

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи освітнього ступеня «магістр»
за спеціальністю 121 «Інженерія програмного
забезпечення»(освітня програма «Інженерія програмного
забезпечення»)

на тему:

**«Прогнозування результатів тенісних матчів на
основі статичних даних та аналізу постів
спортсменів із соцмереж»**

Виконав студент групи ЗІПЗм-22-1
ЗИЧКОВ Дмитро Леонідович

Керівник роботи:
МОРОЗОВ Андрій Васильович

Рецензент:
ЄФРЕМОВ Юрій Миколайович

Житомир – 2023

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
 ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
 ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
 КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«ЗАТВЕРДЖУЮ»

зав. кафедри інженерії
 програмного забезпечення

Тетяна ВАКАЛЮК

«18» вересня 2023 р.

ЗАВДАННЯ на кваліфікаційну роботу

Здобувач вищої освіти: **ЗИЧКОВ Дмитро Леонідович**

Керівник роботи: **МОРОЗОВ Андрій Васильович**

Тема роботи: **«Прогнозування результатів тенісних матчів на основі статичних даних та аналізу постів спортсменів із соцмереж»**, затверджена Наказом закладу вищої освіти від «18» вересня 2023 р. № 481/с

Вихідні дані для роботи: предметом дослідження є методи та технології аналізу даних та машинного навчання, зокрема, використання нейронних мереж для прогнозування результатів спортивних змагань. Об'єктом дослідження є програмний модуль, розроблений для збору та аналізу даних для прогнозування результатів тенісних матчів.

Консультанти з магістерської кваліфікаційної роботи із зазначенням розділів, що їх стосуються:

Розділ	Консультант	Завдання видав	Завдання прийняв
1	Морозов А. В.	22.09.2023р.	22.09.2023р.
2	Морозов А. В.	22.09.2023р.	22.09.2023р.
3	Морозов А. В.	22.09.2023р.	22.09.2023р.

РЕФЕРАТ

Індивідуальне завдання кваліфікаційної магістерської роботи зосереджено на вивченні та розробці методів прогнозування результатів тенісних матчів. Основна увага приділяється аналізу статичних даних про минулі ігри, а також аналізу постів спортсменів у соціальних мережах. Робота спрямована на підвищення точності прогнозувань та ефективності використання інформації з відкритих джерел для оцінки поточної форми гравців.

Кваліфікаційна магістерська робота на тему «Прогнозування результатів тенісних матчів на основі статичних даних та аналізу постів спортсменів із соцмереж» включає теоретичний огляд, практичну частину, та пояснювальну записку. Пояснювальна записка містить вступ, 3 розділи, список літератури з 42 найменувань, 18 рисунків. Загальний обсяг роботи складає 62 сторінок.

КЛЮЧОВІ СЛОВА: ТЕНІС, ПРОГНОЗУВАННЯ, СТАТИСТИЧНІ ДАНІ, СОЦІАЛЬНІ МЕРЕЖІ, МАШИННЕ НАВЧАННЯ, АНАЛІЗ ДАНИХ, СПОРТИВНА АНАЛІТИКА.

ABSTRACT

The individual task of the master's thesis focuses on studying and developing methods for predicting the outcomes of tennis matches. The primary attention is given to the analysis of static data on past games and the examination of athletes' posts on social media. The work aims to enhance the accuracy of predictions and the effectiveness of using information from open sources to assess the current form of players.

The master's thesis on the topic "Predicting tennis match outcomes based on static data and analysis of athletes' social media posts" includes a theoretical review, a practical part, and an explanatory note. The explanatory note contains an introduction, 3 chapters, a bibliography with 42 entries, 18 figures. The total volume of the work is 62 pages.

KEYWORDS: TENNIS, PREDICTION, STATISTICAL DATA, SOCIAL MEDIA, MACHINE LEARNING, DATA ANALYSIS, SPORTS ANALYTICS.

ЗМІСТ

ВСТУП	7
РОЗДІЛ 1. МЕТОДИ АНАЛІЗУ ДАНИХ ТА ПРОГНОЗУВАННЯ В СПОРТІ	10
1.1. Огляд сучасних підходів до аналізу даних у спортивній індустрії	10
1.2. Методи машинного навчання та їх застосування у спорті.....	17
1.3. Аналіз соціальних мереж як інструмент прогнозування в спорті	25
1.4. Мета та сфера застосування машинного навчання в спортивній аналітиці: постановка завдання для подальшого розвитку	31
ВИСНОВКИ ДО РОЗДІЛУ 1	31
РОЗДІЛ 2. МЕТОДИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТЕНІСНИХ МАТЧІВ	32
2.1. Аналіз статистичних даних тенісних матчів та спортсменів	32
2.2. Розробка прогнозової моделі	37
2.3. Оцінка продуктивності моделі та значущості даних	43
ВИСНОВКИ ДО РОЗДІЛУ 2	47
РОЗДІЛ 3. РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ТЕНІСНИХ МАТЧІВ З ВИКОРИСТАННЯМ ЧАТ-БОТА	48
3.1. Розробка та налаштування чат-бота для Telegram	48
3.2. Інтеграція з базою даних та веб-API для збору даних	51
3.3. Розробка алгоритмів для обробки запитів користувачів	52
3.4. Тестування функціональності та ефективності чат-бота	55
ВИСНОВКИ ДО РОЗДІЛУ 3	56
ВИСНОВКИ	58
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	59

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

API – Application Programming Interface

ML – Machine Learning

NLP – Natural Language Processing

DB – Database

SQL – Structured Query Language

JSON – JavaScript Object Notation

HTTP – Hypertext Transfer Protocol

ВСТУП

Актуальність теми. У сучасному світі спорту, зокрема в тенісі, застосування аналітичних та прогнозних інструментів стає все більш поширеним як у професійних, так і в аматорських лігах. Використання статичних даних про минулі ігри та аналізу соціальних мереж спортсменів відкриває нові можливості для глибокого розуміння ігрових стратегій та психологічного стану гравців. Важливість прогнозування в спорті зростає з кожним роком, оскільки воно має велике значення не тільки для тренерів та спортсменів, а й для аналітиків, букмекерів та фанатів.

Розвиток інструментів для прогнозування результатів спортивних змагань, особливо використання машинного навчання та аналізу великих даних, стає ключовим фактором у сучасному спортивному аналізі. Прогнозування результатів на основі об'єктивних даних та аналізу поведінки спортсменів у соціальних мережах дозволяє отримувати цінну інформацію, яка може вплинути на рішення щодо стратегії та підготовки до матчів.

Актуальність даної теми також посилюється зростаючим попитом на кваліфікованих аналітиків в галузі спортивних прогнозів та аналізу даних. Розвиток цифрових технологій та доступність великих обсягів даних створюють нові можливості для аналізу та прогнозування в спорті, відкриваючи шляхи для створення більш точних і об'єктивних прогнозних моделей.

Впровадження сучасних методів аналізу даних у спортивній індустрії дозволяє підвищити якість і точність прогнозувань, а також сприяє покращенню загальної стратегії та підготовки спортсменів до змагань. Це веде до зростання вимог до аналітичних інструментів у спорті та підвищення їхньої важливості в сучасній машинобудівній галузі, зокрема у розвитку тенісу як спорту:

- **Точність та обґрунтованість прогнозів:** використання статичних даних та аналіз соціальних мереж забезпечує високу точність прогнозів, дозволяючи отримувати глибокий аналіз ігрової стратегії та психологічного стану спортсменів.

- **Економія ресурсів:** автоматизація аналізу даних та прогнозування може зменшити витрати часу та ресурсів, які традиційно витрачаються на ручний аналіз та оцінку.
- **Колаборація та обмін даними:** технології аналізу даних полегшують співпрацю між різними спеціалістами, такими як тренери, аналітики та спортсмени, а також сприяють обміну важливою інформацією.
- **Інновації в спортивному аналізі:** застосування передових методів машинного навчання та обробки великих даних відкриває нові перспективи для інновацій у спортивній аналітиці.
- **Стандартизація та якість аналізу:** використання структурованих методів аналізу та однакових критеріїв оцінювання допомагає дотримуватися високих стандартів якості в спортивному прогнозуванні.

Загалом, методи прогнозування результатів тенісних матчів є невід'ємною частиною сучасного спортивного аналізу. Вони не лише сприяють підвищенню точності прогнозів, але й вносять вагомий внесок у розвиток спортивної стратегії та психології, сприяючи інноваціям та покращенню загальної якості спортивних змагань.

Мета і завдання роботи. Метою роботи є розробка програмного модуля для прогнозування результатів тенісних матчів, який використовує статичні дані та аналіз соціальних мереж спортсменів.

Досягнення поставленої мети передбачає розв'язання таких **завдань**:

- обґрунтувати актуальність теми роботи та значення прогнозування в спортивному аналізі;
- здійснити аналіз сучасних методів прогнозування в спорті, зокрема в тенісі;
- провести аналіз існуючих моделей машинного навчання для прогнозування спортивних результатів;
- розробити алгоритм роботи програмного модуля для аналізу статичних даних і соціальних мереж;

- реалізувати програмний модуль, який інтегрує методи машинного навчання та обробки великих даних для прогнозування результатів тенісних матчів;
- розробити інструкцію для користування програмним модулем.

