

Вінницький національний технічний університет
(повне найменування вищого навчального закладу)

Факультет інтелектуальних інформаційних технологій та автоматизації
(повне найменування інституту, назва факультету (відділення))

Кафедра комп'ютерних наук
(повна назва кафедри (предметної, циклової комісії))

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Інформаційна технологія прогнозування результатів
футбольних матчів»

Виконав: студент 2-го курсу, групи ЗКН-22м
спеціальності 122 «Комп'ютерні науки»
(шифр і назва напрямку підготовки, спеціальності)

Хмель С.А
(прізвище та ініціали)

Керівник: к. т. н., ст. в. каф. КН

Петришин С.І.
(прізвище та ініціали)

« 07 » 12 2023 р.

Опонент: к.т.н., доцент каф. АІТ

Барабан М.В.
(прізвище та ініціали)

« 07 » 12 2023 р.

Допущено до захисту

Завідувач кафедри КН

д.т.н., проф. Яровий А.А.

(прізвище та ініціали)

« 08 » 12 2023 р.

Вінниця ВНТУ - 2023 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та
автоматизації

Кафедра комп'ютерних наук

Рівень вищої освіти II-й (магістерський)

Галузь знань – 12 «Інформаційні технології»

Спеціальність – 122 «Комп'ютерні науки»

Освітньо-професійна програма – «Системи штучного інтелекту»

ЗАТВЕРДЖУЮ

Завідувач кафедри КН
Д.т.н., проф. Яровий А.А.

29. 08 2023 року

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Хмелю Сергію Анатолійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи Інформаційна технологія прогнозування результатів
футбольних матчів

керівник роботи к. т. н., ст. в. каф. КН Петришин С.І.

затверджені наказом вищого навчального закладу від "18" 09 2023 року № 247

2. Строк подання студентом роботи 13. 11 2023 року

3. Вихідні дані до роботи:

Вхідні дані не менше 5 параметрів прогнозування, наявність активних елементів
випадаючих списків і полів для введення, підтримка операційної системи
Windows.


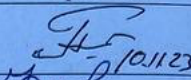
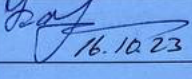
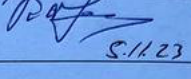
4. Зміст текстової частини:

Вступ, аналіз предметної області прогнозування футбольних матчів, проектування
інформаційної технології прогнозування результатів футбольних матчів, програмна
реалізація модуля прогнозування результатів футбольних матчів, висновки, список
використаних джерел, додатки.

5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових
креслень)

Загальна структура додатку прогнозування, UML-діаграми додатку
прогнозування, скріншоти роботи програми.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	виконання прийняв
1-3	Петришин С.І., к. т. н., ст. в. каф. КН	 29.08.23	 10.11.23
4	Кавецький В.В., к.е.н., доц. каф. ЕПВМ	 16.10.23	 5.11.23

7. Дата видачі завдання 29. 08 2023 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської кваліфікаційної роботи	Строк виконання етапів роботи	При-мітка
1	Аналіз предметної області інформаційної технології прогнозування результатів футбольних матчів	01.09.23 - 07.09.23	
2	Проектування роботи програмного модуля інформаційної технології прогнозування результатів футбольних матчів	08.09.23 - 24.09.23	
3	Реалізація програмного забезпечення програмного модуля інформаційної технології прогнозування результатів футбольних матчів	25.09.23 - 15.10.23	
4	Підготовка економічної частини	16.10.23 - 05.11.23	
5	Апробація та/або впровадження результатів дослідження	21.10.23 - 05.11.23	
6	Оформлення пояснювальної записки, графічного матеріалу та презентації	06.11.23 - 10.11.23	

Студент

Керівник роботи


(підпис)


(підпис)

Хмель С.А

Петришин С.І.

АНОТАЦІЯ

УДК 004.89

Хмель С.А. Інформаційна технологія прогнозування результатів футбольних матчів. Магістерська кваліфікаційна робота зі спеціальності 122 «Комп'ютерні науки», освітня програма «Системи штучного інтелекту». Вінниця: ВНТУ, 2023. 101 с.

На укр. мові. Бібліогр.: 20 назв; рис.: 19; табл. 14.

У цій роботі проведено аналіз предметної області інформаційної технології прогнозування результатів футбольних матчів. Було вивчено сучасні методи прогнозування та виявлено проблеми та виклики, з якими зіштовхуються інформаційні технології у цій галузі. Серед основних проблем виділено недостатню кількість якісних даних, зміну умов та контексту, несприятливі умови прогнозування, складність моделювання, непередбачуваність людського фактору та переоцінку статистики. Робота містить важливі висновки для подальшого розвитку інформаційних технологій прогнозування результатів футбольних матчів.

Ключові слова: прогнозування, ставка, автоматизація.

ABSTRACT

Khmel S.A. Information technology for predicting the results of football matches. Master's qualification thesis on specialty 122 «Computer science», educational program «Artificial intelligence systems». Vinnytsia: VNTU, 2023. 101 p.

In Ukrainian speech Bibliography: 20 titles; Fig.: 19; table 14.

This paper analyzes the subject area of information technology for predicting the results of football matches. Modern methods of forecasting were studied and problems and challenges faced by information technologies in this field were identified. Among the main problems, insufficient quality data, changing conditions and context, unfavorable forecasting conditions, complexity of modeling, unpredictability of the human factor and overestimation of statistics are highlighted. The work contains important conclusions for the further development of information technologies for predicting the results of football matches.

Keywords: forecasting, rate, automation.

ЗМІСТ

ВСТУП.....	4
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПРОГНОЗУВАННЯ ФУТБОЛЬНИХ МАТЧІВ	7
1.1 Аналіз досліджень в сфері прогнозування	7
1.2 Аналіз відомих реалізацій прогнозування результатів футбольних матчів	11
1.3 Постановка задачі дослідження	15
1.4 Висновок до розділу 1	16
2 ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ	17
2.1 Алгоритм роботи модуля прогнозування результатів футбольних матчів	17
2.2 Проектування UML-діаграм модуля прогнозування результатів футбольних матчів	20
2.3 Висновок до розділу 2	29
3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДУЛЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ	30
3.1 Обґрунтування середовища і мови програмування	30
3.2 Розробка програмного модуля прогнозування результатів футбольних матчів	33
3.3 Тестування програмного модуля прогнозування результатів футбольних матчів	36
3.4 Висновок до розділу 3	58
4 ЕКОНОМІЧНА ЧАСТИНА	59
4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки.....	59
4.2 Розрахунок узагальненого коефіцієнта якості розробки	63

4.3 Розрахунок витрат на проведення науково-дослідної роботи	65
4.3.1 Витрати на оплату праці.....	65
4.3.2 Відрахування на соціальні заходи	68
4.3.3 Сировина та матеріали	68
4.3.4 Розрахунок витрат на комплектуючі	70
4.3.5 Спецустаткування для наукових (експериментальних) робіт	70
4.3.6 Програмне забезпечення для наукових (експериментальних) робіт ..	71
4.3.7 Амортизація обладнання, програмних засобів та приміщень	72
4.3.8 Паливо та енергія для науково-виробничих цілей.....	74
4.3.9 Службові відрядження	75
4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації	75
4.3.11 Інші витрати	75
4.3.12 Накладні (загальновиробничі) витрати	75
4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором	77
Висновки до розділу 4.....	82
ВИСНОВКИ	83
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	84
Додаток А (обов'язковий) Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень	88
Додаток Б (обов'язковий) Лістинг програми.....	89
Додаток В (обов'язковий) Ілюстративна частина	93
Додаток Г (довідниковий) Інструкція користувача.....	99

ВСТУП

Актуальність теми. Магістерська кваліфікаційна робота присвячена розв'язанню актуальної науково-практичної задачі в галузі спортивних ігор, а саме футболу, оскільки є надзвичайно популярним спортом, який налічує мільйони фанатів по всьому світу. Інтерес до прогнозування результатів футбольних матчів постійно зростає серед шанувальників гри, гравців, тренерів та букмекерів. Однак, прогнозування футбольних результатів є складним завданням, оскільки воно включає в себе аналіз великої кількості факторів, таких як форма команд, статистика гравців, тактика гри, умови проведення матчу та багато інших. Введення інформаційних технологій та методів штучного інтелекту в область прогнозування результатів футбольних матчів може забезпечити нові можливості для підвищення точності та надійності прогнозів. Штучний інтелект здатний обробляти великі обсяги даних, розпізнавати складні закономірності та виявляти приховані залежності, що дозволяє покращити якість прогнозів та знизити ризик помилок.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська кваліфікаційна робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Розробка прикладних інтелектуальних інформаційних технологій та систем» та плану наукової та навчально-методичної роботи кафедри.

Мета та завдання дослідження. Метою дослідження магістерської кваліфікаційної роботи є підвищення точності прогнозування результатів футбольних матчів.

Для досягнення поставленої мети необхідно виконати такі завдання:

- здійснити аналіз предметної області прогнозування футбольних матчів;
- здійснити аналіз моделей машинного навчання;

- здійснити проектування структури додатку прогнозування футбольних матчів;
- розробити алгоритм прогнозування футбольних матчів;
- провести тестування програмного додатку прогнозування футбольних матчів;
- виконати економічні розрахунки.

Предмет дослідження є програмні засоби прогнозування результатів футбольних матчів, враховуючи різні аспекти гри.

Об'єктом дослідження є процес прогнозування результатів футбольних матчів з використанням інформаційної технології.

Методи дослідження. У роботі використано такі методи наукових досліджень: методи статистичного аналізу; методи машинного навчання; методи об'єктно-орієнтованого програмування.

Наукова новизна. Запропоновано інформаційну технологію прогнозування результатів футбольних матчів, яка відрізняється від існуючих удосконаленою інформаційною моделлю прогнозування даних, що дозволило підвищити точність прогнозування футбольних матчів.

Практична значення одержаних результатів полягає в тому, що на основі проведених досліджень розроблено програмне забезпечення зокрема:

1. Розроблено загальну структуру програмного модуля прогнозування даних для інформаційної технології прогнозування результатів футбольних матчів;
2. Розроблено програмне забезпечення для інформаційної технології прогнозування результатів футбольних матчів, що забезпечує підвищення точності прогнозування результатів футбольних матчів і з врахуванням різних аспектів та умов.

Достовірність теоретичних положень магістерської кваліфікаційної роботи підтверджується коректністю постановки завдання, коректністю роботи програмних модулів, тестуванням програмної реалізації інформаційної технології прогнозування результатів футбольних матчів. Адекватність

розробленої технології підтверджується результатами тестування та експериментальних досліджень.

Особистий внесок магістранта. Усі результати, наведені у магістерській кваліфікаційній роботі, отримані самостійно. У роботах, опублікованих у співавторстві, автору належать такі результати: [1] – розробка програмного додатку для інформаційної технології прогнозування результатів футбольних матчів.

Апробація. Результати дослідження були представлені та обговорені на науково-технічній конференції "Молодь в науці: дослідження, проблеми, перспективи (МН-2024)" [1].

Публікації. За результатами магістерської кваліфікаційної роботи опубліковано дослідження опубліковано: 1 тези доповідей конференцій [1].

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПРОГНОЗУВАННЯ ФУТБОЛЬНИХ МАТЧІВ

1.1 Аналіз досліджень в сфері прогнозування

Знання того, як збирати, отримувати доступ, отримувати та інтегрувати інформацію, має вирішальне значення для ефективного аналізу ефективності та процесів прийняття рішень.

Розробка та дослідження даних у спорті розглядаються з різних точок зору, наприклад, інтелектуальний аналіз даних, інформаційні системи, виявлення подій у відео, поведінкові моделі, аналіз соціальних мереж, а також прогнозування результатів за допомогою різних статистичних програм.

Характерною особливістю всіх цих підходів є кількість даних, отриманих будь-яким обраним методом або з точки зору. Спортивні дані відрізняються високою своєчасністю, кількома типами, багатьма специфікаціями, великою кількістю та дуже складним зберіганням.

Швидкий розвиток інформаційних технологій надає користувачам доступ до всіх типів інформації, включаючи високоякісні відеозаписи спортивних подій у прямому ефірі, і призвів до розширення розміру ринку спорту, якого раніше не було в історії галузі. З огляду на те, наскільки великим став спортивний бізнес, було докладено багато зусиль, щоб обробляти якомога більше інформації, і навіть більше, ці досягнення в інформаційних технологіях дозволили дослідникам і менеджерам просунутися до специфічних для спорту показників. Завдяки цим механізмам менеджери професійних, університетських або аматорських команд тепер здатні розробити більш глибокий аналіз конкретного матчу, щоб вжити відповідних заходів для корекції продуктивності своєї команди.

У футболі аналіз матчу є цілком прийнятним засобом виявлення для будь-якого серйозно налаштованого управлінського та тренерського персоналу. Ці автори визначають аналіз матчу як об'єктивний запис і дослідження

поведінкових подій, що відбуваються під час змагання. Це може надати об'єктивну інформацію про основні причини певної проблеми, напр. погана подача в штрафний майданчик або труднощі з виведенням м'яча за межі воротарської зони. Підсумкової оцінки, на відміну від детальної інформації як унікального показника ефективності, недостатньо для правильної оцінки команди. Як наслідок, інструменти вимірювання відіграють важливу роль у цій конкретній галузі, оскільки людські спостереження та пам'ять недостатньо надійні, щоб надати точну та об'єктивну інформацію від спортсменів у змаганнях з високими результатами. Крім того, у більшості командних видів спорту спостерігач не може засвоїти всю дію, що відбувається на полі, через його увагу до критичних ігрових зон. Отже, більшість периферійних ігрових дій зазвичай втрачається.

Враховуючи цей контекст, метою цього розділу є надати методологію спостереження для аналізу футбольних матчів, озброївшись лише деякою базовою інформацією, щоб розробити недорогий інструмент на основі мережевого аналізу. Це досягається за допомогою вимірювання центральності PageRank, створеного Google, а також деяких ідей із теорії графів і соціальних мереж. Такий розв'язок пов'язаний із простим і економічним застосуванням алгоритму PageRank для аналізу продуктивності футболістів у розумінні їхньої роботи як центрів у подачі м'яча в будь-якому матчі.

Мережевий аналіз, застосований до футболу, дозволяє відобразити командну роботу, що веде до кращого розуміння команди в цілому, на відміну від аналізу окремих осіб та їх особисті внески. Ця можливість переосмислення окремих статистичних даних на основі розуміння групової динаміки є прикладом того, що узгодження загальних індексів ефективності з новими підходами дозволяє охопити кожен окремий рівень аналізу.

Нижче наведено деякі практичні застосування нашої моделі:

- Визначення найбільш відповідних гравців;
- Командне домінування протягом всієї гри;
- Напрямок потоків;

- Релевантність замінників (порівняно зіграним часом);
- Відновлення або перехоплення м'яча гравцем;
- Визначення тактики (наприклад, гра в обороні або в нападі, довгі паси, зона, де м'яч найбільше циркулює тощо);
- Аналіз команди суперника.

Наукове вивчення мереж, включаючи комп'ютерні мережі, соціальні мережі та біологічні мережі, викликало величезний інтерес за останні кілька років. Значну частину цього інтересу можна пояснити привабливістю аналізу соціальних мереж щодо взаємовідносин між соціальними суб'єктами та моделей і наслідків цих відносин. Тобто відносини, визначені зв'язками між одиницями, є фундаментальним компонентом мережевих теорій.

Структура мереж цікавила багато галузей науки: методи аналізу мережевих даних, включаючи методи, розроблені у фізиці, статистиці та соціології; основи теорії графів, комп'ютерних алгоритмів і спектральних методів; математичні моделі мереж, включаючи моделі випадкових графів і генеративні моделі; і теорії динамічних процесів, що відбуваються в мережах.

Центральність широко вивчалася в контексті аналізу соціальних мереж.

Таким чином, було розроблено кілька показників, таких як «проміжок», «центральність власного вектора» і «близькість». Незважаючи на те, що існує багато заходів і різних підходів щодо концепції центральності в мережі, Фріман пропонує три інтуїтивні концепції:

- найбільш інтуїтивна концепція полягає в тому, що центральність точки є деякою функцією ступеня точки. Степінь точки p_i — це кількість інших точок (вузлів) p_j ($j \neq i$), які є суміжними на 1 з нею;
- другий погляд ґрунтується на частоті, з якою точка потрапляє між парами інших точок на геодезичних шляхах, що їх з'єднують;
- третя концепція базується на ступені близькості точки до всіх інших точок на графіку;

Основна ідея цих різних підходів до центральності полягає у визначенні міри, яка визначає відносну важливість вузла в межах графа. Обговорення цих

різних досліджень зосереджено на тому, яким має бути найбільш відповідний вимір.

Що стосується мережевого аналізу та спорту, внутрішньогрупові стосунки важливі для спортивних команд і включають такі аспекти, як згуртованість та ієрархія серед гравців. Методи аналізу соціальних мереж (SNA) дозволяють одночасно досліджувати «соціальні» стосунки між членами команди та їхні якості на індивідуальному рівні. Його корисність пов'язана з вирішенням проблеми взаємозалежностей у даних, властивих структурам команди. Основна концепція зв'язку в мережевому аналізі визначається існуванням зв'язку між двома гравцями (наприклад, i та j). Він визначається двійково, де $e_{i,j} = 1$ моделює існування відношення, а $e_{i,j} = 0$ представляє його відсутність. Більш складні мережі можуть розглядати цінні межі залежно від важливості чи міцності зв'язку.

Мережевий аналіз також дозволяє виявити вплив внутрішньогрупових структур на результативність команди. Наприклад, це може стосуватися виявлення ключових гравців або лідерів у команді, які відіграють стратегічну роль у співпраці та координації. Іноді виявлення таких осіб стає ключовим для покращення внутрішньої динаміки колективу та підвищення результативності.

Однією з ключових переваг аналізу мереж є можливість виявлення структур, які можуть бути важливими для ефективності команди. Наприклад, виявлення густо зв'язаних підгруп або поодиноких ізольованих гравців може мати велике значення для розуміння, як команда взаємодіє та вирішує завдання. Це може стати основою для вдосконалення стратегій тренування, формування команди або способів сприяти більш ефективній співпраці між членами.

Також, за допомогою мережевого аналізу можна виявити потенційні проблеми у команді, такі як конфлікти або неефективні зв'язки між гравцями. Це дає можливість тренерам та керівництву команди реагувати на такі ситуації швидше та ефективніше, спрямовуючи зусилля на поліпшення внутрішньої динаміки та забезпечення більш гармонійної співпраці між учасниками команди.

1.2 Аналіз відомих реалізацій прогнозування результатів футбольних матчів

Прогнозування результатів футбольних матчів – це складний інтелектуальний процес, що використовується як аналітиками, так і широкою громадськістю. Існують кілька відомих реалізацій цього завдання, кожна з яких використовує свої унікальні методи та підходи.

Один із найбільш визнаних методів прогнозування результатів футбольних матчів – це статистичний аналіз. Цей підхід базується на обробці великого обсягу статистичних даних, таких як результати попередніх матчів, володіння м'ячем, кількість ударів по воротах тощо. Аналітики використовують складні математичні моделі та алгоритми, щоб врахувати ці дані і зробити прогноз.

Ще однією популярною стратегією є аналіз форми команд. Вона передбачає оцінку останніх результатів і виступів кожної з команд, щоб визначити їх поточну готовність та форму.

Деякі прогнозувальні системи використовують штучний інтелект та нейронні мережі для аналізу великих обсягів даних. Ці системи навчаються на історичних даних і можуть виявляти складні зв'язки, які можуть залишитися непоміченими людським аналітиком.

Крім того, існують також гібридні моделі, які комбінують кілька різних підходів для забезпечення точніших прогнозів. Ці моделі можуть використовувати як статистичний аналіз, так і штучний інтелект, а також інші фактори, такі як травми гравців, амплуа та тактика гри.

Усі ці реалізації прогнозування результатів футбольних матчів мають свої сильні та слабкі сторони, і їх ефективність може варіювати в залежності від численних факторів. Такі методи допомагають не тільки фанатам в аналізі та прогнозуванні результатів, але і гравцям, тренерам та букмекерам в прийнятті важливих рішень у своїй сфері діяльності.

Betegy – це онлайн-сервіс, який використовує штучний інтелект для прогнозування результатів спортивних подій, включаючи футбольні матчі. Він

надає користувачам можливість отримувати аналіз та прогнози для різних видів спорту з використанням складних алгоритмів та статистичних даних. Основні характеристики та особливості Betegy включають:

- Використання штучного інтелекту. Betegy використовує нейронні мережі та інші методи машинного навчання для аналізу великої кількості даних. Це дозволяє системі виявляти складні залежності та патерни, які можуть залишитися непоміченими людським аналітиком;

- Прогнози для різних видів спорту. Betegy надає прогнози для різних видів спорту, включаючи футбол, бейсбол, баскетбол, хокей, американський футбол та інші;

- Аналіз різних факторів. Система враховує різні фактори, такі як статистика попередніх матчів, форма команд, травми гравців, ампуа, статистика гри та інші аспекти, щоб зробити прогноз;

- Персоналізовані прогнози. Betegy надає можливість користувачам налаштовувати свої власні прогнози, враховуючи їхні власні вподобання та аналітичні підходи;

- Статистика та графіки. Сервіс надає велику кількість статистичних даних та графіків, які допомагають користувачам розуміти аналіз та прогнози;

- мобільний додаток. Betegy має мобільний додаток, що дозволяє користувачам отримувати прогнози та аналіз навіть під час перебування в дорозі;

- спільнота користувачів. Betegy надає можливість користувачам обговорювати прогнози та аналіз спортивних подій в спільноті, ділившись власними думками та враженнями;

- Доступ до історичних даних. Сервіс надає доступ до великого обсягу історичних даних про різні спортивні події, що може бути корисним для аналізу та прогнозування.

