими двома завданнями пов'язаний з тим, що в багатьох випадках патерни результативності є одними з найбільш очевидних елементів інформації, які визначають вимоги до спортсменів для участі в конкретних спортивних змаганнях. Наприклад, якщо в триатлоні модель перемоги передбачає завершення бігової частини з найкращим результатом, то основна вимога цього виду спорту (тобто ключовий фактор успіху в цьому виді спорту) - бути сильним бігуном.

Аналіз попиту на певні види спорту або змагань може також проводитися з точки зору іншої інформації, такої як необхідна попередня підготовка, харчування та фізична сила; вимоги, які не обов'язково виявляються лише в структурі результатів.

1.4 Висновки. Постановка задачі

Методи кластеризації, класифікації і навіть видобування правил, як правило, дають результати, які краще використовувати до великих подій, наприклад, моделі перемог, асоціації успішних подій і послідовності, а також певні ймовірності результатів. Моделювання взаємозв'язків, навпаки, може застосовуватися в режимі реального часу для інтеграції наявних даних з урахуванням умов і специфіки поточної події та отримання ймовірностей певних результатів, на основі яких можна коригувати дії.

Отже, метою даного дослідження є розроблення методу створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних, для цього необхідно вирішити наступні задачі:

- 1) провести огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах;

2) розробити метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних;

3) запропонувати концепцію інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

РОЗДІЛ 2 ІНФОРМАЦІЙНІ ПОТОКИ ІНФОРМАЦІЙНОЇ СИСТЕМИ КЕРУВАННЯ ДАНИМИ СПОРТСМЕНІВ

2.1 Керування даними в спортивній індустрії

Керування великими спортивними даними в основному застосовує методи керування даними, інструменти та платформи для роботи зі спортивними великими даними, включаючи зберігання, попередню обробку та безпеку. Однак управління великими даними є складним процесом, який виникає через неоднорідність і неструктурованість джерел даних [28]. Управління великими спортивними даними має вирішальне значення для успіху національної спортивної індустрії, команд та окремих осіб [15, 16, 29].

Основна мета управління великими спортивними даними полягає в тому, щоб отримати потенційну цінність спортивних великих даних і підвищити якість даних і доступність для прийняття рішень. Практики спорту в сучасну епоху, значною мірою покладаються на докази, отримані з даних та джерел інформації, щоб керувати професійними спортсменами (West та ін., 2019). Базуючись на початкових структурах управління інформацією та знаннями, таких як ієрархія мудрості, такі дослідники, як Dammann (2018), проілюстрували процес генерування знань із даних як чотирирівневу ієрархічну модель (рис. 2.1).

Експерти визначили дані, інформацію та знання з різних точок зору, тому важко дати єдине дійсне визначення даних. Однак передбачається, що дані самі по собі не мають значення, що призвело до визначення інформації як даних зі значенням. Дослідники також продемонстрували, що інформація є потоком, який веде до створення запасу знань.

Таким чином, інформаційний потік може початися з даних і тривати до створення знань. Це демонструє, що дані та інформація є основною для отримання доказів і знань. Отже, у спортивному середовищі, де інформація відіграє важливу роль у створенні доказів і знань, необхідних для керування спортсменами (Рисунок 2.1), життєво важливо, щоб практикуючі спортсмени мали доступ до високоякісних даних та джерел інформації, що допомагають

підтримувати прийняття рішень.



Рисунок 2.1 – Ієрархія «Дані-Інформація-Докази-Знання» за Дамманном, 2018

Крім того, неоптимальний інформаційний потік на мікрорівні може призвести до того, що спеціалісти-практики генеруватимуть неправильні або неповні рішення, що стосуються управління гравцями (наприклад, спеціалісти-практики роблять неправильні оцінки тренувального навантаження гравців через недоступну інформацію). Це, у свою чергу, може призвести до того, що результати управління гравцями будуть відрізнятися від загальних організаційних цілей (наприклад, підвищений ризик отримання травми для спортсмена або неоптимальна адаптація тренувань).

Таким чином, мінімізація шуму та підвищення якості інформаційних потоків, пов'язаних із процесами управління гравцями, шляхом оптимізації, дозволили б створити чіткі системи підтримки прийняття рішень у професійному спортивному середовищі (Шеллінг та Робертсон, 2020) і підвищити загальну ефективність прийняття рішень, пов'язаних з керуванням гравцями. Розширюючи цю лінію мислення, зростає інтерес до впровадження сучасних стратегій, таких як цифрова трансформація, у спортивне середовище (Ströbel та ін., 2021). Хоча архітектури інформаційних систем для спортивних середовищ обговорювалися на макрорівні (Vlobel and Lames, 2020), існують обмежені практичні приклади, щоб

підкреслити, як такі спроби оцифрування можуть бути реалізовані спортсменами в професійних спортивних середовищах на мікрорівні. Крім того, у поточному контексті впровадження систем візуалізації даних (рис. 2.2) для ілюстрації конкретної спортивної інформації є перспективним методом метод оптимізації інформаційних потоків у спортивному середовищі (Перін та ін., 2018; Ду та Юань, 2021).

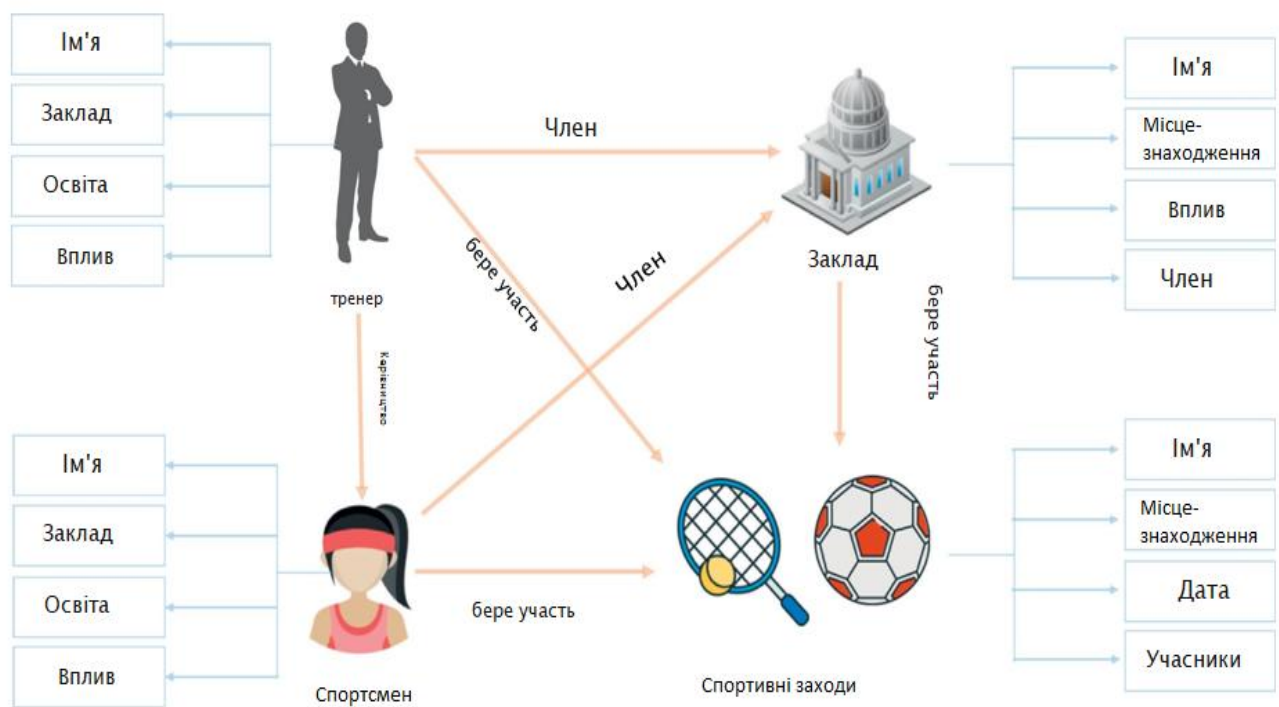


Рисунок 2.2 – Приклад об'єктів і їхніх зв'язків, пов'язаних із спортивними даними

Однак, розробка системи візуалізації даних буде лише частиною повного проекту цифрової оптимізації. Наприклад, для проектів оцифрування спортивна література наразі не містить достатніх практичних прикладів, щоб уточнити, як створювалися вимоги до таких систем візуалізації даних у спортивному середовищі, як оцінювали зручність використання цих систем (наприклад, візуалізації даних) і яке оцінювання проводиться для того, щоб визначити, чи справді такі системи підвищили продуктивність певного спортивного середовища (стосовно параметрів ефективності часу, вартості, якості чи гнучкості).

2.2 Інформаційні потоки в інформаційній системі керування даними спортсменів

У дослідженні було взято за основу управління інформаційними потоками (BPM) (Dumas et al., 2018; Kirchmer et al., 2019), які використовуються в широкому спектрі галузей послуг, таких як фінанси, охорона здоров'я, банківська справа та інформаційні технології. (Anand et al., 2013; Fernández et al., 2019), для визначення поточного стану та можливостей покращення інформаційного потоку.

Як інструмент управління змінами BPM досяг успіху в оптимізації інформаційних потоків, пов'язаних із сервісно-орієнтованими (нематеріальними) процесами, такими як управління інфекціями (Cánovas-Segura et al., 2017). Оскільки природа процесів управління гравцями є сервісно-орієнтованою (наприклад, фізіотерапевт забезпечує лікування м'яких тканин, тренер надає технічні знання для підвищення продуктивності гравців) і підтримуються інформаційними шляхами, BPM розглядається як відповідна структура для використання в поточному дослідженні (Ranaweera та ін., 2021). Було розглянуто управління бізнес-процесами (BPM) як структура управління змінами для оптимізації розглянутого потоку інформації в інформаційній системі. Застосовність BPM до спортивних контекстів обговорювалася багатьма дослідниками (Ранавіра та ін., 2021) і, пропонується прийняти шестифазний життєвий цикл BPM, представлений Дюма та ін. [40].

Шестифазний життєвий цикл BPM – це методологія управління бізнес-процесами, яка включає шість ключових етапів. Кожен етап орієнтований на конкретні завдання та дозволяє систематично розглядати та оптимізувати бізнес-процеси:

1. Аналіз бізнес-процесів. На початку циклу проводиться докладний аналіз існуючих бізнес-процесів. Визначаються цілі, завдання, ресурси, та інші важливі аспекти. Це дозволяє зрозуміти, як процеси функціонують та ідентифікувати можливості для оптимізації.

2. Дизайн бізнес-процесів. На основі результатів аналізу розробляється новий дизайн бізнес-процесів. Цей етап включає в себе визначення оптимальних кроків, розподіл завдань, визначення ролей та взаємодій між учасниками процесів.
3. Моделювання бізнес-процесів. Створюються моделі бізнес-процесів, що відображають їхню структуру та логіку. Використовуються графічні зображення для кращого розуміння процесів усіма учасниками команди.
4. Впровадження та автоматизація. На цьому етапі процеси переносяться з моделі в реальний бізнес-середовище. Це може включати в себе впровадження технологій BPM, які автоматизують виконання завдань та контроль за процесами.
5. Вимірювання та оцінка. Визначаються ключові показники ефективності (KPI), які дозволяють вимірювати продуктивність впроваджених змін. Проводяться систематичні вимірювання та оцінки результатів для забезпечення відповідності цілям та виявлення можливостей для подальших удосконалень.
6. Оптимізація бізнес-процесів. На основі отриманих результатів вживаються заходи для подальшої оптимізації бізнес-процесів. Це може включати в себе внесення коректив у дизайн, впровадження нових технологій або розвиток персоналу.

На рисунку 2.3 представлено методи оптимізації інформаційних потоків в інформаційній системі.

Важливим питанням є оцінка якості розглянутого інформаційного потоку до оптимізації. Цей останній крок є складовою етапу аналізу процесу.

Ідентифікація процесу – генерує архітектуру організаційного процесу, показники продуктивності, зв'язки та систематично визначає, які з цих процесів потребують втручання BPM, щоб допомогти досягти організаційних цілей.

Виявлення процесу – для ідентифікованих процесів інформації про поточний стан (як є) збирається та документується за допомогою методів моделювання процесу.

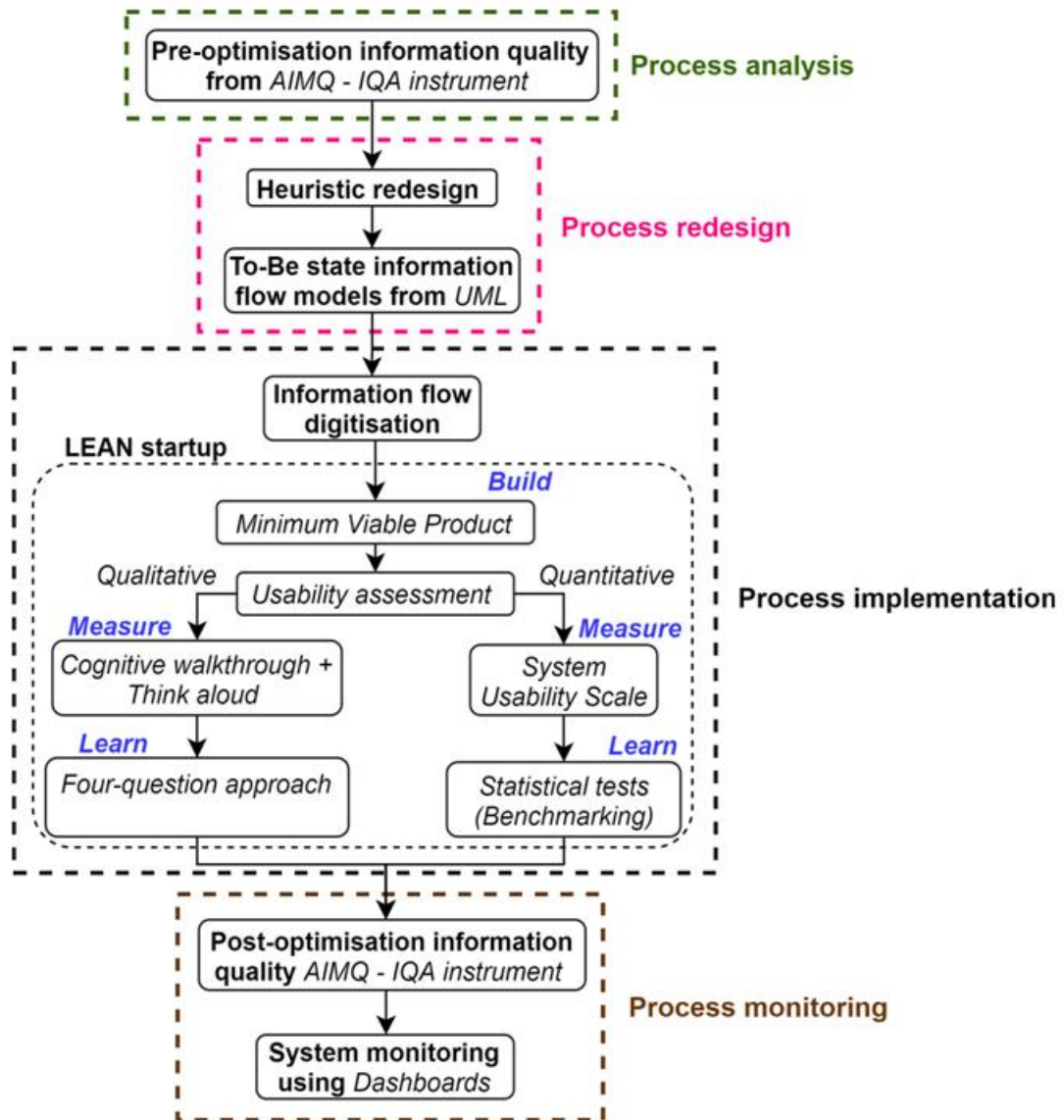


Рисунок 2.3 – Методи оптимізації інформаційних потоків в інформаційній системі

Аналіз процесу – проблеми в межах задокументованого стану процесів «як є» визначаються для оптимізації, потенційно створюючи список переважних проблем. Перепроєктування процесу – процес перепроєктовано (оптимізовано), щоб подолати проблеми, виявлені на етапі аналізу процесу, щоб визначити найкращий майбутній стан (бути).

Впровадження процесу – необхідні зміни для переміщення процесу зі стану «як є» до стану «бути» виконуються шляхом управління організаційними змінами.

Моніторинг процесу – реалізований процес контролюється, щоб визначити ефективність змін. Цикл повторюється до етапу виявлення, якщо є додаткові проблеми або необхідні подальші постійні вдосконалення.

Розглянемо детальніше оцінку якості інформації перед оптимізацією. Існують дослідження, де розглядалось питання якості даних, як результат систем підтримки прийняття рішень, запропонованих для високопродуктивних спортивних середовищ, враховуючи вплив на різні виміри (наприклад, надійність, доступність, своєчасність) якості даних (Шеллінг і Робертсон, 2020). Більш широке дослідження комп'ютерної техніки ілюструє різні системи для оцінки якості даних (Січі та Расс, 2019), включаючи комбіновані суб'єктивні та об'єктивні методи.

В даній роботі пропонуємо посилатися на оцінку якості інформації, а не на якість даних. Тому пропонується використовувати модель ефективності продукту та послуги для якості інформації (Модель PSP/IQ), щоб оцінити якість потоку інформації.

Це зосереджено на оцінці якості інформації за 16 вимірами, які відображаються в чотирьох квадрантах: надійна, корисна, надійна та корисна інформація. У рамках PSP/IQ погляд на інформацію як на продукт (генерація даних для зберігання, зазвичай у базі даних) або послугу (перетворення даних на інформацію) тісно пов'язаний з ієрархією DIEK та її ілюстрацією у спорті. Наприклад, можна переглядати збір, зберігання та підтримку щоденних даних. Процес доступу до збережених даних та візуалізації їх як джерела інформації для персоналу для керування гравцями можна проаналізувати з точки зору інформація як послуга. Можна використовувати стандартний інструмент IQA (опитувальник з 11-бальною шкалою), введений через AIMQ-метод оцінки якості інформації для суб'єктивного визначення інформаційної якості інформаційного потоку, який розглядається для оптимізації.

Редизайн процесу – для поточного дослідження використовувався евристичний редизайн, щоб перетворити потік інформації у покращений майбутній стан. Евристичний редизайн підходить для редизайну процесів керування гравцями, оскільки він зосереджується на зміні поточного (як є) стану поступово (транзакційно), працюючи в контексті процесу «як є» (внутрішнє звернення) за допомогою визначеного набору евристик редизайну (аналітичний). Більшість орієнтованих назовні стратегій редизайну (наприклад, перепроєктування найкращих практик подібних процесів інших подібних спортивних організацій) не були практичними, оскільки спортивні організації, як правило, працюють у закритих межах і не часто розглядають такі стратегії, як відкриті інновації. На конкретному рівні було обрано відповідні евристики редизайну з 29 представлених Дюма та ін. трансформувати розглянутий інформаційний потік в оптимізований майбутній стан (To-Be). Отримані процеси після редизайну та системи, необхідні для перетворення процесів, були змодельовані за допомогою UML.

Реалізація процесу – після переробки інформаційного потоку його було реалізовано за допомогою методів оцифрування. У літературі загальним способом реалізації процесу в життєвому циклі BPM є автоматизація завдань за допомогою системи управління бізнес-процесами (BPMS) [48]. Хоча можуть існувати можливості інтеграції BPMS із системою керування спортсменами (AMS) у спортивних контекстах, для поточного дослідження замість експериментів із BPMS впроваджувався перероблений потік інформації, розробивши цифрові системи для щоденного збору, керування та візуалізації тренування з опором і прогресивні базові дані тестування гравців. Це сталося тому, що, по-перше, запроваджена автоматизація системи мала відповідати унікальним вимогам користувачів, узгоджуючи їх із бізнес-цілями розглянутої організації. По-друге, і це найважливіше, на відміну від деяких інших професійних видів спорту (наприклад, футболу), багато спортивних організацій, не мають фінансової потужності, необхідної для

автоматизації переробленого процесу з використанням існуючої технології BPM, як-от BPMS.

Досить часто через динамічний характер певних видів спорту (наприклад, вимоги, що постійно змінюються), розробки систем повинні бути гнучкими та гнучкими за своєю природою. Тому, як підкреслюють інші дослідники спорту, такий метод, як Lean Startup, який зосереджений на більш адаптованому процесі до змін, що складається з коротшої тривалості проектування з раннім прототипуванням протягом життєвого циклу розробки продукту, був більш придатним для розробки пропонованих систем збору даних, управління та візуалізації. Зокрема, керуючись підходом дизайнерського мислення для збору вимог користувачів, як показано на рисунку 2.3, ми використали ітерацію будувати-вимірювати-навчатися циклу, запропонованого в Lean Startup для оцифрування потоку інформації.

Збір є важливим завданням для обробки даних. Крім того, безпечний збір спортивних даних є життєво важливим кроком для всіх типів програм даних, які можуть забезпечити результат аналізу даних. Оскільки підозрілі джерела даних дозволяють збирати дані для дослідження різноманітних зловмисних атак і обробок, безпечна методологія збору спортивних даних має вирішальне значення для різних програм даних. Щоб забезпечити безпечний збір даних, дослідники досягли певного прогресу в цій галузі, наприклад, забезпечивши енергоефективний збір даних і захист даних у розподіленому середовищі. Пропонується ефективна структура для подолання цих проблем, спираючись на блокчейн і глибоке підкріплення навчання [30]. Блокчейн-платформу Ethereum можна використовувати для забезпечення безпеки даних, коли мобільні термінали обмінюються даними.), цей фреймворк може вирішувати різні атаки, такі як атака більшості, збій пристрою та атака затемнення [30].

Після того, як буде отримано достатньо спортивних даних, наступна обробка полягає в позначенні окремих прикладів. Наприклад, враховуючи набір даних баскетбольної гри, позначте різні результати з метою прогнозування майбутніх результатів індивідуума. Зазвичай збір даних виконується разом із маркуванням

даних. Під час отримання інформації з Інтернету та створення бази знань кожна інформація вважається правильною і, таким чином, неявно позначається як правдива. Під час обговорення літератури з маркування даних легше відокремити її від збору даних, оскільки використовувані методи можуть бути різними.