Об'єкт дослідження є програмний модуль для прогнозування результатів тенісних матчів.

Предметом дослідження є аналіз статичних даних та соціальних мереж для прогнозування в спорті.

Методи дослідження. В кваліфікаційній роботі було застосовано методи машинного навчання та обробки великих даних для розробки та валідації програмного модуля.

Наукова новизна отриманих результатів полягає в такому:

- розроблено алгоритм для інтеграції статичних даних та аналізу соціальних мереж для прогнозування результатів;
- створено прототип програмного модуля, який використовує новітні методи аналізу даних для прогнозування в тенісі.

Практичне значення роботи полягає у такому:

- підвищення точності прогнозів результатів тенісних матчів;
- надання тренерам та аналітикам ефективного інструменту для стратегічного планування та підготовки спортсменів.

Апробація результатів роботи. Матеріали кваліфікаційної роботи доповідалися на XVII Міжнародній науково-практичній конференції магістрантів та аспірантів: “Теоретичні та практичні дослідження молодих вчених” 28-30 листопада 2023 р.

РОЗДІЛ 1. МЕТОДИ АНАЛІЗУ ДАНИХ ТА ПРОГНОЗУВАННЯ В СПОРТІ

1.1. Огляд сучасних підходів до аналізу даних у спортивній індустрії

Сфера спорту завжди була благодатним ґрунтом для конкуренції, стратегії та людської досконалості. Однак в останні роки додатковий компонент постійно змінює ландшафт спорту: аналіз даних. Цей перехід до підходу, що керується даними, у спорті, починаючи від аматорських ліг і закінчуючи професійними змаганнями, був кардинальним, змінивши спосіб гри, формулювання стратегій і підготовку спортсменів.

Аналіз даних у спорті не зовсім новий. Десятиліттями статистики обробляли цифри, щоб отримати розуміння продуктивності гравців і командних стратегій. Однак поява передових технологій і складного програмного забезпечення катапультиувала цю практику в нову еру. Сьогодні спортивна аналітика передбачає всебічний аналіз величезних обсягів даних, щоб отримати інформацію, яка раніше була невидима неозброєним оком або традиційним тренерським посібником. У традиційних спортивних умовах рішення часто приймалися на основі інтуїції, досвіду та «інтуїції». Хоча ці елементи залишаються важливими, тепер вони доповнені статистичними даними. Тренери та менеджери мають доступ до широкого діапазону даних, включаючи стан здоров'я гравців, показники продуктивності в грі та навіть прогнозу аналітику щодо майбутніх виступів. Ця інформація дає змогу приймати більш обґрунтовані рішення, від розробки стратегії гри до керування втомою гравців.

Аналіз даних істотно змінив стратегії гри в різних видах спорту [8]. У баскетболі, наприклад, революція в аналітиці призвела до популяризації стратегії триочкових кидків. Дані показали, що триочкові кидки пропонують вищу очікувану віддачу порівняно з довгими двоочковими, що призвело до стратегічних змін у спорті. У бейсболі концепція саберметрики [11] змінила спосіб оцінювання гравців і прийняття рішень у грі.

Одне з найважливіших застосувань аналізу даних у спорті стосується продуктивності та здоров'я гравців [2]. Носима технологія тепер дозволяє командам відстежувати фізичний стан гравців у режимі реального часу, відстежуючи все від частоти серцевих скорочень до пройденої дистанції під час гри. Ці дані допомагають створювати персоналізовані режими тренувань, оптимізувати продуктивність гравців і, що важливо, запобігати травмам. Наприклад, футбольні команди [3] використовують дані для моніторингу навантаження гравців і коригування інтенсивності тренувань, щоб мінімізувати ризик травм. Окрім фізичних показників, аналітика даних також заглиблюється в психологічні аспекти спорту. Аналізуючи ігрові ситуації, реакції гравців і навіть поведінку в соціальних мережах, команди отримують уявлення про психічний стан і стійкість спортсменів [17,23]. Розуміння командної динаміки та психології гравців може мати вирішальне значення в ситуаціях високого тиску, таких як серія пенальті у футболі [35] або остання чверть гри в баскетболі.

Сплеск збору даних [12] приносить свої власні труднощі. Обсяг даних може бути величезним і вміння полягає в ідентифікації відповідних даних та їх ефективній інтерпретації. Крім того, довіра до даних має бути збалансована з людським аспектом спорту, де непередбачуваність та емоційні фактори часто відіграють значну роль.

З розвитком аналітики даних у спорті також з'явилися етичні міркування. Питання, пов'язані з конфіденційністю даних, особливо щодо персональних даних про здоров'я спортсменів, є першочерговими. Крім того, є питання підтримки рівних умов гри. Оскільки деякі команди отримують доступ до більш розширених інструментів аналізу даних, ніж інші, забезпечення справедливості та конкурентного балансу стає проблемою.

Вирішальний аспект у спорті представляє статистичний аналіз сучасних спортивних змагань і менеджменту [1]. Це передбачає використання історичних даних для виявлення закономірностей, тенденцій і розуміння, які можуть суттєво вплинути на процес прийняття рішень у спорті. Цей підхід стає все

більш витонченим із появою передових методів аналітики та програмного забезпечення. За своєю суттю спортивна статистика [15] — це цифри — забиті голи, набрані ярди, здобуті очки та багато іншого. Однак ці цифри — не просто цифри. Вони є життєво важливими показниками продуктивності, стратегії та потенціалу. Аналізуючи ці показники, тренери, менеджери та аналітики можуть отримати значущу інформацію про ефективність команди та внесок окремого спортсмена [24]. Одним із найпростіших застосувань статистичного аналізу в спорті є перевірка результатів матчів. Для цього потрібно дивитися не тільки на рахунок, щоб зрозуміти, як розгорталася гра. Аналітики аналізують такі аспекти, як відсоток володіння м'ячем у футболі [9,28], точність кидка в баскетболі чи невимушені помилки в тенісі [21,29,30]. Такий аналіз може виявити ефективність стратегії команди та виділити області, які потребують покращення. В індивідуальних видах спорту, таких як теніс [42], або командних видах спорту, таких як баскетбол і футбол, статистика гравців ретельно записується та аналізується. Це може включати точність кидка гравця, точність передач, пройдену дистанцію під час гри, підкати та багато іншого [7,20,32]. Аналізуючи ці показники, команди можуть визначити ключових гравців, зрозуміти сильні та слабкі сторони гравців і навіть прийняти обґрунтовані рішення під час трансферів гравців або драфтів.

Статистичний аналіз — це не лише розуміння минулих показників; все частіше йдеться про прогнозування майбутніх результатів. Прогнозні моделі використовують історичні дані для прогнозування майбутніх подій, як-от ймовірність отримання гравцем травми або перемоги команди в чемпіонаті. Крім того, виявлення тенденцій з часом може мати вирішальне значення для розробки довгострокової стратегії. Наприклад, тенденція може вказувати на те, що оборонна тактика команди є менш ефективною проти певних типів нападів, що спонукає до стратегічного коригування.

Комплексний аналіз даних про продуктивність команди може значно вплинути на стратегію команди. У таких видах спорту, як бейсбол [5,14] і крикет, статистичний аналіз призвів до впровадження інноваційних стратегій,

таких як підхід Moneyball [4], який зосереджується на використанні статистичного аналізу для виявлення недооцінених гравців. У баскетболі, як вказувалось раніше, аналітика привела до збільшення уваги до триочкових кидків, змінивши традиційні стилі гри [22].

Статистичний аналіз також революціонував пошук талантів. Тепер скаути мають доступ до великої кількості даних про потенційних гравців, починаючи від показників їх продуктивності та закінчуючи історією травм. Цей підхід до розвідки на основі даних допомагає командам приймати більш обґрунтовані рішення під час підписання нових талантів, забезпечуючи відповідність стилю гри та стратегічним потребам команди [10].

Незважаючи на переваги, статистичний аналіз у спорті не позбавлений проблем. Однією з важливих проблем є інтерпретація даних: цифри можуть вводити в оману, якщо їх не правильно контекстуалізувати. Крім того, існує ризик надмірної довіри до даних, потенційно пропускаючи такі нематеріальні елементи спорту, як командний дух і мораль гравців.

Також, у динамічному та конкурентному світі спорту поява технологій великих даних спричинила зміну парадигми. Аналіз великих обсягів даних, відомих як BigData [31], став ключовим інструментом у спортивній аналітиці, пропонуючи розуміння та розкриваючи складні закономірності, які традиційні методи можуть не помітити. Ця можливість включає аналіз даних соціальних мереж, показників фізичної продуктивності спортсменів і багато іншого.