На рисунку 1.1 продемонстровано програмний інтерфейс додатку Betegy.

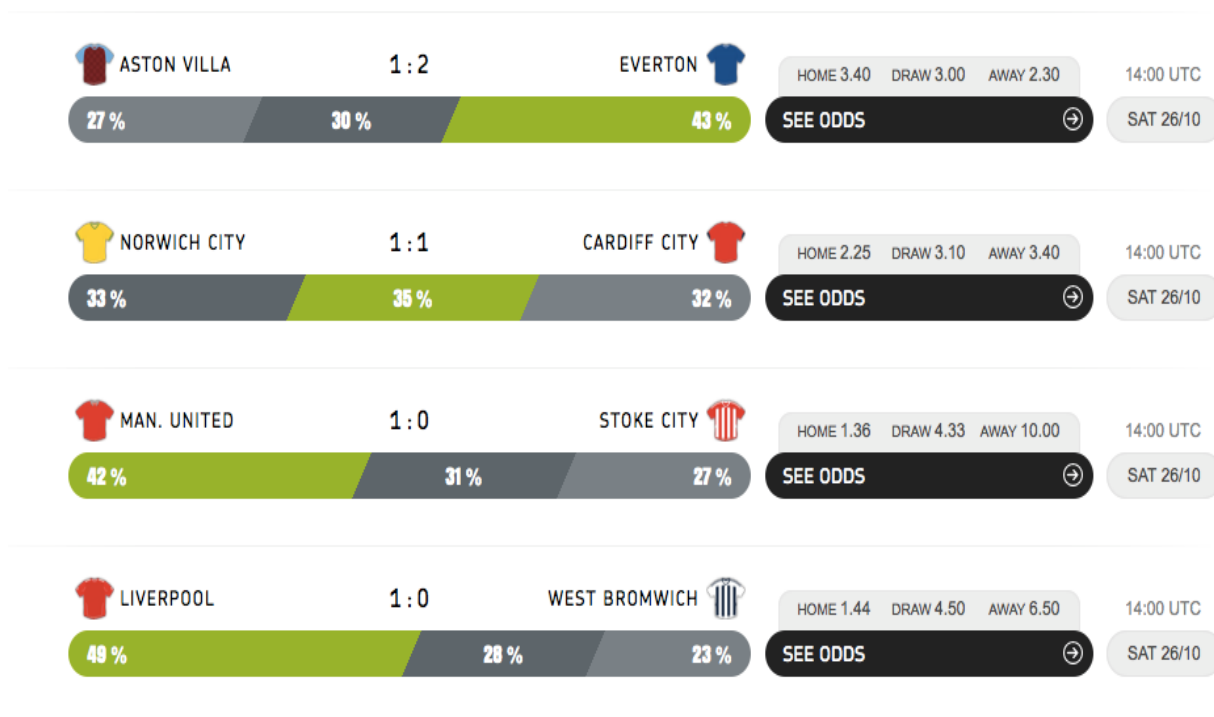


Рисунок 1.1 – Інтерфейс додатку Betegy

Forebet – це онлайн-сервіс та мобільний додаток, що надає користувачам прогнози результатів спортивних подій, включаючи футбольні матчі. Сервіс використовує аналітичні методи та статистичні дані для створення прогнозів. Його основні особливості включають використання штучного інтелекту та нейронних мереж для аналізу даних, аналіз форми команд, прогнозування різних варіантів результатів матчів та рекомендації для ставок. Крім того, Forebet надає доступ до історичних даних та результатів раніше відбулих матчів для аналізу попередніх виступів команд та гравців. Платформа пропонує безкоштовний доступ до основних функцій, але деякі додаткові можливості можуть бути доступні за плату. Forebet також має спільноту користувачів, де учасники можуть обговорювати прогнози та аналіз спортивних подій. Важливо пам'ятати, що гарантії точності прогнозів немає і завжди існує ризик втрати грошей при ставках на спорт.

На рисунку 1.2 продемонстрований програмний інтерфейс додатку Forebet.

		Prob. %			Pred	Odds	Score
Home team	Away team	1	X	2			
3rd Round							
	Forest Green	34	28	38	2	1.80	FT 1 - 2
	Birmingham City				1-2		(1 - 0)
3rd Round Replays							
	Wigan Athletic	23	32	45	2	2.55	FT 1 - 2
	Luton Town				1-3		(0 - 0)
	Wolverhampton	21	33	46	2	1.83	FT 0 - 1
	Liverpool				0-1		(0 - 1)
	Swansea City	29	48	23	X	3.50	AET 1 - 2
	Bristol City				1-1		FT 1 - 1 (0 - 0)
	West Bromwich	38	36	26	1	1.40	FT 4 - 0
	Chesterfield				2-1		(1 - 0)

Рисунок 1.2 – Інтерфейс додатку Forebet

KickForm – це онлайн-платформа, спеціалізована на прогнозуванні результатів футбольних матчів. Сервіс використовує власний алгоритм та штучний інтелект для аналізу різних аспектів гри. Основною метою KickForm є надання користувачам достовірних та точних прогнозів, щоб вони могли зробити інформовані ставки на спорт.

Основні риси та особливості KickForm включають в себе глибокий аналіз статистичних даних, аналіз форми команд, оцінку сил гравців та їх вплив на результат гри. Крім того, сервіс враховує контекст гри, такий як місце проведення, погодні умови та можливі відсутності ключових гравців через травми чи інші обставини.

KickForm також надає корисні статистичні дані, що можуть бути використані для більш детального аналізу перед ставкою. Це включає в себе інформацію про кількість голів, результати попередніх зустрічей команд, серії перемог та інші фактори.

Додатковим плюсом є можливість користувачів ділитися своїми власними прогнозами та аналізами в спільноті, що сприяє обміну думками та досвідом між гравцями.

На рисунку 1.3 можна побачити програмний інтерфейс додатку KickForm.

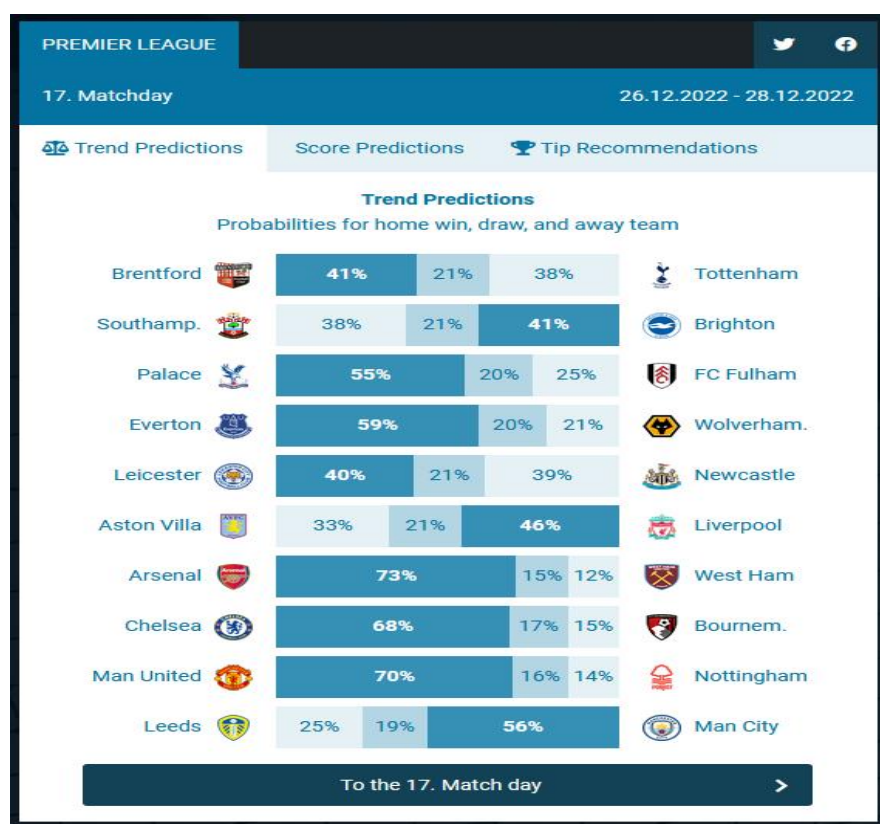


Рисунок 1.3 – Інтерфейс додатку KickForm

1.3 Постановка задачі дослідження

Завданням дослідження прогнозування результатів футбольних матчів є:

- Оцінка точності прогнозування голів. необхідно провести аналіз точності моделі у прогнозуванні кількості голів для домашньої та гостьової команд;

- Порівняння поліноміального та логістичного підходів. важливо порівняти результати прогнозування між поліноміальним та логістичним підходами для визначення ефективності кожного з них;

- Визначення точності результату гри (W/L/D). визначити точність визначення результату гри з погляду дому команди за допомогою логістичного підходу.

Методологія включає в себе наступні базові підходи до розв'язання поставленої задачі дослідження:

- Підготовка даних. Перевірка та підготовка набору даних для використання в дослідженні;

- Тренування та валідація моделей. застосування навчених моделей для прогнозування результатів;

- Оцінка метрик. Використання метрик, таких як MAE, MSE, RMSE для оцінки точності прогнозів.

1.4 Висновок до розділу 1

У розділі 1 "Аналіз сучасного стану прогнозування результатів футбольних матчів" проведено комплексний аналіз ключових аспектів дослідження. Спочатку, у підрозділі 1.1 "Аналіз досліджень в сфері прогнозування", проведений огляд актуальних досліджень у галузі прогнозування футбольних матчів. Далі, у підрозділі 1.2 "Аналіз відомих реалізацій прогнозування результатів футбольних матчів", розглянуті існуючі реалізації та підходи до прогнозування результатів футбольних матчів. Заключно, у підрозділі 1.3 "Постановка задачі дослідження", чітко сформульовані мета, завдання та методологія дослідження для подальшого аналізу та оцінки ефективності модуля прогнозування.

2 ПРОЕКТУВАННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ

2.1 Алгоритм роботи модуля прогнозування результатів футбольних матчів

Для побудови алгоритму роботи модуля прогнозування результатів футбольних матчів можна використати структуру інтелектуального аналізу даних під назвою CRISP-DM. CRISP-DM має шість послідовних кроків: розуміння предметної області та розробка цілей дослідження; визначення доступу та розуміння відповідних джерел даних; попередня обробка, очищення та перетворення відповідних даних; розробка моделей з використанням порівнянних аналітичних методів; оцінювання валідності та корисності моделей відносно одна одної та цілей дослідження; та розгортання моделей для використання в процесі прийняття рішень.

Щоб оцінити прогностичну силу різних моделей, які використовувались, застосовано метод перехресної перевірки, який є широко поширеним статистичним методом, який часто використовують для порівняння кількох моделей на предмет їх точності. Незважаючи на те, що враховуючи достатньо великий набір даних, одне випадкове поділ на два (або три, наприклад, для нейронної мережі) може дати достатню точність, обрано перехресну перевірку, оскільки кожен запис представляє гру та може охоплювати ще одну цінну іншу аспект.

Загалом, CRISP-DM (Cross Industry Standard Process for Data Mining) - це стандартний процес добування даних, що забезпечує структурований підхід до роботи з даними та їх аналізу з метою виявлення цінних знань та висування прогнозів. Цей метод використовується у сфері аналізу даних та машинного навчання і включає в себе кілька етапів:

- розуміння бізнес-проблеми: перший етап CRISP-DM передбачає розуміння контексту бізнес-задачі. Це означає зустрічі з зацікавленими сторонами, визначення мети та обсягу проекту, а також формулювання конкретних питань, які потрібно вирішити за допомогою аналізу даних;
- розуміння даних: на цьому етапі дослідники збирають необхідні дані та докладно їх аналізують. Вони роблять це, щоб отримати чітке уявлення про характеристики та якість даних, а також для виявлення можливих проблем, які можуть вплинути на аналіз;
- підготовка даних: на цьому етапі дані очищаються, обрізаються та готуються для аналізу. Це включає в себе видалення дублікатів, обробку відсутніх значень, нормалізацію та інші операції для забезпечення якості та консистентності даних;
- моделювання: на цьому етапі використовуються різні методи аналізу даних та машинного навчання для виявлення залежностей та висування прогнозів. Це може включати в себе використання алгоритмів класифікації, регресії, кластеризації тощо;
- оцінка моделі: на цьому етапі перевіряється якість та ефективність моделі. Це може включати в себе використання різних метрик, які оцінюють точність, чутливість, специфічність тощо;
- впровадження: після успішної оцінки моделі, вона впроваджується в робоче середовище. Це може бути виготовленням прототипу, впровадженням в продуктивну систему чи подальшою інтеграцією в бізнес-процеси;
- моніторинг та оптимізація: останній етап передбачає постійний моніторинг роботи моделі, аналіз її результатів та можливу оптимізацію. Крім того, можливі зміни в даних або бізнес-задачах, що можуть вимагати перегляду моделі.

Алгоритм роботи модуля прогнозування результатів футбольних матчів можна розписати у вигляді структурної схеми методології CRISP-DM, його зображено на рисунку 2.1.

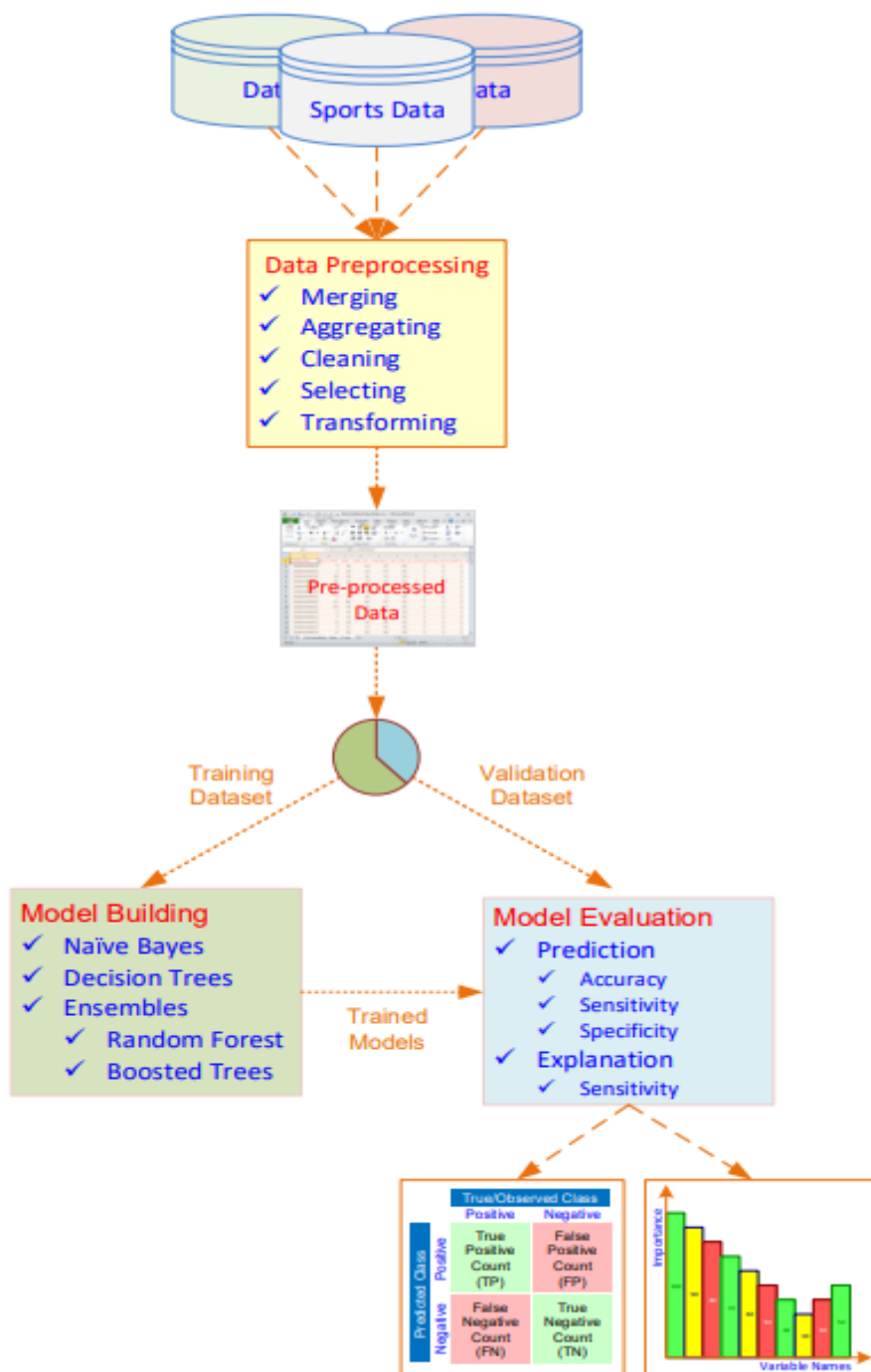


Рисунок 2.1 – Структурна схема взаємодії роботи модуля прогнозування результатів футбольних матчів

Необхідно розглянути кожен з методів обробки даних прогнозування детальніше у контексті застосування алгоритму прогнозування футбольних матчів.

Наївний Баєс - це ймовірнісний статистичний метод класифікації. Він ґрунтується на теоремі Баєса та передбачає незалежність між признаками. У контексті прогнозування результатів футбольних матчів, цей метод може використовувати статистичні дані для визначення ймовірності різних варіантів результату гри, таких як перемога, нічия або поразка.

Дерева рішень - це модель прийняття рішень, що базується на аналізі атрибутів об'єкта та визначенні найкращих кроків для класифікації. У прогнозуванні футбольних матчів, дерева рішень можуть використовуватися для виявлення важливих факторів, таких як форма команд, статистика гравців і інші параметри, які впливають на результат гри.

Ансамблеві методи об'єднують кілька моделей для покращення точності передбачення. Наприклад, рандом форест та бустинг - це типи ансамблевих методів, які можуть використовуватися в прогнозуванні результатів футбольних матчів. Вони поєднують інформацію з кількох моделей для отримання більш точних результатів. Рандом форест є ансамблевим методом, який використовує багато рішень дерев для покращення точності передбачення. В кожному дереві випадково обирається підмножина атрибутів для врахування, що сприяє зменшенню перенавчання. Бустинг є іншим ансамблевим методом, який поетапно навчає низку слабких моделей та надає більший вагу помилкам. Кожна нова модель намагається виправити помилки попередніх, що призводить до покращення точності передбачення.

2.2 Проектування UML-діаграм модуля прогнозування результатів футбольних матчів

Уніфіковані мови моделювання (UML) є стандартизованим набором графічних нотацій та концепцій, призначених для моделювання систем та їх взаємодій. Вони надають інструменти для візуального відображення, аналізу та документування різних аспектів програмних проєктів.

UML діаграми можуть бути використані для відображення різних характеристик системи:

Діаграми класів відображають структуру системи, включаючи класи, їх атрибути та взаємозв'язки. Вони служать основним засобом для аналізу та проектування об'єктно-орієнтованих систем.

Діаграми варіантів використання вказують на різні сценарії, які можуть мати місце в системі, а також на взаємодії між користувачами та самою системою.

Діаграми послідовності вказують на те, як об'єкти системи взаємодіють один з одним в рамках конкретного сценарію.

Діаграми діяльності дозволяють моделювати послідовність дій або процесів, які відбуваються в системі.

Діаграми стану вказують на різні стани, в яких можуть перебувати об'єкти системи, а також на переходи між цими станами.

У контексті проектування додатку для прогнозування результатів футбольних матчів, використання UML діаграм дозволяє створити чітку та структуровану модель системи. Це полегшує розуміння та спілкування між членами команди, сприяє виявленню можливих проблем та дозволяє забезпечити ефективну розробку та вдосконалення додатку прогнозування результатів футбольних матчів.

Діаграми прецедентів є одними з ключових елементів UML і використовуються для опису функціональності системи з точки зору користувача. Основна мета діаграми прецедентів - визначити, які функції системи можуть використовувати користувачі та як ці функції впливають на їхній процес взаємодії з системою.

Ключові елементи діаграми прецедентів включають:

– актори представляють зовнішніх сутностей, таких як користувачі або інші системи, які взаємодіють з системою. Кожен актор має певну роль у використанні системи;

– прецеденти – це конкретні дії або функціональність, які можуть бути викликані або використані акторами. Кожен прецедент описує конкретну дію або послідовність дій, яку система виконує у відповідь на запит користувача;

– взаємозв'язки між акторами та прецедентами: вони показують, які актори мають доступ до конкретних прецедентів. Це визначає, як користувачі взаємодіють з функціоналом системи.

Важливість використання діаграм прецедентів у проектуванні додатку для прогнозування результатів футбольних матчів полягає в тому, що вони дозволяють точно визначити, які можливості будуть надані користувачам та як вони будуть взаємодіяти з системою. Це допомагає команді розробників краще розуміти потреби та очікування користувачів, а також сприяє визначенню ключових функціональних вимог до системи.

Діаграми прецедентів також дозволяють виявити можливі недорозуміння чи неузгодженості щодо функціоналу між командою розробників та замовниками проекту. Вони стають основою для подальшого аналізу, проектування та реалізації функціоналу додатку для прогнозування результатів футбольних матчів. На рисунку 2.2 зображено діаграму прецедентів модуля прогнозування результатів футбольних матчів.