Мітки даних можна розділити на три категорії:

1. Існуючі мітки. Ці існуючі мітки можна використовувати, щоб навчитися з них прогнозувати решту міток.
2. На основі натовпу. Останнім часом багато методів краудсорсингу можуть бути використані, щоб допомогти гравцям стати більш ефективними у маркуванні.
3. Слабкі мітки. Хоча бажано постійно генерувати правильні мітки, цей процес впровадження може бути дуже дорогим. Щоб генерувати менше ніж ідеальні мітки, слабка мітка є альтернативним методом, який використовується в багатьох програмах як дані з мітками.

Зазвичай потрібно багато робочої сили, щоб позначити дані, але можна створити лише невелику кількість позначених даних. Техніка напівконтрольованого навчання досліджує позначені та немарковані дані для прогнозів [31]. Менша галузь дослідження називається самомаркуванням, що є широкою темою. Однією з переваг техніки самопозначення є те, що вона може генерувати більше міток, довіряючи власним прогнозам [32]. Крім того, існують методи розповсюдження міток на основі графів, які спеціалізуються на даних спортивних графіків).

Методи напівконтрольованого навчання можна використовувати в класифікації, регресії та завданнях розповсюдження міток на основі графів. Мета використання методів напівконтрольованого навчання в класифікаційному завданні полягає в тому, щоб навчити модель, яка повертає один із кількох можливих класів для кожного прикладу, використовуючи позначені та немарковані набори даних.) Метою використання методів напівконтрольованого навчання в регресійному завданні є навчання моделі, яка може передбачити дійсне число на прикладі. Розповсюдження міток на основі графів має

застосування в комп'ютерному зорі, пошуку інформації, соціальних мережах та обробці природної мови [33–35]. Кращий спосіб позначити приклади — зробити це вручну.

Однак для завершення великого проекту знадобилися би роки, що більшість користувачів машинного навчання не можуть собі дозволити для власних програм. Традиційно активне навчання було важливою технікою в спільноті машинного навчання для ретельного вибору справжніх прикладів для позначення та мінімізації витрат.

Нещодавно в задачі маркування були запропоновані методи краудсорсингу. Тому більше уваги приділяється тому, як призначати завдання, щоб забезпечити високу якість етикеток [36]. Модель програмування даних досягла прогресу в двох аспектах: точності та зручності використання. Порівняно з навчанням із меншою кількістю міток вручну, навчання дискримінаційної моделі на великій кількості слабких міток може призвести до вищої точності. Було розроблено декілька систем для програмування даних, таких як DeepDive, DDLite та Snorkel [37–39].

Покращення наявних даних. Технологію машинного навчання можна використовувати для роботи з шумними даними та невиправленими мітками. Існує багато літератури щодо покращення якості даних [40, 41]. Представницька система очищення HoloClean створює імовірнісну модель, яка використовує правила якості, співвідношення цінностей і довідкові дані, щоб відобразити, як дані були згенеровані [42]. Крім того, були розроблені інші засоби очищення даних для перетворення необроблених даних у кращу форму для подальших досліджень. Деякі моделі очищення призначені для підвищення точності. Модель ActiveClean розглядає навчання та очищення як форму стохастичного градієнтного спуску для покращення ефекту очищення даних. TARS можна використовувати для вирішення проблеми очищення краудсорсингових міток за допомогою оракулів) е TARS дає дві поради. З одного боку, враховуючи тестові дані з шумними мітками, вони передбачають, наскільки добре модель може працювати на справжніх мітках за допомогою методу оцінки. З іншого боку, враховуючи навчальні дані з

шумовими мітками, TARS може визначити, які приклади надсилати оракулу, щоб максимізувати очікувану модель, що може підвищити точність кожного шуму [43]. Щоб отримати високоякісні мітки даних, покращення якості наявних міток є хорошим рішенням [44]. Вони досліджують покращення (або його відсутність) якості даних шляхом повторного маркування та зосереджуються на вдосконаленні навчальних міток для ввідного навчання під наглядом.

Результати їхнього експерименту показують такі переваги:

- 1) повторне маркування може покращити якість мітки та якість моделі;
- 2) для шумних міток повторне маркування може покращити якість мітки;
- 3) пропонується надійна техніка для покращення якості мітки.

Було використано перші три фази (ідентифікація процесу, перепроєктування та аналіз) життєвого циклу BPM, як основу для виявлення проблем, які існують у поточному інформаційному потоці в підрозділі високої ефективності. Розглянемо їх детальніше.

Ідентифікація процесу генерує архітектуру організаційного процесу, показники ефективності, взаємозв'язки та систематично визначає процеси, які вимагають втручання BPM для досягнення цілей організації.

Відкриття процесу призначено для визначених процесів інформація про поточний стан (як є) збирається та документується за допомогою методів моделювання процесу.

Аналіз процесу необхідний для виявлення проблеми в рамках задокументованого стану процесів «як є», що визначаються для оптимізації, потенційно створюючи список переважних проблем.

Перепроєктування процесу необхідно для того, щоб процес було перепроєктовано (оптимізовано) для подолання проблем, виявлених на етапі аналізу процесу, щоб визначити найкращий майбутній стан (To-Be).

При реалізація процесу необхідні зміни для переміщення процесу зі стану «Як є» до стану «Бути» виконуються шляхом управління організаційними змінами.

Моніторинг процесу потрібний, щоб запроваджений процес контролювався для визначення ефективності змін. Цикл повторюється до етапу виявлення, якщо є додаткові проблеми або необхідні подальші постійні вдосконалення.

Моделювання інформаційних потоків в спортивних інформаційних системах може бути виконане за допомогою різних методів та інструментів. Ось кілька кроків та підходів, які можна використовувати:

1. Визначення процесів. Визначаємо основні процеси та дії, які відбуваються в спортивних інформаційних системах. Це може включати реєстрацію спортсменів, планування тренувань, організацію змагань, обробку результатів та інші.

2. Ідентифікація джерел та призначень інформації. Необхідно визначити, де виникає інформація (джерела) та куди вона направляється (призначення). Наприклад, результати тренувань можуть бути джерелом інформації для тренера та призначенням для бази даних.

3. Створення діаграм потоків даних (DFD). Діаграми потоків даних використовуються для моделювання потоків інформації між різними частинами системи. Вони показують, як дані переходять від одного компонента до іншого та обробляються на кожному етапі.

4. Розгляд діаграм станів. Для конкретних об'єктів або процесів розгляньте діаграми станів. Вони показують різні стани та переходи між ними в залежності від подій.

5. Моделювання бізнес-процесів. На цьому етапі створюємо моделі бізнес-процесів, які відображають послідовність операцій в системі, починаючи від введення даних та завершуючи виведенням результатів.

6. Використання діаграм секвенцій та комунікацій. Якщо потрібно детальніше описати взаємодію між об'єктами чи компонентами, використовуйте діаграми послідовності та комунікацій. Вони демонструють взаємодію та передачу повідомлень між різними елементами системи.

7. Уточнення з використанням мови UML. Для розробки діаграм класів, компонентів, розгортання та інших, які можуть допомогти в деталізації структури

та взаємодії компонентів системи будемо використовувати мову моделювання UML.

8. Оцінка та оптимізація. Оцінюємо ефективність інформаційних потоків, ідентифікуючи можливість оптимізації та удосконалення процесів.

Моделювання інформаційних потоків допомагає розуміти та вдосконалювати функціональність спортивних інформаційних систем, забезпечуючи ефективну роботу та оптимальний обіг даних.

2.3 Концепція інформаційної системи керування даними спортсменів

Створення інформаційної системи керування даними спортсменів зазвичай включає в себе наступні кроки і аспекти.

Збір інформації. Цей етап включає в себе збір різноманітних даних про спортсменів, такі як біографічні дані, статистика виступів, травми, тренувальний процес, харчування тощо.

Зберігання даних. Опис методів зберігання даних, таких як реляційні бази даних, NoSQL рішення, а також обговорення схеми бази даних для забезпечення ефективного доступу до інформації.

Аналіз даних. Використання інтелектуального аналізу даних (Data Analytics) для виявлення цінних зв'язків і патернів в даних. Оцінка впливу різних факторів на результати спортсменів.

Візуалізація даних. Розробка графічних інтерфейсів та візуалізацій для представлення даних і аналізу результатів.

Прогнозування та прийняття рішень. Використання аналітики для прогнозування результатів та оптимізації стратегій тренування і конкуренції.

Захист даних і конфіденційність. Розгляд питань безпеки та конфіденційності даних спортсменів, особливо з урахуванням регуляторних вимог і стандартів (наприклад, GDPR).

Інтеграція з іншими системами. Обговорення процесів і технологій інтеграції інформаційної системи керування даними з іншими системами, такими як системи реєстрації на змаганнях, медичні системи тощо.

Приклади впровадження. Реальні приклади впровадження подібних систем у спортивних організаціях або командах.

Переваги і виклики. Обговорення переваг і викликів створення інформаційних систем для управління даними спортсменів.

Розглянемо проблеми, які можуть виникати в поточному інформаційному потоці.

Тематичний аналіз даних, зібраних у фокус-групах для аналізу поточного стану процесу управління щоденним тренувальним навантаженням виявляє проблеми в поточному інформаційному потоці, що стосуються двох ключових тем, а саме: створення знань і гнучкість процесу. Кожна з основних тем має містити підтеми, які висвітлюють конкретні проблеми в поточному інформаційному потоці, пов'язаному з головним процесом.

Створення знань. Ця тема була зосереджена на визначенні поточних проблем у головному процесі при створенні знань на основі даних, інформації та доказів. Зокрема, вона висвітлює поточні проблеми, пов'язані з інформаційним потоком, коли тренери з сили і кондиціонування, фізіотерапевти, спортивні вчені і лікарі отримують доступ до джерел даних і перетворюють їх на знання для прийняття рішень, що стосуються управління гравцями (наприклад, виставляння прапорців).

Недоступні дані. Хоча певні джерела даних про гравців, такі як силові тренування та дані базового тестування, були зібрані під час тренувань, у деяких випадках вони не були доступними для тренерів-практиків. Ця остання недоступність щоденних даних про силові тренування гравців була очевидною з діаграми процесу BPMN збору даних про поточний стан. У поточному стані гравець завершує тренування і записує свої найкращі результати. Потім, раз на 6 тижнів, тренер фотографує дошку і розраховує найкраще значення кожного гравця протягом 6-тижневого блоку і вносить ці дані в онлайн-таблицю Microsoft

Excel. Таким чином, єдиний спосіб отримати доступ до щоденних даних про тренування гравця в поточному стані – це спостерігати за дошкою в тренажерному залі. Натомість, варто використовувати щоденні GPS-метрики тренувань гравців, які були б доступними як джерела даних/інформації в кожному конкретному випадку, оскільки дані обробляються в інтерфейсі Power BI, до якого має доступ персонал.

Недоступні дані. Ця проблема підкреслила недоступність певних параметрів даних гравців для прийняття рішень. Однак фахівці усвідомлювали, що ці джерела даних були скоріше з категорії необов'язкових, але тих, які бажано, щоб були, а не критично важливими для управління гравцями в поточному контексті.

Інформаційний шум. Ця проблем в основному зосереджувалася на тому, що стейкхолдери вказували на те, що інформація, яка використовується в процесі прийняття рішень щодо щоденного управління тренувальним навантаженням, може бути неповною, непослідовною або застарілою. Однак, слід пам'ятати той факт, що результат головного процесу може знадобитися тренерам для створення остаточного плану тренувань та відбору на змагання. Тому це додатково висвітлило проблему, пов'язану з неналежною синхронізацією процесів управління щоденним тренувальним навантаженням і проектування кондицій спортсменів.

Відсутня інформація. Вказує на відсутність щотижневих стратегічних цілей, визначених з точки зору тренерів, щоб керувати практичними та макрорівневими стратегічними цілями і завданнями індивідуального розвитку гравців.

Відсутність консенсусу у тренерському штабі. Важливо досягати консенсусу серед практиків щодо різних прапорців (повідомлень), які генеруються під час роботи з гравцями. Крім того, конкретні умови введення даних/інформації, пов'язані з цими прапорцями, наразі також були відсутні. Це ілюструє наявність шуму в прийнятті рішень, пов'язаних із позначенням гравців у поточному контексті.

Гнучкість процесу. Ця основна тема визначила проблеми в поточному інформаційному потоці, пов'язані з динамічним і непередбачуваним характером

процесів управління гравцями. Зокрема, основною підтемою були проблеми в поточному інформаційному потоці, пов'язані з незавершеними рішеннями, що стосуються управління спортсменами.

Незавершеність рішень. Поточна інформаційна архітектура не містить стислого інформаційного потоку для управління сценаріями, коли фахівці не можуть прийняти конкретні рішення щодо щоденних тренувальних категорій гравця в рамках процесу відбору на змагання. Випадки, коли тренерському штабу потрібен додатковий час для спостереження за гравцем, щоб рекомендувати йому ідеальну щоденну тренувальну категорію, не підкріплені належним інформаційним потоком.

На рисунку 2.5 показано підсистему інформаційної системи керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків. Рівень джерела даних в основному включає історичні дані спортсменів, траєкторію поведінки спортсменів, відеодані та джерело даних інтернету. Рівень джерела даних є основою для реалізації різноманітних додатків для аналізу та прогнозування спортивних великих даних. Наступний рівень – це рівень збору даних, який збирає дані з рівня джерела даних і виконує таку обробку: збір даних, зберігання даних, обмін даними, імпорт вручну та веб-сканер. Потім зібрані дані очищаються, а необхідна обробка виконується відповідно до різних вимог програми, де дані класифікуються та зберігаються. Оброблені дані зберігатимуться на рівні центрального сховища, включаючи сховище структурованих даних, сховище неструктурованих даних і сховище файлів. Рівень аналізу даних виконує вибір функцій, аналіз зв'язків, статистичний аналіз і аналіз соціальних мереж відповідно до потреб конкретних програм з метою виявлення потенційних знань, законів і закономірностей у спортивних великих даних. Виходячи з наведених вище результатів аналізу, інтеграція машинного навчання та технології великих даних може сприяти розвитку додатків великих даних у спорті.

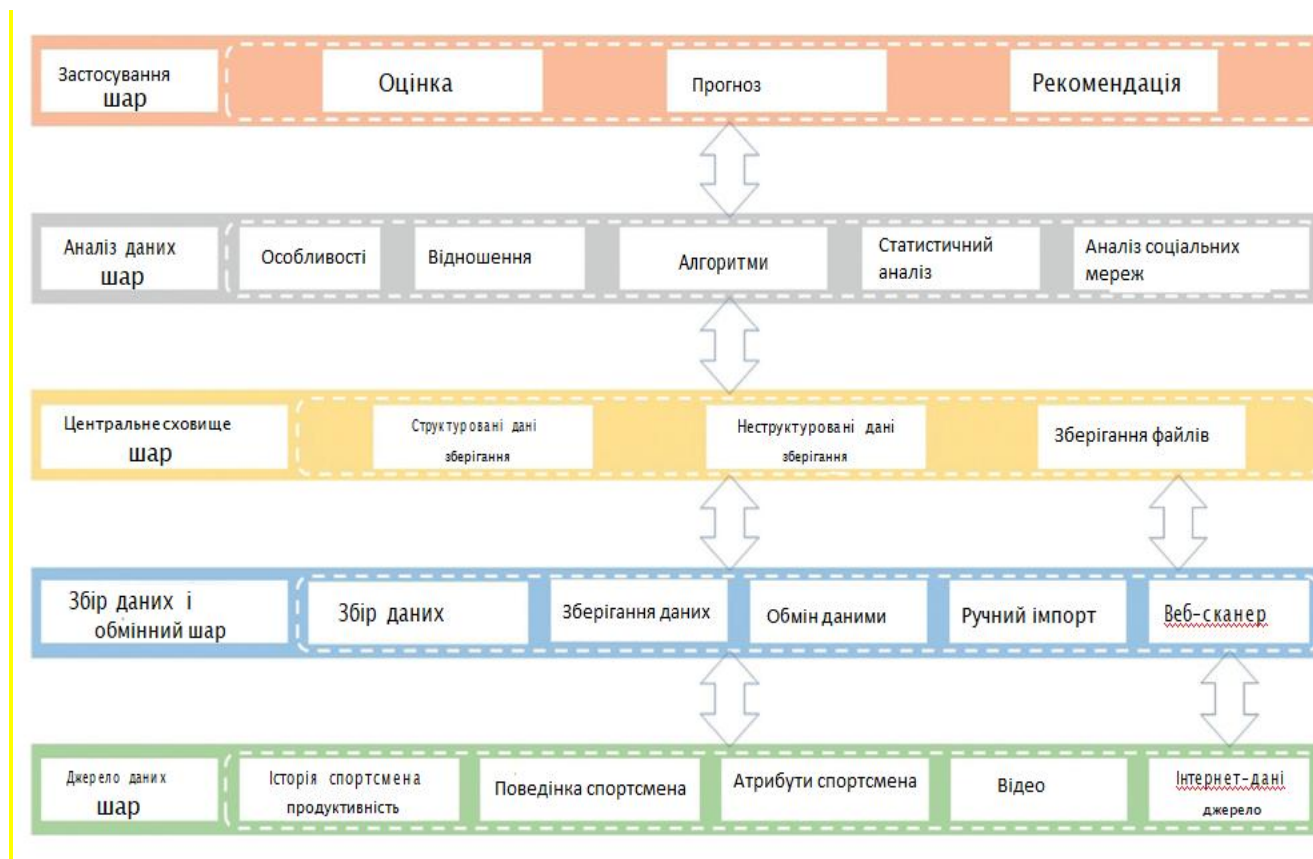


Рисунок 2.5 – Підсистема інформаційної системи керування даними спортсменів

Оцінка є важливим застосуванням спортивних великих даних. Оцінка в спортивних колах є цінним фактором оцінки результатів виступу спортсмена. Розглянемо фактори, що впливають на продуктивність гравців. Наприклад, Брукс та ін. витягує ознаку з кожного володіння м'ячем, щоб побудувати вектори ознак із початковим і кінцевим розташуванням у володінні. Щоб отримати вектор ознак для володіння, вони усереднюють вектори ознак за всіма завершеними передачами у володінні. Крім того, кожен вектор ознаки позначається відповідно до того, як закінчилося володіння. Паппалардо та ін. [55] використовують базу даних футбольних журналів, включаючи 31 496 332 події, 19 619 матчів, 296 клубів і 21 361 гравця. Кожна подія містить унікальний ідентифікатор події, тип події, мітку часу, гравця, пов'язаного з подією, команду гравця, матч, позицію на футбольному полі, підтип події та список тегів. Тип події складається з пасу, фолу, удару, дуелі, штрафного удару, офсайду та дотику. Тип фолу включає чотири особливості: фол без картки, фол жовтого кольору, фол червоного та фол

2-го жовтого. Щоб обчислити вектори продуктивності гравців, вони витягують 76 характеристик із футбольних журналів Wyscout для обчислення векторів продуктивності гравців. Лі та ін. [56] виділяють 22 функції, пов'язані з ефективністю атаки, передач і захисту на основі набору даних Китайської футбольної суперліги з 2014 по 2018 роки, щоб ранжувати команди.

Аналіз великих даних відноситься до техніки, яка може швидко отримати цінну інформацію з усіх типів даних [45]. Техніка аналізу великих даних може використовувати різні алгоритми для статистичного обчислення великих даних і вилучення важливих аналітичних даних для задоволення фактичних потреб. Наприклад, у змагальному спорті технологія аналізу великих даних може не тільки допомогти тренерам і спортсменам проаналізувати попередні тренування та спортивну поведінку на змаганнях, але також може визначити рух і фізичний стан спортсмена та скоригувати тренувальну діяльність спортсмена для покращення змагань. продуктивність. Крім того, технологія аналізу великих даних також може допомогти тренерам і спортсменам зрозуміти сильні та слабкі сторони своїх опонентів, щоб досягти відмінних результатів у масштабних змаганнях.

Статистичний аналіз. На основі статистичної теорії запропоновано методику статистичного аналізу, яка належить до розділу прикладної математики. Статистичний аналіз може надати висновки щодо великих даних. У дослідженнях спортивної галузі метод статистичного аналізу зазвичай використовується для обробки спортивних наборів даних. Грубий аналіз деяких статистичних характеристик наборів спортивних даних, таких як середнє значення, дисперсія, ентропія та максимальне/ мінімальне значення, дослідники можуть досліджувати модель руху спортсмена, і на основі цього статистичного аналізу тренери можуть розробити ефективний план тренувань [46].

В роботі [47] авторами запропоновано інструмент інтелектуального аналізу спортивних даних, який допоможе покращити аналіз техніки і тактики спортивних змагань [47]. У цьому дослідженні автор створює дві статистичні бази даних: одна – технічний набір даних, а інша – тактичний набір даних, включаючи

аркуші, пов'язані з інформацією про змагання з бадмінтону: команди, гравці, тренери, тип технічної дії та траєкторія бадмінтону. Наприклад, за допомогою аналізу статистичних даних технічних рухів, які використовує суперник у змаганнях, поведінку суперника можна заздалегідь оцінити, щоб скласти ефективний план відповіді. Базуючись на англійській Прем'єр-лізі, пропонується статистично значуща модель шляхом вимірювання ентропії передач м'яча для прогнозування конкурентної позиції команди [48]. Їхні експериментальні результати показують, що ентропія може краще визначити важливу роль захисників.

Незважаючи на те, що технологія статистичного аналізу відіграє важливу роль у дослідженні спортивних великих даних, з розвитком спортивної індустрії та технології великих даних все більше технологій, таких як машинне навчання, інтелектуальний аналіз даних та прогнозний аналіз, використовуються в дослідженні спортивних даних [49, 50]. Ці технології зазвичай спираються на спортивні соціальні мережі.