BigData у спорті охоплюють широкий спектр інформації, зібраної з різних джерел. Він включає не лише основну статистику, як результати та показники ефективності гравців, але також поширюється на складніші дані, наприклад рухи гравців під час гри, біометричні дані та навіть настрої вболівальників, висловлені на платформах соціальних мереж. Здатність обробляти й аналізувати цей масивний і різноманітний набір даних відкрила нові шляхи для розуміння та підвищення спортивних результатів. Основне застосування BigData у спорті полягає в покращенні аналізу продуктивності. Аналізуючи дані, зібрані за допомогою датчиків і переносних пристроїв, команди можуть

отримати глибоке уявлення про продуктивність, витривалість і фізичний стан спортсмена. Ці дані допомагають пристосувати програми тренувань до індивідуальних потреб спортсменів, оптимізувати їх продуктивність і мінімізувати ризик травм.

Прогностична аналітика є ще одним важливим аспектом, де BigData відіграють вирішальну роль. Аналізуючи історичні дані, команди можуть прогнозувати майбутні результати, визначати потенційні ризики травм і навіть прогнозувати результати ігор або сезонів. Ця можливість прогнозування дозволяє тренерам і менеджерам приймати стратегічні рішення, такі як заміни гравців і стратегії гри, на основі прогнозів, керованих даними. Тобто, BigData виходить за межі поля чи суду. Аналіз даних про залучення шанувальників із соціальних мереж та інших цифрових платформ дає цінну інформацію про вподобання та поведінку шанувальників. Спортивні організації та маркетологи використовують ці дані, щоб покращити досвід уболівальників, адаптувати маркетингові стратегії та підвищити залученість уболівальників, що зрештою сприяє популярності та фінансовому успіху спорту [40].

Пошук талантів у спорті також революціонізував BigData. Аналізуючи комплексні набори даних, які включають не лише статистику продуктивності, але й записи про здоров'я та навіть психологічні профілі, скаути можуть точніше визначити перспективні таланти. Цей підхід на основі даних дозволяє приймати більш об'єктивні рішення під час залучення нових гравців.

Аналітика BigData дозволяє командам оптимізувати свої стратегії на основі конкретних даних, а не просто інтуїції. Наприклад, баскетбольні команди аналізують шаблони кидків гравців, щоб розробити стратегії нападу, тоді як футбольні команди використовують позиційні дані, щоб покращити свою тактику захисту [18,35,37]. Ця стратегічна оптимізація має вирішальне значення для отримання конкурентної переваги.

Незважаючи на свої переваги, обробка великих даних у спорті пов'язана з труднощами. Величезний обсяг і різноманітність даних може бути надзвичайною, що потребує складних інструментів і досвіду для ефективного

аналізу. Крім того, забезпечення точності та конфіденційності даних має першочергове значення, особливо при роботі з конфіденційною інформацією про спортсменів. А в міру того, як великі дані стають все більш вкоріненими у спорті, також виникають етичні міркування. Проблеми, пов'язані з конфіденційністю даних, особливо для спортсменів, і можливість неправомірного використання даних для досягнення конкурентних переваг є проблемними сферами. Забезпечення того, щоб використання великих даних у спорті залишалося етичним і справедливим, має вирішальне значення для його стійкої інтеграції.

Також критично важливим інструментом для розуміння та покращення фізичних і тактичних аспектів продуктивності спортсмена у світі спорту стала аналітика продуктивності. Ця сфера передбачає збір і аналіз даних, пов'язаних із фізичною підготовленістю та продуктивністю спортсмена під час тренувань і змагань [13]. Ці знання дозволяють тренерам розробляти персоналізовані програми тренувань і вдосконалювати тактику та стратегію гри.

Аналітика результативності у спорті стосується вимірювання, аналізу та інтерпретації різних аспектів продуктивності спортсмена. Вона охоплює широкий спектр точок даних, від основних показників, таких як швидкість і пройдена відстань, до більш складних, таких як витрати енергії, залучення м'язів і швидкість відновлення. Основною метою є максимізація потенціалу кожного спортсмена шляхом адаптації програм тренувань і відновлення відповідно до їхніх конкретних потреб.

Носимі технології зіграли ключову роль у вдосконаленні аналітики продуктивності. Такі пристрої, як GPS-трекери, монітори пульсу та акселерометри, надають дані про фізичний стан спортсмена в реальному часі під час тренувань і змагань. Ця інформація допомагає контролювати робоче навантаження, оцінювати рівень фізичної підготовки та виявляти ранні ознаки потенційних травм. Наприклад, бігуни на дистанції використовують носіння, щоб відстежувати пульс і відповідно регулювати інтенсивність тренувань.

Одне з найбільш важливих застосувань аналітики ефективності —

розробка навчальних програм, керованих даними. Аналізуючи зібрані дані, тренери можуть визначити сильні та слабкі сторони у виконанні кожного спортсмена. Це дозволяє їм створювати персоналізовані режими тренувань, які зосереджуються на конкретних сферах, таких як покращення витривалості, швидкості чи сили, тим самим підвищуючи загальну ефективність.

Запобігання травмам є ще одним важливим аспектом, на який аналітика ефективності справила значний вплив. Відстежуючи дані спортсменів, тренери та медичний персонал можуть визначити закономірності, які можуть призвести до травм. Раннє виявлення за допомогою аналізу даних дозволяє своєчасно втручатися, зменшуючи ризик травм і сприяючи швидшому одужанню. Окрім фізичної підготовки, аналітика продуктивності пропонує цінну інформацію про тактику та стратегію гри. Тренери та аналітики використовують дані, щоб зрозуміти, як різні стратегії відтворюються в реальних сценаріях. Вони аналізують такі аспекти, як позиціонування гравців, шаблони рухів і виконання гри, щоб удосконалити стратегії для майбутніх ігор. У таких командних видах спорту, як футбол і баскетбол, цей аналіз має вирішальне значення для оптимізації формування команд і тактики під час гри [6,39].

Хоча аналітика продуктивності пропонує численні переваги, вона не позбавлена проблем. Однією з головних проблем є інтерпретація величезних обсягів отриманих даних. Тренери та аналітики повинні розрізняти, які дані є найбільш релевантними та як вони перетворюються на практичні ідеї. Крім того, існує ризик перевантаження даними, коли величезний обсяг інформації може стати непосильним і контрпродуктивним.

Як і в будь-якому підході, що керується даними, етичні міркування мають першорядне значення в аналітиці ефективності. Це включає занепокоєння щодо конфіденційності спортсмена та можливого зловживання особистими даними про продуктивність. Забезпечення того, щоб збір і аналіз даних проводилися прозоро та за згодою спортсменів, має важливе значення для підтримки довіри та чесності в спорті.

1.2. Методи машинного навчання та їх застосування у спорті

Машинне навчання, підмножина штучного інтелекту, стає все більш важливим інструментом у сфері спортивної аналітики. Використовуючи різні методи машинного навчання, спортивні професіонали можуть отримати глибше розуміння продуктивності, оптимізації стратегії та запобігання травмам. У цьому розділі розглядаються методи машинного навчання та їх застосування в спортивній індустрії.

Машинне навчання (ML) у спорті означає радикальний перехід від традиційного аналізу даних до більш складного прогнозного моделювання. По суті, ML навчає алгоритми розпізнавати закономірності та робити прогнози з величезних наборів даних. У спортивному контексті ці алгоритми аналізують усе, починаючи від показників ефективності гравців і закінчуючи факторами навколишнього середовища, забезпечуючи детальне розуміння та передбачливі стратегії.

У спорті поширені дві основні методики ML: контрольоване та неконтрольоване навчання. Контрольоване навчання включає моделі, навчені на позначених наборах даних, де відомий результат кожної точки даних, що дозволяє алгоритму навчатися та робити прогнози. Цей підхід особливо корисний у прогнозуванні продуктивності, коли історичні дані можуть вказувати на майбутні тенденції. З іншого боку, неконтрольоване навчання працює з немаркованими даними, виявляючи приховані закономірності та кореляції. Цей метод важливий для виявлення нетрадиційних стратегій або нових атрибутів гравців, які можуть бути не відразу очевидними. Впровадження ML у спортивній аналітиці є новаторським. Обробляючи складні та багатовимірні дані, алгоритми машинного навчання пропонують розуміння, що виходить за межі людських можливостей. Наприклад, вони можуть проаналізувати всю кар'єру футболіста, враховуючи різні фактори, такі як вік, історія травм і ігрові умови, щоб передбачити майбутню продуктивність або вразливість до травм [16,34]. Однак ефективність цих моделей залежить від якості та кількості даних. Неточні або упереджені дані можуть призвести до

помилкових висновків, а недостатня кількість даних може зробити модель неефективною. Таким чином, спортивна індустрія все більше зосереджується на зборі повних високоякісних даних для введення в ці складні моделі машинного навчання.

Машинне навчання суттєво змінює аналіз продуктивності гравців, надаючи передбачувану інформацію, яка раніше була недоступна. Використовуючи історичні дані та дані в реальному часі, моделі ML можуть прогнозувати продуктивність гравців, ризики травм і навіть траєкторії відновлення. Одне вражаюче застосування – це прогнозування та лікування травм. Наприклад, футбольні клуби використовують алгоритми машинного навчання для аналізу даних про травми гравців, інтенсивність гри та фізичне навантаження для прогнозування ризиків травм. Такий підхід дозволяє застосовувати проактивні стратегії запобігання травмам, забезпечуючи довговічність гравців і стабільність командної роботи. Інше застосування полягає в оптимізації продуктивності. У баскетболі моделі ML оцінюють шаблони кидків, визначаючи зони, де гравець найбільш ефективний. Ці відомості дають змогу тренерам розробляти плани гри, які використовують ці сильні сторони. Подібним чином у таких видах спорту, як плавання чи легка атлетика, алгоритми ML аналізують техніку та фізіологічні дані спортсмена, щоб запропонувати покращення та підвищити їхню конкурентну перевагу.