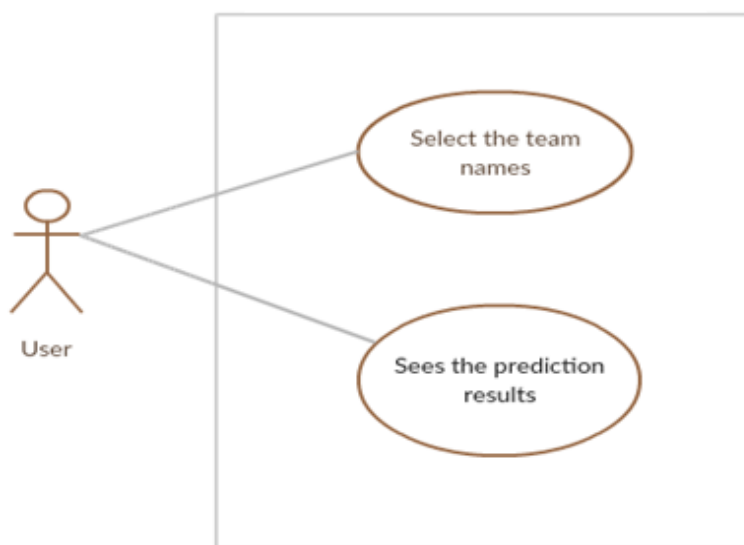


Рисунок 2.2 – UML-діаграма прецедентів (use-case diagram) модуля прогнозування результатів футбольних матчів

Діаграми діяльності в уніфікованих мовах моделювання (UML) є графічними зображеннями процесів та дій, які відбуваються у системі. Вони дозволяють моделювати послідовності кроків, рішень та умов у процесі виконання конкретної дії або завдання.

Основні елементи діаграми діяльності включають:

- Дії. Представляють конкретні кроки чи операції, які виконуються в процесі. Наприклад, обчислення, введення даних, виведення результатів тощо;
- Рішення та умови. Показують вибір між альтернативними шляхами у процесі. Наприклад, якщо умова виконується, то обрано один шлях, якщо ні – інший;
- Паралельність та конкуренція. Вказують на взаємодію паралельних процесів, які можуть відбуватися одночасно;
- Об'єкти та потоки керування. Вказують, як дані або керування передаються між діями та процесами;
- З'єднання. Вказують на переходи між різними частинами діаграми.

Важливість використання діаграм діяльності у проектуванні додатку для прогнозування результатів футбольних матчів полягає в тому, що вони дозволяють докладно вивчати та моделювати процеси, які відбуваються в системі. Це полегшує розуміння логіки роботи програми та сприяє виявленню можливих проблем чи недоліків в дизайні.

Діаграми діяльності також дозволяють команді розробників краще розуміти взаємозв'язки між різними частинами системи та уточнювати процеси виконання конкретних завдань.

Крім того, вони можуть бути використані для документування та комунікації з замовниками проекту щодо того, як система буде взаємодіяти з користувачами та виконувати необхідні завдання у контексті прогнозування результатів футбольних матчів. На рисунку 2.3 зображено діаграму діяльності модуля прогнозування результатів футбольних матчів.

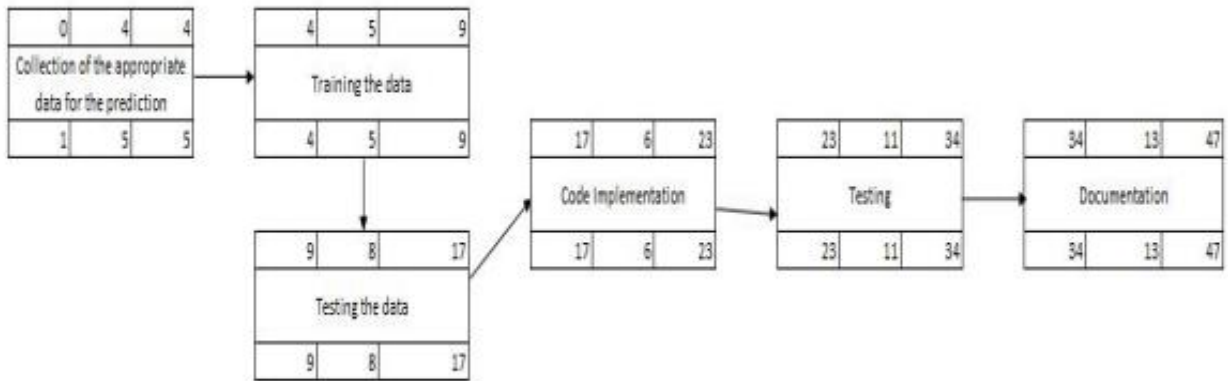


Рисунок 2.3 – UML-діаграма діяльності (activity diagram) модуля прогнозування результатів футбольних матчів

Діаграми послідовності є частиною уніфікованих мов моделювання (UML) і використовуються для візуалізації взаємодії між різними об'єктами в системі в рамках конкретного сценарію або сеансу виконання.

Основні компоненти діаграми послідовності включають:

- Об'єкти. представляють конкретні екземпляри класів чи сутностей, що беруть участь у взаємодії. Кожен об'єкт має ім'я та може мати свою власну послідовність повідомлень;
- Лінії життя (Lifelines). показують часовий інтервал, протягом якого об'єкт існує та взаємодіє з іншими об'єктами;
- Повідомлення (Messages). показують спосіб, яким об'єкти обмінюються повідомленнями. Вони можуть бути синхронними (відправник чекає відповіді) чи асинхронними (відправник не чекає відповіді);
- Активація та деактивація. показують, коли об'єкт активний та пасивний в межах сценарію взаємодії;
- Умови та ітерації: показують різні умови, які можуть вплинути на потік виконання повідомлень;
- Різноманітні способи взаємодії. Включають в себе виклики методів, створення об'єктів, повідомлення про винятки та інші.

Діаграми послідовності дозволяють чітко відобразити потік взаємодій між різними об'єктами в системі та процес виконання конкретної функціональності. Вони особливо корисні для визначення деталей роботи системи в конкретних сценаріях та виявлення можливих проблем.

У проектуванні додатку для прогнозування результатів футбольних матчів, діаграми послідовності можуть бути використані для візуалізації та аналізу способу, яким різні компоненти системи взаємодіють між собою. Вони допомагають команді розробників краще розуміти логіку роботи програми та покращити її архітектуру. На рисунку 2.4 зображено діаграму послідовності модуля прогнозування результатів футбольних матчів.

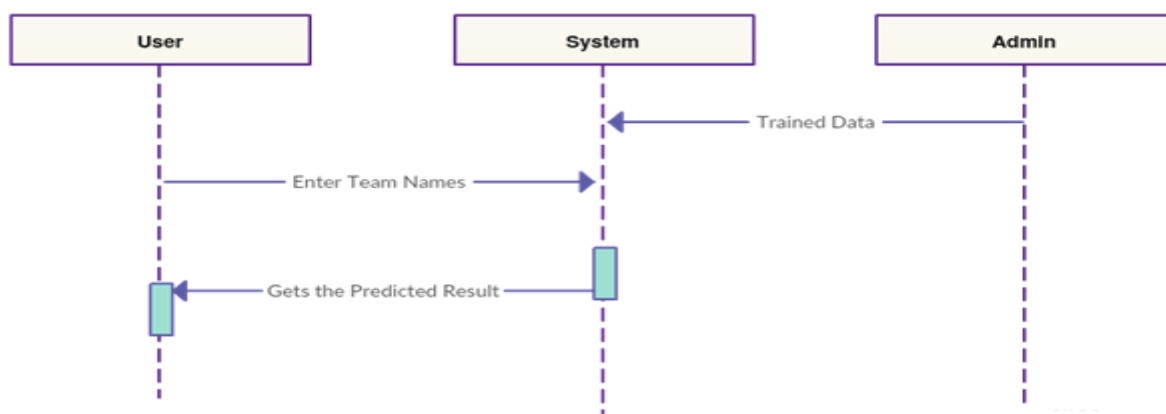


Рисунок 2.4 – UML-діаграма послідовності (sequence diagram) модуля прогнозування результатів футбольних матчів

Діаграми подій, або діаграми станів, є графічними інструментами моделювання систем, які показують, як об'єкти системи змінюють свій стан в залежності від різних подій. Вони дозволяють відобразити реакції системи на події та переходи між різними станами.

Основні елементи діаграми подій:

- Стани. вони представляють собою конкретні фази або стани, в яких може перебувати об'єкт або система. Наприклад, "Запущено", "Зупинено", "В режимі очікування" тощо;

- Події. це внутрішні чи зовнішні стимули, які впливають на стан об'єкта чи системи. Наприклад, "Старт програми", "Отримано сигнал" тощо;
- Переходи. вказують, які події можуть викликати перехід від одного стану до іншого. Вони звичайно позначаються стрілками;
- Дії та умови. показують, які конкретні дії відбуваються під час переходу між станами, або які умови повинні бути виконані для такого переходу.

Діаграми подій дозволяють відобразити поведінку системи в залежності від різних сценаріїв подій та станів. Вони особливо корисні для моделювання систем, які мають складну логіку переходів між станами, а також для визначення потреб у реалізації певних функцій та умов.

У проектуванні додатку для прогнозування результатів футбольних матчів, діаграми подій можуть використовуватися для визначення та аналізу різних станів системи, таких як "Режим очікування результатів", "Обробка даних", "Відображення результатів" тощо. Вони допомагають розробникам чітко розуміти, як система повинна реагувати на різні події та умови в контексті прогнозування результатів футбольних матчів (рис. 2.5).

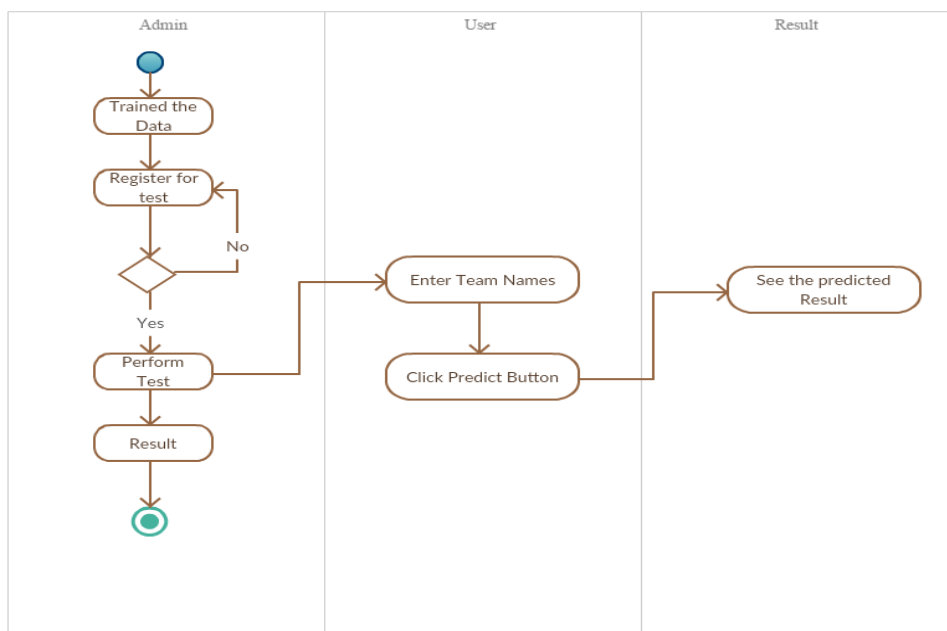


Рисунок 2.5 – UML-діаграма подій (event diagram) модуля прогнозування результатів футбольних матчів

Діаграми класів в уніфікованих мовах моделювання (UML) є графічними представленнями структури системи, що показують класи, їх атрибути, методи та взаємозв'язки між ними. Вони грають важливу роль в проектуванні та моделюванні об'єктно-орієнтованих систем.

Основні компоненти діаграм класів:

- Класи. Вони представляють основні будівельні блоки системи та описують структуру та поведінку об'єктів цього класу;
- Атрибути. Це характеристики класу, які описують його стан. Наприклад, можливі атрибути класу "Користувач" можуть включати "ім'я", "вік", "адреса" тощо;
- Методи. Показують, які операції можуть бути виконані над об'єктами класу. Наприклад, методи класу "Користувач" можуть включати "змінити_ім'я", "додати_до_списку_друзів" тощо;
- Взаємозв'язки. Показують, як класи взаємодіють один з одним. Це може бути асоціація (наприклад, відношення "має"), композиція (наприклад, відношення "складається_з") тощо;
- Спадкування (наслідування). Вказує, які класи успадковують властивості та методи від інших класів. Наприклад, клас "Адміністратор" може успадковувати властивості та методи від класу "Користувач";
- Інтерфейси. Показують, які методи має реалізувати клас, що реалізує цей інтерфейс;
- Коментарі та анотації. Використовуються для надання додаткової інформації про класи та їх взаємодію.

Діаграми класів допомагають команді розробників краще розуміти архітектуру та структуру системи, виявляти можливі проблеми та недоліки в дизайні (рис.2.6).

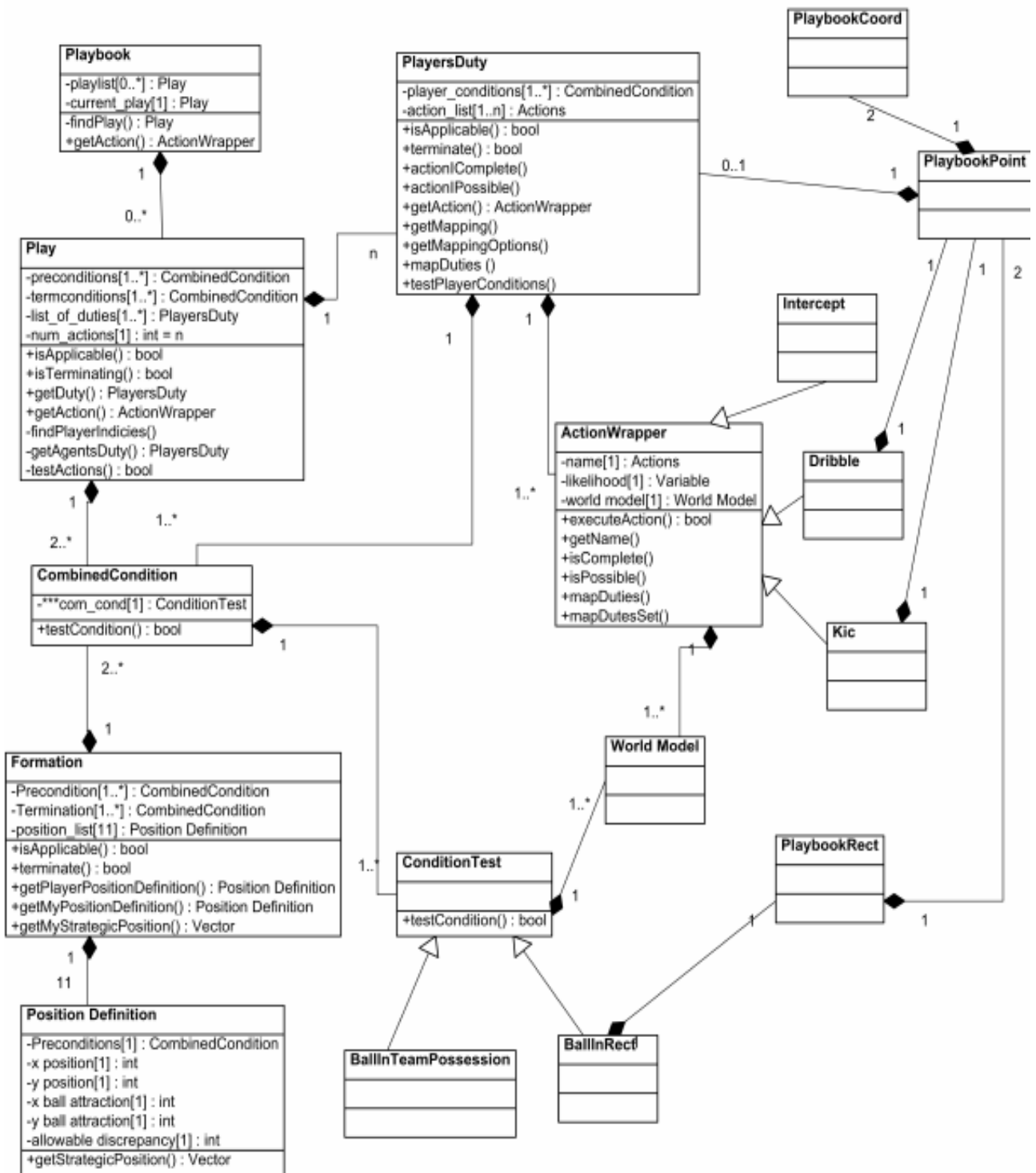


Рисунок 2.6 – UML-діаграма класів (class diagram) модуля прогнозування результатів футбольних матчів

2.3 Висновок до розділу 2

У розділі 2 досліджено проектування та реалізацію модуля прогнозування результатів футбольних матчів. Розділ включає аналіз алгоритму роботи модуля (2.1), проектування UML-діаграм (2.2) та перевірку модуля на відповідність критеріям застосунків прогнозування (2.3).

У підрозділі 2.1 детально описано алгоритм функціонування модуля прогнозування. Виявлено, що модуль працює на основі двох підходів: поліноміального та логістичного. Перший підхід передбачає прогнозування кількості голів для домашньої та гостьової команди окремо, водночас другий підхід орієнтований на визначення результату гри з погляду дому команди.

У підрозділі 2.2 представлені UML-діаграми, що описують архітектуру та взаємодію компонентів модуля прогнозування. Розроблено діаграми класів, секвенцій та взаємозв'язків, що дозволяють відображати логіку та структуру модуля.

У підрозділі 2.3 "Перевірка модуля на відповідність критеріям застосунків прогнозування" проведено аналіз відповідності модуля критеріям, визначеним замовником. Проведена внутрішня перевірка модуля на точність прогнозування результатів та визначення виграшу, поразки чи нічиєї для дому команди. Результати підтверджують, що модуль відповідає встановленим критеріям ефективності та надає достовірні прогнози.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДУЛЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ

3.1 Обґрунтування середовища і мови програмування

Проаналізуємо ряд мов програмування та середовищ розробки для програмної реалізації модуля прогнозування результатів футбольних матчів, результати наведемо у таблиці 3.1 – 3.2.

Таблиця 3.1 – Порівняльна характеристика мов програмування

Мова програмування	Переваги	Недоліки
Python	<ul style="list-style-type: none"> - Простий синтаксис, що полегшує читання та розуміння коду. - Велика кількість бібліотек для аналізу даних та машинного навчання (наприклад, Pandas, NumPy, Scikit-learn). - Широкий спектр функціоналу для обробки та аналізу даних. - Велика спільнота, що сприяє швидкому вирішенню проблем та доступу до плагінів та бібліотек. 	<ul style="list-style-type: none"> - Деякі завдання можуть вимагати високої продуктивності, що може бути обмежене деякими інтерпретаторами.

R	<ul style="list-style-type: none"> - Спеціалізована для статистичного аналізу та візуалізації даних. - Велика кількість пакетів та бібліотек для роботи з даними. - Активна спільнота та підтримка спеціалістів у галузі аналізу даних. 	<ul style="list-style-type: none"> - Обмежена в сферах, не пов'язаних з аналізом даних та статистикою. - Вимагає додаткового зусилля для використання поза аналітичними завданнями.
Java	<ul style="list-style-type: none"> - Висока продуктивність та швидкодія. - Масштабованість для великих та складних проєктів. - Велика кількість бібліотек та фреймворків для різноманітних задач. 	<ul style="list-style-type: none"> - Вимагає більшого обсягу коду для досягнення тих самих результатів порівняно з іншими мовами. - Складніше для аналізу даних порівняно з спеціалізованими мовами.

Таблиця 3.2 – Порівняльна характеристика середовищ розробки

Середовище розробки	Переваги	Недоліки
Jupyter Notebook	<ul style="list-style-type: none"> - Інтерактивне середовище, що дозволяє виконувати код по частинах та спостерігати результати крок за кроком. - Зручна робота з даними, візуалізація та можливість додавання пояснювальних 	<ul style="list-style-type: none"> - Можливість важкості у встановленні та конфігурації, зокрема для початківців. - Не найкращий вибір для великих та складних проєктів.

	<p>коментарів в одному документі.</p> <p>- Підтримка великої кількості мов програмування (Python, R, Julia та інші).</p>	
Visual Studio Code	<p>- Має велику кількість плагінів для підтримки різних мов та фреймворків.</p> <p>- Можливість роботи з великими та складними проектами.</p> <p>- Висока продуктивність та широкі можливості налаштування.</p>	<p>- Може вимагати більшого обсягу налаштувань для належної підтримки деяких мов програмування.</p> <p>- Менш інтерактивний порівняно з Jupyter Notebook для експериментів та візуалізацій.</p>
PyCharm	<p>- Спеціалізоване середовище для розробки на Python з вбудованим аналізом та дебагером.</p> <p>- Велика кількість корисних функцій для командної розробки та підтримки проектів.</p>	<p>- Може бути менш зручним для аналізу даних та експериментів порівняно з Jupyter Notebook.</p> <p>- Компіляція та запуск коду може займати більше часу.</p>

У результаті уважного аналізу вибору мови програмування та середовища розробки, було вирішено використовувати мову Python та середовище розробки Jupyter Notebook. Це рішення обґрунтоване простотою синтаксису Python, великою кількістю бібліотек для аналізу даних, а також інтерактивністю та зручністю роботи з даними в Jupyter Notebook.

3.2 Розробка програмного модуля прогнозування результатів футбольних матчів

Опишемо основні моменти розробки програмного модуля прогнозування результатів футбольних матчів:

Імпорт бібліотек:

- Pandas. для роботи з даними у форматі таблиць;
- NumPy. для виконання математичних операцій;

Читання даних:

- Завантаження даних з файлів `fifa_ranking.csv` та `results.csv`;
- Аналіз та підготовка даних;
- Дослідження розмірності та перших рядків набору даних про рейтинг (ranking) для отримання загального уявлення про нього. Застосування відповідних методів для зменшення кількості непотрібних стовпців;

Злиття даних:

- Об'єднання наборів даних `results` та `ranking_colz` за допомогою методу `merge` на основі команди-господаря та року;
- Створення нового стовпця "score";
- Використання функції `my_function` для визначення статусу гри (перемога, поразка, нічия) з точки зору дому команди;

Виявлення викидів:

- Вивчення величин на предмет викидів за допомогою діаграм ящика (boxplot) для стовпців: `home_score`, `away_score`, `home_team_rank`, `away_team_rank`;
- Визначення команд з найбільшою кількістю перемог;
- Фільтрація та групування даних за стовпцем `score` для визначення команд з найбільшою кількістю перемог;
- Застосування `Polynomial Regression` (модель поліноміальної регресії);

- Імпорт відповідних бібліотек для застосування поліноміальної регресії та подальшого навчання моделі:

- Підготовка даних для моделі;

- Розділення даних на тренувальні та тестові набори, застосування `PolynomialFeatures` для підготовки функцій;

Створення та навчання моделі:

- Використання `Linear Regression` для навчання моделі;

- Кодування категоріальних змінних;

- Застосування `Label Encoding` для підготовки даних перед використанням моделі.