Моделі оцінювання на основі даних. Оцінка продуктивності гравців привернула інтерес наукової та спортивної спільноти завдяки наявності масивних спортивних даних [54–59]. Брукс та ін. [54] пропонують структуру рейтингу гравців на основі вартості виконаних передач, яка виходить із співвідношення місць передач під час володіння та згенерованих можливостей для кидка. Паппалардо та ін. [55] розробили керовану даними структуру, яка забезпечує принципову багатовимірну та рольову оцінку продуктивності футболіста. Базуючись на величезному наборі даних футбольних журналів і мільйонах матчів чотирьох сезонів 18 відомих футбольних змагань, вони порівнюють запропонований PlayeRank і відомі алгоритми, і результати показують, що PlayeRank перевершує конкурентів. Лі та ін. [56] використовують модель класифікатора лінійного опорного вектора для ранжування продуктивності команд. Їх експериментальні результати показують, що прогнозна точність запропонованої моделі на основі даних становить до 0,83, а продуктивність команд, які займають рейтинг, сильно корелює з їхнім фактичним рейтингом. Крім

того, рейтинги різних команд сильно корелюють з їхнім підсумковим рейтингом у лізі. Пелехрініс та ін. [57] пропонують алгоритм ранжування, заснований на аналізі команд відповідних ліг, які фіксують співвідношення вигрантів, і алгоритм PageRank. Результати показують, що цикли в мережі значною мірою пов'язані з продуктивністю. Ghosh та ін. [58] пропонують керований даними підхід для оцінки продуктивності гравця на основі стійки або пози гравця. У своїх експериментах вони використовують алгоритми поверхневого та глибокого навчання для класифікації ударів, які використовуються для аналізу стійки. Виходячи з цього, вони порівнюють позицію гравця середнього рівня або новачка з позицією професійного гравця. Крім того, вони дізнаються про помилку між професійним гравцем і учасником. Для оцінки їх наближення використовується сенсорна мережа. Лю та ін. [59] пропонують покращений метод оцінки продуктивності футболістів. У їх дослідженні використовується текстова інформація післяматчевих звітів, і результати свідчать про те, що запропонований метод є більш ефективним і обґрунтованим з точки зору оцінки гравця.

2.4 Висновки

На основі проведеного аналізу літературних джерел було запропоновано підсистему інформаційної системи керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків. Рівень джерела даних є основою для реалізації різноманітних додатків для аналізу та прогнозування спортивних великих даних. Наступний рівень – це рівень збору даних, який збирає дані з рівня джерела даних і виконує таку обробку: збір даних, зберігання даних, обмін даними, імпорт вручну та веб-сканер. Потім зібрані дані очищаються, а необхідна обробка виконується відповідно до різних вимог програми, де дані класифікуються та зберігаються. Оброблені дані зберігатимуться на рівні центрального сховища, включаючи сховище структурованих даних, сховище

неструктурованих даних і сховище файлів. Рівень аналізу даних виконує вибір функцій, аналіз зв'язків, статистичний аналіз і аналіз соціальних мереж відповідно до потреб конкретних програм з метою виявлення потенційних знань, законів і закономірностей у спортивних великих даних.

Представлена підсистема є основою для створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

3 МЕТОД СТВОРЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ КЕРУВАННЯ ДАНИМИ СПОРТСМЕНІВ

3.1 Формування вимог до інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Вимоги до інформаційних систем в спорті визначаються різноманітними аспектами, які охоплюють інтеграцію технологій у спортивні процеси та покращення продуктивності.

Основні вимоги можуть включати:

1. Ефективність та продуктивність. Забезпечення швидкості та надійності системи під час обробки та передачі даних. Оптимізація роботи для підвищення ефективності тренувань та спортивних подій.

2. Інтеграція зі смарт-технологіями. Сумісність з різноманітними сенсорами та пристроями для збору даних про фізичну активність, здоров'я, інтенсивність тренувань тощо.

3. Аналіз та звітність. Можливість проводити аналіз великого обсягу даних для вивчення результатів тренувань, покращення стратегій та прийняття інформованих рішень. Створення звітів та аналітичних матеріалів для тренерів, лікарів та спортсменів.

4. Безпека даних. Забезпечення конфіденційності та цілісності особистої та медичної інформації спортсменів. Використання заходів для уникнення несанкціонованого доступу до даних.

5. Мобільність. Розробка мобільних додатків для зручного моніторингу та обміну даними між тренерами та спортсменами.

6. Співпраця та комунікація. Можливість обміну даними та комунікації між різними членами тренувальної групи чи команди.

7. Підтримка навчання та розвитку. Інтеграція з системами навчання та розвитку для підвищення професійних навичок тренерів і спортсменів.

8. Сумісність зі стандартами. Дотримання стандартів та протоколів для спрощення інтеграції з іншими системами та обміном даними.

Наступні компоненти зосереджені на дослідженні, необхідному перед початком розробки такої системи, включаючи розуміння конкретних організаційних потреб і ресурсів, а також очікуваної інформації (релевантність та засоби), що надається.

Техніко-економічне обґрунтування. Аналіз здійсненності забезпечує структурований спосіб вивчення факторів і ризиків, що впливають на потенціал успішної розробки та впровадження інформаційної системи.

Доцільність може бути розглянута шляхом вивчення організаційного контексту та вимог до системи, а також шляхом аналізу системи, а також аналізуючи технічні, операційні та економічні обставин. Таким чином, успішне впровадження інформаційної системи керування даними спортсменів можна інтерпретувати як функцію від її передбачуваних витрат (операційних та економічних) та переваг. Вони можуть бути узагальнені на основі чотирьох аспектів.

Організаційний аспект. Розробник інформаційної системи керування даними спортсменів повинен враховувати організаційний контекст (попередні практики та процеси прийняття рішень), щоб зрозуміти, наскільки організація готова до нового впровадження. Ступінь відповідності запропонованої системи відповідає існуючій культурі та цілям організації, також є важливим. У деяких випадках розробнику також може знадобитися визначити, які зміни в організації будуть потрібні для сприяння впровадженню інформаційної системи. Наприклад, доцільність системи з метою інформування вибору гравців на драфті може бути підвищена, якщо буде доведено, що вона систематично покращує шанси на вибір «несподіваних» найкращих гравців. Ступінь життєздатності інформаційної системи в рамках поточної практики організації можна розглядати як рівень її «операційної сумісності».

Технічний аспект. Оцінка технічних вимог до інформаційної системи керування даними спортсменів з точки зору інформаційних систем (апаратне та

програмне забезпечення), а також з точки зору ефективності процесу (апаратне та програмне забезпечення), а також з точки зору ефективності процесу (час і ресурси, необхідні для збору даних, та час аналітичної обробки), є критично важливими, та ресурсів, необхідних для збору даних, і часу аналітичної обробки), є критично важливими для того, щоб визначити, чи буде система ефективно і задовільно вирішувати поставлені завдання. У цьому сенсі одна з найбільших технічних проблем полягає в тому, як зберігати дані та обмінюватися ними між відділами в межах однієї організації, або, за необхідності, між різними організаціями чи платформами. Це є ключовим моментом у спорті, де фронт-офіси можуть бути відокремлені як структурно, так і географічно від спортивних або тренувальних відділів. Щоб оптимізувати цей процес, наявність комплексного документообігу має першорядне значення. Хоча існує безліч варіантів, найпоширенішим прикладом є Cross Industry Standards Process for Data Mining (CRISP-DM). Як методологія, CRISPDM включає опис звичайних фаз проекту, завдань, пов'язаних з кожною фазою, а також пояснення взаємозв'язків між цими завданнями. Як модель процесу вона надає огляд життєвого циклу інтелектуального аналізу даних.

Модель життєвого циклу складається з шести фаз зі стрілками, що вказують на найбільш важливі і часті залежності між фазами. Послідовність фаз не є суворою. Насправді, більшість проектів переходять між фазами за необхідності (Рисунок 3.1).

Фінансовий. Стосується прогнозованих витрат на впровадження запропонованої інформаційної системи, а також того, чи виправдовують ці витрати потенційні вигоди.

Юридичний. Розробник повинен гарантувати, що заходи безпеки даних, права особи на приватність, а також процеси доступу до системи відповідають законодавству.

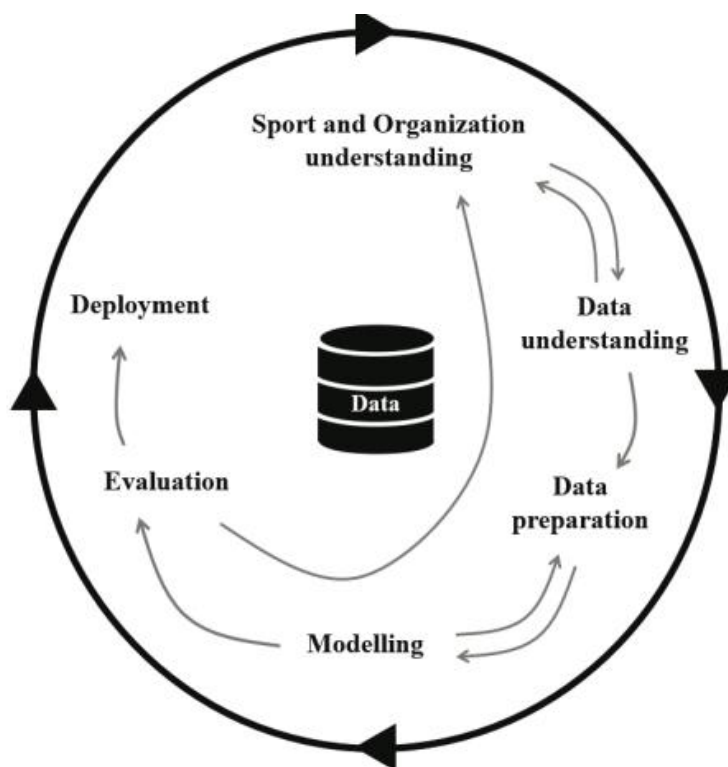


Рисунок 3.1 – Процес розробки міжгалузевих стандартів для інтелектуального аналізу даних (CRISP-DM) згідно Келлера, 2015

Під час аналізу даних спортивних результатів є три основні атрибути, які цікавлять спортивних вчених і професіоналів, а саме рейтинги, час і результати (RTS). Ці три показники представляють продуктивність і використовуються для оцінки спортсменів у більшості спортивних подій. Існуючий розрив між прикладним спортом і інтелектуальним аналізом даних можна вирішити за допомогою відображення, як показано на рисунку 3.2, від спортивної області до області інтелектуального аналізу даних. Таке відображення дозволяє вибрати відповідну техніку аналізу даних для спортивного питання. У той час як спортивна область включає основні правила, положення, тактику, стратегію, виступи, умови та здібності, пов'язані з конкретними видами спорту, область аналізу даних включає репрезентативні показники ефективності, а саме показники RTS. Методи попередньої обробки та аналізу даних, які можна використовувати в просторі інтелектуального аналізу даних, можуть інтерпретувати доступні дані лише у формі показників ефективності. Глибше розуміння (спортивної) сфери знань, а також краще розуміння доступних і відповідних інструментів

інтелектуального аналізу даних слугує для мінімізації проблем під час аналізу спортивних результатів, які можуть виникнути через недостатню точність, правильний підхід або обґрунтованість аналізу. Відображення, яке відбувається від спортивної області до області інтелектуального аналізу даних для аналізу продуктивності, слугуватиме відправною точкою для більш ефективного поєднання досліджень у галузі спортивної науки та інформатики. Поєднання досліджень між дисциплінами вимагає, щоб відображення мало здатність успадковувати необхідні знання (спортивної) області. Ці знання необхідні аналітикам даних не тільки для оцінки результатів свого аналізу, але й для підтвердження процесів, які вони виконують для підготовки даних і вилучення інформації з цих даних. Знання про спорт існують на різних рівнях, таких як формальні правила (наприклад, знання про рейтинги та підрахунок балів) і неформальні правила (наприклад, етичні питання) окремих видів спорту чи спортивної сфери загалом.

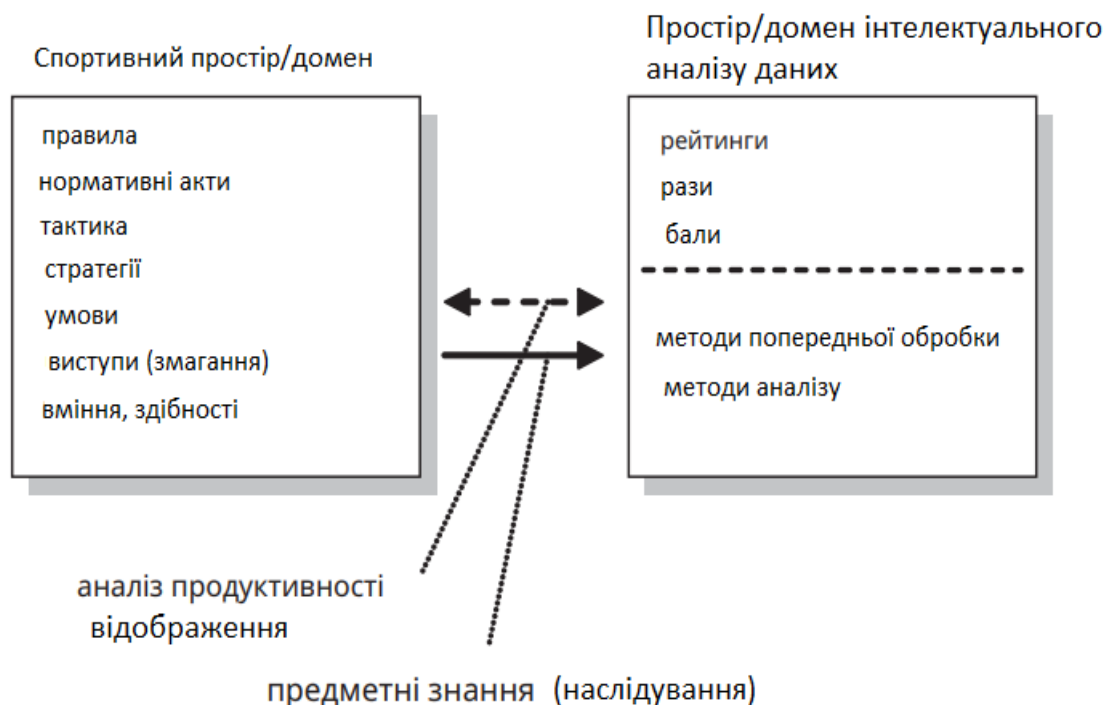


Рисунок 3.2 – Схема аналізу спортивних результатів, що включає спортивну область і область аналізу даних

Попередня обробка даних, пов'язаних зі спортом, у просторі інтелектуального аналізу даних передбачає підготовку та сортування даних для аналізу. Залежно від поточних і необхідних форматів даних і досліджуваної проблеми попередня обробка може складатися з одного або кількох із таких завдань.

1. Фільтрування. Фільтрування записів даних за певними категоріями змагань (наприклад, класифікація результатів веслування на швидкі, середні та повільні траси).

2. Перетворення формату. Перетворення даних у формат, який можна інтерпретувати спеціальним аналітичним програмним забезпеченням та інструментами, які використовуються для проведення фактичного аналізу даних (наприклад, перетворення часу гг:хх:сс у сукупні секунди).

3. Вилучення. Пошук нових даних, явно недоступних, на основі зібраних даних (наприклад, вилучення абсолютних і кумулятивних рейтингів човнів щодо різних секторів перегонів з веслування на 2000 м на основі часу човнів).

4. Структурне перетворення. Перетворення частин даних у формат, який забезпечує точніші аналітичні процеси даних (наприклад, узагальнення підсумкової таблиці в діапазоні від 1 до кількості учасників на призера [позиції 1–3] і не медаліста [позиції вище, ніж 3]).

5. Описове перетворення. Перетворення окремих частин даних у формат, який краще описує природу конкретного виду спорту/проблема (наприклад, перетворення абсолютного часу у відносний/ диференціальний час, який показує різницю в часі між лідером та іншими спортсменами).

Хоча перші три типи попередньої обробки даних є простими, різниця між останніми двома (структурним і описовим перетворенням) може бути делікатнішою та менш відчутною. У структурному перетворенні семантика даних не зміниться (наприклад, рейтинги до і після узагальнення все ще мають ту саму семантику), тоді як в описовому перетворенні значення нових перетворених даних не буде таким самим, як їх вихідна форма (наприклад, час проти різниці в часі або мітки часу та впорядковані позиції проти необроблених позицій).

Завдання попередньої обробки даних у фільтрації, перетворенні формату та вилученні не залежать від специфіки певного виду спорту.

Але структурне перетворення зазвичай залежить від:

- специфіки аналітичного методу даних, за допомогою якого деякі аналітичні методи даних можуть більш ефективно обробляти номінальні значення порівняно з числовими типами даних (наприклад, системи класифікації);

- обсягу даних, які доступні, завдяки чому часто невеликі обсяги даних, що стосуються конкретних параметрів, можуть вимагати узагальнення значень до більш грубих значень (наприклад, узагальнених місць на фініші, таких як категорії, визначені Офогі, Желєзніковим, МакМагоном і Дваєром для велоспорту на треку, які охоплюють більшу кількість записів даних).

Описове перетворення тісно пов'язане з розумінням виду спорту та його особливостей. Відповідна техніка попередньої обробки даних може бути обрана або керована комбінацією особливостей виду спорту, таких як кількість подій, кількість спортсменів, тривалість або критерії перемоги. У випадках, коли необхідно враховувати низку спортивних особливостей, доречним може бути поєднання завдань попередньої обробки.

Критерії перемоги в спорті демонструють свою дію з точки зору попередньої обробки даних, особливо описового перетворення, у спосіб збирання, підтримки та звітності даних під час і після великих подій. У видах спорту, де результати визначають переможця (наприклад, карате та дзюдо), у більшості випадків повідомляються лише результати. У деяких випадках, наприклад у хокеї, результати позначаються мітками часу. Інша статистика ігор (наприклад, форс/невимушені передачі та програш/ здобуті позиції) також повідомляється. Загалом, однак, якщо залежні від часу аспекти ігор під питанням, то певні дані потрібно перетворити або витягнути. У футболі, наприклад, якщо хтось хоче зрозуміти ймовірність перемоги в грі, враховуючи, що команда програвала в першому таймі, тоді аналітики повинні маркувати інші попередні ігри щодо результатів перших таймів, завдання, яке може не бути частиною вихідного збору даних.

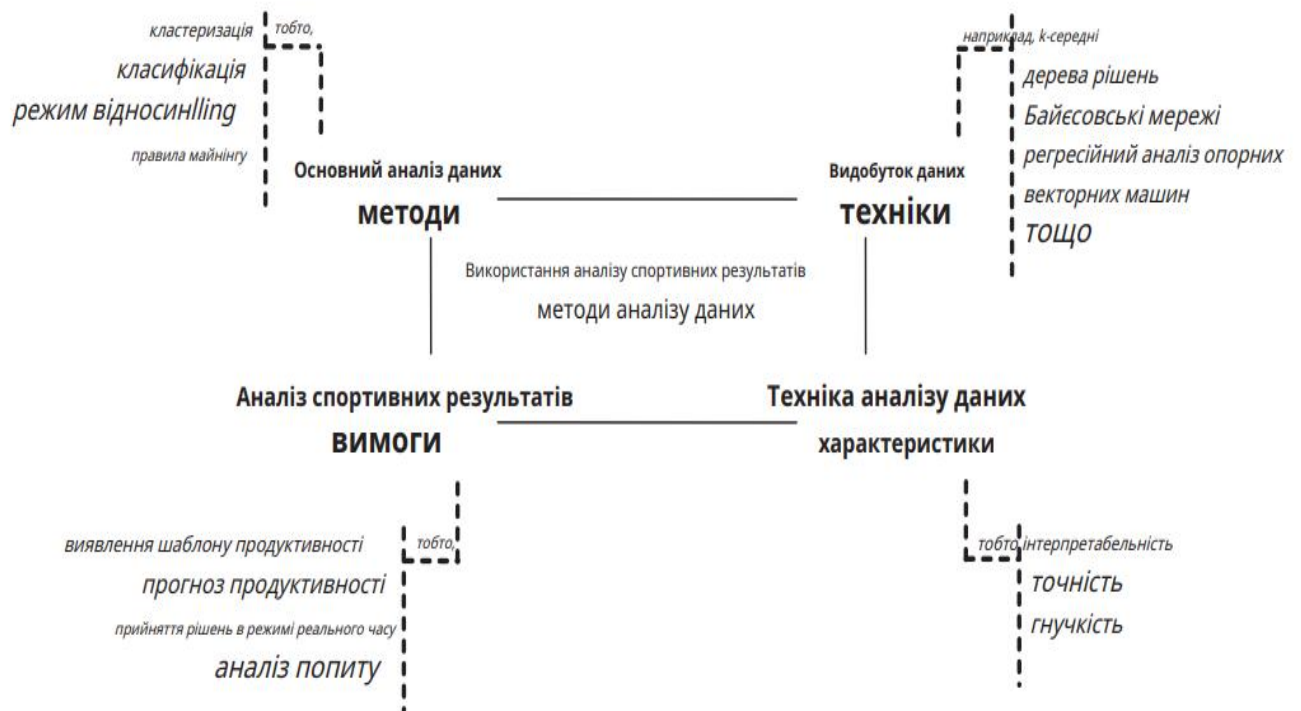


Рисунок 3.3 – Структурна модель, що характеризує підхід інтелектуального аналізу даних до аналізу спортивних результатів

У видах спорту, де час є критерієм перемоги, таких як марафон і біг на середні дистанції, вимірювання часу, що відповідають кожному спортсмену разом з його особистими характеристиками та деякими фізіологічними даними, в основному вимірюються та повідомляються. У деяких випадках це вимагає ретельного пошуку даних про спортсмена, які не повідомляються явно. Вони можуть включати місця народження та дати народження досліджуваних спортсменів. Попередня обробка необхідна для аналізу більш складних аспектів цих видів спорту з точки зору швидкості та/або відстань, пов'язану з часом. Прикладом такого дослідження є робота Джонса та Віппа (2002), яка розрахувала всі прив'язані до часу швидкості та відстані від шляхів, які пройшли досліджувані бігуни. Вони використовували сповільнене відтворення відео для золотих і срібних медалістів у забігах на 800 і 5000 метрів на Олімпійських іграх у Сіднеї, щоб визначити загальну дистанцію, яку вони подолали під час забігів. Потім вони використали офіційний час перегонів, щоб обчислити середню швидкість бігунів

на дистанціях, які вони подолали під час перегонів. Природа попередньо встановленої кількості очок, яких необхідно отримати в певних видах спорту зі змінним часом, наприклад у волейболі та тенісі, робить кінцеві результати, необхідні для перемоги, менш важливими порівняно з іншими видами спорту. У цьому випадку акцент робиться на заходах під час гри, таких як дані, пов'язані з виступами окремих спортсменів, індивідуальними сценаріями ігор або тактикою, яка може призвести до успіху. Деякі з цих даних частково вимірюються та повідомляються під час великих змагань. Деякі дані можна отримати лише з наявних і звітних даних. З точки зору волейболу, аналіз ймовірності перемоги в грі (враховуючи перемогу в певних сетах, наприклад, у перших двох сетах) вимагає позначення гри щодо виступів команд у першому та другому сетах, а також результат у фіналі. Аналіз можливостей певних тенісистів-чоловіків перемагати в тенісних іграх турніру Великого Шолома, наприклад, коли вони відстають на два сету, також вимагає попередньої обробки наявних даних щодо результатів кожного сету в їхніх іграх. Це можна зробити, позначивши їхні ігри кількістю сетів, які вони програли (0, 1 або 2), а також їхнім кінцевим результатом (виграш, програш). Індивідуальні результати в кожному сеті не мають значення, значення має лише те, хто виграв сет. У багаторазових видах спорту, де повідомляються лише бали за кожну спробу та остаточні бали, будь-яка обробка даних може залежати від конкретної досліджуваної проблеми. Подібно до видів спорту з попередньо встановленою кількістю очок (наприклад, теніс і волейбол), у видах спорту з кількома зусиллями (наприклад, стрибки у воду) може знадобитися вилучення вимірювань під час гри з наявних даних перед аналізом даних. проведено. Аналіз ймовірності завершення події на першому місці враховуючи те, що спортсмен не показав найкращих результатів у своїй першій спробі, є прикладом такого аналізу, який потребує додаткового тегування даних щодо перших спроб і підсумкової позиції спортсменів.