Ці прогностичні моделі не обмежуються лише фізичними атрибутами, а й охоплюють психічні та психологічні фактори. Наприклад, вони можуть оцінити виступ тенісиста під тиском, щоб адаптувати тренування психологічної стійкості [41].

Потенціал ML у підвищенні продуктивності гравців величезний. Однак це також викликає етичні міркування, особливо щодо використання особистих даних про здоров'я. Забезпечення конфіденційності та безпеки цих даних має першочергове значення. Крім того, довіра до рішень, керованих даними, повинна бути збалансована з традиційною мудрістю тренера, щоб уникнути надмірної залежності від технологій.

Машинне навчання (ML) стало незамінним інструментом у сучасному спорті, особливо у сфері прийняття тактичних рішень. Ця складна гілка штучного інтелекту допомагає тренерам і аналітикам аналізувати складність ігрового процесу, пропонуючи інформацію, яка може суттєво вплинути на результат спортивних подій.

У командних видах спорту, таких як футбол і регбі, алгоритми машинного навчання використовуються для аналізу тонкощів ігрового процесу [25,26,27]. Ці алгоритми відсіюють величезну кількість даних, зібраних під час матчів, включаючи розташування гравців, володіння м'ячем, послідовність передач і оборонні домовленості. Мета полягає в тому, щоб виділити значущі шаблони, які можуть обґрунтувати стратегічні рішення. Наприклад, у футболі модель ML може аналізувати слабкі сторони захисту суперника [19,33,38], такі як сприйнятливість до контратак або вразливість у повітряних дуелях. Озброївшись цими знаннями, коуч може розробити конкретні стратегії використання цих слабких місць.

Крім того, ML допомагає оптимізувати формування команд і стилі гри. Аналізуючи історичні дані про продуктивність, алгоритми можуть рекомендувати найефективніші схеми проти конкретних суперників. Це може включати пропозиції щодо того, чи слід застосувати агресивний стиль пресингу чи більш консервативний підхід, заснований на володінні м'ячем. У регбі подібний аналіз може включати оптимізацію стратегій раку або прийняття рішення про те, коли застосовувати тактичні удари ногами.

Аналіз у реальному часі, який пропонують моделі ML, забезпечує додаткову стратегічну перевагу. Під час гри тренери отримують інформацію на основі даних, яка може спрямувати для негайного коригування тактики. Наприклад, якщо модель ML визначає зміну в моделі гри суперника, тренер може швидко адаптувати стратегію своєї команди, щоб протистояти цій зміні.

Окрім внутрішньоігрової тактики, ML також допомагає у довгостроковому стратегічному плануванні. Це дає змогу тренерам оцінювати ефективність певних тактик протягом сезону, керуючи майбутніми

тренуваннями та тактичними змінами.

Однак впровадження ML у процес прийняття тактичних рішень не позбавлене проблем. Якість аналізу значною мірою залежить від якості та обсягу даних, що вводяться в алгоритми. Існує також ризик надмірної залежності від даних, що підкреслює потребу у збалансованому підході, який поєднує в собі розуміння даних із досвідом тренера та розумінням динаміки команди.

На закінчення можна сказати, що інтеграція машинного навчання в процес прийняття тактичних рішень у спорті є значним прогресом. Він дає командам детальне розуміння власних стилів гри та стилів гри суперників, що веде до більш обґрунтованих стратегічних рішень, які можуть змінити хід ігор і навіть сезонів. У міру того як технологія продовжує розвиватися, її вплив на спортивну стратегію поглиблюється, далі впроваджуючи аналітику даних як важливий компонент спортивного менеджменту та стратегії.

Інтеграція носимих технологій у спорт ознаменувала значний прогрес у моніторингу та аналізі продуктивності спортсменів. Ці пристрої, оснащені безліччю датчиків, збирають широкий спектр фізіологічних і біомеханічних даних від спортсменів у режимі реального часу. Ці дані, оброблені за допомогою алгоритмів машинного навчання, дають безцінне уявлення про продуктивність, здоров'я та загальний фізичний стан спортсмена, революціонізуючи підхід до тренувань і відновлення.

Носимі технології у спорті зазвичай включають такі пристрої, як фітнес-трекери, розумні годинники та спеціальні датчики, які можна прикріпити до тіла або спорядження спортсмена. Ці пристрої відстежують різні показники, такі як частота серцевих скорочень, прискорення, пройдена відстань, швидкість і навіть більш складні дані, такі як насичення м'язів киснем і біомеханічні моделі. Природа збору цих даних у режимі реального часу дозволяє отримати негайний зворотний зв'язок і аналіз, що має вирішальне значення під час тренувань і змагань.

Алгоритми машинного навчання відіграють ключову роль в інтерпретації

цього величезного обсягу даних. Вони можуть виявити закономірності та аномалії, які можуть вказувати на рівень втоми спортсмена, ризик травми або області потенційного покращення. Наприклад, раптовий стрибок частоти серцевих скорочень або зміна бігової ходи, виявлені цими датчиками, можуть сигналізувати про перетренованість або початок потенційної травми.

Однією з найбільш значних переваг носимих технологій у спорті є можливість адаптувати програми тренувань для окремих спортсменів. Аналізуючи зібрані дані, тренери та спортивні науковці можуть розробляти персоналізовані режими тренувань, які відповідають унікальним фізіологічним характеристикам і цілям продуктивності кожного спортсмена. Це налаштування допомагає оптимізувати ефективність тренувань, підвищити продуктивність і, найголовніше, мінімізувати ризик травм.

Носимі технології також відіграють вирішальну роль у запобіганні травмам і відновленні. Постійний моніторинг може допомогти вчасно виявити потенційні проблеми зі здоров'ям, дозволяючи швидко втручатися. Процеси відновлення подібним чином інформуються за допомогою даних, отриманих від пристроїв для носіння, що гарантує, що спортсмени не повертаються до тренувань передчасно, таким чином уникаючи ризику повторної травми.

Незважаючи на переваги, використання носимих технологій у спорті також створює проблеми, зокрема в управлінні даними та конфіденційності. Забезпечення безпеки конфіденційних даних про здоров'я та їх етичне використання має першочергове значення. Заглядаючи вперед, можна сказати, що майбутнє носимих технологій у спорті багатообіцяюче, а постійний прогрес, ймовірно, дасть ще глибше розуміння. Цей прогрес вказує на майбутнє, де підходи, що керуються даними, ставатимуть все більш центральними для оптимізації спортивних результатів і захисту здоров'я спортсменів.

У динамічному світі спорту роль уболівальників важко переоцінити. З появою ML здатність взаємодіяти з шанувальниками та розуміти їх досягла нових висот. Окрім самих спортивних змагань, ML тепер відіграє ключову роль у розшифровці поведінки, уподобань і моделей залучення вболівальників,

таким чином революціонізувавши аналіз ринку та стратегії залучення вболівальників.

Застосування ML для залучення шанувальників передусім передбачає аналіз великих наборів даних, щоб отримати розуміння поведінки та вподобань шанувальників. Ці дані можуть надходити з різних джерел, включаючи взаємодію в соціальних мережах, продаж квитків, купівлю товарів і навіть прямі відгуки шанувальників. Обробляючи цю інформацію, алгоритми машинного навчання можуть ідентифікувати закономірності та тенденції, які не відразу видно за допомогою традиційних методів аналізу. Наприклад, алгоритм може аналізувати соціальні медіа, щоб оцінити настрої вболівальників щодо нещодавньої гри або підписання нового гравця, надаючи організаціям зворотній зв'язок у реальному часі щодо реакції вболівальників.

Однією з найважливіших переваг використання ML у спорті є можливість покращити загальний досвід уболівальників. Спортивні організації можуть використовувати інформацію, отриману в результаті аналізу даних, щоб адаптувати свої маркетингові стратегії, персоналізувати взаємодію вболівальників і навіть покращити враження від ігрового дня. Наприклад, розуміючи вподобання своїх шанувальників, команда може пропонувати цільові рекламні акції, персоналізовані товари та вміст, який більше резонує з їхньою аудиторією.

ML також забезпечує більш ефективні маркетингові та рекламні стратегії. Розуміючи демографічні показники та поведінку вболівальників, спортивні організації можуть створювати цільові маркетингові кампанії, які з більшою ймовірністю викликать відгук у їхньої аудиторії. Це може включати персоналізований маркетинг електронною поштою, цільову рекламу в соціальних мережах та рекламні акції, які адаптовані до конкретних сегментів шанувальників.