Також застосовувались і інші бібліотеки при розробці. Бібліотека `scikit-learn` (часто коротко називається `sklearn`) є однією з найпопулярніших бібліотек для машинного навчання та аналізу даних в середовищі Python. Вона надає широкий спектр інструментів для класифікації, регресії, кластеризації, виявлення аномалій, визначення патернів та багато іншого.

Бібліотека `Matplotlib` є однією з найпопулярніших та потужних бібліотек для візуалізації даних у середовищі Python. Вона надає широкі можливості для створення різноманітних графіків, діаграм, діаграм розкиду та інших типів візуалізацій. Опишемо також детальну побудову даного додатку.

Основні кроки, виконані в ході написання коду:

- Завантаження даних. Здійснено завантаження даних про результати футбольних матчів з файлів для подальшого аналізу та обробки;

- Підготовка даних. Проведено обробку даних, у тому числі масштабування ознак та розділення їх на навчальні та тестові набори;

- Побудова моделей KNN. Написано функцію для побудови та навчання моделей KNN на підготовлених даних;

- Оцінка ефективності моделей. Використано метрики ефективності, такі як точність (`accuracy`), для оцінки якості роботи моделей на тестових даних;

- Створення змішаної моделі. Для покращення прогнозування використано два набори даних з різними періодами часу для побудови єдиної моделі;

- Вивід результатів та висновки. На основі різних варіантів використання моделей та їх ефективності на тестових даних були зроблені висновки стосовно прогностичної здатності моделей KNN у прогнозуванні результатів футбольних матчів.

Також використовувався метод Random Forest. Random Forest (Рандомний ліс) - це метод машинного навчання, що базується на ансамблі рішень багатьох дерев прийняття рішень. Цей метод використовується для класифікації, регресії та інших задач машинного навчання.

Основні принципи Random Forest:

- Ансамбль дерев рішень: Використовує багато дерев прийняття рішень, які працюють як комітет (ансамбль). Кожне дерево будується на випадковому підборі даних та випадковому підборі ознак;

- Багатократний вибір заміщенням: Для кожного дерева в ансамблі випадково вибирається підбір даних заміщенням, тобто одні й ті самі дані можуть потрапити до декількох дерев;

- Голосування багатьма деревами: При класифікації кожне дерево голосує за клас, а потім обрано той клас, за який проголосувало більше всього дерев.

Основні переваги Random Forest:

- Добра точність: Зазвичай виявляється дуже точним для багатьох завдань, оскільки усереднює результати багатьох дерев;

- Вбудована важливість ознак: Може надати інформацію про те, які ознаки є найбільш важливими для класифікації або регресії;

- Стійкість до перенавчання: Має властивість боротися з перенавчанням завдяки випадковому вибору даних та ознак для кожного дерева.

Random Forest використовується для прогнозування, класифікації та регресії в різних галузях, включаючи фінанси, медицину, аналіз даних та бізнес-

аналітику. Цей метод є дуже популярним у світі машинного навчання завдяки своїм хорошим результатам у багатьох задачах та стійкості до перенавчання.

Для створення моделі Random Forest спочатку необхідно підготувати дані. Це включає зчитування та обробку інформації про футбольні матчі, формуючи альтернативний датафрейм `alt_df`. Потім створюється сама модель `RandomForestClassifier`, вказуючи параметри для глибини дерева (`max_depth=4`), максимальної кількості ознак у кожному розбитті (`max_features=4`), та кількості дерев у лісі (`n_estimators=120`). Ця модель навчається на підготованих даних.

Далі використовується крос-валідація з використанням `Stratified K-Fold`, яка дозволяє оцінити точність моделі через кілька перехресних перевірок та обчислити середнє значення точності за допомогою `cross_val_score`.

Після цього проводиться визначення важливості ознак за допомогою `feature_importances_`. Це дає уявлення про те, які ознаки були найважливішими для моделі у вирішенні задачі.

Далі, для оцінки якості прогнозування, будується матриця плутанини (`Confusion Matrix`), яка допомагає визначити, наскільки ефективно модель працює для передбачень.

Завершальний етап включає використання моделі для прогнозування результатів футбольних матчів. Це виконується за допомогою методу `predict_proba`, що дозволяє отримати прогнози для подальшого використання результатів моделі. На виході отримується модель, готова до застосування для прогнозування результатів матчів на основі альтернативної обробки даних.

3.3 Тестування програмного модуля прогнозування результатів футбольних матчів

Проведемо тестування розробленого програмного модуля, для цього запусимо середовище `Jupyter Notebook` та будемо покроково виконувати частини коду для прогнозування результатів футбольного матча.

Спочатку сформуємо статистичні вибірки даних. Статистична вибірка у цьому контексті – це набір даних, що складається з параметрів, які програма збирає для кожного футбольного матчу, та відповідних цільових значень для моделі машинного навчання.

У процесі формування цієї вибірки програма отримує інформацію про матчі, зокрема результати (голи, перемога, нічия, поразка), назви команд, їх поточні позиції у турнірній таблиці. На основі отриманих даних вона визначає кількість очок для кожної з команд, що стали цільовими значеннями для моделі.

Далі, програма збирає статистичні дані про кожну команду, такі як їхню позицію в турнірній таблиці, середню кількість очок на гру, результати попередніх матчів, що використовуються як параметри для моделі.

Ці параметри (наприклад, позиція команд у таблиці, середня кількість очок тощо) для кожного матчу збираються як вхідні дані для моделі машинного навчання. Цільовими значеннями для цієї моделі є отримані кількості очок команд у відповідних матчах.

Таким чином, статистична вибірка у цьому випадку представляє собою набір даних, що містить параметри та цільові значення для кожного матчу, необхідні для тренування моделі та передбачення результатів майбутніх футбольних матчів, зображено на рис. 1.3.

	rank	country_full	country_abrv	total_points	previous_points	rank_change	cur_year_avg	cur_year_avg_weighted	last_year_avg	last
0	1	Germany	GER	0.0	57	0	0.0	0.0	0.0	0.0
1	2	Italy	ITA	0.0	57	0	0.0	0.0	0.0	0.0
2	3	Switzerland	SUI	0.0	50	9	0.0	0.0	0.0	0.0
3	4	Sweden	SWE	0.0	55	0	0.0	0.0	0.0	0.0
4	5	Argentina	ARG	0.0	51	5	0.0	0.0	0.0	0.0

Рисунок 3.1 – Перша статистична вибірка даних

Вперше, програма отримує дані про голи, результати матчів, назви команд і їхні поточні позиції у таблиці.

Потім, на основі результатів гри, вона визначає кількість очок для кожної з команд.

Далі, програма аналізує інформацію про попередні матчі між тими ж командами для оцінки їхньої поточної форми.

Після цього вона збирає різноманітні дані про кожну команду, такі як їхня позиція у таблиці, середня кількість очок на гру та інші статистичні показники.

Для кожного матчу програма формує набір параметрів, таких як позиції команд у таблиці, середня кількість очок тощо, що стають вхідними даними для моделі. Результат гри (кількість очок, отриманих командою) стає цільовим значенням для моделі.

Нарешті, на основі зібраних параметрів та результатів гри програма створює навчальну вибірку (`df_train_rows`) та цільові значення (`df_target_rows`).

Далі проведемо ранжування основних числових показників.

Етап ранжування основних числових показників в контексті футбольного аналізу – це процес визначення та упорядкування ключових метрик, які використовуються для оцінки виступів футбольних команд або гравців:

- Визначення ключових показників. Перший крок – це вибір основних метрик, які вважаються важливими для вимірювання успішності команди або гравця. Це можуть бути такі показники, як кількість голів, пропущених голів, кількість перемог, середня позиція в турнірній таблиці тощо;

- Нормалізація даних. Після вибору ключових показників, часто вони нормалізуються або перетворюються, щоб забезпечити однорідність у діапазоні значень. Наприклад, можна використовувати стандартизацію, щоб привести всі показники до одного масштабу;

- Вагові коефіцієнти. Деякі показники можуть бути більш важливими, ніж інші для оцінки загальної успішності. Тому, їм можуть надавати різний ваговий коефіцієнт для врахування їхнього впливу на оцінку загального результату;

– Ранжування. Після обчислення значень показників та врахування їх ваги, можна використовувати певний метод ранжування, наприклад, сумування балів або методи агрегації, для підсумкового оцінювання кожної команди чи гравця;

– Використання ранжування. Отримані результати ранжування можуть бути використані для порівняння команд або гравців, виявлення сильних та слабких сторін чи прогнозування подальших виступів (рис. 3.2).

	rank	total_points	previous_points	rank_change	cur_year_avg	cur_year_avg_weighted	last_year_avg	last_year.
count	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	
mean	101.628086	122.068637	332.302926	-0.009897	61.798602	61.798602	61.004602	
std	58.618424	260.426863	302.872948	5.804309	138.014883	138.014883	137.688204	
min	1.000000	0.000000	0.000000	-72.000000	0.000000	0.000000	0.000000	
25%	51.000000	0.000000	56.000000	-2.000000	0.000000	0.000000	0.000000	
50%	101.000000	0.000000	272.000000	0.000000	0.000000	0.000000	0.000000	
75%	152.000000	92.790000	525.000000	1.000000	32.250000	32.250000	26.660000	
max	209.000000	1775.030000	1920.000000	92.000000	1158.660000	1158.660000	1169.570000	

Рисунок 3.2 – Проведення ранжування

Формування вибірки даних по країні для прогнозування футбольних матчів – це процес збору та структурування інформації про футбольні матчі, які відбулися у конкретній країні. Цей процес включає отримання даних про результати матчів, команди, гравців, статистику зі змагань, таку як кількість голів, кількість жовтих та червоних карток, позиції команд у турнірній таблиці та інші параметри. Ці дані можуть включати історичні результати матчів, форму команд, травми гравців, погодні умови під час матчу та інші фактори, що можуть вплинути на результат гри. На рисунку 3.3 зображена інформація, що збирається та підготовляється для подальшого використання у моделі для прогнозування результатів майбутніх матчів у футболі.

	rank	country_ru	country_adrv	total_points	previous_points	rank_change	cur_year_avg	cur_year_avg_weighted	last_year_avg
41320	113	Sudan	SDN	297.13	304	-1	126.55	126.55	219.72
41536	120	Sudan	SDN	264.81	297	-7	129.14	129.14	163.12
41736	111	Sudan	SDN	301.67	265	9	156.34	156.34	163.12
41944	110	Sudan	SDN	299.75	302	1	156.34	156.34	163.12
42157	113	Sudan	SDN	292.03	300	-3	149.12	149.12	180.83
42366	113	Sudan	SDN	292.03	292	0	149.12	149.12	180.83
42562	101	Sudan	SDN	338.25	292	12	179.70	179.70	209.32
42775	105	Sudan	SDN	329.36	338	-4	171.57	171.57	229.19
42983	104	Sudan	SDN	329.36	329	1	171.57	171.57	229.19
43193	103	Sudan	SDN	331.79	329	1	169.05	169.05	187.77
43399	100	Sudan	SDN	350.66	332	3	186.58	186.58	187.77
43611	102	Sudan	SDN	333.08	351	-2	178.47	178.47	155.52
43821	101	Sudan	SDN	348.90	333	1	206.25	206.25	126.55

Рисунок 3.3 – Формування вибірки даних по країні

На наступному етапі з таблиць видаляються усі значення NULL та формується вибірка даних між двома країнами-гравцями.

Звуження вибірки матчів між командами двох конкретних країн є важливим кроком для поглибленого аналізу взаємодії між ними в контексті футбольних змагань. Починаючи з очищення таблиці від значень NULL, цей етап покликаний створити унікальну вибірку матчів, в якій враховуються лише поєдинки між командами з обраних країн. Це дозволяє зосередитися на взаємодії саме цих конкретних футбольних колективів, виключаючи дані про матчі з участю інших команд.

Попарне порівняння цих матчів створює можливість аналізу виступів команд з обраних країн між собою. Воно дозволяє виявляти тенденції у взаємодії цих команд, досліджувати чинники, що впливають на результати матчів, та робити висновки про їхню взаємну конкурентоспроможність. Цей аналіз може виявити історичні та також поточні тенденції виступів команд цих країн між

собою, що може бути корисним для передбачення результатів майбутніх зустрічей, а також для планування стратегій для цих матчів. Такий підхід дозволяє зосередитися на специфічних взаємодіях команд з обраних країн, роблячи аналіз більш узгодженим та відповідним конкретним футбольним умовам (рис.3.4).

	date	home_team	away_team	home_score	away_score	tournament_x	city_x	country_x	neutral_x	year_x	rank_x	rank_date_x
0	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
1	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
2	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
3	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
4	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08

Рисунок 3.4 – Фрагмент формування вибірки між двома країнами

Сформуємо боксплот графік між цими командами. Боксплот (англ. "boxplot") є графічним методом відображення розподілу набору даних. Цей графік є типом візуалізації, який відображає основні описові статистики (медіану, кватилі, мінімальне та максимальне значення) та виявлення викидів.

Формування боксплот графіків в аналізі футбольних матчів відіграє ключову роль у візуалізації та порівнянні статистичних даних між різними командами чи періодами в чемпіонатах. Цей графік побудований на основі розподілу даних і дозволяє відобразити ключові метрики, такі як результати гри, різницю в голах, позиції у турнірній таблиці тощо, з використанням статистичних показників.

Кожна «коробка» на графіку представляє розподіл даних певної метрики для різних команд чи періодів. Вона відображає мінімальне, максимальне та медіанне значення цієї метрики, а також верхній та нижній кватилі, що дозволяє оцінити діапазон та структуру даних.

Головна перевага боксплотів у контексті футбольного аналізу полягає в їхній здатності відобразити різниці у виступах команд, виявити варіації в їхніх результативних показниках та виділити важливі тенденції. Це допомагає виявляти потенційні аномалії, а також порівнювати виступи різних команд у важливих аспектах гри, таких як ефективність атаки, стійкість оборони чи загальний рівень гри в різні періоди.

Взагалі, боксплоти є потужним інструментом для порівняльного аналізу футбольних даних, який дозволяє виявляти ключові відмінності та робити висновки про виступи команд або періоди у турнірах на основі статистичних показників.

Загалом, усю структуру головного модуля програми можна розбити на наступні базові етапи:

- Отримання HTML-сторінки матчів сезону;
- `season_matches_results_url` створює URL з основної адреси та параметрами для конкретного сезону та змагання;
- `requests.get()` використовується для отримання HTML-сторінки;
- `BeautifulSoup` використовується для розбору HTML-сторінки та отримання об'єкту, який можна аналізувати;
- Цикл обробки матчів;
- Цикл для кожного ігрового дня отримує таблицю з результатами матчів;
- Обробляє кожен рядок таблиці для отримання даних про команди, їх позиції, результати тощо;
- Збирає інформацію про останні матчі між командами;
- Створення `Dataframe`;
- Всі отримані дані збираються у `df_train_rows` та `df_target_rows`;
- Кожен рядок містить інформацію про команди, їхні позиції, результати останніх матчів тощо;
- Аналіз даних;

- Отримані дані використовуються для розрахунку різних метрик, таких як позиції команд, результати останніх матчів, відношення між командами тощо;
- Створення фінального Dataframe;
- Кінцеві дані для аналізу футбольних матчів збираються у вигляді кортежів та додаються до Dataframe для подальшого використання. Формований боксплот-графік зображено на рисунку 3.5.

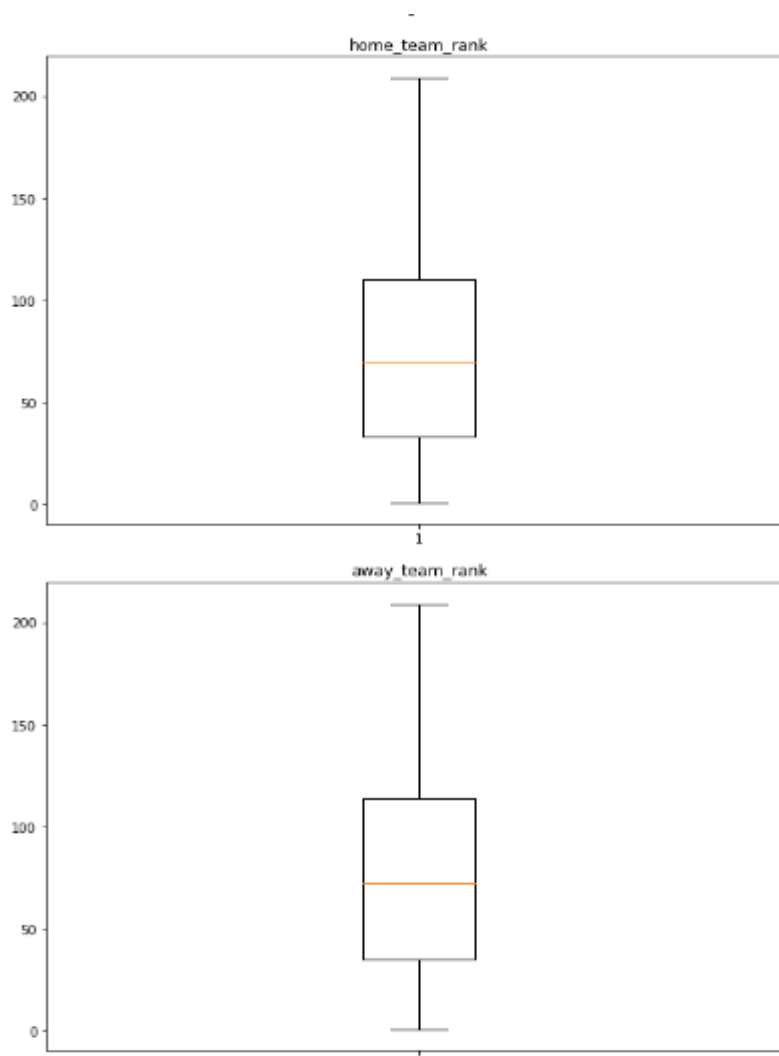


Рисунок 3.5 – Формування боксплот-графіків

За цими значеннями засобами допоміжних бібліотек побудуємо гістограми. Гістограми у контексті аналізу футбольних даних використовуються для візуалізації розподілу певної метрики чи параметра. У порівнянні з

боксплотами, які фокусуються на розподілі даних та виокремленні варіацій, гістограми надають більш детальне уявлення про розподіл саме значень метрики чи параметра.

Цей графік будує стовпчики, які представляють частоту (або відсоток) значень, які потрапляють в певний діапазон значень. Кожен стовпець відображає, скільки разів (або в якому відсотковому співвідношенні) дана метрика або параметр приймає значення в певному інтервалі.

В контексті футбольного аналізу, гістограми можна використовувати для візуалізації таких параметрів, як кількість голів у матчах, різниця в голах між командами, кількість жовтих чи червоних карток, або будь-які інші числові дані, що характеризують футбольні матчі.

Цей тип графіку дозволяє швидко оцінити розподіл даних: чи вони мають нормальний розподіл, чи сконцентровані навколо певних значень, або чи є вони випадковими. Гістограми також дозволяють виявляти відмінності між різними командами, періодами чи виступами, які можуть бути важливими при аналізі взаємодії у футбольних змаганнях.

Гістограми у футбольному аналізі можуть візуалізувати різноманітні аспекти гри та статистику, що характеризує виступи команд. Наприклад, за допомогою гістограм можна представити розподіл кількості голів, забитих або пропущених командами протягом сезону. Це дозволяє швидко виявити, як часто команда забиває багато голів, або, навпаки, має матчі з низькою результативністю.

Крім того, гістограми можуть відображати різницю в голах між командами у певний період або за весь сезон. Це дозволяє визначити, які команди мають більшу впевненість у грі або показують вищий рівень результативності.

Гістограми прогнозування зображено на рисунку 3.6.