На аналіз даних, який надходить після попередньої обробки даних, більшою мірою впливає тип проблеми, яка розглядається під час аналізу спортивних результатів. Проблеми виявлення шаблонів продуктивності, прогнозування

продуктивності, прийняття рішень у реальному часі та аналізу попиту часто краще виконувати за допомогою різних методів інтелектуального аналізу даних, що вимагають різних заходів щодо інтерпретації, точності та гнучкості. Щоб охопити всі аспекти аналізу спортивних результатів у загальній структурі, було представлено структуру, яка об'єднує вимоги до аналізу результатів, методи аналізу даних, методи аналізу даних і характеристики техніки. Ця взаємопов'язана структура вимагає достатньої уваги перед виконанням практичних і корисних завдань аналізу ефективності. Відображення показують, які методи та прийоми інтелектуального аналізу даних підходять для тих чи інших проблем аналізу спортивних результатів. Ця структурна модель підтверджує, що інтелектуальний аналіз даних у порівнянні з альтернативними інструментами може допомогти спортсменам і тренерам працювати краще. Крім того, адаптація та розробка інструментів інтелектуального аналізу даних, які відповідають конкретним потребам сфери спорту, дозволяють швидко та точно оцінювати попередню, поточну та майбутню поведінку. Ця діяльність потребує мереж досліджень і практики, що дозволяє поєднувати знання про спорт і інтелектуальний аналіз даних, дотримуючись етичних правил, багаторічних архівів і спілкування між зацікавленими сторонами.

3.2 Метод створення інформаційної системи керування даними спортсменів формування

Метод розробки інформаційної системи (ІС) - це систематичний підхід до створення, розширення або підтримки інформаційних систем. Цей метод визначає процеси, етапи, методи та інструменти, які використовуються під час розробки системи з метою досягнення поставлених цілей.

Створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних включає кілька ключових етапів. Запропоновано метод, який можна використовувати для цієї мети, який складається з наступних етапів.

Етап 1. Визначення вимог. Необхідно визначити функціональні та технічні вимоги до системи.

Етап 2. Збір та обробка даних. На цьому етапі необхідно визначити типи, формати та кількість даних, які будуть збиратися та аналізуватися. А також необхідно забезпечити механізми для збору даних про спортсменів, тренування, результати змагань та інші важливі параметри.

Етап 3. Очищення та підготовка даних. На цьому етапі потрібно провести очистку та підготовку даних, включаючи обробку відсутніх значень, видалення дублікатів та нормалізацію даних.

Етап 4. Застосування інтелектуального аналізу даних. Цей етап в свою чергу складається з таких підпунктів, як вибір методів аналізу (наприклад, машинне навчання, статистичний аналіз, класифікація тощо), розробка моделей для прогнозування та аналізу результатів на основі вибраних методів; навчання моделей, де потрібно використати існуючі дані для тренування моделей.

Етап 5. Розробка інформаційної системи, де потрібно розробити архітектуру інформаційної системи, враховуючи розподілену базу даних, модулі аналізу та інтеграцію з іншими системами. При цьому необхідно створити структуру бази даних, що враховує потреби системи та методів аналізу даних та реалізувати модулі для збору, зберігання та аналізу даних.

Етап 6. Тестування та вдосконалення. На етапі тестування проводиться ряд тестів для перевірки функціональності та надійності системи. А також досить важливий момент – проведення оцінки точності моделей інтелектуального аналізу даних. Також при необхідності вносяться виправлення.

Етап 7. Впровадження та підтримка. На цьому етапі виконується розгортання інформаційної системи та моделі аналізу даних в реальному середовищі. Також на цьому етапі здійснюється навчання користувачів якщо потрібно, щодо використання нової системи та аналізу результатів. Підтримка та оновлення має бути забезпечене постійно на основі змін в потребах.

Цей метод дозволяє поєднати ефективність інформаційної системи з потужністю інтелектуального аналізу даних для покращення керування даними

спортсменів у системі організації змагань, який на відміну від відомих базується на інтелектуальному аналізі даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо.

Метою інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є підвищення ефективності роботи тренерів для спостереження та відбору спортсменів на змагання. Завданнями інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є збір, обробка, зберігання, опрацювання даних спортсменів та на основі цих даних прогнозування результативності.

Головною частиною інформаційної системи керування даними спортсменів є підсистема прогнозування результативності спортсмена представлена на рисунку 3.4.

Підсистема прогнозування складається з трьох частин: входу, моделі та виходу. У вхідній частині отримуються деякі необроблені спортивні дані, такі як результати спортсмена, його тренування та поведінка на змаганнях, вроджені атрибути спортсмена, психологічний стан спортсмена. Модельна частина включає два етапи: тренування та тестування. Зазвичай, для навчання прогностичної моделі використовуються деякі важливі характеристики прогностичної моделі, в тому числі історичні результати спортсмена результати, соціальна поведінка спортсмена, стать спортсмена та інші особливості. На основі цих важливих характеристик використовуються деякі алгоритми машинного алгоритми машинного навчання використовуються для навчання моделі, включаючи нейронну мережу, машину опорних векторів глибоке навчання. Після навчання, нові дані і вибрані функції використовуються для тестування і прогнозування результатів спортсмена.

Фактори, що впливають на результати спортсменів. Прогностична модель виступу спортсмена враховує більшість факторів, які впливають на зміну його/її результатів. Ці важливі фактори поділяються на наступні категорії:

- 1) спортивні історичні результати;

- 2) спортивна поведінка та атрибути спортивної поведінки, включаючи процес руху спортсмена, ігрову продуктивність і тренувальну продуктивність;
- 3) вроджені атрибути, включаючи статичні атрибути статі, сім'ї та освіти спортсмена освіти;
- 4) психологічний стан здоров'я.

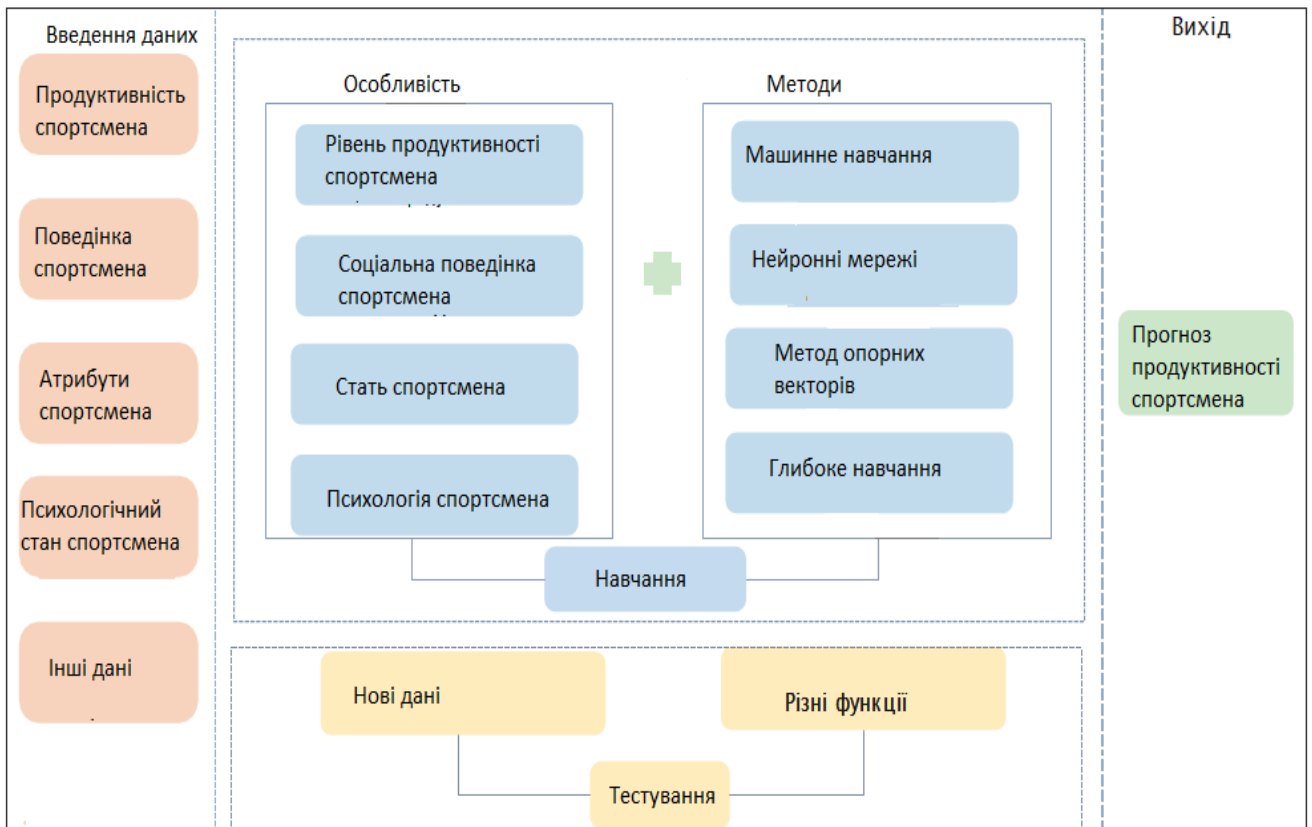


Рисунок 3.4 – Підсистема прогнозування результативності спортсмена

Деякі дослідники використовують ознаки, пов'язані з матчами, та зовнішні ознаки, такі як результати історичних матчів, показники ефективності гравців інформацію про суперника, нещодавню форму та гравців, доступних для гравця, доступного для матчу. Constantinou та Fenton [60] використовують такі особливості: змагання, кваліфікація до змагань, стрес і втома команди, стрес і втома команди, і командний стрес і втома, матчі, досвід участі в змаганнях, управлінські зміни (новий менеджер, довголіття попереднього, управлінська

нестабільність попереднього, нестабільність керівництва та управлінські здібності), рівень травматизму, профіль гравця (команда), здатність справлятися з травмами, здатність команди справлятися з травмами, очки ліги та різниця в очках ліги між сезонами, заробітна плата команди, заробітна плата суперників, середня заробітна плата команди, середня зарплата команди суперників, заробітна плата команди у % різниці від середньої заробітної плати суперника, та різниця в зарплатах команд у % від середньої зарплати супротивника та чисті трансферні витрати для прогнозування точності довгострокових результатів футбольної команди. Частина науковців прогнозують результати ігор НБА, використовуючи впливові такі характеристики, як кількість підбирань у захисті, відсоток триочкових кидків, відсоток штрафних кидків та загальну кількість підбирань, що може підвищити точність прогнозованої моделі. Лі та ін. [29] пропонують метод прогнозування результатів спортивної команди, використовуючи історичні дані гравців та команди, включаючи кількість ігор, ігровий час, двоочкові та триочкові кидки, штрафні, штрафні кидки, підбирання в захисті, підбирання в нападі, передачі, перехоплення, блок-шоти, втрати та персональні фоли. Для реалізації прогнозованої моделі використовуються вищезгадані функції та методи машинного навчання.

Оскільки обсяг структурованих і неструктурованих даних у спортивній індустрії, прогнозування результатів спортсменів є важливим застосуванням спортивних великих даних, а технологія машинного навчання зазвичай використовується для прогнозування результатів спортсменів. Однак прогнозування результатів спортивних змагань є складним завданням. Щоб дізнатися про навички команд, автори в роботі [62] пропонують імовірнісну графічну модель, і на основі цієї моделі вони розкладають відносні ваги везіння та майстерності в кожній грі. Їхні експериментальні результати показують, що везіння суттєво присутнє в найбільш чемпіонатах і дають пояснення, чому складні функціональні моделі, навряд чи виграють у простих моделей для прогнозування результатів спортивних змагань. Бункер та інші пропонують прогностичну структуру для прогнозування спортивних результатів за допомогою штучної

нейронної мережі. Точність прогнозування отриманої ними моделі є вищою, ніж результати прогнозування традиційних математичних і статистичних моделей. Constantinou та Fenton покращують точність прогнозування довгострокових результатів футбольної команди. Їхня модель може передбачити загальну кількість очок у лізі, яку команда, як очікується, здобуде протягом сезону. У їхній моделі використовується байєсівська мережева модель. Для прогнозування результатів гри НБА пропонують інтелектуальний фреймворк машинного навчання, фреймворк наївного Байєса, штучну нейронну мережу та алгоритми дерева рішень. Лі та ін. [29] використовують багатовимірний логістичний регресійний аналіз для виявлення зв'язку між ймовірністю виграшу та результатами гри на командному рівні. Їхні експериментальні результати свідчать про ефективність моделі на основі даних Національної баскетбольної асоціації та команди "Голден Стейт Ворріорс". Чжу та Сунь [57] прогнозують результати спортсменів, використовуючи допоміжний вектор машинного алгоритму. Їхні експериментальні результати показують що, порівняно з поточною моделлю прогнозування результатів спортсменів, точність прогнозування їхня модель має більш високу точність прогнозування, ніж поточна надійніша модель.

Хоча дослідники досягли безпрецедентних результатів у прогнозуванні результатів спортсменів, більшість їхніх прогнозів моделі зосереджені на вилученні ознак та машинному навчанні. Проблема полягає в тому, що взаємозв'язок знань між результатами спортсменів та спортсменами, тренерами та подіями ігнорується. Існуючі дослідники приділяють більше уваги статистичним зв'язкам між ними. Як точніше прогнозувати результати спортсменів? Одним з можливих рішень є прогнозування результатів спортсменів на основі спортсменів на основі графа знань зі спортивних великих даних.

Отже, як побудувати граф знань про спортивні спортивних результатів та сутностей, пов'язаних з ними, є важливим завданням. Крім того, як використовувати побудований граф знань зі спортивних великих даних для прогнозування результатів спортсменів – дуже складне завдання.

Успіх спортивної кар'єри залежить не лише від особистих здібностей спортсмена, але також пов'язаний з командою та країною спортсмена. Для команди або країни, щоб виховати видатного спортсмена, потрібно багато людських і матеріальних ресурсів. Висхідна зірка спорту відноситься до спортсменів, які не є видатними серед спортсменів, які не є видатними серед однолітків, і знаходяться на початковому етапі своєї спортивної кар'єри, але вони мають тенденцію стати зірками спорту в майбутньому. Пошук висхідних зірок спорту не тільки забезпечує конструктивні рекомендації щодо інвестування національних коштів, але й надає необхідну допомогу спортсменам, щоб вони могли показати відмінні результати раніше. Однак, мало що відомо про те як знайти висхідну зірку спорту. Сучасні дослідження здебільшого використовують статистичні методи для оцінки спортсменів. Побудова графу знань для пошуку висхідної зірки спорту є складним завданням.

У традиційній спортивній системі різні спортивні установи створюють незалежні платформи спортивних даних відповідно до потреб своїх клубів або команд, що призводить до утворення "острівців даних". Інтеграція різних спортивних систем є необхідною і важливо інтегрувати різні спортивні системи та створити єдину велику сервісну платформу для спортивних великих даних. На основі цієї платформи дослідники можуть аналізувати різні дослідники можуть аналізувати різні взаємозв'язки між спортивними суб'єктами. Крім того, за допомогою платформи тренери, спортсмени та команди можуть отримувати цінну інформацію про ігри. На основі цієї платформи можна створити багатовимірний можна скласти багатовимірний портрет спортсменів, тренерів, команд і країн, а також надавати точні послуги спортсменів, тренерів, команд і країн, а також надавати їм точні послуги, такі як рекомендувати тренерів і клуби та виявляти висхідних висхідних зірок спорту.

Великі спортивні дані привертають увагу вітчизняних і зарубіжних дослідників та розбудовників спортивної індустрії. Великі спортивні дані не тільки надають любителям спорту різноманітні точні сервіси, але й допомагають керівникам спортивних організацій розподіляти кошти на спортивних керівників,

які приймають рішення, щодо розподілу коштів для покращення спортивних рішень, щодо виділення коштів на покращення результатів спортсменів. Однак, існує дуже мало спортивних великих даних для дослідників, на відміну від наукових даних ресурсів, таких як Google Scholar, Mendeley та Web of Science. Крім того, хоча різні спортивні установи та клуби мають різні спортивні дані, ці дані переважно ізольовані. Тому великі спортивні дані відкриті для дослідників, щоб допомогти та сприяти активному розвитку спорту та досягти мети надання зручного, ефективного та точного сервісу для спортсменів та любителів спорту.

В епоху великих даних, в той час як спортивні великі дані приносять велику цінність, вони також створюють певні проблеми в захисті приватності спортсменів. Питання захисту приватності спортсменів і запобігти витoku конфіденційної інформації в процесі розробки та застосування спортивних великих даних стало новим викликом. З одного боку, особисте життя спортсменів спортсменів вимагає від міжнародних спортивних організацій створення незалежної агенції із захисту приватності; з іншого боку, необхідно створити спеціальну систему конфіденційності, метою якої є забезпечення пріоритету недоторканності приватного життя спортсменів. Для захисту приватності спортсменів необхідно впровадити ретельний контроль з боку органів влади та співпрацювати з відповідними стратегіями десенсибілізації даних для кращого захисту конфіденційності спортсменів.

3.3 Висновки

Було розглянуто особливості роботи з спортивними даними, керування спортивними даними, методах аналізу спортивних великих даних та застосунках спортивних великих даних. У сфері великих даних у спорті відбулося кілька змін: від простого статистичного оцінювання до опрацювання спортивних великих даних; від простого статистичного оцінювання до оцінювання на основі моделей; від простого статистичного аналізу до прогнозування результатів спортсменів на

основі даних; від аналізу соціальних мереж до аналізу графів знань; від явних спортивних особливостей до неявних спортивних особливостей.

Запропоновано метод створення інформаційної системи керування даними спортсменів в системі організації змагань, що дозволяє поєднати ефективність інформаційної системи з потужністю інтелектуального аналізу даних для покращення керування даними спортсменів у системі організації змагань, який на відміну від відомих базується на інтелектуальному аналізі даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо. Також було проаналізовано та сформовано вимоги до інформаційної системи керування даними спортсменів та розроблено і представлено підсистему прогнозування результативності спортсмена, яка є основною частиною. Запропонованої підсистеми.

Завданнями інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є збір, обробка, зберігання, опрацювання даних спортсменів та на основі цих даних прогнозування результативності.

4 ІНФОРМАЦІЙНА СИСТЕМА КЕРУВАННЯ ДАНИМИ СПОРТСМЕНІВ В СИСТЕМІ ОРГАНІЗАЦІЇ ЗМАГАНЬ НА ОСНОВІ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

4.1 Розробка архітектури інформаційної системи керування даними спортсменів

Архітектура інформаційної системи керування даними спортсменами повинна бути розроблена з урахуванням специфіки вимог і потреб в спортивній галузі. Загальний огляд архітектурних складових такої системи:

Фронтенд (користувацький інтерфейс). Розробка інтуїтивного та зручного в користуванні інтерфейсу для введення та відображення даних спортсменів. Це може включати в себе веб-інтерфейс для тренерів та апаратів, а також мобільний додаток для спортсменів для введення і відстеження особистої інформації.

Бекенд (серверна складова). Реалізація серверної частини для обробки запитань користувачів, зберігання та управління базою даних спортсменів. Важливо враховувати швидкодію та масштабованість системи.

База даних. Створення ефективної бази даних для зберігання різноманітної інформації про спортсменів, такої як біографічні дані, медична інформація, результати змагань, статистика тренувань тощо.

Системи аналізу даних та інтелектуального аналізу. Інтеграція інтелектуальних засобів аналізу даних для виявлення тенденцій, прогнозування результатів та надання рекомендацій тренерам для оптимізації тренувального процесу.

Модуль аутентифікації та безпеки. Забезпечення безпеки даних спортсменів за допомогою надійних методів аутентифікації, а також контроль доступу до конфіденційної інформації.