Соціальні мережі — це золота жила для даних шанувальників, і алгоритми машинного навчання вправно аналізують ці дані, щоб зрозуміти настрої та тенденції шанувальників. Аналізуючи дописи, уподобання,

поширення та коментарі, моделі ML можуть надати уявлення про те, про що говорять уболівальники, їхні емоційні реакції на події та те, як вони взаємодіють зі спортивним контентом в Інтернеті. Ця інформація є безцінною для спортивних організацій, які прагнуть ефективно взаємодіяти з уболівальниками на платформах соціальних мереж.

Однак використання ML для залучення шанувальників і аналізу ринку не позбавлене проблем. Точність аналізу залежить від якості даних і алгоритмів, що використовуються. Існують також етичні міркування, зокрема щодо конфіденційності даних шанувальників. Забезпечення того, щоб дані шанувальників використовувалися відповідально та етично, має вирішальне значення для збереження довіри та позитивних стосунків із фанатами.

Хоча ML і глибоко вплинуло на спортивну аналітику, воно породжує значні проблеми та етичні міркування. Однією з головних проблем є точність прогнозів, яка залежить від якості та кількості даних. Моделі ML настільки хороші, наскільки хороші дані, які в них вводяться. Низька якість або недостатня кількість даних може призвести до неточних прогнозів, потенційно вплинувши на стратегічні рішення в іграх і керування гравцями. Крім того, існує ризик зміщення даних. Зібрані дані можуть містити упередження залежно від того, як і звідки вони збираються. Якщо їх не перевірити, ці упередження можуть спотворити результати моделей машинного навчання, що призведе до помилкових висновків і рішень.

Ще одним серйозним викликом є забезпечення конфіденційності та безпеки конфіденційних даних, особливо щодо спортсменів. Зі збільшенням використання носимих технологій і відстеження біометричних даних створюється величезна кількість даних про особисте здоров'я. Етичне поводження з цими даними має першочергове значення, оскільки воно стосується прав особи на конфіденційність. Спортивні організації та аналітичні команди повинні обережно орієнтуватися в цих етичних водах, забезпечуючи дотримання законів про захист даних і поважаючи межі особистої конфіденційності.

Крім того, етичне використання ML у спорті також поширюється на міркування чесної гри. Технологія не повинна створювати нерівні ігрові умови або використовуватися таким чином, що підриває дух конкуренції. Це включає забезпечення того, щоб усі команди мали рівноправний доступ до розширених інструментів аналітики та щоб дані не використовувалися для отримання несправедливої переваги над суперниками.

У спорті майбутнє машинного навчання обіцяє ще більше новаторських програм і ідей. З розвитком технологій ми можемо очікувати появи більш досконалих моделей машинного навчання, здатних до глибшого та деталізованого аналізу. Ймовірно, ці вдосконалення стосуватимуться різних аспектів спорту, від пошуку талантів до внутрішньоігрової стратегії, підвищення ефективності гравців, залучення вболівальників і маркетингових стратегій.

У пошуках талантів ML може змінити спосіб виявлення та розвитку молодих спортсменів. Аналізуючи дані про продуктивність на різних рівнях, алгоритми ML можуть допомогти скаутам точніше та ефективніше визначити перспективні таланти. Це включає не лише фізичні якості та навички, але й потенційні траєкторії розвитку, що дозволяє командам розумно інвестувати в майбутніх зірок.

Підвищення продуктивності гравців — ще одна область, яка дозріла для інновацій. Майбутні програми ML можуть включати більш просунуті прогнозні моделі для запобігання травмам, персоналізовані програми навчання та харчування, а також оптимізацію продуктивності в реальному часі. Наприклад, ML міг би надавати внутрішньоігрові рекомендації тренерам, пропонуючи заміни гравців або тактичні зміни на основі аналізу даних у реальному часі.

Залучення шанувальників і маркетинг також отримають значну користь від ML. Майбутні програми можуть запропонувати ще більш персоналізований досвід шанувальників, використовуючи дані для адаптації контенту, товарів і взаємодії відповідно до індивідуальних уподобань. У маркетингу ML може призвести до більш складної сегментації та стратегій націлювання, підвищення

ефективності рекламних кампаній і підвищення лояльності та залученості шанувальників.

Крім того, інтеграція ML у спортивну аналітику, ймовірно, стимулюватиме нові види збору та аналізу даних. Це може включати більш складний біометричний моніторинг, розширене відстеження рухів і навіть аналіз психологічних і емоційних даних, що забезпечує цілісне уявлення про продуктивність і стан спортсмена.

Підсумовуючи, зазначимо, що майбутнє машинного навчання у спорті – це не лише складніші алгоритми, а й цілісна інтеграція технологій у структуру спорту. Ця інтеграція обіцяє покращити кожен аспект спортивного досвіду, від того, як спортсмени тренуються та змагаються, до того, як уболівальники взаємодіють із улюбленими видами спорту та насолоджуються ними.

1.3. Аналіз соціальних мереж як інструмент прогнозування в спорті

У цифрову еру платформи соціальних мереж стали безцінними ресурсами для вимірювання громадської думки та тенденцій, у тому числі у сфері спорту. Використання даних соціальних мереж для оцінки популярності спортсменів і команд стало важливим аспектом спортивної аналітики, пропонуючи інформацію, що виходить за рамки традиційних показників.

Платформи соціальних медіа забезпечують прямий зв'язок із настроями та вподобаннями шанувальників. Аналізуючи вподобання, поширення, коментарі та згадки, спортивні організації можуть оцінити популярність і рівень залучення вболівальників своїх спортсменів і команд. Ці дані пропонують динамічне розуміння в реальному часі того, як уболівальники взаємодіють зі спортивними організаціями. Наприклад, збільшення активності в соціальних мережах після видовищної гри або під час великої спортивної події може вказувати на підвищений інтерес і залученість.

Популярність спортсменів і команд у соціальних мережах часто відображається на кількості їхніх підписників, рівні взаємодії з їхніми публікаціями та вірусності їхнього контенту. Такі відомі спортсмени, як

Кріштіану Роналду чи Леброн Джеймс, мають мільйони підписників у всьому світі, що робить їх не лише іконами спорту, а й впливовими особистостями в соціальних мережах. Ці показники мають вирішальне значення для спортивних маркетологів і спонсорів, оскільки вони вказують на товарність і цінність бренду спортсменів і команд.

Розширені інструменти аналітики та алгоритми машинного навчання можуть просіювати величезну кількість даних соціальних мереж, щоб визначити тенденції та закономірності. Ці інструменти можуть виявляти стрибки популярності чи залученості, які можуть бути пов'язані з певними подіями, такими як перемога в чемпіонаті, підписання нового гравця або випуск нової форми. Команди та бренди можуть використовувати цю інформацію, щоб узгодити свої маркетингові стратегії з цими тенденціями, максимізуючи залучення шанувальників і комерційні можливості.

Крім кількісних показників, аналіз настроїв дає якісну інформацію про те, як вболівальники ставляться до певних спортсменів або команд. Аналізуючи тон і контекст публікацій і коментарів у соціальних мережах, алгоритми можуть визначити, чи є настрій позитивним, негативним чи нейтральним. Цей аналіз допомагає спортивним організаціям зрозуміти сприйняття громадськості та відповідним чином реагувати, чи то використання позитивних настроїв, чи вирішення проблемних питань.

Популярність соціальних медіа безпосередньо впливає на угоди про спонсорство та підтримку. Спортсмени та команди з високою активністю в соціальних мережах і позитивними настроями часто залучають вигідні угоди, оскільки бренди прагнуть асоціюватися з популярними та улюбленими спортивними діячами. Наприклад, спортсмен, який постійно спілкується з уболівальниками в соціальних мережах і має позитивний суспільний імідж, стає привабливим амбасадором для брендів.

Аналітика соціальних медіа також дозволяє зрозуміти демографічні показники шанувальників, такі як вік, стать, місцезнаходження та інтереси. Ця інформація є безцінною для адаптації контенту, товарів і маркетингових

кампаній до певних сегментів аудиторії, тим самим покращуючи досвід і лояльність шанувальників.

Однак використання даних із соціальних мереж для оцінки популярності не позбавлене проблем. Достовірність показників соціальних мереж (таких як підписники та лайки) іноді може бути сумнівною через наявність ботів і підроблених облікових записів. Крім того, величезний обсяг даних вимагає складних аналітичних інструментів і досвіду для отримання значущої інформації.

У сфері спорту платформи соціальних мереж стали динамічними аренами для вболівальників, де вони можуть висловлювати свою думку, святкувати перемоги та співчувати поразкам. Аналіз настроїв і реакції вболівальників на спортивні події в соціальних мережах став невід'ємною частиною розуміння ширшого досвіду та залученості вболівальників.

Аналіз настроїв у контексті спорту передбачає обробку та інтерпретацію даних із соціальних мереж, щоб оцінити емоції та думки вболівальників. Це може варіюватися від радості та хвилювання до розчарування та розчарування, залежно від результату гри чи конкретного випадку під час спортивної події. Використовуючи обробку природної мови (NLP) і алгоритми машинного навчання, аналітики можуть класифікувати публікації в соціальних мережах як позитивні, негативні або нейтральні, надаючи оцінку настрою, яка відображає загальний настрій фан-бази.