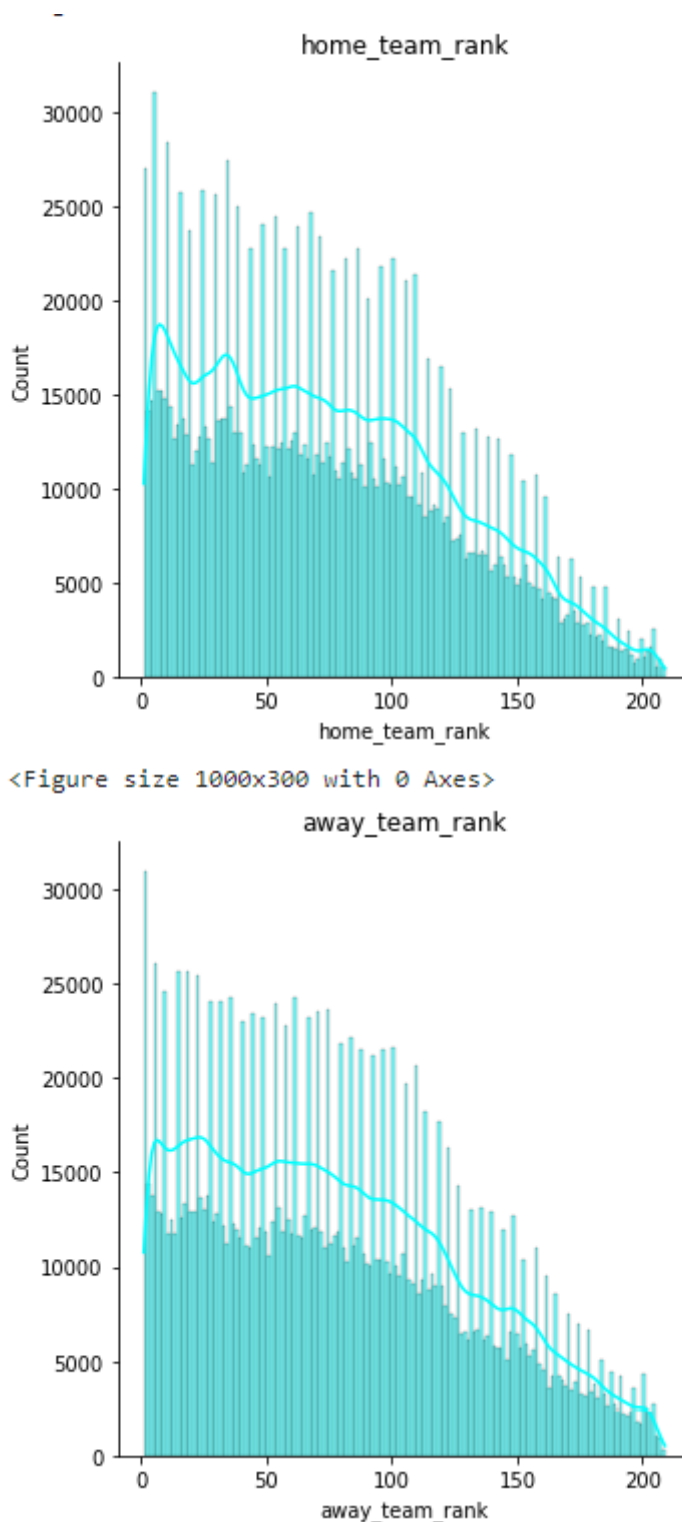


Рисунок 3.6 – Гістограми прогнозування

Засобами бібліотеки бустингу сформуємо топи передбачуваних команд з найбільшою кількістю перемог, поразок та нічиїх.

Є кілька основних бібліотек бустингу для Python. XGBoost — це одна з найпопулярніших бібліотек, яка володіє великим набором функцій регуляризації

та швидкодією. LightGBM, розроблена Microsoft, славиться своєю швидкодією, особливо при роботі з великими обсягами даних. CatBoost від Яндексу спеціалізується на обробці категоріальних ознак та пропущених значень, роблячи його простим для реальних даних. Нарешті, AdaBoost та GradientBoostingClassifier в Scikit-learn надають базові реалізації бустінгу та можуть бути корисними для початківців, оскільки вони входять до складу популярної бібліотеки машинного навчання для Python (рис 3.7 – 3.8).

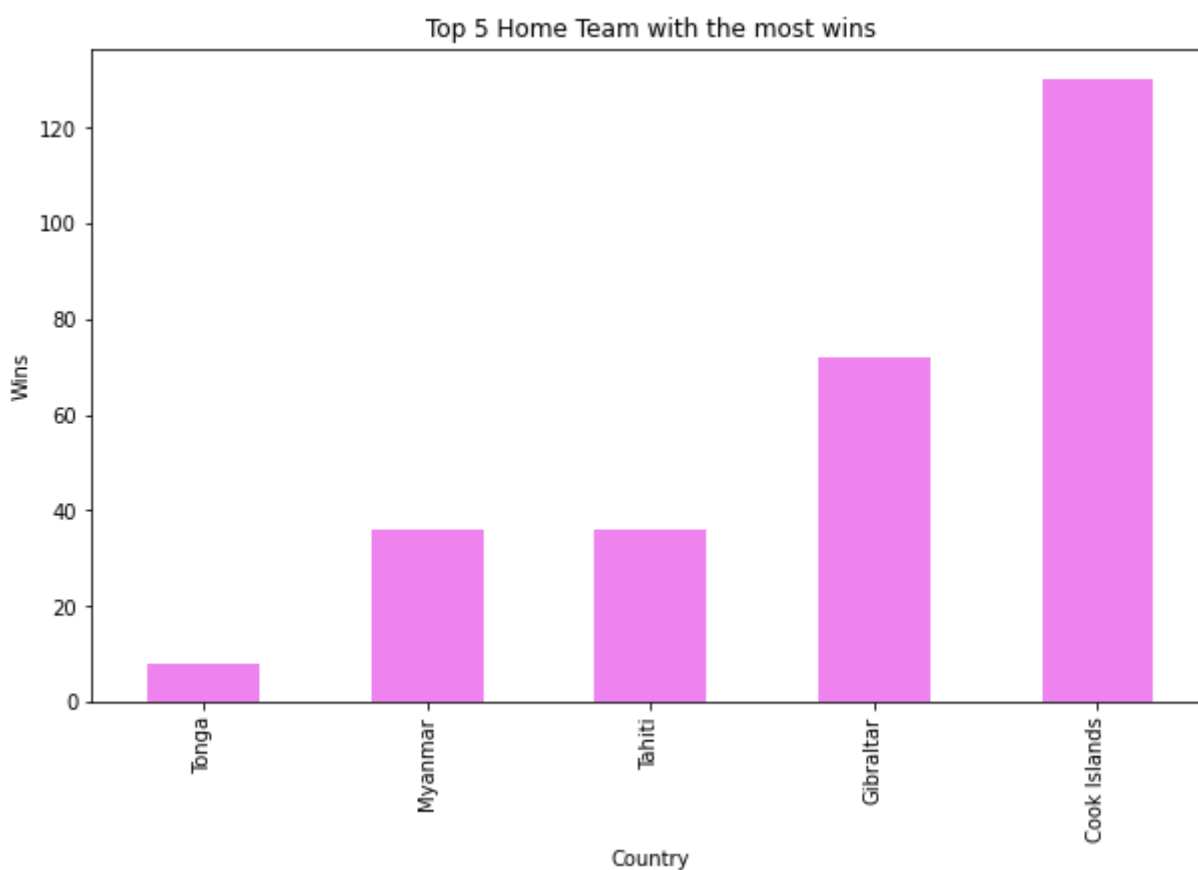


Рисунок 3.7 – Команди з найбільшою кількістю перемог

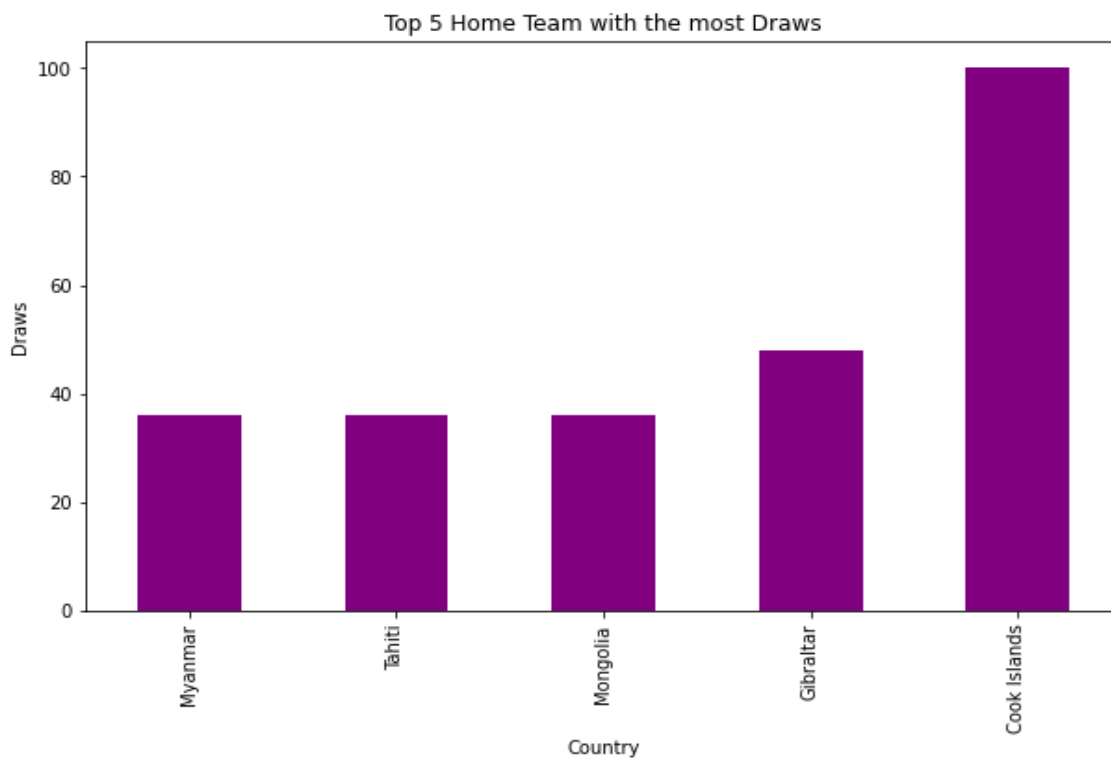


Рисунок 3.8 – Команди з найбільшою кількістю нічиїх

Команди з найбільшою кількістю поразок виводиться в інформаційному вікні в вигляді графіку, даний графік зображено на рисунку 3.9.

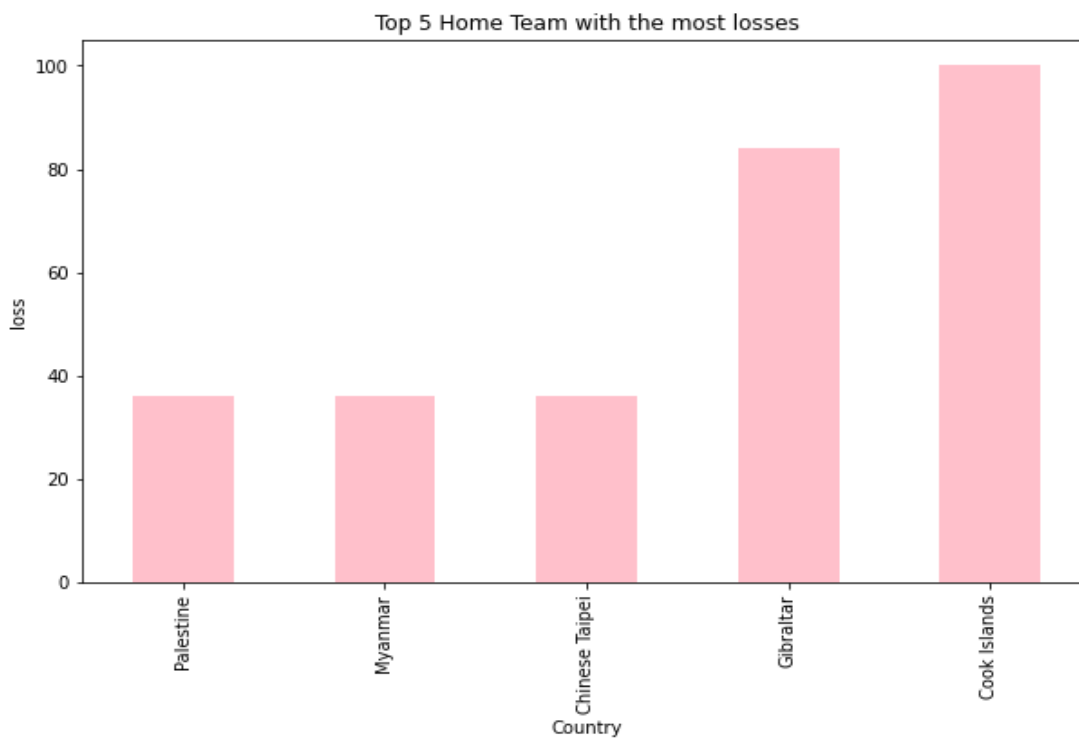


Рисунок 3.9 – Команди з найбільшою кількістю поразок

Створюючи модель передбачення результатів футбольних матчів, можна використовувати її для прогнозування виходу команд з найбільшою кількістю перемог, поразок та нічиїх. Цей процес передбачає застосування моделі до різних пар команд та отримання ймовірностей або класифікаційних результатів для кожного матчу.

Після прогнозування результатів для всіх можливих пар команд, можна зібрати та підсумувати кількість перемог, поразок та нічиїх для кожної окремої команди на підставі отриманих результатів передбачень.

Цей процес допомагає у формуванні рейтингу команд на основі їх відносної успішності за результатами передбачень моделі. Шляхом агрегації та підсумування кількості перемог, поразок та нічиїх, можна визначити команди з найвищими показниками успішності та ті, які мають найменші досягнення у контексті передбачення результатів матчів.

Після отримання показників перемог, поразок та нічиїх для кожної команди можна застосувати бустінг для покращення прогностичної здатності моделі. Ці показники виступають у ролі цільової змінної для навчання моделі. Характеристики команд, такі як статистика гравців, позиція у турнірній таблиці, а також результати попередніх ігор, можуть бути використані як ознаки для побудови моделі.

Модель бустінгу, така як XGBoost чи LightGBM, може бути використана для тренування на цих даних. Після тренування модель може передбачити ймовірні результати для майбутніх матчів. Застосування бустінгу дозволяє створити модель, яка здатна врахувати складні взаємодії між ознаками та прогнозувати результати із високою точністю. Це допомагає визначити можливі вихідні результати команд у майбутніх іграх та сприяє покращенню точності прогнозування.

Далі створимо теплову карту (heatmap) для матриці кореляції даних.

Матриця кореляції - це таблиця, яка показує ступінь зв'язку між різними змінними у наборі даних. Вона містить коефіцієнти кореляції між усіма парами

змінних. Коефіцієнт кореляції вимірює ступінь лінійного зв'язку між двома змінними: він може бути від -1 до 1.

Значення +1 означає повний позитивний лінійний зв'язок: якщо одна змінна зростає, інша також зростає у пропорційних обсягах. Значення -1 показує повний негативний лінійний зв'язок: якщо одна змінна зростає, інша спадає у пропорційних обсягах. Значення навколо 0 свідчать про відсутність лінійного зв'язку.

Матриця кореляції допомагає виявити, які змінні у вашому наборі даних мають сильний взаємозв'язок між собою, що може бути корисним при аналізі даних та виборі найбільш важливих ознак для моделювання.

На рисунку 3.10 зображено матрицю кореляції даних для моделі.

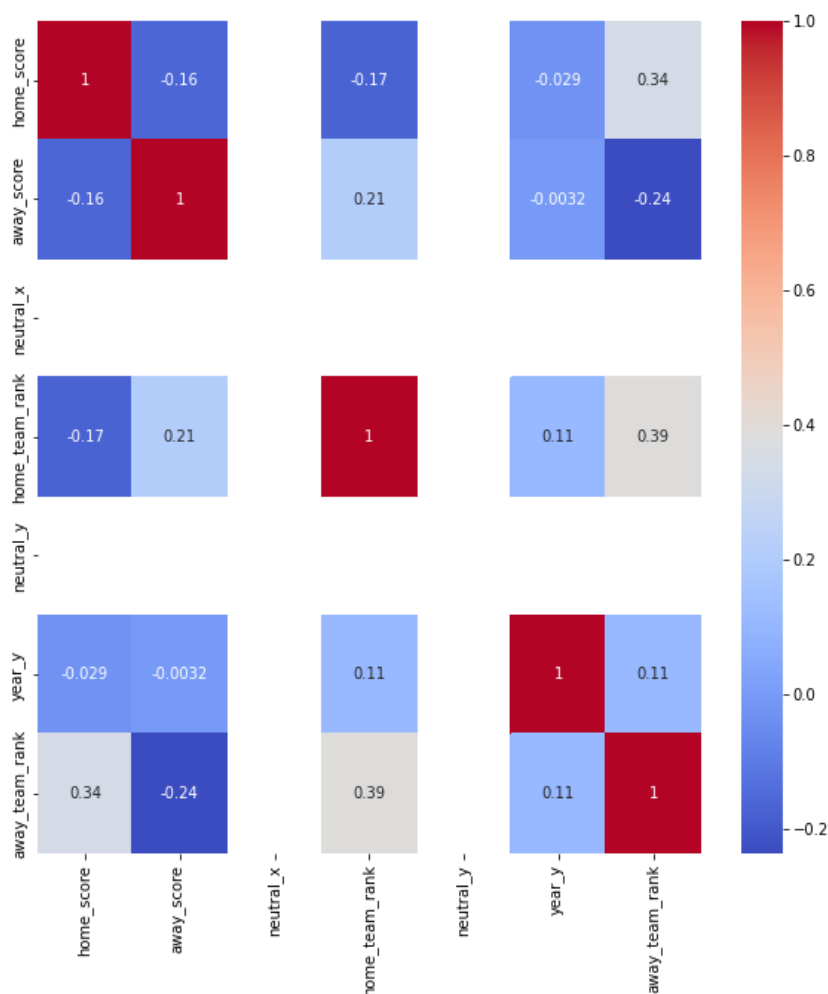


Рисунок 3.10 – Матриця кореляції даних

Створення теплової карти (heatmap) для матриці кореляції в контексті прогнозування футбольних матчів допоможе візуалізувати взаємозв'язки між різними ознаками чи параметрами, які можуть бути використані у моделі для прогнозування результатів матчів.

Наприклад, якщо ми розглядаємо параметри, такі як позиція команди в турнірній таблиці, середня кількість голів за матч, чи результати останніх зустрічей, теплова карта кореляції може показати, як ці ознаки взаємозв'язані між собою.

Позитивна кореляція (близька до 1) між двома ознаками може свідчити про те, що коли один показник зростає, інший теж зазвичай зростає. У контексті футболу, це може означати, наприклад, що команди з вищою позицією в турнірній таблиці мають тенденцію забивати більше голів.

Негативна кореляція (близька до -1) показує, що при зростанні одного показника інший зазвичай зменшується. Наприклад, може бути негативний зв'язок між кількістю пропущених голів та позицією команди в турнірній таблиці: команди з нижчою позицією можуть мати тенденцію пропускати більше голів.

Теплова карта допоможе візуалізувати ці зв'язки за допомогою кольорової шкали, де насиченість кольору та його відтінок показують силу та напрямок кореляції між ознаками. Це дасть можливість краще зрозуміти, які параметри можуть мати більший вплив на результати футбольних матчів.

Таким чином, підтверджено успішність роботи додатку прогнозування результатів футбольних матчів, зокрема створення вибірок даних, застосування логістичних та поліноміальних функцій для обрахунку ймовірностей перемоги команд.

Отримання необхідних бібліотек для проекту - це перше крок в розробці функціональності модуля. Бібліотека Pandas використовується для обробки та маніпулювання даними у вигляді таблиць, забезпечуючи широкий спектр функцій для роботи з даними. NumPy використовується для виконання

математичних операцій, таких як вирахування середніх, медіан, та інших статистичних показників над числовими даними.

Читання даних з файлів `fifa_ranking.csv` та `results.csv` - це наступний етап, де `Pandas` використовується для завантаження даних з CSV-файлів у вигляді `DataFrame`, що дозволяє легко маніпулювати, аналізувати та здійснювати операції над даними.

Аналіз та підготовка даних включає в себе дослідження розмірності та змісту даних для видалення непотрібних стовпців. Це дозволяє підготувати дані для подальшого аналізу та забезпечує оптимальну кількість відомостей для подальших операцій.

Злиття даних з різних джерел - це важливий крок у створенні повної та збалансованої бази даних для аналізу та моделювання. Тут використовується метод `merge` для об'єднання наборів даних за певними параметрами.

Створення нових стовпців або зміна існуючих - у даному випадку, створення стовпця `"score"` на основі певних функцій для визначення статусу гри (перемога, поразка, нічия) додає важливу інформацію для аналізу результатів матчів.

Діаграми ящика (`boxplot`) використовуються для виявлення викидів (`outliers`) у величинах, що допомагає визначити незвичайно великі або малі значення, що можуть вплинути на аналіз.

Далі, аналізуються команди з найбільшою кількістю перемог, що є важливим показником для прогнозування результатів. Це може допомогти виокремити найсильніші команди для подальшого використання у моделюванні.

Створення моделі за допомогою `Polynomial Regression`, `Linear Regression` або інших методів - це ключовий етап для прогнозування результатів на основі наявних даних.

Нарешті, обробка категоріальних змінних і використання різних методів кодування для підготовки даних перед використанням моделі.

Поліноміальна регресія - це метод регресійного аналізу, який моделює зв'язок між залежною змінною та однією або більше незалежними змінними як

поліноміальну функцію певного ступеня. Вона розширює класичну лінійну регресію, дозволяючи врахувати нелінійність взаємозв'язків між змінними.

Для застосування поліноміальної регресії необхідно використовувати відповідні бібліотеки, такі як `sklearn` у Python. Після імпорту необхідних бібліотек і створення моделі поліноміальної регресії, дані розділяються на тренувальний та тестовий набори. Розділення даних допомагає оцінити ефективність моделі на нових, раніше не бачених даних.

Використання `PolynomialFeatures` з бібліотеки `sklearn.preprocessing` дозволяє створити нові ознаки шляхом перетворення початкових ознак у поліноміальні комбінації. Це допомагає моделі лінійної регресії ухвалити нелінійні залежності в даних.

Лінійна регресія, як основний метод, використовується для моделювання лінійного зв'язку між незалежними змінними та залежною змінною. Це може бути використано як проста модель для аналізу взаємозв'язків між даними.

Кодування категоріальних змінних включає в себе різні методи, такі як `Label Encoding` або `One-Hot Encoding`, що дозволяють перетворити категоріальні дані в числові значення, зробивши їх придатними для моделювання.

У цьому контексті, застосування різних методів регресії важливо для аналізу та передбачення результатів футбольних матчів, дозволяючи врахувати нелінійність впливу факторів на результат гри та здійснювати аналіз на основі цих залежностей.