Інтеграція з зовнішніми джерелами даних. Можливість інтеграції з зовнішніми системами, такими як медичні бази даних, системи вимірювання показників фізичного стану, платформи для взаємодії з фанатами та інші.

Модуль запитань та звітності. Розробка системи для створення запитань та генерації звітів для тренерів, менеджерів та інших зацікавлених сторін.

Модуль сповіщень. Впровадження системи сповіщень для тренерів та спортсменів для швидкого сповіщення про важливі події та зміни.

Модуль аналізу продуктивності. Розробка функціоналу для аналізу ефективності спортсменів під час тренувань та змагань.

Модуль резервного копіювання та відновлення. Забезпечення можливості регулярного резервного копіювання та відновлення даних для запобігання втратам інформації.

Це лише загальний огляд можливих компонентів архітектури інформаційної системи керування даними спортсменами. Фактична архітектура буде залежати від конкретних вимог та специфікацій проекту.

В роботі представлено базову архітектуру системи для керування та підтримки всієї діяльності, пов'язаної зі спортивними подіями, в одній онлайн-системі, де весь центр керування має логіни, як-от спортивні координатори, працівники спортивних офісів, а також спортсмени також мають свої логіни для керування та перегляду оновлень власної спортивної кар'єри (рисунок 4.1).

Уся інформація, пов'язана з усіма подіями, які проводять відеосесії експертів, і все це можна проводити в цій системі, оскільки ми можемо керувати всім з одного облікового запису адміністратора та можемо відобразитися для кожного району, а також для входу користувачів. Усі дати, результати та функції спортивних подій також керуються цим веб-сайтом. Управління всіма звітами, аудитами та роботою комітетів з усіма резервними копіями та обслуговуванням даних здійснюватиметься за допомогою цієї системи. Інформаційна система класифікує своїх відвідувачів за різними категоріями та надає їм відповідні функції відповідно до рівня доступу. Новий користувач може зареєструватися в системі.

Усі новини, пов'язані з усіма подіями, відеосесіями, лекціями експертів, все це можна використовувати в цій системі, оскільки є можливість керувати всім з одного облікового запису адміністратора, а також є вхід для користувачів. Усі

дати спортивних подій, результати та функції також керуються можуть керуватися даною системою. Управління всіма звітами, аудитами та роботою комітетів з резервним копіюванням даних і обслуговуванням здійснюватиметься за допомогою цієї системи.

Проект переносить весь ручний процес керування спортивними подіями в інтернеті, створений за допомогою Android як інтерфейсу та MySQL Server як бекенда. Основною метою цього проекту є спрощення процесу обробки кожної спортивної події шляхом надання веб-інтерфейсу для адміністратора.

Підсистема адміністрування складається з кількох модулів, щоб ініціювати спортивну подію шляхом додавання типу спорту (у приміщенні чи на відкритому повітрі), додавання користувачів, які зацікавлені в певній спортивній діяльності, додавання менеджерів, які будуть проводити конкретну спортивну діяльність, яка призначена сам адмін і, нарешті, перегляд результатів спортивної події, що проводиться в клубі. Адміністратор виконує різноманітні завдання, як-от перевірку відвідуваності спортсменів, зареєстрованих на певну спортивну подію, перегляд списку гравців для позначення переможця кожного раунду, генерацію результатів на основі кількох виграних спортсменом раундів.

4.2 Побудова UML-діаграм інформаційної системи керування даними спортсменів

Для інформаційної системи керування даними спортсменів у системі організації змагань можна використовувати різні види UML-діаграм. Ось декілька можливих типів діаграм та їх короткий опис.

Use Case Diagram (Діаграма випадкового аналізу). Визначає функціональність системи з точки зору користувача (наприклад, адміністратора, тренера, спортсмена). Включає випадки використання, такі як реєстрація спортсмена, планування та ведення тренувань, організація змагань тощо.

Class Diagram (Діаграма класів). Відображає основні класи та їх атрибути в системі. Може включати класи, такі як «Спортсмен», «Тренер», «Змагання», «Система керування даними» тощо.

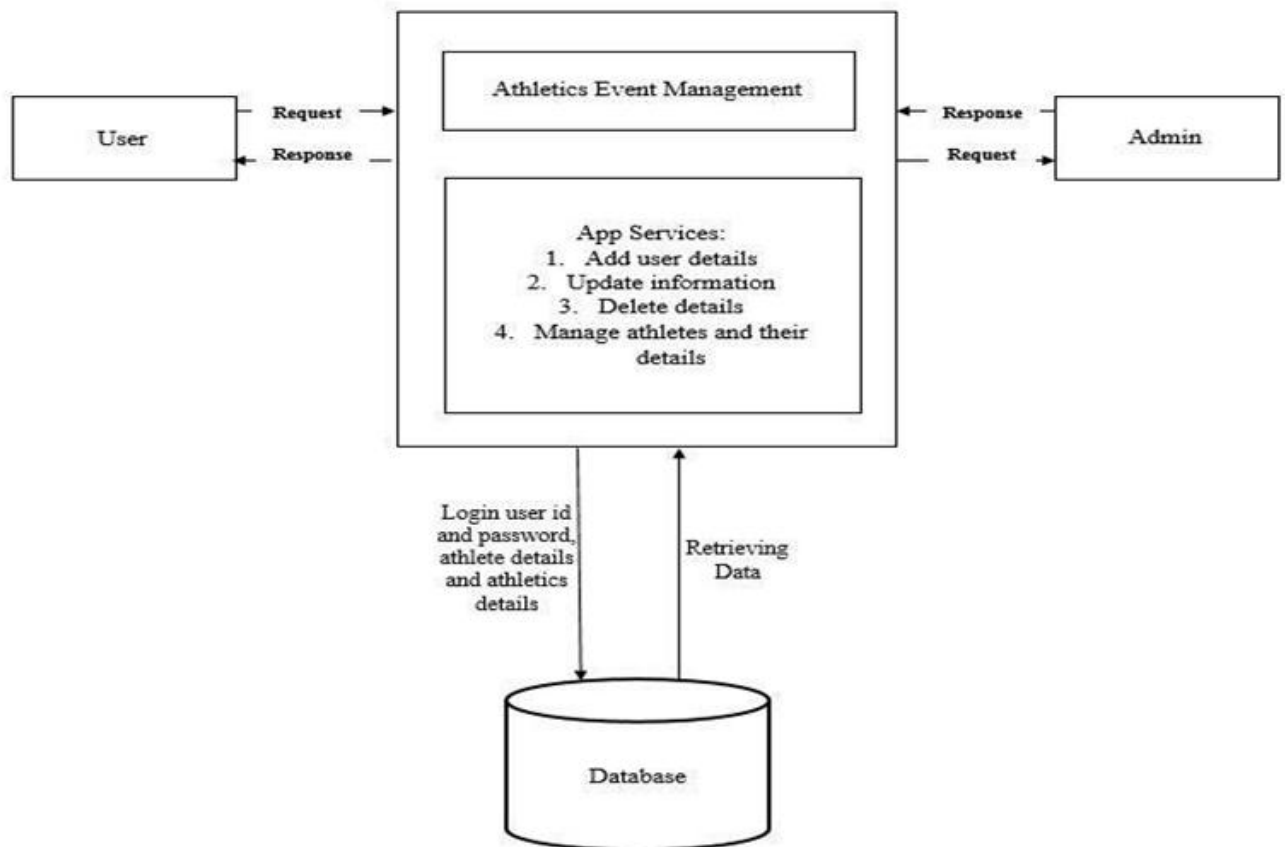


Рисунок 4.1 – Базова архітектура інформаційної системи керування даними спортсменів

Activity Diagram (Діаграма активності). Показує послідовність та взаємодію процесів у системі. Може включати активності, такі як «Реєстрація спортсмена», «Планування тренувань», «Організація змагань» і т. д.

Sequence Diagram (Діаграма послідовності). Визначає взаємодію об'єктів та компонентів системи в часі. Показує послідовність повідомлень між об'єктами, наприклад, процес реєстрації спортсмена чи організації змагань.

Component Diagram (Діаграма компонентів). Показує архітектуру системи та компоненти, з яких вона складається. Може включати компоненти, такі як «База даних спортсменів», «Модуль тренувань», «Модуль змагань» тощо.

Deployment Diagram (Діаграма розгортання). Визначає фізичне розташування компонентів системи та їх зв'язок. Може включати сервери, бази даних, апаратне забезпечення для забезпечення функціональності.

Ці діаграми можуть використовуватися для моделювання та розробки інформаційної системи керування даними спортсменів в системі організації змагань, щоб забезпечити зрозумілість та ефективність розробки проекту.

Діаграма випадкового аналізу (Use Case Diagram)

Головні актори: Адміністратор, Тренер, Спортсмен.

Випадки використання: Реєстрація спортсмена, Планування тренувань, Організація змагань.

Діаграма класів (Class Diagram). Класи: Спортсмен, Тренер, Змагання, Система керування даними. Атрибути: Ім'я, Вік, Результати тренувань тощо.

Діаграма активності (Activity Diagram).

Активності: Реєстрація спортсмена, Планування тренувань, Організація змагань.

Діаграма послідовності (Sequence Diagram): Об'єкти: Спортсмен, Тренер, Система керування даними. Послідовність повідомлень: Реєстрація, Введення результатів, Запит на план тренувань тощо.

Діаграма компонентів (Component Diagram). Компоненти: База даних спортсменів, Модуль тренувань, Модуль змагань.

Діаграма розгортання (Deployment Diagram):

Вузли: Сервер, База даних, Клієнтські додатки.

На рисунку 4.2 представлено діаграму збору даних з тренування (стан «як є»).

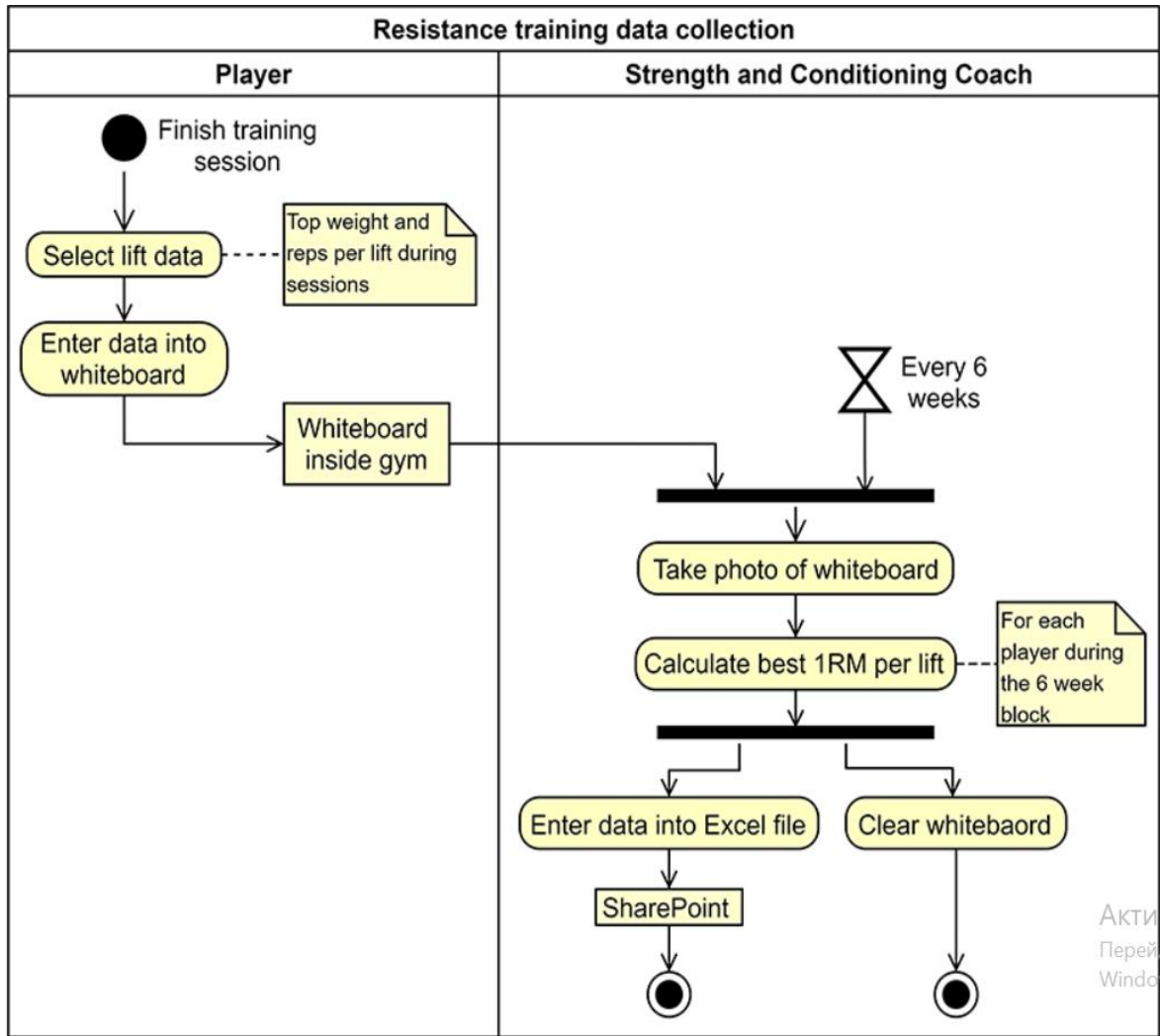


Рисунок 4.2 – Діаграма збору даних з тренування (стан «як є»)

Ілюстрації результатів спроб впровадження системи високого рівня, для оптимізації потоку інформації. У зв'язку з цим, відповідно до методів, цифрова система була реалізована з використанням двох (внутрішнього та зовнішнього) циклів побудови-вимірювання навчання.

Ця модель методології намагається вирішити проблема великої затримки між фазою аналізу та доставкою системи. Замість послідовного проектування та впровадження він виконує загальний проект для всієї системи, а потім розділяє проект на серію окремих підпроектів, які можна розробляти та впроваджувати паралельно. Після завершення всіх підпроектів відбувається остаточна інтеграція окремих частин і система поставляється.

Діаграма варіантів використання представлена на рисунку 4.3.

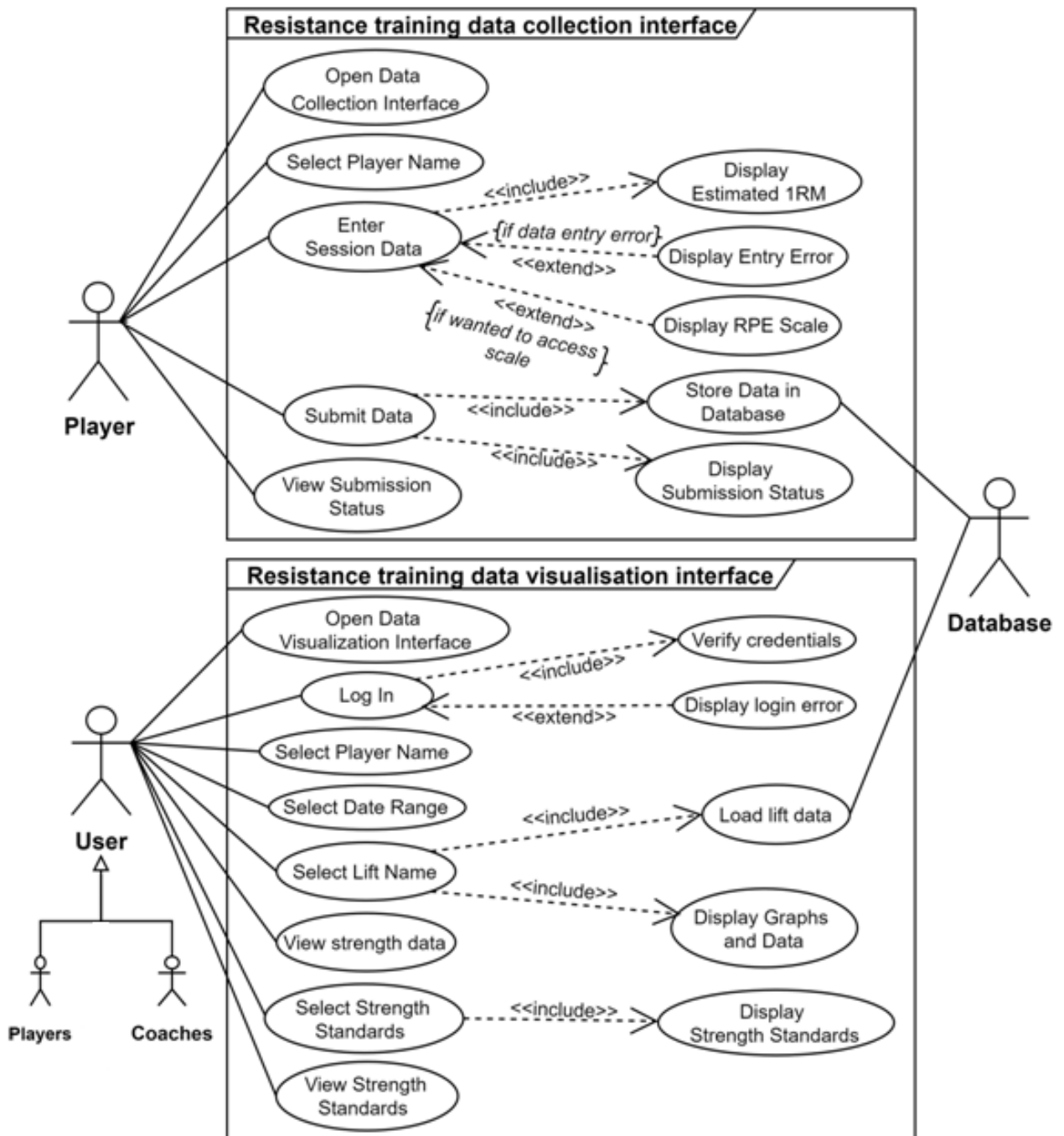


Рисунок 4. 3 – Діаграма варіантів використання

Android Studio – це офіційне інтегроване середовище розробки (IDE) для операційної системи Google Android, створене на основі програмного забезпечення IntelliJ IDEA від JetBrains і розроблене спеціально для розробки Android. Він доступний для завантаження в операційних системах Windows, macOS і Linux. Це заміна Eclipse Android Development Tools (ADT) як основної

IDE для розробки програм Android. Android Studio підтримує ті самі мови програмування, що й IntelliJ (і Clion), напр. Java, C++ та інші з розширеннями, як от Go та Android Studio 3.0 або новішої версії, підтримують Kotlin і всі функції мови Java 7, а також підмножину функцій мови Java 8, які відрізняються залежно від версії платформи.

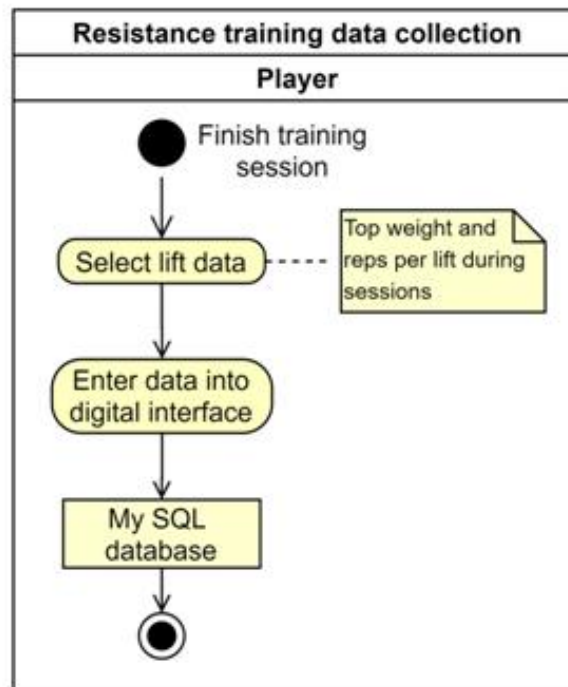


Рисунок 4.4 – Процес збору даних тренування (стан «як має бути»).

Зовнішні проекти портують деякі функції Java 9. Хоча IntelliJ, на якому побудовано Android Studio, підтримує всі випущені версії Java та Java 12, незрозуміло, до якого рівня Android Studio підтримує версії Java до Java 12 (у документації згадується часткова підтримка Java 8). Принаймні деякі нові мовні функції до Java 12 можна використовувати в Android. MySQL є безкоштовним програмним забезпеченням із відкритим вихідним кодом згідно з умовами GNU General Public License, а також доступним за різними пропріетарними ліцензіями. MySQL корпорація Oracle).

У 2010 році, коли Oracle придбала Sun, Widenius розгалужив проект MySQL з відкритим кодом, щоб створити MariaDB. MySQL є компонентом програмного стека веб-додатків LAMP (та інших), що є акронімом для Linux, Apache, MySQL, Perl/PHP/Python. MySQL використовується багатьма веб-додатками, керованими базами даних, включаючи Drupal, Joomla, phpBB і WordPress. MySQL також використовується багатьма популярними веб-сайтами, включаючи Facebook, Flickr, MediaWiki, Twitter і YouTube.