Результати ігор часто безпосередньо впливають на настрої вболівальників у соціальних мережах. Наприклад, драматична перемога у вирішальному матчі може викликати виплиск позитивних емоцій, а суперечливе рішення арбітра чи карколомна поразка – негативні настрої. Цей миттєвий зворотний зв'язок надає спортивним організаціям і командам інформацію про емоційний стан їхніх уболівальників, дозволяючи їм відповідним чином адаптувати свої стратегії спілкування та залучення вболівальників.

Природа соціальних мереж у режимі реального часу дозволяє спортивним організаціям відстежувати реакцію вболівальників у міру розвитку подій. Під

час гри вболівальники часто звертаються до соціальних мереж, щоб прокоментувати ігровий процес, відзначити ключові моменти або критикувати рішення. Відстеження цих реакцій у режимі реального часу дає цінну інформацію про те, які аспекти гри резонують з аудиторією, що може мати вирішальне значення для мовників, коментаторів і маркетингових команд.

Розуміння настроїв шанувальників також має вирішальне значення для формування стратегій маркетингу та зв'язків з громадськістю. Позитивну реакцію вболівальників можна використати для рекламних кампаній, тоді як негативні настрої можуть вказувати на сфери, які потребують уваги, будь то продуктивність команди, управлінські рішення чи навіть продаж товарів і квитків. Цей цикл зворотного зв'язку дозволяє спортивним організаціям залишатися гнучкими та реагувати на потреби та вподобання своїх шанувальників.

Аналіз настроїв також допомагає зрозуміти демографічні показники та поведінку шанувальників. Сегментуючи дані про настрої за такими демографічними ознаками, як вік, місцезнаходження та стать, спортивні організації можуть отримати уявлення про те, які сегменти найбільше залучені або який вміст резонує з різними групами. Ця сегментація є безцінною для персоналізації вражень шанувальників і розробки цільових маркетингових кампаній.

У світі спорту, де емоції вирують, аналіз настроїв у соціальних мережах має вирішальне значення для управління кризою та контролю репутації. Тенденції негативних настроїв у соціальних мережах можуть діяти як сигнали раннього попередження про можливі проблеми чи кризи, дозволяючи спортивним організаціям вживати проактивних заходів для вирішення проблем, керувати своєю репутацією та підтримувати позитивні стосунки зі своїми вболівальниками.

Незважаючи на свої переваги, аналіз настроїв соціальних мереж у спорті стикається з проблемами. Точність визначення настроїв може змінюватися залежно від складності мови, сарказму та культурних нюансів у публікаціях

шанувальників. Крім того, величезний обсяг даних потребує складних інструментів і досвіду для ефективного синтаксичного аналізу.

На завершення аналіз настроїв уболівальників і реакцій на спортивні події в соціальних мережах дає величезну кількість уявлень про емоційний ландшафт спортивного світу. Це дає змогу спортивним організаціям глибше зв'язуватися зі своїми фан-базами, адаптувати свої стратегії та підтримувати позитивну та привабливу спортивну спільноту. Оскільки соціальні медіа продовжують залишатися основною платформою для висловлювання вболівальників, роль аналізу настроїв у розумінні та реагуванні на динаміку вболівальників лише зростатиме.

Поява вдосконаленої аналітики у спорті відкрила нові можливості для прогнозування результатів спортивних змагань. Одним із інноваційних підходів, який набув популярності, є аналіз тенденцій на платформах соціальних мереж. Оцінюючи настрої, дискусії та прогнози, якими діляться вболівальники та експерти в соціальних мережах, аналітики тепер можуть прогнозувати результати спортивних подій з надзвичайною точністю.

Соціальні медіа-платформи є багатими джерелами громадської думки та настроїв, пропонуючи різноманітні погляди на майбутні спортивні події. Від пристрасних уболівальників і досвідчених спортивних журналістів до випадкових спостерігачів, ці платформи охоплюють широкий спектр поглядів і прогнозів. Алгоритми машинного навчання можуть агрегувати та аналізувати ці точки зору, щоб визначити переважаючі настрої та тенденції, які можуть вказувати на потенційні результати спортивних змагань.

Настрої вболівальників у соціальних мережах можуть слугувати цінним інструментом прогнозування. Ентузіастична підтримка та впевненість у команді, відображені в позитивних публікаціях, публікаціях і коментарях, іноді можуть корелювати з продуктивністю команди. Подібним чином широка критика чи песимізм можуть натякати на потенційну недостатню ефективність. Однак, незважаючи на те, що настрої шанувальників можуть дати розуміння, вони також схильні до упереджень і емоційних коливань, що вимагає ретельної

інтерпретації та перехресних посилань з іншими джерелами даних.

Прогностична аналітика в спорті виходить за рамки простого аналізу настроїв, залучаючи кореляцію тенденцій соціальних медіа з історичними даними та іншими показниками ефективності. Аналітики поєднують інформацію про соціальні мережі зі статистикою гравців, формою команди, минулими показниками та навіть погодними умовами, щоб розробити комплексні прогностичні моделі. Наприклад, зростання активності в соціальних мережах навколо певного гравця або стратегії можна порівняти з останніми показниками продуктивності цього гравця, щоб оцінити його потенційний вплив на майбутню гру.

Крім настроїв уболівальників, промовистими є також тенденції в соціальних мережах, керовані інфлюенсерами та спортивними експертами. Схвалення або критика видатних діячів може вплинути на громадську думку, а в деяких випадках навіть може вплинути на моральний стан і сприйняття команд і гравців громадськістю. Аналіз цих тенденцій потребує тонкого розуміння динаміки впливу в спортивній спільноті.

Прогнозувати спортивні результати на основі аналізу соціальних медіа не без труднощів. Платформи соціальних медіа іноді можуть поширювати nereпрезентативні думки або створювати ехокамери, які можуть неточно відображати ширші суспільні настрої. Крім того, спонтанний і швидко мінливий характер дискусій у соціальних мережах вимагає аналізу та реагування в режимі реального часу, що робить це складним і ресурсомістким завданням.

Існують також етичні міркування щодо використання соціальних медіа для прогностичної аналітики у спорті. Забезпечення конфіденційності даних і навігація щодо законності використання публічних публікацій і обговорень є надзвичайно важливими. Крім того, існує відповідальність за те, щоб прогнози та аналізи не сприяли отриманню несправедливих переваг або впливали на цілісність спорту.

Підсумовуючи, використання аналізу тенденцій соціальних медіа для

прогнозування результатів спортивних змагань є значним нововведенням у сфері спортивної аналітики. Хоча він пропонує захоплюючі можливості, він вимагає обережного та етичного застосування. Оскільки соціальні медіа продовжують залишатися невід'ємною частиною глобальної спортивної дискусії, їх роль у формуванні та прогнозуванні динаміки спортивних змагань, ймовірно, стане більш виразною та витонченою.

1.4. Мета та сфера застосування машинного навчання в спортивній аналітиці: постановка завдання для подальшого розвитку

Пропонується використовувати алгоритми машинного навчання та аналітику соціальних мереж для підвищення ефективності спортивної аналітики, таким чином оптимізуючи процеси прийняття рішень у різних аспектах управління спортом.

Пропонований аналітичний модуль має на меті:

- поєднати основні дані про продуктивність спортсмена з тенденціями в соціальних мережах, щоб створити зручний інструмент аналітики;
- розробити спрощену модель машинного навчання, яка може обробляти й аналізувати ключові показники ефективності та настрої спортсменів;
- створити інтерфейс, який дозволить спортивним користувачам легко вводити дані та інтерпретувати результати аналізу машинного навчання.

ВИСНОВКИ ДО РОЗДІЛУ 1

Оглядаючи поточний статус спортивної аналітики, стає очевидним, що інтеграція машинного навчання з акцентом на зручність і практичність є ключем до просування спортивних стратегій і рішень. Основними критеріями успіху цього проекту є простота інструменту, точність його прогнозів і відповідність потребам спортивних професіоналів. Цей підхід має на меті зробити розширений аналіз даних доступним і корисним для стратегічного планування в спорті, не перевантажуючи користувачів складністю.

РОЗДІЛ 2. МЕТОДИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТЕНІСНИХ МАТЧІВ

2.1. Аналіз статистичних даних тенісних матчів та спортсменів

Цей підрозділ присвячений глибокому аналізу таблиці **matches**, яка є центральною для нашого дослідження в області прогнозування результатів тенісних матчів.