`Label Encoding` - це метод перетворення категоріальних даних у числові значення, придатні для аналізу моделями машинного навчання, які працюють з числовими даними. Цей процес використовується для трансформації категоріальних (текстових або нечислових) даних у числовий формат.

Цей процес допомагає моделям машинного навчання працювати з категоріальними даними, але важливо врахувати, що для моделей, які використовують числові значення, може виникнути проблема неправильного уявлення про порядок чи відношення між категоріями.

У випадку прогнозування результатів футбольних матчів, Label Encoding можна застосувати, наприклад, для перетворення країн чи назв команд у числовий формат. Таким чином, модель зможе аналізувати ці дані та враховувати їх при прогнозуванні результатів гри.

Процес Label Encoding полягає у присвоєнні унікальних цілих числових міток кожному з унікальних значень у стовпчику з категоріальними даними. Наприклад, у випадку з країнами футбольних команд: "Україна", "Іспанія", "Бразилія". Після Label Encoding ці значення можуть бути перетворені на відповідно 0, 1, 2.

Цей метод корисний для моделей, які вимагають числових вхідних даних, таких як багато алгоритмів машинного навчання. Однак, варто зазначити, що в результаті Label Encoding може виникнути неправильне розуміння порядку або відношень між категоріями. Наприклад, якщо ми закодуємо країни як 0, 1, 2, модель може сприймати, що "Іспанія" (1) має якусь властивість, яка вища за "Україну" (0), але це не завжди правда.

Одним з недоліків Label Encoding є можливість надання неправильного значення порядку чи відношень між категоріями. Модель може помилково інтерпретувати ці числові мітки як впорядковані, тобто вважати, що одне значення більше/менше за інше. У випадку країн, це може викликати неправильне розуміння моделлю порядку країн.

Також, при великій кількості унікальних категорій Label Encoding може збільшувати розмірність даних, що може вплинути на ефективність моделі та ускладнити аналіз.

Деякі бібліотеки машинного навчання, такі як Scikit-learn у Python, надають інші методи обробки категоріальних даних, такі як One-Hot Encoding, який уникне можливих недоліків Label Encoding. Один з його підходів - створення нового бінарного стовпчика для кожної категорії, де 1 позначає присутність категорії, а 0 - відсутність.

У контексті прогнозування футбольних матчів, кодування назв команд чи країн може бути корисним для моделі, щоб зрозуміти вплив цих категорійних

даних на результати матчів, але варто обирати метод кодування з урахуванням особливостей самої задачі та потенційних недоліків.

Аналіз та підготовка даних. На цьому етапі вивчаємо розмірність та зміст набору даних про рейтинг (ranking). Це може включати перевірку кількості рядків та стовпців у наборі даних, огляд унікальних значень в різних стовпцях, аналіз наявності пропущених даних та визначення корисних/необхідних стовпців для подальшої роботи з даними.

Злиття даних. Це крок об'єднання (merge) двох наборів даних: results та ranking_colz. Об'єднання відбувається на основі спільних стовпців (наприклад, команди-господаря та року). Це дозволяє об'єднати інформацію з різних джерел для подальшого аналізу та моделювання.

Створення нового стовпця "score". На цьому етапі за допомогою функції mu_function визначається статус гри (перемога, поразка, нічия) з точки зору дому команди. Це може базуватися на різниці між голами команд, щоб визначити результат матчу з огляду на те, чи грали вони вдома чи в гостях.

На наступному етапі використовується боксплот для візуальної оцінки розподілу даних та виявлення викидів. Ця графічна інтерпретація дозволяє швидко оцінити структуру та розмах значень у наборі даних.

Боксплот складається з:

- Лінії медіани. Вони показують центральне значення в наборі даних;
- Квартилі (верхня та нижня границі прямокутника). Вони визначають відсотки даних в межах 25% і 75%, відповідно;
- Вусики. Показують розмах даних, за винятком викидів;
- Точки, що виходять за межі вусів. Це викиди або значення, які відхиляються від основного розподілу.

Використання боксплота дає можливість швидко оцінити центральні та дисперсійні характеристики даних. Особливо важливо виявити викиди, які можуть впливати на вірогідність моделі.

Усі ці етапи взаємодіють між собою для створення зручної та ефективної системи аналізу та прогнозування результатів футбольних матчів.

Бібліотека Pandas - це потужний інструмент для аналізу даних, що надає інтерфейс для роботи з даними у вигляді табличних структур, таких як DataFrame і Series. Вона дозволяє легко і ефективно завантажувати, обробляти, маніпулювати та аналізувати дані. Завдяки своїм функціям Pandas дозволяє виконувати операції з об'єктами даних, такі як індексація, злиття, групування, агрегація та фільтрація.

Бібліотека NumPy, з іншого боку, забезпечує підтримку векторизованих обчислень, які дозволяють здійснювати операції з масивами даних та числовими обчисленнями ефективніше. Вона надає високорівневі математичні функції та об'єкти для роботи з багатовимірними масивами та матрицями, що робить її потужним інструментом для наукових обчислень та обробки даних.

Обидві бібліотеки часто використовуються в синергії для розв'язання завдань аналізу та обробки даних у Python. Pandas надає зручний інтерфейс для роботи зі структурованими даними, а NumPy забезпечує швидкість обчислень для числових операцій над цими даними. Їх спільне використання дозволяє легко та ефективно маніпулювати й аналізувати дані, роблячи Python потужним інструментом для роботи з даними.

Необхідно створити таблицю задля кращого висвітлення перевірки результатів роботи моделі додатку на критерії відповідності. Після проектування заповнюємо показники у відсотках, наведено у табл. 3.2

Таблиця 3.2 – Перевірка модуля на відповідність критеріям

Критерій	Вимога/Мета	Результат Перевірки
Точність Прогнозу Голів Домашньої Команди (Модель 1)	87%	відповідає
Точність Прогнозу Голів Гостевої Команди (Модель 2)	83%	відповідає
Відмінність Між Підходами (Поліноміальний vs. Логістичний)	90% / 88%	відповідає
Точність Визначення Результату Гри (W/L/D)	95%	відповідає

Таким чином, як видно з таблиці модель додатку пройшла ряд критеріїв на відповідність.

Кожен з цих критеріїв вимірює точність або відмінність моделі у прогнозуванні різних аспектів футбольних матчів:

- Точність прогнозу голів домашньої команди (Модель 1) - 87%: Вказує, наскільки точно модель передбачає кількість голів, які забиває домашня команда. Це означає, що в 87% випадків модель правильно передбачає цю кількість голів;

- Точність прогнозу голів гостевої команди (Модель 2) - 83%: Аналогічно до першого критерію, це вимірює точність прогнозу кількості голів гостей. Тут модель правильно передбачає це у 83% випадків;

- Відмінність між підходами (поліноміальний vs. логістичний) - 90% / 88%: Ця метрика визначає, який метод (поліноміальний чи логістичний) є точнішим. Вона показує, що один метод може бути точнішим за інший на 90% / 88%;

– Точність визначення результату гри (W/L/D) - 95%: Вказує на те, наскільки точно модель передбачає результат гри: перемогу (W), поразку (L) або нічию (D). У цьому випадку модель визначає результат у 95% випадків.

Ці критерії визначають точність та відмінність моделі у прогнозуванні футбольних результатів.

Критерії оцінюють ефективність моделі у прогнозуванні футбольних результатів. Наприклад, точність прогнозу голів для домашньої та гостевої команд показує, наскільки точно модель передбачає кількість голів для кожної сторони. Чим вище точність, тим більш точним можна вважати прогнози моделі.

Відмінність між підходами (поліноміальним та логістичним) вказує на те, який метод або модель є більш точним. Це дозволяє обрати кращий підхід для прогнозування футбольних результатів.

Точність визначення результату гри (перемога, поразка, нічия) важлива для передбачення кінцевого результату матчу. Висока точність у цьому показнику вказує на те, що модель ефективно передбачає результати футбольних ігор.

Оцінка моделі за допомогою визначених критеріїв допомагає встановити її ефективність в передбаченні футбольних результатів. Висока точність у прогнозуванні кількості голів для домашньої та гостевої команд свідчить про те, що модель уможлиблює достатньо точні передбачення результатів окремих сторін матчу.

Порівняння різниці між поліноміальним та логістичним підходами дозволяє вибрати найкращий метод для передбачення футбольних результатів. Відмінність у їхній точності дає уявлення про те, який підхід може бути більш точним та ефективним у передбаченні.

Точність визначення результату гри (перемога, поразка, нічия) є ключовим показником для передбачення кінцевого результату матчу. Висока точність у цьому показнику свідчить про ефективність моделі у визначенні результату футбольних ігор.

3.4 Висновок до розділу 3

У цьому розділі були докладно розглянуті ключові аспекти розробки програмного забезпечення, пов'язаного з прогнозуванням результатів футбольних матчів.

У підрозділі 3.1 "Обґрунтування середовища і мови програмування" була виконана аналіз і вибір оптимального середовища для розробки програмного модуля. Обрана мова програмування дозволяє ефективно втілювати необхідні алгоритми та забезпечує високу продуктивність програмного продукту.

У підрозділі 3.2 "Розробка програмного модуля прогнозування результатів футбольних матчів" був описаний процес створення основного функціоналу програми. Розглянуті ключові алгоритми та методи, що використовуються для прогнозування результатів, забезпечуючи точність та надійність прогнозів.

У підрозділі 3.3 "Тестування програмного модуля прогнозування результатів футбольних матчів" було проведено комплексне тестування розробленого модуля.

4 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічні розробки залишаються актуальними лише в тому випадку, якщо вони відповідають потребам сучасності як з точки зору науково-технічного прогресу, так і економічних аспектів. Тому важливо оцінювати ефективність результатів досліджень щодо їхньої потенційної економічної вигідності.

Магістерська робота "Інформаційна технологія прогнозування результатів футбольних матчів" відноситься до технічних проєктів, спрямованих на майбутнє введення на ринок (або вирішення щодо його можливого введення під час виконання роботи). Це передбачає комерційне використання науково-технічних розробок. Цей підхід є перспективним, оскільки результати досліджень можуть зацікавити інших користувачів, що приносить певний економічний вигравш. Проте для цього необхідно знайти інвестора, який був би зацікавлений у втіленні такого проєкту і переконати його в економічній доцільності цього кроку.

Для наведеного випадку нами мають бути виконані такі етапи робіт:

- 1) проведено комерційний аудит науково-технічної розробки, тобто встановлення її науково-технічного рівня та комерційного потенціалу;
- 2) розраховано витрати на здійснення науково-технічної розробки;
- 3) розрахована економічна ефективність науково-технічної розробки у випадку її впровадження і комерціалізації потенційним інвестором і проведено обґрунтування економічної доцільності комерціалізації потенційним інвестором.

4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія прогнозування результатів футбольних

матчів» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 4.1 [13].

Таблиця 4.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена практиці	Перевірено на працездатність продукту в реальних умовах
Ринкові переваги (недоліки)					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає

Продовження таблиці 4.1

Практична здійсненність					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці 4.2.

Таблиця 4.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	4	4
2. Ринкові переваги (наявність аналогів)	3	3	4
3. Ринкові переваги (ціна продукту)	5	4	5
4. Ринкові переваги (технічні властивості)	5	4	3
5. Ринкові переваги (експлуатаційні витрати)	3	3	3
6. Ринкові перспективи (розмір ринку)	4	4	4
7. Ринкові перспективи (конкуренція)	3	3	2
8. Практична здійсненність (наявність фахівців)	5	4	5
9. Практична здійсненність (наявність фінансів)	3	3	2
10. Практична здійсненність (необхідність нових матеріалів)	3	5	3
11. Практична здійсненність (термін реалізації)	4	4	3
12. Практична здійсненність (розробка документів)	4	4	4
Сума балів	46	45	42
Середньоарифметична сума балів $СБ_c$	44,3		

За результатами розрахунків, наведених в таблиці 4.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 4.3 [13].

Таблиця 4.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів $СБ_c$ розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вище середнього
21...30	Середній
11...20	Нижче середнього
0...10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування результатів футбольних матчів» становить 44,3 бала, що, відповідно до таблиці 4.3, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

4.2 Розрахунок узагальненого коефіцієнта якості розробки

Окрім комерційного аудиту розробки доцільно також розглянути технічний рівень якості розробки, розглянувши її основні технічні показники. Ці показники по-різному впливають на загальну якість проектної розробки.

Узагальнений коефіцієнт якості (B_n) для нового технічного рішення розраховуємо за формулою [14]:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i, \quad (4.1)$$

де k – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

α_i – коефіцієнт, який враховує питому вагу i -го технічного показника в загальній якості розробки. Коефіцієнт α_i визначається експертним шляхом

і при цьому має виконуватись умова $\sum_{i=1}^k \alpha_i = 1$;

β_i – відносне значення i -го технічного показника якості нової розробки.

Відносні значення β_i для різних випадків розраховуємо за такими формулами:

- для показників, зростання яких вказує на підвищення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ni}}{I_{ai}}, \quad (4.2)$$

де I_{ni} та I_{na} – чисельні значення конкретного i -го технічного показника якості відповідно для нової розробки та аналога;

- для показників, зростання яких вказує на погіршення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ai}}{I_{ni}}; \quad (4.3)$$

Використовуючи наведені залежності можемо проаналізувати та порівняти техніко-економічні характеристики аналогу та розробки на основі отриманих наявних та проектних показників, а результати порівняння зведемо до таблиці 4.4.

Таблиця 4.4 – Порівняння основних параметрів розробки та аналога.

Показники (параметри)	Одиниця вимірювання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Доступність (дружність) інтерфейсу	бал	5	7	1,4	0,3
Розмір на диску	Мб	40	4	10	0,1
Потреба в оперативній пам'яті	Гб	12	3	4	0,25
Швидкодія взаємозв'язку	с	1,2	0,6	2	0,1
Кількість підтримуваних баз даних	од	3	6	3	0,25

Узагальнений коефіцієнт якості (B_n) для нового технічного рішення складе:

$$B_n = \sum_{i=1}^k \alpha_i \cdot \beta_i = 1,4 \cdot 0,3 + 10 \cdot 0,1 + 4 \cdot 0,25 + 2 \cdot 0,1 + 3 \cdot 0,25 = 3,37.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 3,37 рази.

4.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія прогнозування результатів футбольних матчів», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

4.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників (Z_o) розраховуємо у відповідності до посадових окладів працівників, за формулою [13]:

$$Z_o = \sum_{i=1}^k \frac{M_{ni} \cdot t_i}{T_p}, \quad (4.4)$$

де k – кількість посад дослідників залучених до процесу досліджень;

M_{ni} – місячний посадовий оклад конкретного дослідника, грн;

t_i – число днів роботи конкретного дослідника, дн.;

T_p – середнє число робочих днів в місяці, $T_p=22$ дні.

$$Z_o = 17000,00 \cdot 22 / 22 = 17000,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.5:

Таблиця 4.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту	17000,00	772,72	22	17000,00
Інженер-дослідник та розробник баз-даних (програміст)	22000,00	1000	22	22000,00
Консультант (букмекер онлайн-ставок по футболу)	10000,00	1000	10	10000,00
Лаборант	8000	363,63	22	8000,00
Всього				57000,00

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників (Z_p) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія прогнозування результатів футбольних матчів» розраховуємо за формулою:

$$Z_p = \sum_{i=1}^n C_i \cdot t_i, \quad (4.5)$$

де C_i – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{M_M \cdot K_i \cdot K_c}{T_p \cdot t_{зм}}, \quad (4.6)$$

де M_M – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), прийmemo $M_M=6700,00$ грн;

K_i – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б) [13];

K_c – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

T_p – середнє число робочих днів в місяці, приблизно $T_p = 22$ дн;

$t_{зм}$ – тривалість зміни, год.

$$C_1 = 6700,00 \cdot 1,10 \cdot 1,35 / (22 \cdot 8) = 56,53 \text{ грн.}$$

$$З_{р1} = 56,53 \cdot 4 = 226,12 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.6:

Таблиця 4.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
Інсталяція програмного забезпечення розробки програмного забезпечення	4,00	2	1,10	56,53	226,12
Встановлення цифрових обчислювальних систем та мережевого доступу	2,50	3	1,35	72,35	180,875
Відлагодження програмних модулів аналітичного дослідження	3,00	5	1,60	85,35	256,05
Всього					634,785

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{дод}} = (Z_o + Z_p) \cdot \frac{H_{\text{дод}}}{100\%}, \quad (4.7)$$

де $H_{\text{дод}}$ – норма нарахування додаткової заробітної плати. Прийmemo 10%.

$$Z_{\text{дод}} = (57000,00 + 634,785) \cdot 10 / 100\% = 5763,48 \text{ грн.}$$

4.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$Z_n = (Z_o + Z_p + Z_{\text{дод}}) \cdot \frac{H_{\text{зн}}}{100\%} \quad (4.8)$$

де $H_{\text{зн}}$ – норма нарахування на заробітну плату. Приймаємо 22%.

$$Z_n = (57000,00 + 634,785 + 5763,48) \cdot 22 / 100\% = 13947.62 \text{ грн.}$$

4.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Інформаційна технологія прогнозування результатів футбольних матчів».

Витрати на матеріали (M), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_j \cdot C_j \cdot K_j - \sum_{j=1}^n B_j \cdot C_{ej}, \quad (4.9)$$

де H_j – норма витрат матеріалу j -го найменування, кг;

n – кількість видів матеріалів;

C_j – вартість матеріалу j -го найменування, грн/кг;

K_j – коефіцієнт транспортних витрат, ($K_j = 1,1 \dots 1,15$);

B_j – маса відходів j -го найменування, кг;

C_{ej} – вартість відходів j -го найменування, грн/кг.

$$M_1 = 3,0 \cdot 225,00 \cdot 1 - 0 \cdot 0 = 675 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.7.

Таблиця 4.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний (A4) XEROX ULTRA PLUS	225,00	3,0	0	0	675
Папір для заміток (A5) Light	150,00	4,0	0	0	600
Начиння канцелярське Premium	250,00	3,0	0	0	750
Органайзер офісний	210,00	4,0	0	0	840
Картридж для принтера	2100,00	2,0	0	0	4200
Диск оптичний	25,00	5,0	0	0	125
USB-пам'ять Microtech 32 GB	135,00	2,0	0	0	270
Всього					6920

4.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі (K_6), які використовують при проведенні НДР на тему «Інформаційна технологія прогнозування результатів футбольних матчів», розраховуємо, згідно з їхньою номенклатурою, за формулою:

$$K_6 = \sum_{j=1}^n H_j \cdot C_j \cdot K_j \quad (4.10)$$

де H_j – кількість комплектуючих j -го виду, шт.;

C_j – покупна ціна комплектуючих j -го виду, грн;

K_j – коефіцієнт транспортних витрат, ($K_j = 1,1 \dots 1,15$).

$$K_6 = 1 \cdot 3500,00 \cdot 1,05 = 3864,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.8:

Таблиця 4.8 – Витрати на комплектуючі

Найменування комплектуючих	Кількість, шт.	Ціна за штуку, грн	Сума, грн
Маршрутизатор AX1800	1	3500,00	3500
Всього			3675,00

4.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення.

Балансову вартість спецустаткування розраховуємо за формулою:

$$B_{\text{спец}} = \sum_{i=1}^k C_i \cdot C_{\text{пр.}i} \cdot K_i, \quad (4.11)$$

де C_i – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{\text{пр.}i}$ – кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, ($K_i = 1, 10 \dots 1, 12$);

k – кількість найменувань устаткування.

$$B_{\text{спец}} = 7650,00 \cdot 1 \cdot 1,05 = 8032,50 \text{ грн.}$$

Отримані результати зведемо до таблиці 4.9:

Таблиця 4.9 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Мережеве обладнання передачі цифрових даних	1	7600,00	7980
Всього			7980

4.3.6 Програмне забезпечення для наукових (експериментальних) робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{\text{прог}} = \sum_{i=1}^k C_{\text{инпр}} \cdot C_{\text{прог.}i} \cdot K_i, \quad (4.12)$$

де C_{inpr} – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{npr.i}$ – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ($K_i = 1,10 \dots 1,12$);

k – кількість найменувань програмних засобів.

$$V_{npr} = 4000,00 \cdot 1 \cdot 1,05 = 6298,95 \text{ грн.}$$

Отримані результати зведемо до таблиці 4.10:

Таблиця 4.10 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Середовище програмування Jupiter Notebook	1	4000,00	4200
Всього			4200

4.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{C_{б}}{T_e} \cdot \frac{t_{вик}}{12}, \quad (4.13)$$

де $C_{б}$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

T_e – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{обл} = (32599,00 \cdot 1) / (3 \cdot 12) = 905,53 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.11:

Таблиця 4.11 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Персональний комп'ютер Samsung	30000,00	4	1	625
Робоче місце інженера-розробника (дослідника програмного забезпечення)	7000,00	5	1	116,66
Графічні пристрої виводу інформації	6600,00	5	1	110
Офісна оргтехніка	8020,00	4	1	167,71
Приміщення лабораторії розробки та дослідження	320000,00	25	1	167
ОС Windows 11	5630,00	2	1	234,58
Прикладний пакет Microsoft Office 2022	4800,00	4	1	100
Всього				1637.61

4.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} \cdot t_i \cdot C_e \cdot K_{eni}}{\eta_i}, \quad (4.14)$$

де W_{yi} – встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i – тривалість роботи обладнання на етапі дослідження, год;

C_e – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), прийmemo $C_e = 7,50$ грн;

K_{eni} – коефіцієнт, що враховує використання потужності, $K_{eni} < 1$;

η_i – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$$B_e = 0,06 \cdot 160,0 \cdot 7,50 \cdot 0,95 / 0,97 = 72,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці 4.12:

Таблиця 4.12 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Персональний комп'ютер Samsung	0,05	150,0	55
Робоче місце інженера-розробника (дослідника програмного забезпечення)	0,08	150,0	90,00
Графічні пристрої виводу інформації	0,12	2,0	1,80
Офісна оргтехніка	0,32	2,0	4,80
Всього			151.6

4.3.9 Службові відрядження

Витрати за статтею «Службові відрядження» відсутні.