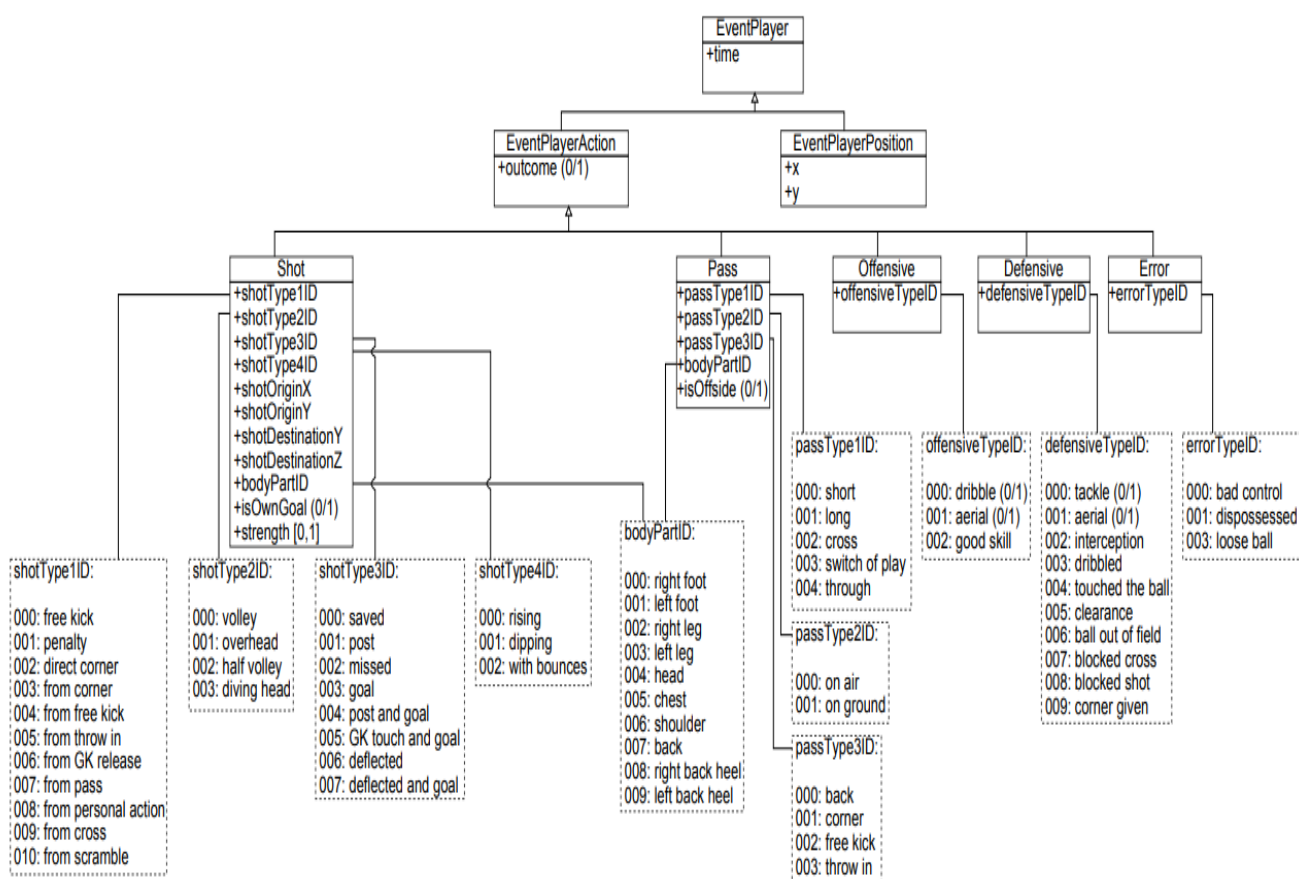


Рисунок 4.5 – Структурна схема даних, що використовуються для збору та структурування просторово-часових подій

MySQL пропонується у двох різних версіях: MySQL Community Server з відкритим кодом і пропріетарний Enterprise Server. Сервер MySQL Enterprise Server відрізняється низкою власних розширень, які встановлюються як серверні

плагіни, але в іншому випадку має спільну систему нумерації версій і побудовано з тієї самої кодової бази.

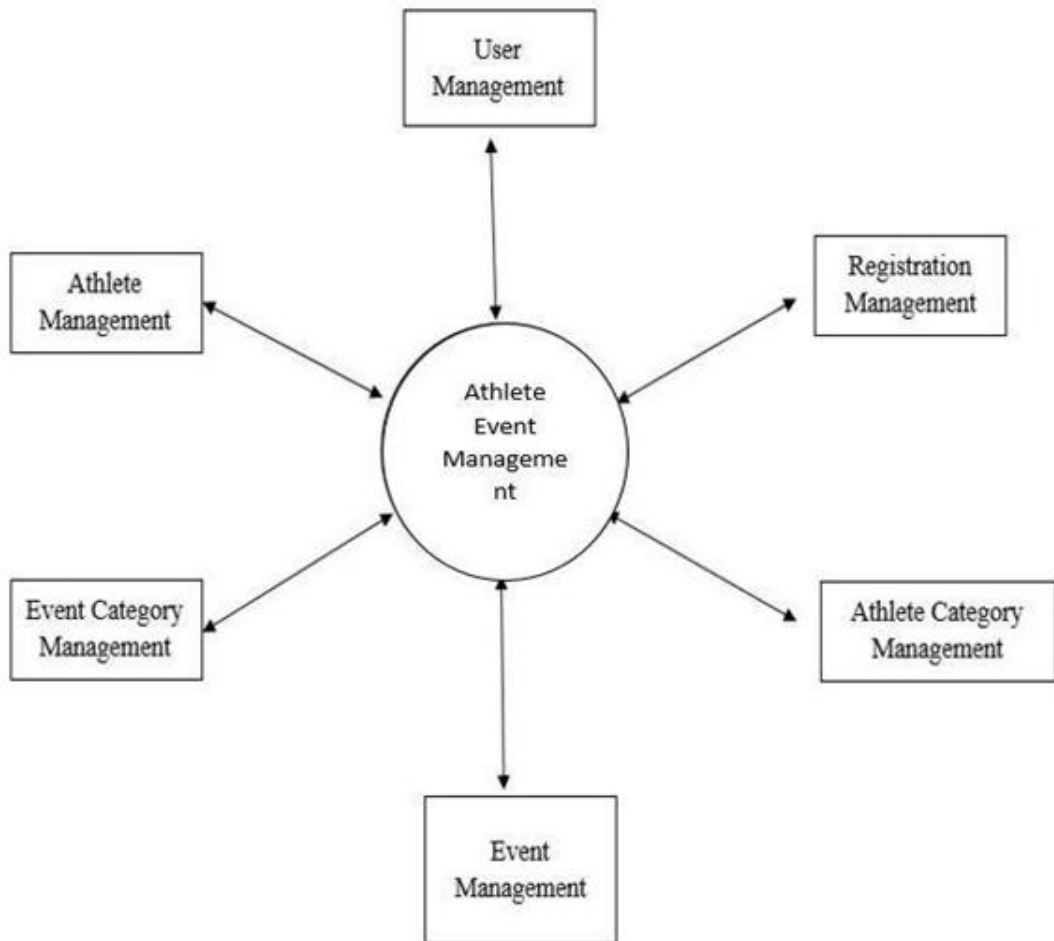


Рисунок 4.6 – Діаграма потоків даних 1-го рівня

Діаграми рівня 1 ілюструють процеси, пов'язані з системою, з відповідними вхідними даними, вихідними перетвореними даними та сховищами даних. Наступні рівні діаграми потоків даних (рівень 2, рівень 3 і так далі) розбивають основні процеси з діаграми рівня 1. Джерела інформації (зовнішні сутності) породжують інформаційні потоки (потоки даних), що переносять інформацію до підсистем або процесів. Ті у свою чергу перетворюють інформацію і породжують нові потоки, які переносять інформацію до інших процесів або підсистем, накопичувачам даних або зовнішнім сутностям – споживачам інформації. Таким

чином, основними компонентами діаграм потоків даних є: зовнішні сутності; системи/підсистеми; процеси; накопичувачі даних; потоки даних.

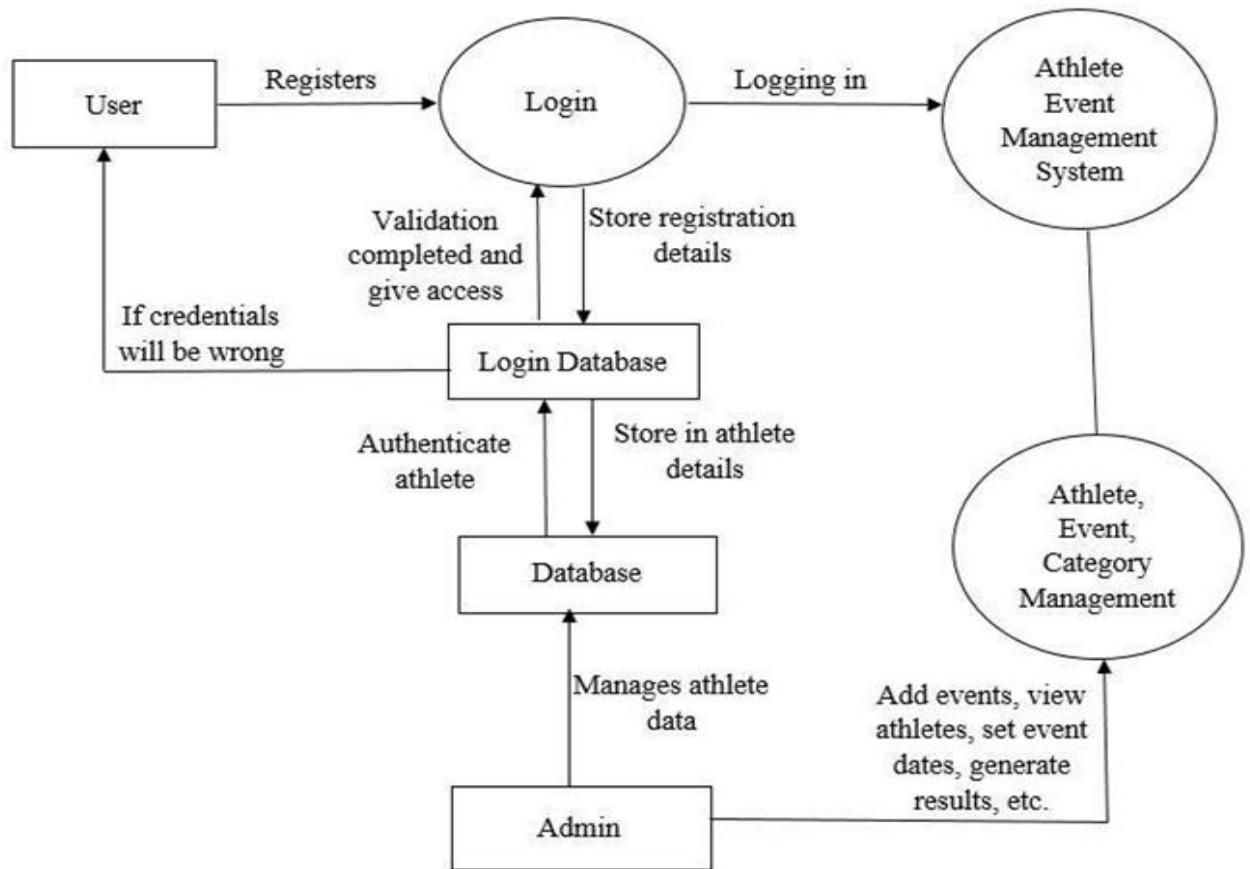


Рисунок 4.7 – Діаграма потоків даних 2-го рівня

Перевагами такої системи є забезпечення доступу для кількох користувачів для реєстрації з будь-якого місця та будь-коли. Полегшує рутинну роботу тренерського штабу. Обмеженнями є те, що потрібне підключення до інтернету.

4.3 Оцінка ефективності інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Оцінка ефективності інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних включає в себе кілька ключових аспектів:

1. Перевірку точності аналітичних моделей. Необхідно оцінити точність моделей, які використовуються для інтелектуального аналізу даних. Це може бути оцінено за допомогою метрик, таких як точність прогнозувань та відхилення від реальних результатів.

2. Покращення рішень. Необхідно визначити, наскільки інформаційна система допомагає у покращенні стратегій тренувань, визначенні індивідуальних потреб спортсменів та оптимізації підготовки для змагань.

3. Швидкість та продуктивність. Проводиться оцінка швидкості та продуктивності системи при обробці та аналізі великого обсягу даних. Швидке реагування системи може значно покращити її ефективність.

4. Взаємодія з користувачами. Оцінюється взаємодія системи з користувачами, зокрема тренерами та спортсменами. Визначається, наскільки легко вони можуть отримати доступ до аналітичної інформації та використовувати її для прийняття рішень.

5. Відповідність вимогам. Відбувається перевірка, наскільки система відповідає вимогам і потребам організації та спортсменів. Це включає в себе функціональні та технічні вимоги.

6. Захист даних. Необхідно оцінити рівень захисту конфіденційної інформації про спортсменів та інші дані, які обробляються системою.

7. Вартість та повернення інвестицій. Оцінюється вартість розробки та впровадження системи в порівнянні з отриманими користями та покращеннями у веденні тренувань та організації змагань.

8. Стабільність та надійність. Визначається наскільки стабільно працює система, та її надійність у випадку інтенсивного використання.

Загальна оцінка ефективності повинна враховувати всі ці аспекти, а результати можуть служити підставою для подальших вдосконалень та оптимізації системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Існує три поняття, які слід враховувати при оцінці похибки інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних: точність, правильність і зміщення.

Точність – це міра дисперсії процедури оцінювання, опис випадкових помилок; тоді як правильність – це міра різниці, опис систематичних помилок, яка представляє загальну відстань між оціненими значеннями та істинним значенням. Крім того, будь-яка інформаційна система матиме певний рівень внутрішньої похибки (тобто похибки системи або алгоритму), яка впливатиме на її точність. Це явище виникає, коли система видає результати, які систематично упереджені через помилкові припущення в аналітичних процесах або алгоритмах. Ті ж самі труднощі, з якими стикаються люди в обробці та розумінні складності, що пояснюються теорією обмеженої раціональності, можуть бути застосовані до алгоритмів. Отже, продуктивність алгоритмів буде обмежена інформацією, яку вони мають у своєму розпорядженні (яка може бути обмеженою і помилковою), а також (яка може бути обмеженою і помилковою), вбудованими процесами (які можуть включати когнітивні упередження розробника), а також обмеженою кількістю часу, який система має для отримання результату. Щоб подолати це, багато алгоритмів використовують евристичні процеси, які дають «достатньо задовільний результат». Таксономія Бейза-Йейтса може бути використана для визначення різних типів обчислювальних упереджень: пов'язаних з діяльністю, даними, вибіркою, алгоритмом, інтерфейсом і само-вибором. У завданнях класифікації помилка системи залежатиме від кількості та частки неправильно класифікованих подій. Популярними методами оцінки помилки в бінарних класифікаторах є чутливість та специфічність, які відповідно представляють

здатність класифікатора ідентифікувати всі релевантні класифікатора ідентифікувати всі релевантні екземпляри та повертати лише релевантні. Найпоширенішими засобами візуалізації чутливості і точності є: матриця помилок, крива робочої характеристики приймача (ROC), та площа під кривою. Іншою популярною метрикою класифікації є показник Log-Loss, який спеціально карає за неправильну класифікацію моделей.

4.4 Висновки

В розділі представлені UML-діаграми інформаційної системи керування даними спортсменів та оцінка ефективності інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

ВИСНОВКИ

У роботі за результатами виконаних теоретичних та практичних досліджень набув подальшого розвитку метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних, в частині інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо. А також запропонована інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних, що дозволяє підвищити ефективності роботи тренерів для спостереження та відбору спортсменів на змагання. Завданнями інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є збір, обробка, зберігання, опрацювання даних спортсменів та на основі цих даних прогнозування результативності.

У першому розділі було проведено огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах, а також сформовано постановку задачі.

У другому розділі на основі проведеного аналізу літературних джерел було запропоновано підсистему інформаційної системи керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків. Рівень джерела даних є основою для реалізації різноманітних додатків для аналізу та прогнозування спортивних великих даних. Наступний рівень – це рівень збору даних, який збирає дані з рівня джерела даних і виконує таку обробку: збір даних, зберігання даних, обмін даними, імпорт вручну та веб-сканер. Потім зібрані дані очищаються, а необхідна обробка виконується відповідно до різних вимог програми, де дані класифікуються та зберігаються. Оброблені дані зберігатимуться на рівні центрального сховища, включаючи сховище структурованих даних, сховище неструктурованих даних і сховище файлів. Рівень

аналізу даних виконує вибір функцій, аналіз зв'язків, статистичний аналіз і аналіз соціальних мереж відповідно до потреб конкретних програм з метою виявлення потенційних знань, законів і закономірностей у спортивних великих даних.

Представлена підсистема є основою для створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

У третьому розділі було розглянуто особливості роботи з спортивними даними, керування спортивними даними, методах аналізу спортивних великих даних та застосунках спортивних великих даних. У сфері великих даних у спорті відбулося кілька змін: від простого статистичного оцінювання до опрацювання спортивних великих даних; від простого статистичного оцінювання до оцінювання на основі моделей; від простого статистичного аналізу до прогнозування результатів спортсменів на основі даних; від аналізу соціальних мереж до аналізу графів знань; від явних спортивних особливостей до неявних спортивних особливостей.

Запропоновано метод створення інформаційної системи керування даними спортсменів в системі організації змагань, що дозволяє поєднати ефективність інформаційної системи з потужністю інтелектуального аналізу даних для покращення керування даними спортсменів у системі організації змагань, який на відміну від відомих базується на інтелектуальному аналізі даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо. Також було проаналізовано та сформовано вимоги до інформаційної системи керування даними спортсменів та розроблено і представлено підсистему прогнозування результативності спортсмена, яка є основною частиною запропонованої підсистеми.

Завданнями інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є збір, обробка, зберігання, опрацювання даних спортсменів та на основі цих даних прогнозування результативності.

У четвертому розділі представлені UML-діаграми інформаційної системи керування даними спортсменів та оцінку ефективності інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

За темою кваліфікаційної роботи магістра опубліковані тези доповіді авторів Швайко В.К., Кузьмін А.А., Шатровський А.О. «Гейміфікація інформаційної системи для вибору виду спорту на основі морфофункціональних показників людини» III Всеукраїнської науково – технічної конференції молодих вчених, аспірантів та студентів «комп'ютерні ігри і мультимедіа як інноваційний підхід до комунікації - 2023» 28-29 вересня 2023 р. Одеса.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Legg P. A, Chung D. H., Parry M. L., Jones M. W., Long R., Griffiths I. W., Chen M. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. In: *Computer graphics forum, Wiley Online Library*. 2012. Vol. 31, pp. 1255–1264.
2. Legg P.A, Chung D.H, Parry M.L, Bown R, Jones M.W, Griffiths I.W, Chen M. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *IEEE Trans Vis Comput Graph*. 2013. Vol. 19(12), pp. 2109–2118.
3. Losada A.G, Thero´n R, Benito A. Bkviz: a basketball visual analysis tool. *IEEE Comput Graph Appl*. 2016. Vol. 36(6). P. 58–68.
4. Perin C., Vuillemot R., Fekete J.D. Soccerstories: a kick-off for visual soccer analysis. *IEEE Trans Vis Comput Graph*. 2016. Vol. 19(12), pp. 2506–2515.
5. Perin C., Boy J., Vernier F. Using gap charts to visualize the temporal evolution of ranks and scores. *IEEE Comput Graph Appl*. 2016. Vol. 36(5). P.38–49.
6. Perin C., Vuillemot R., Stolper C.D, Stasko J.T, Wood J., Carpendale S. State of the art of sports data visualization. In: *Computer graphics forum, Wiley Online Library*.2018. Vol 37, pp. 663–686.
7. Polk T., Yang J., Hu Y., Zhao Y. Tennis: visualization for tennis match analysis. *IEEE Trans Vis Comput Graph*. 2014. Vol. 20(12), PP. 2339–2348.
8. Janetzko H., Sacha D., Stein M., Schreck T., Keim D.A., Deussen O., et al. Feature-driven visual analytics of soccer data. In: *2014 IEEE conference on visual analytics science and technology (VAST), IEEE*. 2014. Pp. 13–22.
9. Janetzko H., Stein M., Sacha D., Schreck T. Enhancing parallel coordinates: statistical visualizations for analyzing soccer data. *Electron Imaging*. 2016. Vol. 1. P.1–8.
10. Goldsberry K. Courtvision: new visual and spatial analytics for the NBA. *MIT sloan sports analytics conference, ACM Comput Surv (CSUR)*. 2017. Vol. 50(2). P. 22.

11. Liu Aijun Mahapatra, Rajendra Prasad Mayuri, A. V. R. Hybrid design for sports data visualization using AI and big data analytics. *Complex & Intelligent Systems Heidelberg*. Vol. 9, Iss. 3, (Jun 2023). P. 2969–2980. DOI:10.1007/s40747-021-00557-w. Complex & Intelligent Systems.
12. Du M., Yuan X. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*. 2021. Vol. 24(1), pp.47–67.
13. Alen Rajšp, Iztok Fister Jr. A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training. *Applied Sciences*. 2020. Vol.10. P. 3013.
14. Liu Y. College students' physical fitness test data analysis, visualization and prediction using data mining techniques. *Journal of Physics: Conference Series*. 2020. Vol. 1631(1). P. 012121.
15. A. Hirsch, M. Bieleke, J. Schuler, and W. Wolff Implicit theories about athletic ability modulate the effects of if-then planning on performance in a standardized endurance task. *International Journal of Environmental Research and Public Health*. 2020. Vol. 17, №. 7. P. 2576.
16. O. Sova, O. Turinskyi, A. Shyshatskyi, Development of an algorithm to train artificial neural networks for intelligent decision support systems. *Eastern-European Journal of Enterprise Technologies*. 2020. Vol. 1, №9. Pp. 46–55.
17. S. W. Chen Research on the integration of basketball course teaching and physical fitness assessment. *Contemporary Sports Technology*. 2017. Vol. 7, №. 4. pp. 129–131.
18. Рашкевич Ю.М., Ткаченко Р.О., Цмоць Г.І., Пелешко Д.Д. Нейроподібні методи, алгоритми та структури обробки зображень у реальному часі. Львів: Вид-во Львів. політехніки, 2014. 256 с.
19. Творошенко І. С. Конспект лекцій з дисципліни «Цифрова обробка зображень». Харків : ХНУМГ ім. О. М. Бекетова, 2017. 75 с.
20. Черняк О. І., Захарченко П.В. Інтелектуальний аналіз даних : підручник. К.: Знання, 2014. 599 с.