Перш за все хочемо сказати що більша частина даних була синтетична, тобто створена штучно, так як не всі потрібні дані є в доступності або в потрібній кількості.



```
df.head()
```

executed in 28ms, finished 01:04:25 2023-12-15

	code	event_id	league_name	starts_date	starts_time	player_1	player_2	coef_1	coef_2	post_1	post_2	answer_1	answer_2	is_p1_win
0	1525	1583297268	ITF Men Wellington - R1	11.12.2023	23:20	Gillian Osmont	Sam Thompson	3.600701	1.228367	#Acelt Anxious	#TennisLife	-0.2500	0.0000	0
1	1526	1583303516	ITF Women Wellington - R1	11.12.2023	23:30	Sahra Dennehy	Haine Ogata	2.863892	1.374539	#GameTime Anxious opponent	#Servelt Thrilled	-0.2500	0.4404	0
2	1527	1583287841	ITF Men Wellington - R1	12.12.2023	00:00	Anton Shepp	Jacob Bradshaw	2.684339	1.632667	#BringItOn	Pumped #PlayHard	0.0000	0.0000	0
3	1528	1583287843	ITF Men Wellington - R1	12.12.2023	00:00	Corban Crowther	Diego Quispe-Kim	1.544386	3.331891	#MatchDay next competition Ready	Anxious upcoming event #WinningMood	0.3612	-0.2500	1
4	1529	1583287849	ITF Men Wellington - R1	12.12.2023	00:00	James Watt	Reece Falck	1.640145	2.750057	Ready next game	tomorrow's match #Acelt Optimistic	0.3612	0.3182	1

Рис. 2.1. Таблиця matches з основними даними

Таблиця **matches** (Рис. 2.1) складається з наступних ключових полів:

- **code**: унікальний ідентифікатор запису, який забезпечує легкий доступ до кожного рядка даних. Використовується для внутрішньої організації бази даних.
- **event_id**: унікальний код події, який вказує на конкретний тенісний матч. Це дозволяє точно ідентифікувати кожен матч у базі даних.
- **league_name**: назва ліги або турніру. Важливий параметр для класифікації матчів по рівнях і значимості змагань.
- **starts_date** та **starts_time**: визначають дату та час початку матчу. Ці дані критично важливі для аналізу сезонності виступів гравців та їхньої ефективності в різний час. А для нас ці дані ще важливіші, так як на їх основі

ми знаходимо твіти спортсменів.

- **player_1** і **player_2**: імена учасників матчу. Забезпечують розуміння про суперництво між конкретними гравцями та їхню історію зіткнень. За відсутності інших майбутніх даних можуть бути ключовими показниками в аналізі.
- **coef_1** і **coef_2**: коефіцієнти ставок, які відображають очікування на результат матчу від ставочних компаній. Це дає уявлення про фаворитів та аутсайдерів у кожному матчі.
- **post_1** і **post_2**: тексти повідомлень із соціальних мереж від обох гравців. Ці дані використовуються для сентиментального аналізу психологічного стану гравців.
- **answer_1** і **answer_2**: результати сентиментального аналізу настрою (sentiment analysis) повідомлень. Числові значення, які відображають позитивні або негативні емоції у повідомленнях.
- **is_p1_win**: показник, що відображає результат матчу для першого гравця (1, якщо виграв; 0, якщо програв). Це основний цільовий параметр для моделей прогнозування.

Методологія аналізу:

1. Вивчення параметрів гри: аналізуються ключові статистичні дані, такі як дата, час, учасники матчу, щоб визначити можливі закономірності у виступах гравців.
2. Оцінка коефіцієнтів: аналізуємо вплив коефіцієнтів ставок на результати матчів, що допомагає зрозуміти вплив громадського сприйняття на ісход змагань.
3. Аналіз соціальних мереж: вивчення повідомлень із соціальних мереж та їхнього аналізу настрою для оцінки впливу психологічного стану спортсменів на їхні виступи.
4. Статистичний аналіз: використовуємо статистичні методи для визначення кореляцій між різними параметрами та результатами матчів.

На наступних графіках (Рис. 2.2) ми можемо бачити що:

- **coef_1** і **coef_2** (коефіцієнти ставок) - розподіл коефіцієнтів ставок **coef_1** і **coef_2** підтверджує, що низькі значення коефіцієнтів (ближче до 1) є більш поширеними для **coef_2**, що може вказувати на вищу ймовірність перемоги другого гравця у більшості матчів у цій вибірці. Водночас, **coef_1** має ширший розкид, із значеннями, що сягають вищих рівнів, що може свідчити про меншу впевненість у перемозі першого гравця.

```
df.hist(figsize=(12, 10))
plt.show()
```

executed in 958ms, finished 01:04:28 2023-12-15

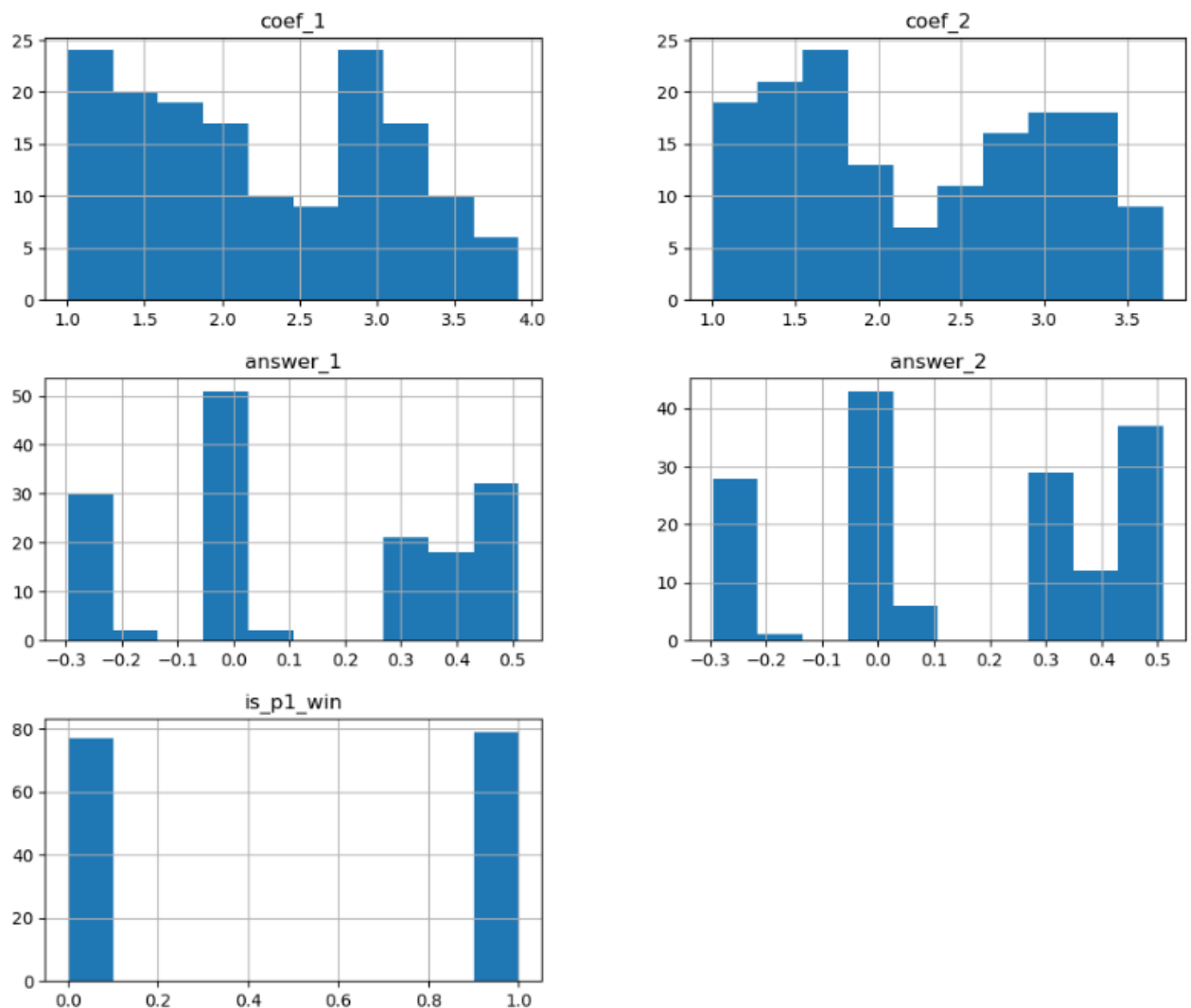


Рис. 2.2. Графіки числових показників

- **answer_1** і **answer_2** (аналіз настрою) - гістограми **answer_1** і **answer_2** показують розподіл аналізу настроїв повідомлень гравців. Можемо спостерігати, що більшість постів мають нейтральний тон, але також є виразні піки для позитивних та негативних значень. Це свідчить про варіативність емоційного забарвлення в повідомленнях, які можуть відігравати роль у

прогнозуванні результатів матчів.

- **is_p1_win** (результат матчу для першого гравця) - з графіку **is_p1_win** видно, що перемоги та поразки першого гравця розподілені майже порівну. Це вказує на збалансованість даних і є корисним для навчання прогностичних моделей, оскільки вони не будуть схильні до переваги одного класу результатів.