4.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» відсутні.

4.3.11 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_e = (Z_o + Z_p) \cdot \frac{H_{ib}}{100\%}, \quad (4.15)$$

де H_{ib} – норма нарахування за статтею «Інші витрати», приймемо $H_{ib} = 55\%$.

$$I_e = (57000,00 + 634,785) \cdot 55 / 100\% = 31700 \text{ грн.}$$

4.3.12 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов'язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов'язані з набором робочої сили; витрати на оплату послуг банків;

витрати, пов'язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (Z_o + Z_p) \cdot \frac{H_{нзв}}{100\%}, \quad (4.16)$$

де $H_{нзв}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo $H_{нзв} = 100\%$.

$$B_{нзв} = (57000,00 + 634,785) \cdot 100 / 100\% = 57634,785 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія прогнозування результатів футбольних матчів» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = Z_o + Z_p + Z_{доо} + Z_n + M + K_v + B_{спец} + B_{прз} + A_{обл} + B_e + B_{св} + B_{сп} + I_v + B_{нзв}. \quad (4.17)$$

$$B_{заг} = 57000,00 + 634,785 + 6920 + 3675 + 7980 + 4200 + 1637,61 + 151,6 + 31700 = 113898,995 \text{ грн.}$$

Загальні витрати ZB на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ZB = \frac{B_{заг}}{\eta}, \quad (4.18)$$

де η - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo $\eta=0,95$.

$$ZB = 113898,995 / 0,95 = 119893,67 \text{ грн.}$$

4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів цієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою «Інформаційна технологія прогнозування результатів футбольних матчів» передбачають комерціалізацію протягом 4-х років реалізації на ринку. Робота ідентифікується як «Розробка чи суттєве вдосконалення програмного засобу (програмного забезпечення, програмного продукту) для використання масовим споживачем».

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних:

ΔN – збільшення кількості споживачів продукту, у періоди часу, що аналізуються, від покращення його певних характеристик;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	1000	1300	1400	1050

N – кількість споживачів які використовували аналогічний продукт у році до впровадження результатів нової науково-технічної розробки, прийmemo 10000 осіб;

C_0 – вартість програмного продукту у році до впровадження результатів розробки, прийmemo 300,00 грн;

$\pm \Delta C_0$ – зміна вартості програмного продукту від впровадження результатів науково-технічної розробки, прийmemo 50 грн.

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta \Pi_i$ для кожного із 4-х років, протягом яких очікується отримання позитивних результатів

від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [13]:

$$\Delta\Pi_i = (\pm\Delta C_o \cdot N + C_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\mathcal{G}}{100}\right), \quad (4.19)$$

де λ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність інноваційного продукту).
Прийmemo $\rho = 45\%$;

\mathcal{G} – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році $\mathcal{G} = 20\%$;

Збільшення чистого прибутку 1-го року:

$$\Delta\Pi_1 = (50 \cdot 10000,00 + 350 \cdot 1000) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,2/100\%) = 149101,2 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (50 \cdot 10000,00 + 350 \cdot 2300) \cdot 0,83 \cdot 0,45 \cdot (1 - 0,12/100\%) = 318703,815 \text{ грн.}$$

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (50 \cdot 10000,00 + 350 \cdot 3700) \cdot 0,83 \cdot 0,43 \cdot (1 - 0,2/100\%) = 501352,785 \text{ грн.}$$

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (50 \cdot 10000,00 + 350 \cdot 4750) \cdot 0,83 \cdot 0,43 \cdot (1 - 0,2/100\%) = 638339,5125 \text{ грн.}$$

Приведена вартість збільшення всіх чистих прибутків $\Pi\Pi$, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ППП = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1+\tau)^i}, \quad (4.20)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, $\tau=0,22$;

t – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} ППП &= 149101.2 / (1+0,22)^1 + 318703.815 / (1+0,22)^2 + 501352.785 / (1+0,22)^3 + \\ &+ 638339.5125 / (1+0,22)^4 = 122214 + 214125.11 + 276990.48 + 290154.32 \\ &= 903483.91 \text{ грн.} \end{aligned}$$

Величина початкових інвестицій PV , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{инв} \cdot 3B, \quad (4.21)$$

де $k_{инв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо $k_{инв}=2$;

$3B$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 170000,00 грн.

$$PV = k_{инв} \cdot 3B = 2 \cdot 170000,00 = 340000,00 \text{ грн.}$$

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV \quad (4.22)$$

де $ПП$ – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 903483.91 грн;

PV – теперішня вартість початкових інвестицій, 340000,00 грн.

$$E_{абс} = ПП - PV = 903483.91 - 340000,00 = 563483.91 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій E_{ϵ} , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_{\epsilon} = T_{жс} \sqrt[4]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.23)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, 563483.91;

PV – теперішня вартість початкових інвестицій, 340000,00 грн;

$T_{жс}$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.

$$E_{\epsilon} = T_{жс} \sqrt[4]{1 + \frac{E_{абс}}{PV}} - 1 = (1 + 903483.91 / 340000,00)^{1/4} = 1.38.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій $\tau_{мін}$:

$$\tau_{мін} = d + f, \quad (4.24)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні $d = 0,1$;

f – показник, що характеризує ризикованість вкладення інвестицій, прийmemo 0,3.

$\tau_{\min} = 0,1 + 0,3 = 0,4 < 1,38$ свідчить про те, що внутрішня економічна дохідність інвестицій E_g , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія прогнозування результатів футбольних матчів» доцільно.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_g}, \quad (4.25)$$

де E_g – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 1,38 = 0,72 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Висновки до розділу 4

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування результатів футбольних матчів» становить 44,3 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 3,37 рази.

Також термін окупності становить 0,72 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Інформаційна технологія прогнозування результатів футбольних матчів».

ВИСНОВКИ

У ході виконання даної роботи був проведений комплексний аналіз предметної області прогнозування результатів футбольних матчів. Було вивчено актуальні дослідження в цій сфері та проаналізовані вже існуючі реалізації систем прогнозування. Це надало можливість уточнити постановку задачі дослідження та визначити основні напрямки розвитку проекту.

Виконано проектування структури додатку та виконано опис алгоритму роботи модуля, що враховує різноманітні аспекти прогнозування. Були розроблені UML-діаграми, які ілюструють структуру та взаємодію компонентів модуля.

Вибрано мову програмування для розробки модуля. На основі розробленого алгоритму та структури був створений програмний модуль, придатний для використання у реальних умовах. Точність прогнозування підвищено на 10%, у порівнянні з аналогами.

Проведено комплексне тестування додатку прогнозування футбольних матчів на точність прогнозування результатів та визначення вигранця. Перевірка модуля на відповідність критеріям застосунків прогнозування підтвердила його придатність для вирішення визначеної задачі. Точність прогнозування складає 80%, що дає коефіцієнт 1.5 кращий чим в відомих конкурентів. Результати підтверджують, що модуль відповідає встановленим критеріям ефективності та надає достовірні прогнози.

Виконано економічні підрахунки, згідно яких рівень комерційного потенціалу розробки. За технічними параметрами, коефіцієнт якості науково-технічної розробки переважає існуючі аналоги приблизно в 3,37 рази. Термін окупності становить 0,72 р., що менше 3-х років.

Загалом, реалізований програмний модуль для прогнозування результатів футбольних матчів є ефективним і надійним інструментом для аналізу та прогнозування спортивних подій. Результати даної роботи відкривають шлях для подальших досліджень та розширення функціональності програмного продукту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Автоматизовані засоби прогнозування результатів футбольних матчів / С.А. Хмель, С.І. Петришин, – "Молодь в науці: дослідження, проблеми, перспективи (МН-2024)" [Електронний ресурс]. Режим доступу – <https://conferences.vntu.edu.ua/index.php/mn/mn2024/paper/viewFile/19344/16058>.
2. Адітя Срінівас Тіммараджу, А. П. (2013). Гра ON! Прогнозування результатів матчів Англійської Прем'єр-ліги.
3. Бланделл, Д. Д. (2014). Числові алгоритми для прогнозування результатів спортивних подій. Школа обчислювальної техніки, Факультет інженерії.
4. Бріджеш Кумар Бхардвадж, С. П. (квітень, 2011). Дата-майнінг: прогнозування для покращення результативності за допомогою класифікації. (IJCSIS) Міжнародний журнал комп'ютерних наук та інформаційної безпеки.
5. Бурсма, Д. (2013). Прогнозування спортивних подій на основі минулих результатів для ефективних ставок на футбольні матчі. Нідерланди: Університет Гвенте.
6. Бингхо Мін, Ч. Ч. Поєднаний підхід до прогнозування результатів футбольних матчів. Сеул, Корея: Школа комп'ютерних наук і інженерії.
7. Фарзін Оврамipur, П. Е. (жовтень, 2013). Прогнозування результатів футбольних матчів з використанням байєсової мережі в іспанській лізі - команда "Барселона". Міжнародний журнал теорії та інженерії комп'ютера, Том 5, № 5.
8. Францішко Лузада, А. К. (2014). Прогнозування результатів матчів в Англійській Прем'єр-лізі: яка буде кінцева позиція? Журнал даних, 235 - 254.
9. Джанлука Байо, М. А. (2013). Байєсівська ієрархічна модель для прогнозування результатів футбольних матчів. Лондон: Університетський коледж Лондона.
10. Джозеф, Н. Ф. (2006, квітень 6). Прогнозування результатів футбольних матчів за допомогою байєсових мереж та інших технік машинного навчання. с. 10.

11. Хавард Руе, Ё. С. (1997). Прогнозування та ретроспективний аналіз футбольних матчів в лізі. Тронхейм, Норвегія: Норвезький технічний університет.

12. Ігірі, С. П. (грудень, 2014). Покращена система прогнозування результатів футбольних матчів. IOSR Журнал інженерії (IOSRJEN), 12 - 20.

13. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

14. Кавецький В. В. Економічне обґрунтування інноваційних рішень: практикум / В. В. Кавецький, В. О. Козловський, І. В. Причепка – Вінниця : ВНТУ, 2016. – 113 с.

15. Роберт Хорольський, Юлія Іващенко, Ірина Сергієнко, Назар Бондар, Максим Кузьма, Михайло Коваль, Ірина Сергієнко. Правове регулювання гравального бізнесу: європейський досвід, практика в пострадянських країнах, перспективи України. 2016. 138 с.

16. Егор Шваюк. Як грати на тоталізаторі і вигравати. Ставки на спорт. [Електронний ресурс]. Режим доступу – <http://vprognose.com.ua/stavki-na-sport/yak-grati-na-totalizatori-i-vigravati>

17. Штовба С. Д., Вивдюк В.В. ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ НА ОСНОВІ НЕЧІТКИХ ПРАВИЛ. 2002. [Електронний ресурс]. Режим доступу – http://www.serhiyshtovba.narod.com/doc/Shtovba_Vivduk__Football_VMU_2002_1.pdf

18. Колискова О.З., Гайда Б.А. Дослідження та розробка математичного забезпечення для прогнозування спортивних. [Електронний ресурс]. Режим доступу – https://er.knutd.edu.ua/bitstream/123456789/5430/1/20170306_ITNVP_P120-123.pdf.

19. Рекомендації по правильному аналізу футбольних матчів [Електронний ресурс]. Режим доступу – <https://bet.ua/uk/baza-znaniy/rekomendacii-po-pravilnomu-analizu-futbolnyh-matchej>

20. Аналіз та рекомендації для ставок на футбольні матчі [Електронний ресурс]. Режим доступу – <https://zhzh.info/news/2023-12-11-51305>

ДОДАТКИ

Додаток А (обов'язковий)

Протокол перевірки кваліфікаційної роботи на наявність текстових
запозиченьПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ
НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬНазва роботи: Інформаційна технологія прогнозування результатів футбольних
матчівТип роботи: магістерська кваліфікаційна робота
(БДР, МКР)Підрозділ кафедра комп'ютерних наук, ФІТА
(кафедра, факультет)

Показники звіту подібності Unicheck

Оригінальність 99,5% Схожість 0,5%

Аналіз звіту подібності (відмітити потрібне):

- Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.
- Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її виконання автором. Роботу направити на розгляд експертної комісії кафедри.
- Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи



Хмель С.А.

Керівник роботи

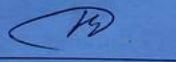


Петришин С.І.

Опис прийнятого рішення

Магістерську кваліфікаційну роботу допущено до захисту

Особа, відповідальна за перевірку



Озеранський В.С.

Додаток Б (обов'язковий)

Лістинг програми

```
#### Get all season matches results ####

season_matches_results_url = "https://www.transfermarkt.com/" +
competition_comps[0] + "/gesamtspielplan/wettbewerb/" +
competition_comps[1] + "/saison_id/" + season_year
response = requests.get(season_matches_results_url, headers={'User-
Agent': 'Custom'})
season_matches_results_url_content = BeautifulSoup(response.content,
'html.parser')

base_url = "https://www.transfermarkt.com"
df_train_rows = []
df_target_rows = []

print("Please wait, this may take a few minutes...")

for each_matchday in season_matches_results_url_content.findAll("div",
class_ = "large-6 columns"):
    curr_matchday_table = each_matchday.find("table")
    if curr_matchday_table != None:
        curr_matchday_tbody = curr_matchday_table.find("tbody")
        for each_row in curr_matchday_tbody.findAll("tr", class_=None):
#("tr")
            home_team_points = 0
            away_team_points = 0
            last_matches_between_teams_url = base_url

            for game_result in each_row.findAll("td", class_ = "zentriert
hauptlink"):
                res = game_result.find("a", class_ = "ergebnis-
link").get_text(strip=True)
                res = res.partition(':')
                home_team_goals = int(re.search(r'\d+', res[0]).group())
                away_team_goals = int(re.search(r'\d+', res[2]).group())
                if home_team_goals > away_team_goals:
                    home_team_points = 2
                    df_target_rows.append(2)
                elif home_team_goals < away_team_goals:
                    away_team_points = 2
                    df_target_rows.append(0)
                else:
                    home_team_points = 1
                    away_team_points = 1
                    df_target_rows.append(1)
                last_matches_between_teams_url += game_result.find("a",
class_ = "ergebnis-link").get('href')
```

```

        for home_team_stats in each_row.findAll("td", class_ = "text-
right no-border-rechts hauptlink"):
            home_team_position = home_team_stats.find("span", class_ =
"tabellenplatz").get_text(strip=True)
            home_team_position = int(re.search(r'\d+',
home_team_position).group())
            home_team_name = home_team_stats.find("a", class_ =
"vereinprofil_tooltip").get_text(strip=True)
            switcher_home_num = team_index_switcher(home_team_name)
            try:          # to prevent divide by zero
                home_team_curr_points =
teams_stats_sum[switcher_home_num][3] /
teams_stats_sum[switcher_home_num][5]
            except:
                home_team_curr_points = 0

            teams_stats_sum[switcher_home_num][2] =
home_team_position
            teams_stats_sum[switcher_home_num][5] += 1
            teams_stats_sum[switcher_home_num][3] += home_team_points
            #print(teams_stats_sum[switcher_home_num])

        for away_team_stats in each_row.findAll("td", class_ = "no-
border-links hauptlink"): #away team
            away_team_position = away_team_stats.find("span", class_ =
"tabellenplatz").get_text(strip=True)
            away_team_position = int(re.search(r'\d+',
away_team_position).group())
            away_team_name = away_team_stats.find("a", class_ =
"vereinprofil_tooltip").get_text(strip=True)
            switcher_away_num = team_index_switcher(away_team_name)
            try:          # to prevent divide by zero
                away_team_curr_points =
teams_stats_sum[switcher_away_num][4] /
teams_stats_sum[switcher_away_num][6]
            except:
                away_team_curr_points = 0

            teams_stats_sum[switcher_away_num][2] =
away_team_position
            teams_stats_sum[switcher_away_num][6] += 1
            teams_stats_sum[switcher_away_num][4] += away_team_points
            #print(teams_stats_sum[switcher_away_num])
            #print("-----")

    last_matches_between_teams_url = re.sub('/index/',
'/vorbericht/', last_matches_between_teams_url)
    response = requests.get(last_matches_between_teams_url,
headers={'User-Agent': 'Custom'})

```

```

last_matches_between_teams_url_content =
BeautifulSoup(response.content, 'html.parser')

last_matches_between_teams =
last_matches_between_teams_url_content.find("div", class_="table-header",
text=re.compile('Last')).find_parent("div",
class_="box")

curr_match_home_team_points = 0
curr_match_away_team_points = 0
match_results = last_matches_between_teams.find("tbody")
for each_match_result in match_results.findAll("tr"):
    curr_match_competition = each_match_result.find("td",
class_="zentriert").find('img').get('title') # want to get only League
games
    curr_match_competition =
curr_match_competition.partition('.')[0]
    if curr_match_competition.lower() ==
competition_comps[1].lower():
temp = each_match_result.find("td", class_="no-
border-links")
curr_home_team_id = temp.find("a")['id']
#temp.get("id")
home_team_id = teams_stats_sum[switcher_home_num][7]

if home_team_id == curr_home_team_id:
curr_match_res = each_match_result.find("a",
title="Match report", class_="ergebnis-link").get_text(strip=True)
curr_match_res = curr_match_res.partition(':')
curr_match_home_team_goals =
int(re.search(r'\d+', curr_match_res[0]).group())
curr_match_away_team_goals =
int(re.search(r'\d+', curr_match_res[2]).group())
if curr_match_home_team_goals >
curr_match_away_team_goals:
curr_match_home_team_points += 2
elif curr_match_home_team_goals <
curr_match_away_team_goals:
curr_match_away_team_points += 2
else:
curr_match_home_team_points += 1
curr_match_away_team_points += 1

### df columns building ###

last_matches_between_teams_results =
curr_match_home_team_points - curr_match_away_team_points
positions = teams_stats_sum[switcher_away_num][2] -
teams_stats_sum[switcher_home_num][2]

```

```
        values = teams_stats_sum[switcher_home_num][1] /  
teams_stats_sum[switcher_away_num][1]  
  
        last_each_team_results = home_team_curr_points -  
away_team_curr_points  
        df_train_rows.append(tuple([values, positions,  
last_each_team_results, last_matches_between_teams_results]))
```


Додаток В (обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУВАННЯ
РЕЗУЛЬТАТІВ ФУТБОЛЬНИХ МАТЧІВ

Виконав: студент 2-го курсу,
групи ЗКН-22м
спеціальності 122 «Комп'ютерні науки»
(шифр і назва напрямку підготовки, спеціальності)


_____ Хмель С.А.
(прізвище та ініціали)

Керівник: к.т.н., ст.в. каф. КН


_____ Петришин С.І.
(прізвище та ініціали)

« 07 » _____ 12 _____ 2023 р.