21. Hans Weber. Big Data and Artificial Intelligence: Complete Guide to Data Science. *AI, Big Data and Machine Learning*. URL: <https://www.amazon.com/dp/B08FP9YZXN?tag=uuid10-20>. (дата звернення: 17.10.2023).
22. Jay M. Patel. Getting Structured Data from the Internet: Running Web Crawlers. *Scrapers on a Big Data Production Scale 1st ed. Edition*. URL: <https://www.amazon.com/dp/1484265750?tag=uuid10-20>. (дата звернення: 15.11.2023).
23. Тимощук П. В. Штучні нейронні мережі. Навчальний посібник. Львів: Видавництво Львівської політехніки, 2011. 441с.
24. Субботін С.О. Нейронні мережі: теорія і практика. Навчальний посібник. Житомир: Вид. О.О. Євєнюк, 2020. 184с.
25. Glassner A. Deep Learning: From Basics to Practice. *Published by The Imaginary Institute, Seattle, WA*. 2018. Volume 1. 872p.
26. Glassner A. Deep Learning: From Basics to Practice. *Published by The Imaginary Institute, Seattle, WA*. 2018. Volume 2. Pp. 873-1747.
27. Tetiana Novorushchenko, Alla Herts, Yelyzaveta Hnatchuk: Concept of Intelligent Decision Support System in the Legal Regulation of the Surrogate Motherhood. *IDDM*. 2019. P. 57-68.
28. Доля В. Г. Комп'ютерні системи штучного інтелекту: навч. посібник. Київ : Університет "Україна", 2011. 296 с.
29. J. Hwang, Cheating in E-sports: a proposal to regulate the growing problem of E-doping. *Northwestern University Law Review*. 2022. Vol. 116, №5, pp. 1283–1318.
30. P. E. Malik Sports: a competitive electronic sports. *International journal of multidisciplinary research and analysis*. 2019. Vol. 4, №1, pp. 87–90.
31. Thornton H., Delaney J., Duthie G., Dascombe B. Effects of pre-season training on the sleep characteristics of professional rugby league players. *International Journal of Sports Physiology and Performance*. 2018. Vol. 13(2). Pp.176-182.

32. Murray N., Gabbett T., Townshend A. The use of relative speed zones in Australian football: are we really measuring what we think we are. *Journal of Sports Physiology and Performance*. 2018. Vol. 13(4). Pp. 442–451.
33. Baumer B., Jensen S., Matthews G. openWAR: An open source system for evaluating overall player performance in major league baseball. *Journal of Quantitative Analysis in Sports*. 2018. Vol. 11(2), pp. 69–84. <https://doi.org/10.1515/jqas-2014-0098>.
34. Becker A., Sun X. A. An analytical approach for fantasy football draft and lineup management. *Journal of Quantitative Analysis in Sports*. 2016. Vol. 12(1), Pp. 17–30.
35. Kolbush J., Sokol J. H. (2017). A logistic regression/Markov chain model for American college football. *International Journal of Computer Science in Sport*. 2017. Vol. 16(3). Pp.185–196.
36. Parmezan Bonidia R., Duilio Brancher J., Marques Busto R. Data mining in sports: A systematic review. *IEEE Latin America Transactions*. 2018. Vol. 16(1). Pp. 232–239.
37. Shih H.C. A survey on content-aware video analysis for sports. *IEEE Transactions on Circuits and Systems for Video Technology*. 2017. Vol. 28(5). Pp.1212–1231.
38. Longo U.G., Sofi F., Candela V., Dinu M., Cimmino M., Massaroni C., Eschena E., Denaro V. Performance activities and match outcomes of professional soccer teams during the 2016/2017 Serie A season. *Medicina*. 2019. Vol. 55(8), p. 469. <https://doi.org/10.3390/medicina55080469>.
39. Mitrotasios M., Gonzalez-Rodenas J., Armatas V., Aranda R. The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. *International Journal of Performance Analysis in Sport*. 2019. Vol. 19, pp. 452–465. <https://doi.org/10.1089/big.2018.0067>.
40. Alomari H., Ramasamy V., Kiper J., and Potvin G. A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in

computer science and software engineering education. *Heliyon* 6. 2020. Volume 6 (Issue 5). doi: 10.1016/j.heliyon.2020.e03917.

41. Blobel T., Lames M. A Concept for Club Information Systems (CIS) - An Example for Applied Sports Informatics. *International Journal of Computer Science in Sport*. 2020. Vol.19 (Issue 1), pp. 102-122. <https://doi.org/10.2478/ijcss-2020-0006>.

42. Cichy C., Rass S. An overview of data quality frameworks. *IEEE Access*. 2019. Vol. 7, pp. 24634–24648. doi: 10.1109/ACCESS.2019.2899751.

43. Ranaweera J. S., Weaving D., Zanin M., and Roe G. Identifying the current state and improvement opportunities in the information flows necessary to manage professional athletes: a case study in rugby union. *Front Sport Act Liv*. 2022. doi: 10.3389/fspor.2022.882516.

44. Yuliawan D., Widyandana D., Nur Hidayah R. Utilization of Nursing Education Progressive Web Application (NEPWA) media in an education and health promotion course using Gagne's model of instructional design on nursing students: quantitative research and development study. *JMIR Nursing*. 2020. Vol. 3. e19780. doi: 10.2196/19780.

45. Wu Y., Xie X., Wang J., Deng D., Liang H., Zhang H., Cheng S., Chen W. Forvizor: visualizing spatio-temporal team formations in soccer. *IEEE Transactions on Visualization and Computer Graphics*. 2018. Vol. 25(1), p.65–75.

46. Legg P.A, Chung D.H, Parry M.L, Bown R., Jones M.W, Griffiths I.W, Chen M. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *IEEE Transactions on Visualization and Computer Graphics*. 2013. Vol. 19(12). Pp.2109–2118.

47. Chiang R. H., Grover V., Liang T. P. and Zhang D. Special Issue: Strategic Value of Big Data and Business Analytics. *Journal of Management Information Systems*. 2018. Vol. 35(2), pp. 383–387.

48. Khakimullin A. Big Data vs. Smart Data, available. 2018. URL: <https://www.dataversity.net/bigdata-vs-smart-data/#>. (дата звернення: 15.11.2023).

49. Lalwani A. K. How Do Consumers' Cultural Backgrounds and Values Influence Their Coupon Proneness? A Multimethod Investigation. *Journal of Consumer Research*. 2018. Vol. 45. Pp. 1037–1050.
50. Гороховатський В. О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних : навч. посіб. М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. Харків : ХНУРЕ, 2021. 92 с.
51. Niyato D., Thai Hoang D. Smart data pricing models for the internet of things: a bundling strategy approach. *IEEE Network*. 2016. Vol. 30(2), pp.18–25.
52. Petty R. E., Cacioppo J. T. Attitudes and persuasion: Classic and contemporary approaches. Routledge. New York. 2018. P. 336.
53. Teece D., Peteraf M., Leih S. Dynamic Capabilities and Organizational Agility. *California Management Review, The World Economic Forum. Big Data, Big Impact: New Possibilities for International Development*. 2016. Vol. 58(4), pp.13–35.
54. Geneva Thorstad R., Wolff P. A big data analysis of the relationship between future thinking and decision-making. *PNAS*. 2018. Vol. 115(8), pp. E1740–E1748.
55. Trkman M., Trkman P. A framework for increasing business value from social media. *Economic Research*. 2018. Vol. 31(1), pp. 1091–1110.
56. Vassakis K., Petrakis E., Kopanakis I. Big Data Analytics: Applications, Prospects and Challenges. *Mobile Big Data, Springer*. 2018. pp. 3–22.
57. Du M, Yuan X A survey of competitive sports data visualization and visual analysis. *Journal of Visualization*. 2021. Vol. 24(1), pp.47–67.
58. Vu D. L., Nguyen T.K., Nguyen T.V., Nguyen T.N., Massacci F., Phung P.H. HIT4Mal: hybrid image transformation for malware classification. *Transactions on Emerging Telecommunications Technologies*. 2020. Vol. 31(11), p.e3789.
59. Brewer B.W, Van Brewer T.P. A visualization template for the graphical representation of sport injury antecedents and consequences models and data. *Journal of Functional Morphology and Kinesiology*. 2020. Vol. 5(4), p.87.

60. Cao Q., Deng Z., Liu J., Li X. SGDB: a sports gene database for visualization of sports effects on human skeletal muscle gene expression. *IEEE Access*. 2020. Vol. 8, pp. 20557–20562.
61. Wu W., Lin W., Hsu C.H., He L. Energy-efficient hadoop for big data analytics and computing: a systematic review and research insights. *Future Generation Computer Systems*. 2018. Vol. 86, pp.1351–1367.
62. Li S., Zhang B., Fei P., Shakeel P.M., Samuel R.D.J. Computational efficient wearable sensor network health monitoring system for sports athletics using IoT. *Aggress Violent Behav* 101541. 2020.
63. Lin Q., Li T., Shakeel P.M., Samuel R.D.J. Advanced artificial intelligence in heart rate and blood pressure monitoring for stress management. *Journal of Ambient Intelligence and Humanized Computing*. 2021. Vol. 12(3), pp. 3329–3340.
64. Швайко В.К., Кузьмін А.А., Шатровський А.О. Гейміфікація інформаційної системи для вибору виду спорту на основі морфофункціональних показників людини. *III Всеукраїнської науково – технічної конференції молодих вчених, аспірантів та студентів «Комп'ютерні ігри і мультимедіа як інноваційний підхід до комунікації - 2023»* 28-29 вересня 2023 р., Одеса. с.104–106.

ДОДАТОК А
(обов'язковий)

ТЕЗИ ДОПОВІДІ

**Міністерство освіти і науки України
Одеський національний технологічний університет
Вінницький національний технічний університет
Інститут комп'ютерної інженерії, автоматизації,
робототехніки та програмування ім.П.Н.Платонова**



ПРОГРАМА

**III ВСЕУКРАЇНСЬКОЇ
НАУКОВО – ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ
МОЛОДИХ ВЧЕНИХ, АСПІРАНТІВ
ТА СТУДЕНТІВ**

**«КОМП'ЮТЕРНІ ІГРИ І МУЛЬТИМЕДІА
ЯК ІННОВАЦІЙНИЙ ПІДХІД
ДО КОМУНІКАЦІЇ - 2023»**

**28-29 вересня 2023 р.
ОДЕСА**

УДК 004.01/08

Комп'ютерні ігри та мультимедіа як інноваційний підхід до комунікації - 2023 / Матеріали III Всеукраїнської науково-технічної конференції молодих вчених, аспірантів і студентів, Одеса, 28-29 жовтня 2023 р. - Одеса, Видавництво ОНТУ, 2023 р. – 270 с.

Збірник включає матеріали доповідей учасників конференції, які об'єднані за тематичними напрямками конференції.

Збірник буде корисним як для фахівців і працівників фірм, зайнятих в області розробки та просування комп'ютерних ігор, так і для викладачів, магістрів і студентів вищих навчальних закладів, які навчаються за напрямками і спеціальностями програмного забезпечення, комп'ютерних наук, комп'ютерної інженерії, прикладної математики та обробки інформації, буде корисним професіоналам у сферах гейміфікації, кіберспорту, стрімінгу, віртуальної реальності, доповненої реальності, штучного інтелекту, машинного навчання, геймдизайну, саунддизайну.

Результати досліджень у збірнику представляють собою своєрідний зріз сучасного стану справ в перерахованих галузях знань, який може допомогти як фахівцям, так і студентам університетів скласти загальну картину розвитку комп'ютерних ігор та мультимедіа та пов'язаних з ними питань.

Наукові праці згруповані за напрямками роботи конференції та наведені в алфавітному порядку прізвищ авторів.

Матеріали (тези доповідей) друкуються в авторській редакції. Відповідальність за якість та зміст публікацій несе автор.

Матеріали подано українською та англійською мовами.
Редактор збірника Котлик С.В.

Одним із прикладів освітньої гейміфікації можна назвати Ribbon Hero: гра, яка вчить як користуватись Microsoft Office.

Ribbon Hero є доповненням, яке доступне при безкоштовному завантаженні пакета програм Microsoft, вона допомагає користувачеві навчитись використовувати основні засоби Office 2007 або 2010. Після установки, гра буде запущена з будь-якої програми Office (Word, Excel або PowerPoint).

У самій грі користувачеві необхідно вирішити проблему, а за її виконання він отримає бали досвіду. Проблеми згруповані в чотири розділи: робота з текстом, дизайн сторінки і макету, художнє представлення і більш узагальнений розділ швидких завдань.

В Ribbon Hero є її здатність відстежувати прогрес користувача під час його навчання коли він використовує інструменти Office, а також відповідний рівень складності завдань [5].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Афузова, Г. В. (2021). Вища освіта в контексті Європейських цінностей різноманітності та інклюзії. Вища освіта України. (3). 67–74
2. Цифрова інклюзія та доступність: соціальна діджиталізація [Електронний ресурс]. – Режим доступу: https://vsei.vn.ua/images/Doc/Nauka/Inclusivna_osvita/cifrova-inklyuziya-ta-dostupnist-socialna-didzitalizaciya.pdf
3. Meet the First Blind Individual to Complete Diablo 4 Without Sighted Assistance. [Електронний ресурс]. – Режим доступу: <https://www.ign.com/articles/meet-the-first-blind-individual-to-complete-diablo-4-without-sighted-assistance>
4. Тестування доступності: яке тестування можна вважати достатнім [Електронний ресурс]. – Режим доступу: https://dou.ua/forums/topic/45294/?from=comment-digest_post&utm_source=digest-comments&utm_medium=email&utm_campaign=19092023
5. Топ 10 прикладів гейміфікації (перетворення у гру) в освіті, які змінять наше майбутнє [Електронний ресурс]. – Режим доступу: <https://osvitanova.com.ua/posts/1143-top-10-prykladiv-heimifikatsii-peretvorennia-u-hru-v-osviti-i-aki-zminiat-nashe-maibutnie>
6. Освіта дітей з особливими потребами (інклюзивне навчання) [Електронний ресурс]. – [https://wiki.legalaid.gov.ua/index.php/Освіта_дітей_з_особливими_потребами_\(інклюзивне_навчання\)](https://wiki.legalaid.gov.ua/index.php/Освіта_дітей_з_особливими_потребами_(інклюзивне_навчання))

УДК 004.4

ГЕЙМІФІКАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ВИБОРУ ВИДУ СПОРТУ НА ОСНОВІ МОРФОФУНКЦІОНАЛЬНИХ ПОКАЗНИКІВ ЛЮДИНИ

ШВАЙКО В.К. (lera.schvajko@gmail.com), КУЗЬМІН А.А. (andriy1731@gmail.com),
ШАТРОВСЬКИЙ А.О. (shatrikua@gmail.com)
Хмельницький національний університет

У сучасному світі все більше обертів набирає малорухливий спосіб життя. Віртуалізація суспільства несе з собою як позитивні наслідки у досягненні нових звершень, так і вкрай негативні щодо рівня фізичного здоров'я.

Найбільше цьому впливу піддається молодь. За статистичними даними, отриманими у результаті соціологічного опитування проведеного Фондом Фрідріха Еберта у 2017 році, 33% молоді віком 14–29 років займалися спортом у вільний час часто та дуже часто, 18% не робили цього ніколи, а 46% зрідка або інколи. До того ж регулярність занять також знижується з віком — від 48% серед підлітків до 25% у 25–29 років [1].

У наших попередніх роботах [2-4] було запропоновано принцип роботи та структуру інформаційної системи для вибору виду спорту на основі морфофункціональних показників з метою розробки дієвого механізму залучення молоді до занять різними видами спорту, які культивуються у місті. На даному етапі нашою цільовою аудиторією є школярі віком 10-11 років.

Тому метою даної роботи є гейміфікація процесу збору вхідних даних, що в подальшому будуть аналізуватись у запропонованій інформаційній системі.

Для збору вхідних даних необхідних для підбору на основі морфофункціональних показників, потрібно здати перелік певних нормативів та пройти певні заміри. Враховуючи, що нашою ЦА є діти віком 10-11 років, цей процес може мати певні ускладнення у зв'язку з проблемами з концентрацією уваги на поставленій задачі. Тому метою даної роботи є підвищення інтересу дітей для проходження необхідних вправ під виглядом ігрового процесу.

Дослідження також показують, що розміщення дітей у конкурентне середовище може сприяти покращенню результатів. Тому одним з можливих варіантів гейміфікації процесу збору даних є поділ дітей на команди. Такий тип ігрової взаємодії можна пристосувати до переліку певних нормативів, які необхідно скласти, наприклад біг 30 метрів, кидок набивного м'яча, човниковий біг та стрибки у довжину чи висоту. Деякі інші нормативи можна також перетворити в гру, додавши якісь певні яскраві елементи. До прикладу норматив швидкості реакції, де потрібно ловити палицю, можна застосовувати з кількома різними, щоб урізноманітнити процес, при можливості яскравого забарвлення, адже згідно з дослідженнями більшість дітей краще сприймає інформацію, коли предмет привабливе одне з їх сприйняття, у цьому випадку візуальне.

Вимірювання параметрів можна зробити яскравим за допомогою допоміжних засобів. Наприклад при вимірі зросту попередньо скласти приблизний список тварин і використовувати їх ріст чи довжину тіла для порівняння. Схожий метод можна використати і при вимірі індексу маси тіла на етапі зважування дітей. З якою тваринкою вони мають схожу вагу, або з кількома іншими тваринами. Наприклад вага дитини 42 кілограми дорівнює вазі шести кроликів (один важить 7 кілограм). Також, для покращення сприйняття процесу підбору вибору спорту для ЦА було обрано форму імплементації запропонованої у [2-4] інформаційної системи у вигляді кросплатформного мобільного застосунку, оскільки за результатами опитування ЦА (рисунок 1), саме цій формі було надано перевагу 71% опитаних учнів.

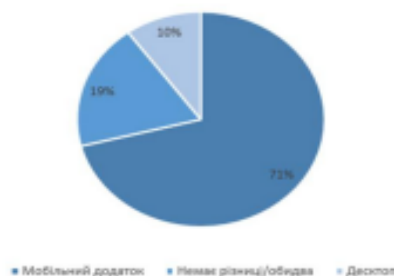


Рисунок 1 – Результати опитування ЦА щодо форми представлення інформаційної системи для вибору виду спорту на основі морфофункціональних показників людини

Також у наших попередніх роботах було запропоновано концепцію інтерфейсу користувача інформаційної системи для вибору виду спорту на основі морфофункціональних показників людини. На рисунку 2 представлено дизайн інтерфейсу користувача для мобільного застосунку, виконаний у брендбуку Хмельницького національного університету. Наступним нашим кроком буде презентація цього дизайну для групи представників ЦА та обговорення його з ними, а також обговорення та вибір тварин-амбасадорів та тварин-помічників для гейміфікації процесу збору морфофункціональних показників у дітей 10-11 років.

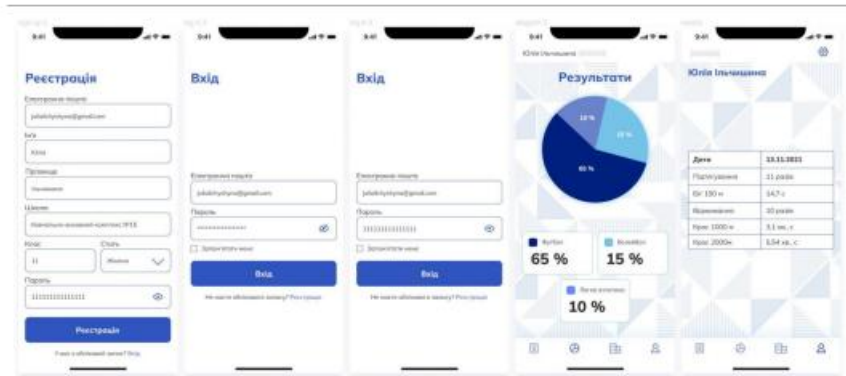


Рисунок 2 – Дизайн інтерфейсу користувача для мобільного застосунку для вибору виду спорту на основі морфофункціональних показників людини

Отже, подальші зусилля авторів будуть спрямовані на аналіз результатів опитувань цільової аудиторії щодо можливості гейміфікації підходу до збору та вимірювання морфофункціональних показників дітей та імплементації запропонованої інформаційної системи у вигляді кросплатформного мобільного застосунку.

ПЕРЕЛІК ПОСИЛАНЬ

1. Спорт як шлях до здоров'я українців URL: <https://www.radiosvoboda.org/a/2227020.html> (доступ 12.03 .2023)
2. Швайко В. К., Фесік З. Ю. Інформаційна система для вибору виду спорту на основі аналізу морфофункціональних показників людини. "Інформаційні технології та інженерія - 2023" (IT&I), 7-10 лютого 2023, Миколаїв, Україна, с. 28-29
3. Швайко В.К., Павлова О.О. Технологія підтримки прийняття рішень щодо вибору виду спорту на основі морфофункціональних показників людини. Актуальні Проблеми Комп'ютерних Наук (АПКН-2022), Хмельницький, Україна, 18-19 листопада 2022. Хмельницький: ХНУ, 2022. с.314-318
4. Pavlova, O., Solytk, O., Shvaiko, V., Ilchyshyna, J., Bouhissi, H.E. Human Morphofunctional Indicators-Based Decision Support System for Choosing Kind of Sport. Materials of the 4th International Conference on Intelligent Information Technologies and Systems of Information Security "IntelITSIS-2023". CEUR Workshop Proceedings, 2023, 3426, pp. 322–333

ДОДАТОК Б
(обов'язковий)
ПРЕЗЕНТАЦІЯ

**Метод створення інформаційної
системи керування даними
спортсменів в системі
організації змагань на основі
інтелектуального аналізу даних**

Виконав: магістр гр.ІСТм-22-1 Шатровський А.О.
Керівник: к.т.н., доцент Гнатчук Є.Г.