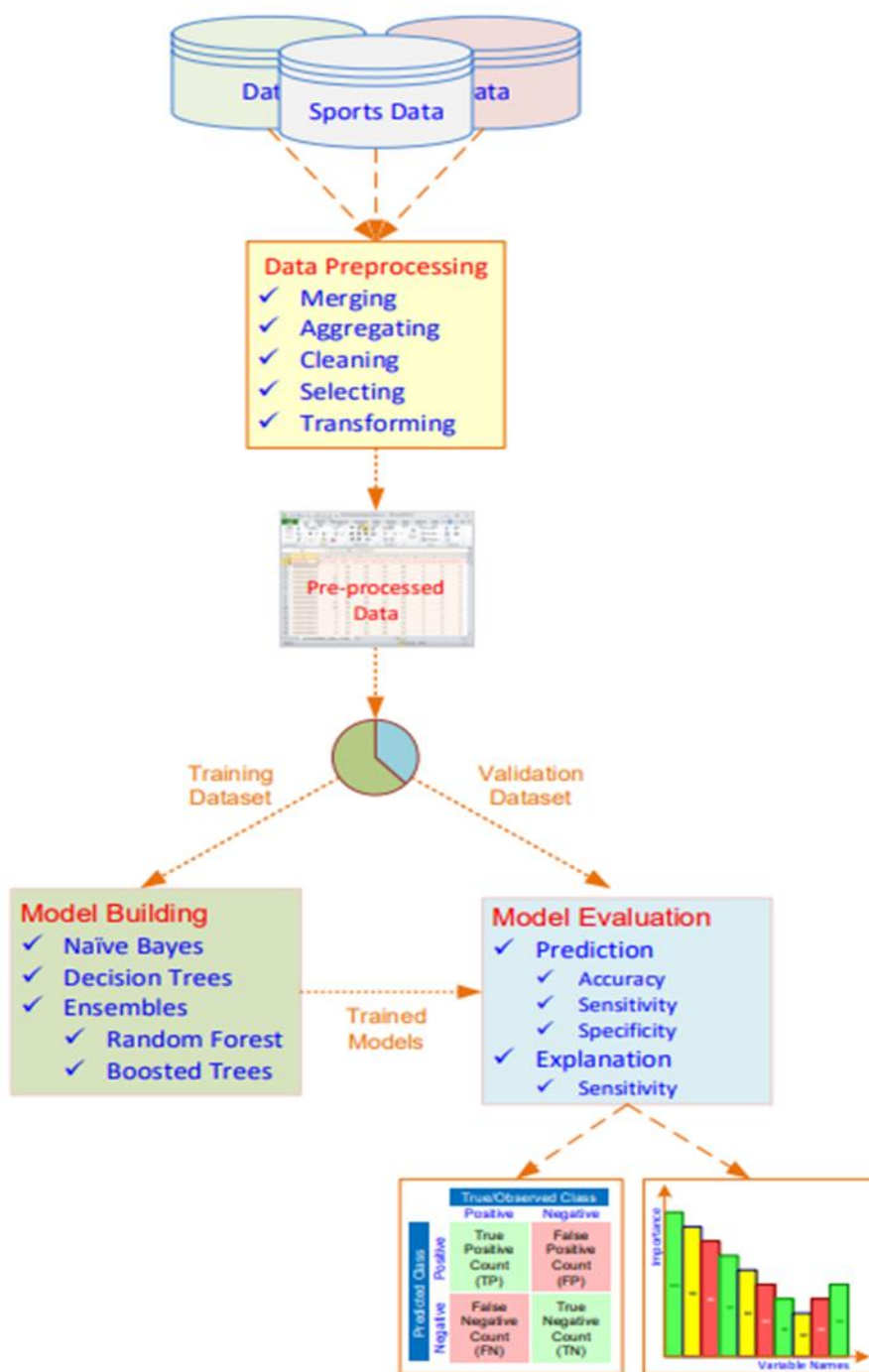


Рисунок В.1 – Структурна схема взаємодії роботи модуля прогнозування результатів футбольних матчів

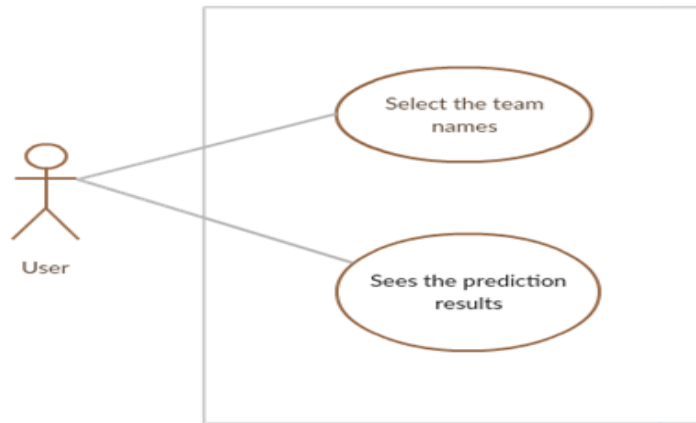


Рисунок В.2 – UML-діаграма прецедентів (use-case diagram) модуля прогнозування результатів футбольних матчів

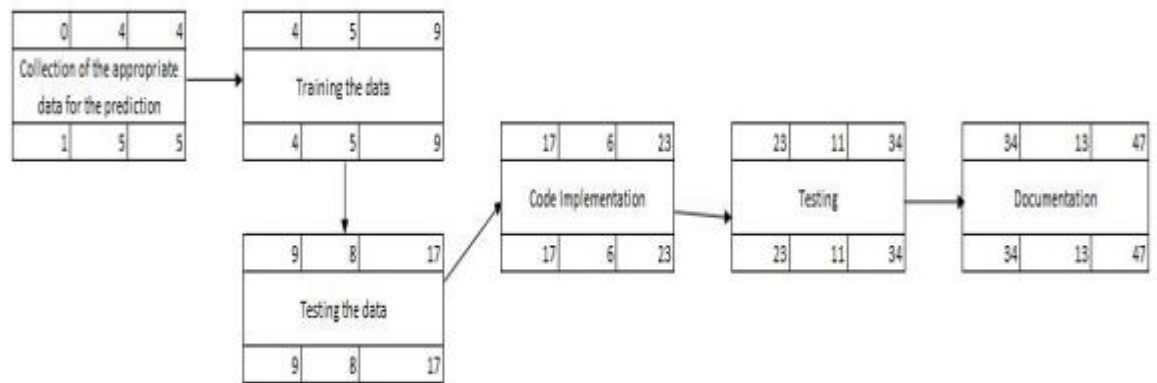


Рисунок В.3 – UML-діаграма діяльності (activity diagram) модуля прогнозування результатів футбольних матчів

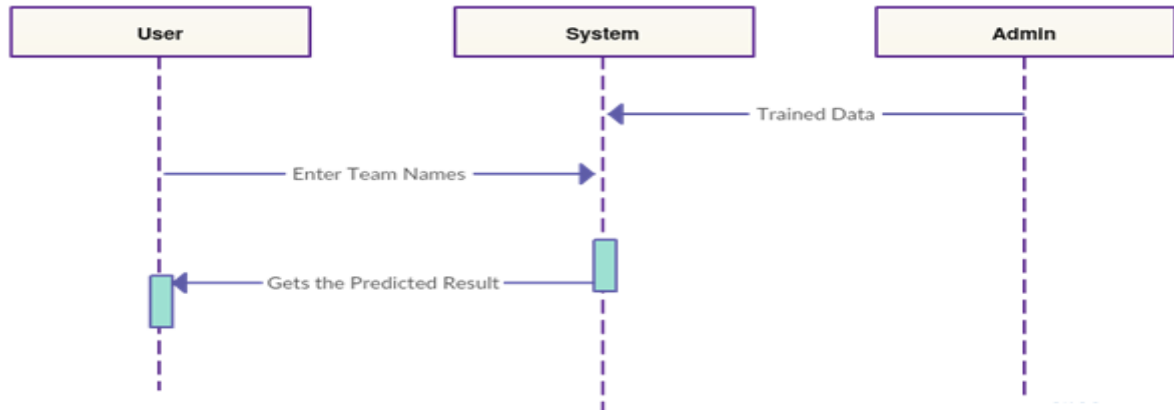


Рисунок В.4 – UML-діаграма послідовності (sequence diagram) модуля прогнозування результатів футбольних матчів

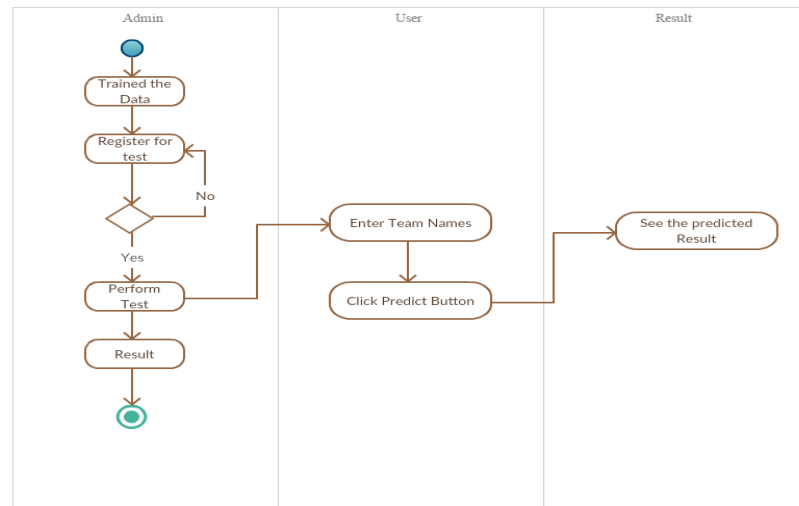


Рисунок В.5 – UML-діаграма подій (event diagram) модуля прогнозування результатів футбольних матчів

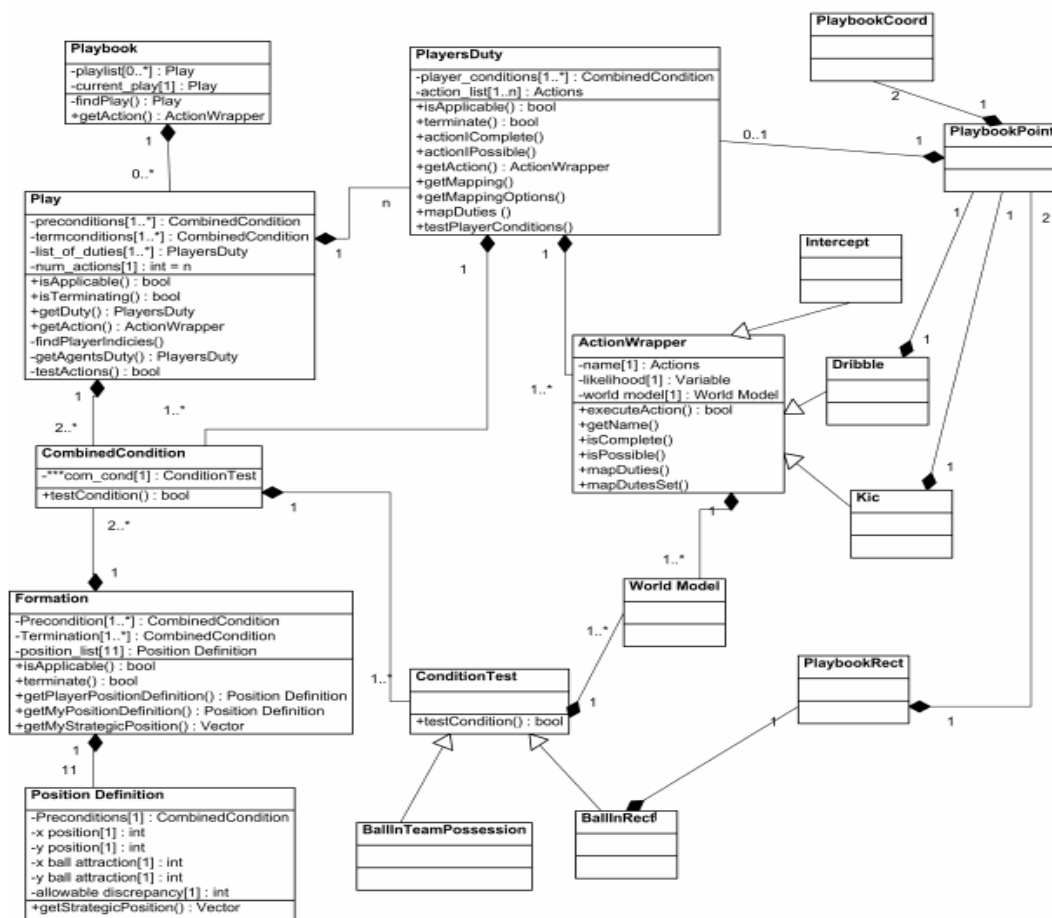
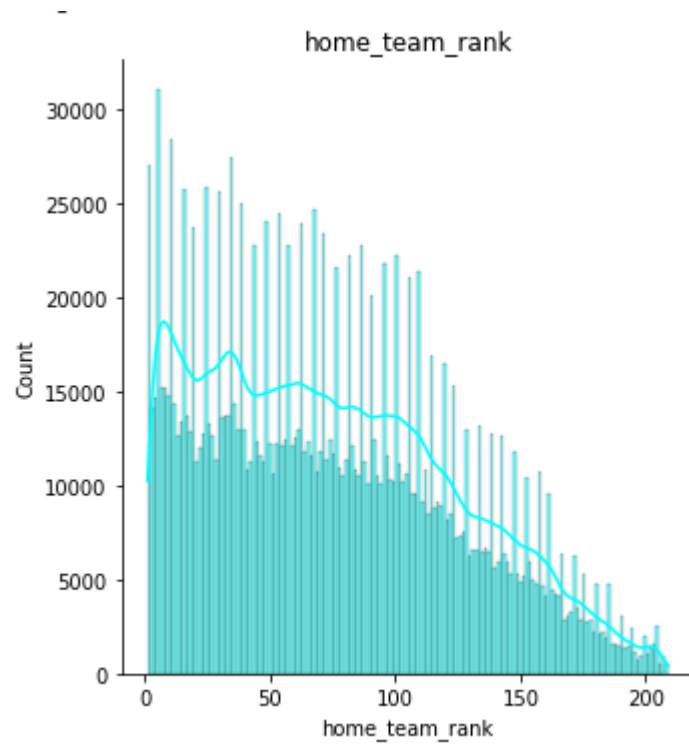


Рисунок В.6 – UML-діаграма класів (class diagram) модуля прогнозування результатів футбольних матчів



<Figure size 1000x300 with 0 Axes>

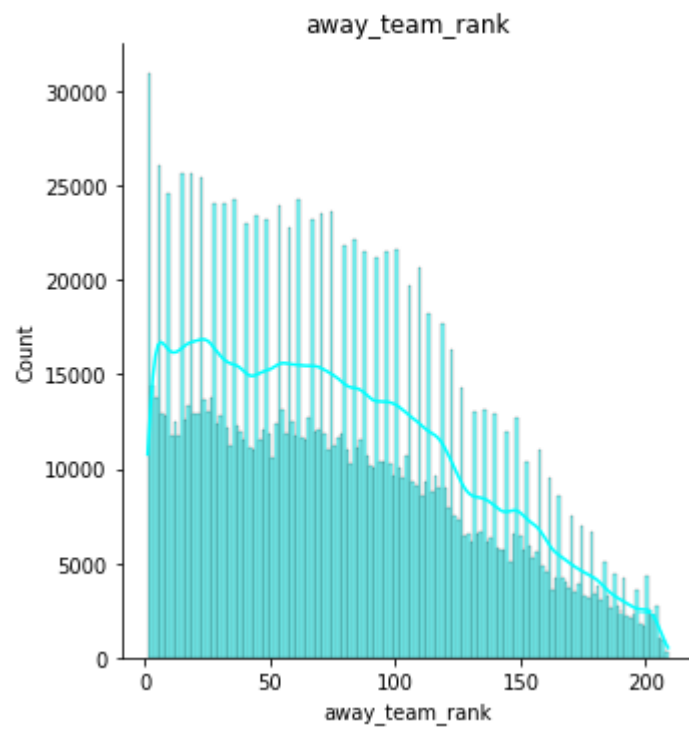


Рисунок В.7 – Скріншот роботи додатку

Додаток Г (довідниковий)

Інструкція користувача

Відкриємо програмний додаток у середовищі Jupiter Notebook, та підвантажимо файл з даними у csv форматі, отримаємо навчальну вибірку даних:

	rank	country_full	country_abrv	total_points	previous_points	rank_change	cur_year_avg	cur_year_avg_weighted	last_year_avg	last
0	1	Germany	GER	0.0	57	0	0.0	0.0	0.0	0.0
1	2	Italy	ITA	0.0	57	0	0.0	0.0	0.0	0.0
2	3	Switzerland	SUI	0.0	50	9	0.0	0.0	0.0	0.0
3	4	Sweden	SWE	0.0	55	0	0.0	0.0	0.0	0.0
4	5	Argentina	ARG	0.0	51	5	0.0	0.0	0.0	0.0

Рисунок Г.1 – Утворення навчальної вибірки даних

Нарешті, на основі зібраних параметрів та результатів гри програма створює навчальну вибірку (`df_train_rows`) та цільові значення (`df_target_rows`).

Далі проведемо ранжування основних числових показників.

	rank	total_points	previous_points	rank_change	cur_year_avg	cur_year_avg_weighted	last_year_avg	last_year
count	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000	57793.000000
mean	101.628086	122.068637	332.302926	-0.009897	61.798602	61.798602	61.004602	61.004602
std	58.618424	260.426863	302.872948	5.804309	138.014883	138.014883	137.688204	137.688204
min	1.000000	0.000000	0.000000	-72.000000	0.000000	0.000000	0.000000	0.000000
25%	51.000000	0.000000	56.000000	-2.000000	0.000000	0.000000	0.000000	0.000000
50%	101.000000	0.000000	272.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	152.000000	92.790000	525.000000	1.000000	32.250000	32.250000	26.660000	26.660000
max	209.000000	1775.030000	1920.000000	92.000000	1158.660000	1158.660000	1169.570000	1169.570000

Рисунок Г.2 – Проведення ранжування

На наступному етапі з таблиць видаляються усі значення NULL та формується вибірка даних між двома країнами-гравцями.

	date	home_team	away_team	home_score	away_score	tournament_x	city_x	country_x	neutral_x	year_x	rank_x	rank_date_x
0	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
1	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
2	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
3	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08
4	1993-01-23	Ghana	Niger	4	0	Friendly	Kumasi	Ghana	False	1993	48	1993-08-08

Рисунок Г.3 – Формування вибірки між країнами-гравцями

Сформуємо боксплот графік між цими командами.

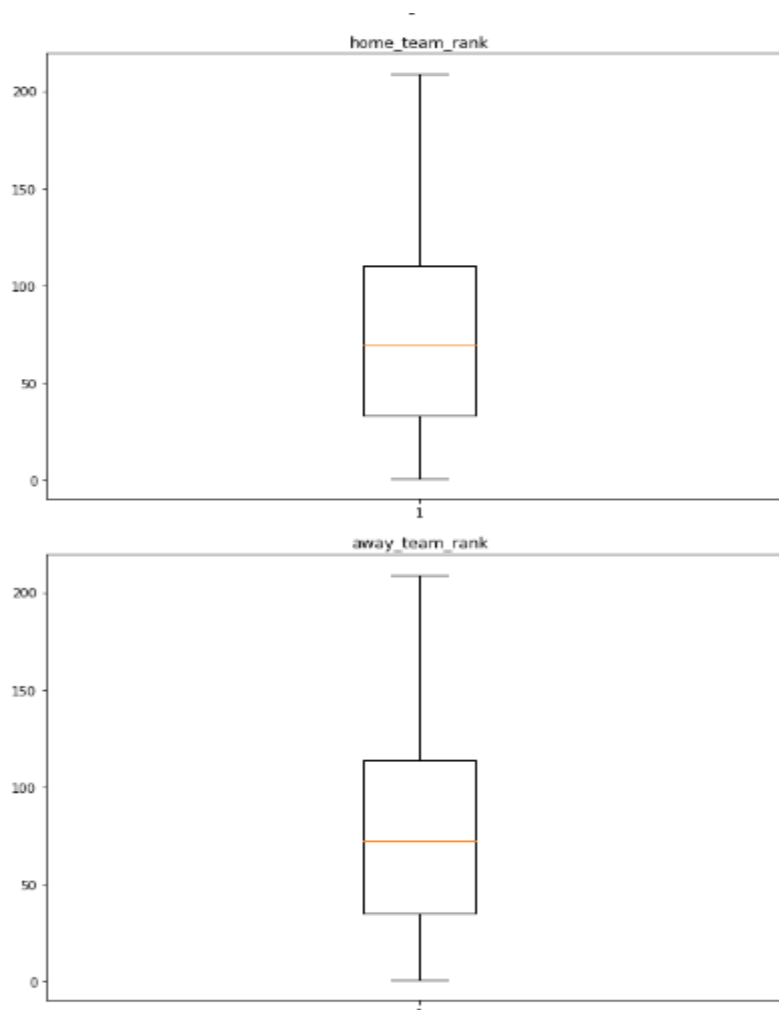
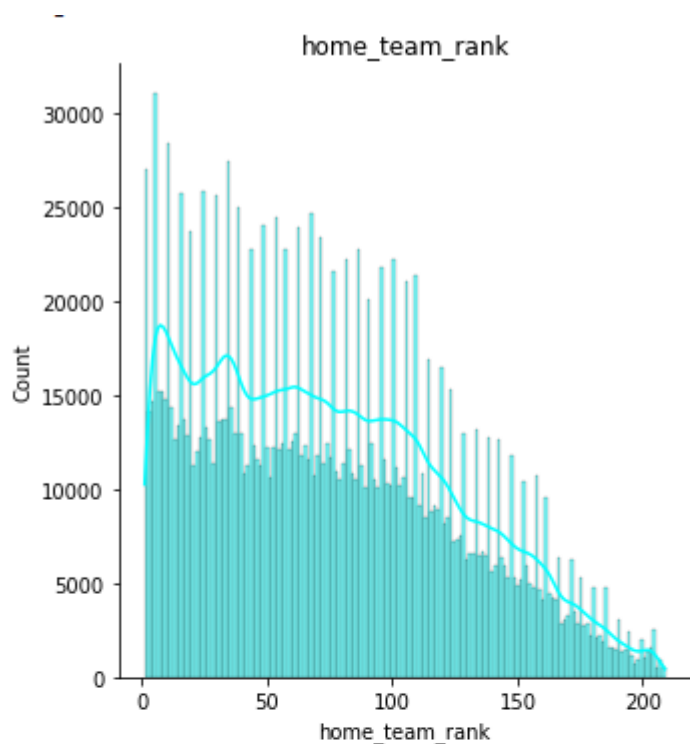


Рисунок Г.4 – Формування боксплот-графіків

Надалі ми можемо формувати іншу аналітику по прогнозуванню матчів у потрібному нам поданні.



<Figure size 1000x300 with 0 Axes>

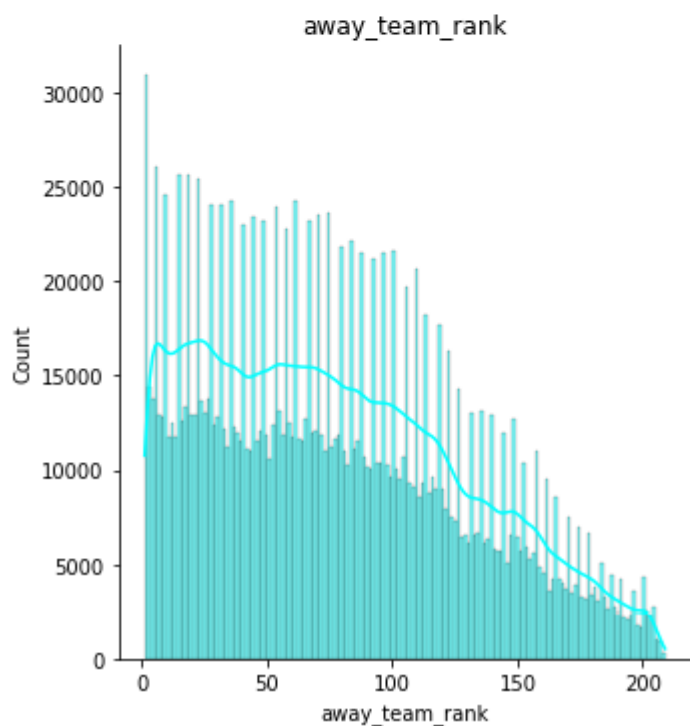


Рисунок Г.5 – Гістограми прогнозування