Актуальність

- ▶ Змагання різного рівня є результатом довгої виснажливої підготовки спортсменів, а отже важливим питанням для тренерів є збір різноманітних даних про стан підготовки спортсмена, аналіз та дослідження цих даних з метою якісного відбору спортсменів для змагань та керування цими даними. Така робота є досить часозатратною та виснажливою. Вирішити таку задачу може розроблення відповідної інформаційної системи, яка полегшить рутинну роботу тренерського штабу при аналізі даних спортсменів.
- ▶ Отже, на сьогоднішній день актуальною задачею є розроблення методу створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

- ▶ Мета дослідження: підвищення ефективності роботи тренера в системі організації змагань для спостереження та відбору спортсменів на змагання за допомогою інформаційної системи керування даними спортсменів на основі інтелектуального аналізу даних.
- ▶ Об'єкт дослідження: інформаційна система керування даними спортсменів в системі організації змагань.
- ▶ Предмет дослідження: метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Задачі дослідження

- ▶ 1) провести огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах;
- ▶ 2) розробити метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних;
- ▶ 3) запропонувати концепцію інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

Наукова новизна

- ▶ набув подальшого розвитку метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині методів інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо;
- ▶ набула подальшого розвитку інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині додавання підсистеми керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків.

5 А

- ▶ Практична значимість отриманих результатів полягає у можливості використовувати результати досліджень для ефективного відбору спортсменів на основі аналізу його попередніх результатів, фізичного стану та готовності, а також наданні тренерам та управлінцям додаткових інструментів для прийняття обґрунтованих та докладних рішень щодо тренувальних стратегій, участі в змаганнях та медичного супроводу.

А
6 П

- ▶ Огляд літературних джерел показав, що на даний момент актуальність використання інформаційних систем в спортивній галузі зростає.
- ▶ Завданнями інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних є збір, обробка, зберігання, опрацювання даних спортсменів та на основі цих даних прогнозування результативності спортсменів.



- ▶ Керування спортивними даними в основному застосовує методи керування даними, інструменти та платформи для роботи зі спортивними великими даними, включаючи зберігання, попередню обробку та безпеку. Однак управління великими даними є складним процесом, який виникає через неоднорідність і неструктурованість джерел даних. Управління великими спортивними даними має вирішальне значення для успіху національної спортивної індустрії, команд та окремих осіб .
- ▶ Основна мета керування спортивними даними полягає в тому, щоб отримати потенційну цінність спортивних великих даних і підвищити якість даних і доступність для прийняття рішень.

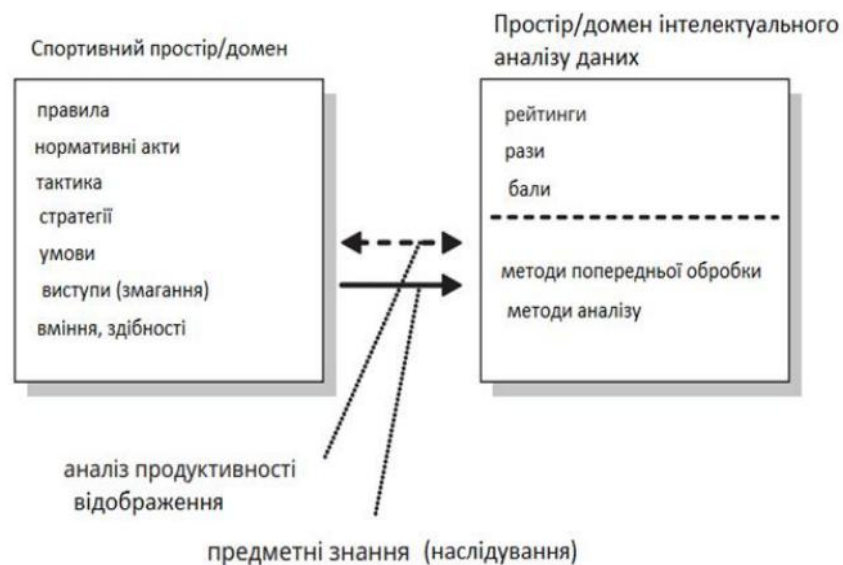


Таблиця 1 – Приклади використання методів інтелектуального аналізу даних для аналізу спортивних досягнень в елітних видах спорту

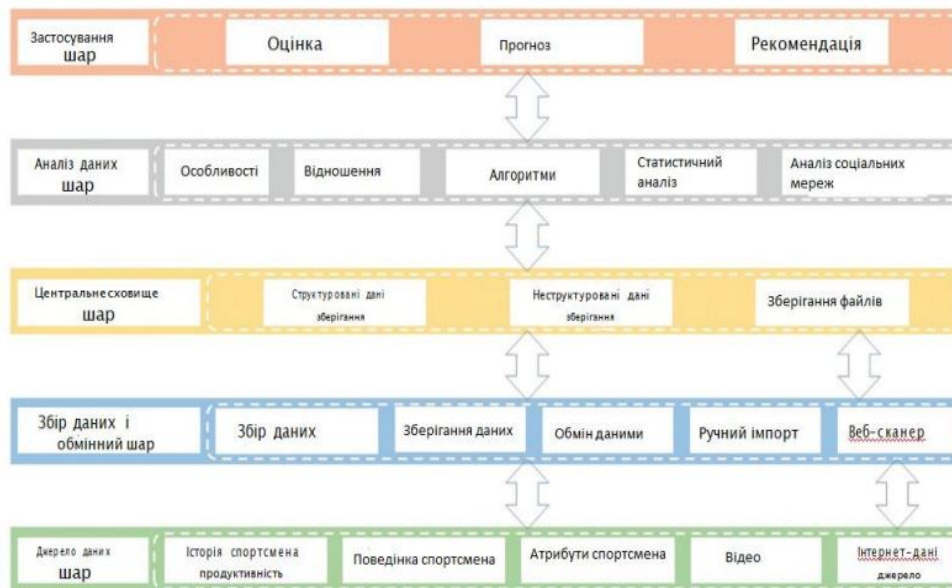
Метод	Алгоритм	Вид спорту
Кластеризація	k-середніх	гольф, велоспорт на треку
	середній зв'язок (ієрархічний)	плавання
	алгоритм Варда	гольф
	змішаний	десятиборство
	карти, що самоорганізуються	баскетбол
Класифікація	наївний Байєс	велоспорт на треку
	лінійний дискримінантний аналіз	регбі, веслування, стрибки в довжину
Моделювання відносин	лінійна регресія	плавання
	лінійна та поліноміальна регресія	плавання
	нейронні мережі	плавання, аеробіка, футбол
Видобуток правил	Правила асоціації	настільний теніс, баскетбол

9 A

Схема аналізу спортивних результатів, що включає спортивну область і область аналізу даних

10 A
Pe
wi

Підсистема керування даними спортсменів



Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних:

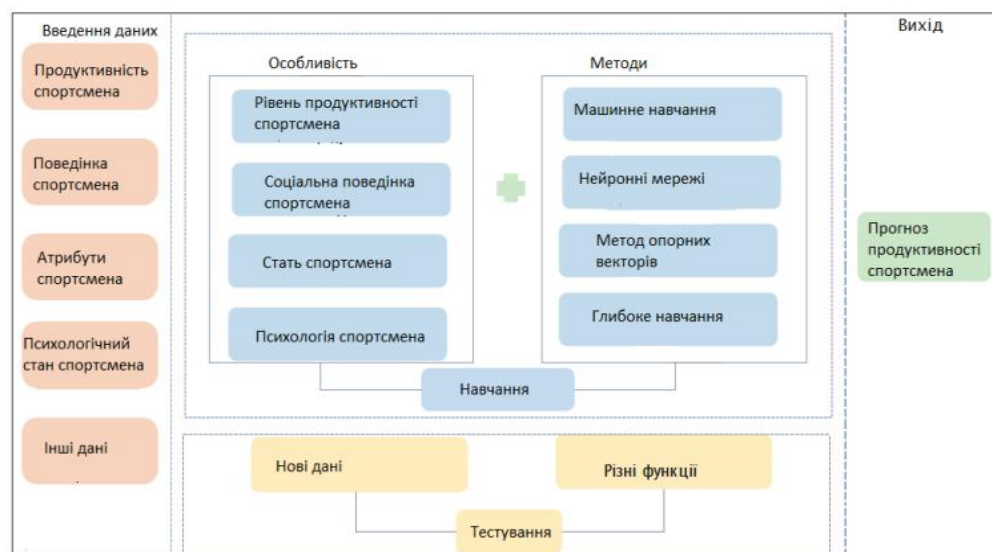
- ▶ Етап 1. Визначення вимог. Необхідно визначити функціональні та технічні вимоги до системи.
- ▶ Етап 2. Збір та обробка даних. На цьому етапі необхідно визначити типи, формати та кількість даних, які будуть збиратися та аналізуватися. А також необхідно забезпечити механізми для збору даних про спортсменів, тренування, результати змагань та інші важливі параметри.
- ▶ Етап 3. Очищення та підготовка даних. На цьому етапі потрібно провести очистку та підготовку даних, включаючи обробку відсутніх значень, видалення дублікатів та нормалізацію даних.
- ▶ Етап 4. Застосування інтелектуального аналізу даних. Цей етап в свою чергу складається з таких підпунктів, як вибір методів аналізу (наприклад, машинне навчання, статистичний аналіз, класифікація тощо), розробка моделей для прогнозування та аналізу результатів на основі вибраних методів; навчання моделей, де потрібно використати існуючі дані для тренування моделей.

- ▶ **Етап 5. Розробка інформаційної системи,** де потрібно розробити архітектуру інформаційної системи, враховуючи розподілену базу даних, модулі аналізу та інтеграцію з іншими системами. При цьому необхідно створити структуру бази даних, що враховує потреби системи та методів аналізу даних та реалізувати модулі для збору, зберігання та аналізу даних.
- ▶ **Етап 6. Тестування та вдосконалення.** На етапі тестування проводиться ряд тестів для перевірки функціональності та надійності системи. А також досить важливий момент – проведення оцінки точності моделей інтелектуального аналізу даних. Також при необхідності вносяться виправлення.
- ▶ **Етап 7. Впровадження та підтримка.** На цьому етапі виконується розгортання інформаційної системи та моделі аналізу даних в реальному середовищі. Також на цьому етапі здійснюється навчання користувачів якщо потрібно, щодо використання нової системи та аналізу результатів. Підтримка та оновлення має бути забезпечене постійно на основі змін в потребах.

Підсистема спортсменів

прогнозування

результативності



Архітектура інформаційної системи керування даними спортсменів

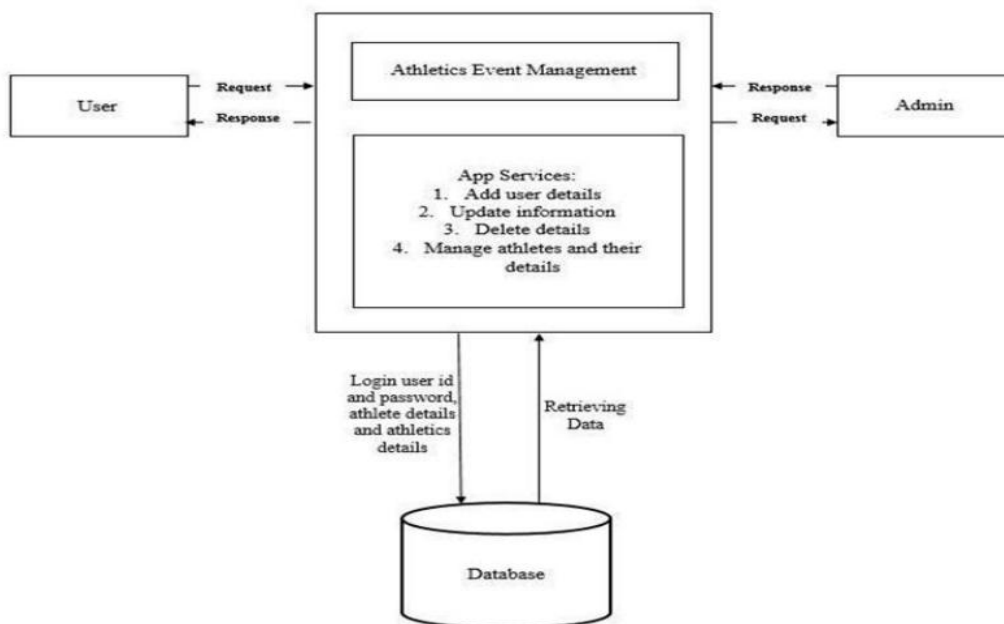
- ▶ Фронтенд (користувацький інтерфейс). Розробка інтуїтивного та зручного в користуванні інтерфейсу для введення та відображення даних спортсменів. Це може включати в себе веб-інтерфейс для тренерів та апаратів, а також мобільний додаток для спортсменів для введення і відстеження особистої інформації.
- ▶ Бекенд (серверна складова). Реалізація серверної частини для обробки запитань користувачів, зберігання та управління базою даних спортсменів. Важливо враховувати швидкодію та масштабованість системи.
- ▶ База даних. Створення ефективної бази даних для зберігання різноманітної інформації про спортсменів, такої як біографічні дані, медична інформація, результати змагань, статистика тренувань тощо.
- ▶ Системи аналізу даних та інтелектуального аналізу. Інтеграція інтелектуальних засобів аналізу даних для виявлення тенденцій, прогнозування результатів та надання рекомендацій тренерам для оптимізації тренувального процесу **(ЗАПРОПОНОВАНІ В РОБОТІ ПІДСИСТЕМИ Є ЇЇ СКЛАДОВИМИ)**



- ▶ Модуль аутентифікації та безпеки. Забезпечення безпеки даних спортсменів за допомогою надійних методів аутентифікації, а також контроль доступу до конфіденційної інформації.
- ▶ Інтеграція з зовнішніми джерелами даних. Можливість інтеграції з зовнішніми системами, такими як медичні бази даних, системи вимірювання показників фізичного стану, платформи для взаємодії з фанатами та інші.
- ▶ Модуль запитань та звітності. Розробка системи для створення запитань та генерації звітів для тренерів, менеджерів та інших зацікавлених сторін.
- ▶ Модуль сповіщень. Впровадження системи сповіщень для тренерів та спортсменів для швидкого сповіщення про важливі події та зміни.
- ▶ Модуль аналізу продуктивності. Розробка функціоналу для аналізу ефективності спортсменів під час тренувань та змагань.
- ▶ Модуль резервного копіювання та відновлення. Забезпечення можливості регулярного резервного копіювання та відновлення даних для запобігання втратам інформації.



Базова архітектура інформаційної системи керування даними спортсменів



17 A

Оцінка ефективності інформаційної системи керування даними спортсменів

- ▶ **Перевірка точності аналітичних моделей.** Необхідно оцінити точність моделей, які використовуються для інтелектуального аналізу даних. Це може бути оцінено за допомогою метрик, таких як точність прогнозувань та відхилення від реальних результатів.
- ▶ **Покращення рішень.** Необхідно визначити, наскільки інформаційна система допомагає у покращенні стратегій тренувань, визначенні індивідуальних потреб спортсменів та оптимізації підготовки для змагань.
- ▶ **Швидкість та продуктивність.** Проводиться оцінка швидкості та продуктивності системи при обробці та аналізі великого обсягу даних. Швидке реагування системи може значно покращити її ефективність.
- ▶ **Взаємодія з користувачами.** Оцінюється взаємодія системи з користувачами, зокрема тренерами та спортсменами. Визначається, наскільки легко вони можуть отримати доступ до аналітичної інформації та використовувати її для прийняття рішень.

18 A

- ▶ Відповідність вимогам. Відбувається перевірка, наскільки система відповідає вимогам і потребам організації та спортсменів. Це включає в себе функціональні та технічні вимоги.
- ▶ Захист даних. Необхідно оцінити рівень захисту конфіденційної інформації про спортсменів та інші дані, які обробляються системою.
- ▶ Вартість та повернення інвестицій. Оцінюється вартість розробки та впровадження системи в порівнянні з отриманими користями та покращеннями у веденні тренувань та організації змагань.
- ▶ Стабільність та надійність. Визначається наскільки стабільно працює система, та її надійність у випадку інтенсивного використання.

A
Пк
19 W

Висновки

- 1) проведено огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах;
- 2) набув подальшого розвитку метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині методів інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо;
- 3) набула подальшого розвитку інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині додавання підсистеми керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків.

Дякую за увагу!





Ім'я користувача:
Кафедра КІ

ID перевірки:
1016008786

Дата перевірки:
15.12.2023 16:10:47 EET

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
15.12.2023 16:11:04 EET

ID користувача:
100005591

Назва документа: Шатровський_v1_перевірка на плагіат

Кількість сторінок: 78 Кількість слів: 16152 Кількість символів: 126857 Розмір файлу: 1.82 MB ID файлу: 1015694440

2.9% Схожість

Найбільша схожість: 0.61% з джерелом з Бібліотеки (ID файлу: 1011002831)



0.63% Цитат



0% Вилучень

Немає вилучених джерел

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 0.0%

Словники перевірки: en_US, ru_RU, ua_UA. Помилки в документах: 6%

ID: 123382 Назва: ДП Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних Додано в БД: 2023-12-15 Автора: Шатровський А.О. Керівники: Гнагчук Є.Г. Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	103573	1528	1333 (1%)	25 (2%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

РЕЦЕНЗІЯ НА ДИПЛОМНУ РОБОТУ

Дипломник: Шатровський Андрій Олександрович

Тема: Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Спеціальність: 126 «Інформаційні системи та технології»

Обсяг дипломної роботи:

Кількість листів креслень —; кількість сторінок записки 77

1. Короткий зміст роботи та прийнятих рішень у роботі запропоновано метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині методів інтелектуального аналізу даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо: також набула подальшого розвитку інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних в частині додавання підсистеми керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків.

2. Висновок про відповідність роботи дипломному завданню Дипломна робота відповідає виданому завданню

3. Характеристика виконання кожного розділу, ступінь використання останніх досягнень науки і техніки і передових методів роботи: У першому розділі було проведено огляд існуючих моделей, методів та засобів інтелектуального аналізу даних в спортивних інформаційних системах, а також сформовано постановку задачі. У другому розділі на основі проведеного аналізу літературних джерел було запропоновано підсистему інформаційної системи керування даними спортсменів, яка включає рівень джерела даних, рівень збору та обміну даними, рівень центрального сховища, рівень аналізу даних і рівень додатків. У третьому розділі було розглянуто особливості роботи з спортивними даними, керування спортивними даними, методах аналізу спортивних великих даних та застосунках спортивних великих даних. У сфері великих даних у спорті відбулося кілька змін: від простого статистичного оцінювання до опрацювання спортивних великих даних; від простого статистичного оцінювання до оцінювання на основі моделей; від простого статистичного аналізу до прогнозування результатів спортсменів на основі даних; від аналізу соціальних мереж до аналізу графів знань; від явних спортивних особливостей до неявних спортивних особливостей.

Запропоновано метод створення інформаційної системи керування даними спортсменів в системі організації змагань, що дозволяє поєднати ефективність інформаційної системи з потужністю інтелектуального аналізу даних для покращення керування даними спортсменів у системі організації змагань, який на відміну від відомих базується на інтелектуальному аналізі даних про спортсменів, їх фізичних показників, показників продуктивності та вмотивованості, їх психологічного стану, тощо. У четвертому розділі представлені UML-діаграми інформаційної системи керування даними спортсменів та оцінку ефективності інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних.

4. Позитивні сторони роботи: Запропонована інформаційна система керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних, що дозволяє підвищити ефективності роботи тренерів для спостереження та відбору спортсменів на змагання.

5. Негативні сторони роботи: В роботі не приділено достатньої уваги практичному застосуванню конкретних методів інтелектуального аналізу даних.

6. Оцінка графічного оформлення та пояснювальної записки роботи: —

7. Відгук про роботу в цілому: В загальному робота виконана на достатньому професійному рівні

8. Інші зауваження: —

9. Оцінка дипломної роботи:

Розглянувши позитивні та негативні сторони представленої дипломної роботи вважаю, що робота заслуговує оцінки «добре» 4,25 (В)

Рецензент (прізвище, ім'я, по батькові, посада, місце роботи)

Богданович Леонід Петрович

“ ” 2023р.



Завідувачу кафедри КПС
д-р.техн.наук, проф. Говорущенко Т. О.

Шатровського Андрія Олександровича

ПІБ здобувача вищої освіти

ФІТ, 2 курсу, групи ІСТМ-22-1

ЗАЯВА

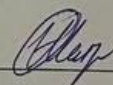
З правилами чинного Положення «Про дотримання академічної доброчесності в Хмельницькому національному університеті» від 26.09.2020 (зі змінами від 26.11.2020), згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування заходів дисциплінарної та академічної відповідальності, ознайомлений (а). Про використання програмно-технічних засобів для перевірки кваліфікаційних робіт здобувачів вищої освіти на плагіатоповіщений (а) та надаю свою згоду на обробку та збереження університетом моєї роботи в інституційному репозитарії університету.

Також надаю університету право на передачу моєї роботи для обробки та збереження в базах даних програмно-технічних засобів (Unicheck та Anti-Plagiarism) та використання роботи для виявлення плагіату в інших роботах, які перевіряються програмно-технічними засобами та користувачами, що мають доступ до цих програмно-технічних засобів, виключно в обмежених цілях для виявлення плагіату в текстах робіт.

Робота для перевірки університетом надається в друкованому та електронному варіанті. Електронна версія моєї роботи збігається (ідентична) з друкованою.

18.12.2023.

дата



підпис

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ
КАФЕДРИ КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ ТА ІНФОМАЦІЙНИХ СИСТЕМ
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: Метод створення інформаційної системи керування даними спортсменів в системі організації змагань на основі інтелектуального аналізу даних

Автор: Шатровський Андрій Олександрович

Спеціальність: 126 – Інформаційні системи та технології

Освітня програма: освітньо-професійна

Науковий керівник: Гнатчук Є.Г., к.т.н, доцент

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

- 1) усі запозичення фрагментарні, або мають належним чином оформленні посилання;
- 2) окремі виявлені збіги є загальноживаними фразами або виразами, про що свідчить посилання системи на збіг з джерелами на один фрагмент речення;

Сумарний обсяг всіх запозичень, визначений системою виявлення збігів/ ідентичності/схожості Unichек, складає 2.9% і адресується до 111 першоджерела; та системою Anti-Plagiarism складає 0%.

Керівник роботи

Гарант ОП

Завідувач кафедри КІС

Є. Г. Гнатчук

О. О. Павлова

Т. О. Говорущенко