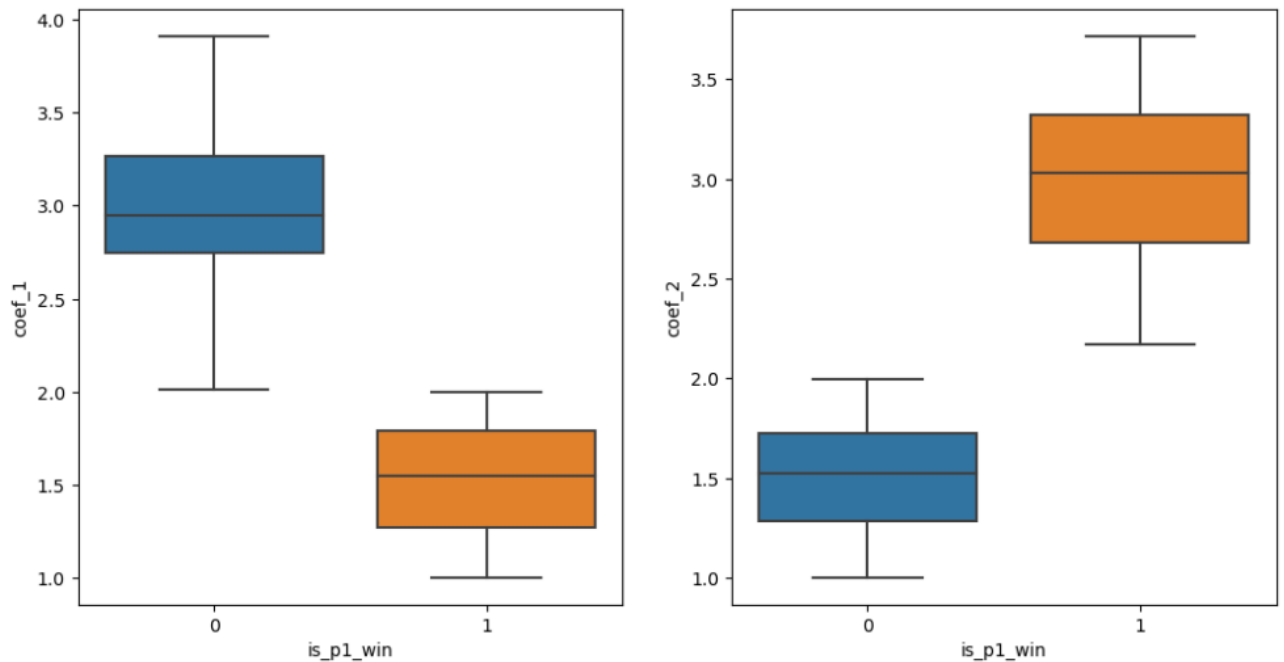


Рис. 2.3. Графіки ключових показників

На графіках з «вусами» (Рис. 2.3) показано розподіл коефіцієнтів ставок (**coef_1** і **coef_2**) у залежності від того, чи виграв перший гравець матч (**is_p1_win**):

1. **coef_1** (коефіцієнти ставок на першого гравця):

- Боксплот для **coef_1** показує, що медіанні значення коефіцієнтів нижчі, коли перший гравець виграє (**is_p1_win = 1**). Це свідчить про те, що нижчі коефіцієнти відповідають більшій імовірності перемоги гравця;
- Широкі «вуса» та розмах боксплоту вказують на значний розкид у коефіцієнтах, особливо коли гравець не виграє (**is_p1_win = 0**), що може означати велику невизначеність у ставках.

2. **coef_2** (коефіцієнти ставок на другого гравця):

- У випадку **coef_2**, медіанні значення вищі для матчів, які перший гравець програв (**is_p1_win = 0**), що відповідає традиційному

розумінню коефіцієнтів ставок - вищий коефіцієнт відповідає меншій імовірності перемоги.

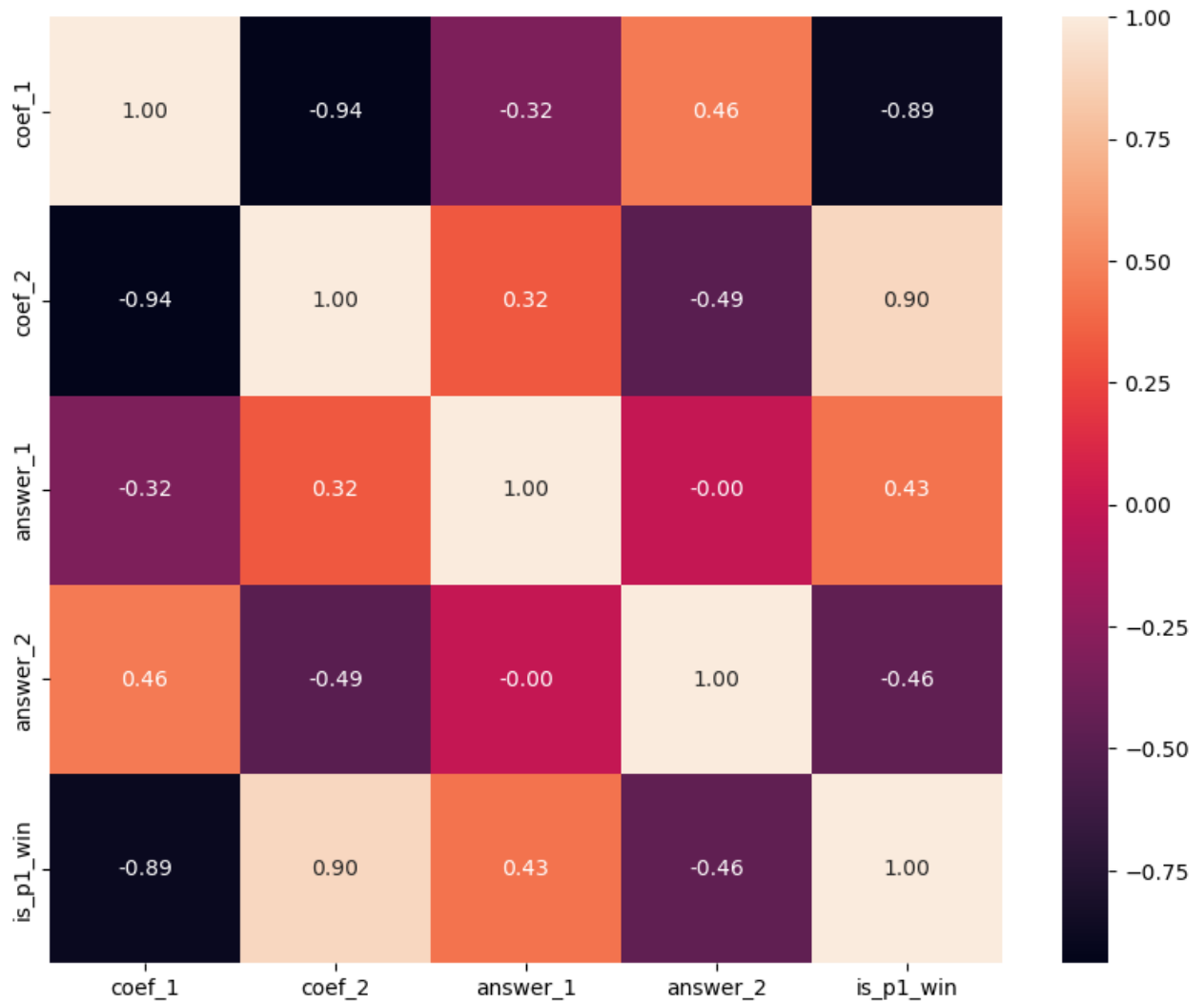


Рис. 2.4. Теплова карта кореляцій

Завдяки тепловій карті ми можемо бачити наступну кореляцію між різними змінними в даних:

- Кореляція близька до -1 між **coef_1** та **is_p1_win** підтверджує, що нижчі коефіцієнти часто відповідають вищій імовірності перемоги першого гравця;
- Позитивна кореляція близько 0.9 між **coef_2** та **is_p1_win** вказує на те, що коли коефіцієнти на другого гравця вищі, перший гравець частіше виграв;
- **answer_1** та **answer_2** мають маленьку або жодної кореляції з імовірністю перемоги, що може свідчити про обмежений вплив аналізу настроїв соціальних мереж на результат матчу;

- Значення кореляції між **coef_1** та **coef_2** дуже близьке до -1, що логічно, оскільки як правило, коли коефіцієнт одного гравця знижується, коефіцієнт його суперника зростає.

2.2. Розробка прогнозної моделі

Для завдання прогнозування результатів тенісного матчу ми обрали TensorFlow, потужну бібліотеку програмного забезпечення з відкритим кодом для машинного навчання, як наш основний інструмент з кількох вагомих причин.

По-перше, TensorFlow пропонує широку екосистему, яка включає інструменти, бібліотеки та ресурси спільноти, які дозволяють дослідникам розвивати ML, а розробникам створювати та розгортати програми на основі ML. TensorFlow забезпечує гнучкість, щоб почати з чистого аркуша, щоб створювати просунуті моделі, а також пропонує Keras API, який ідеально підходить для швидкого початку роботи та може працювати поверх TensorFlow.

Однією з видатних особливостей TensorFlow є його здатність ефективно обробляти алгоритми глибокого навчання. Нейронні мережі, особливо глибоке навчання, продемонстрували чудову продуктивність у складних завданнях прогнозування, які включають багатовимірні дані та нелінійні зв'язки. Враховуючи багатогранність результатів тенісних матчів, на які можуть впливати складні закономірності та тенденції в даних, розширені алгоритми оптимізації TensorFlow роблять його чудовим вибором для фіксації цих складнощів.

Здатність TensorFlow автоматично обчислювати похідні економить значну кількість часу та зменшує ймовірність помилок у зворотному поширенні, що є основою навчання нейронної мережі. Ця функція, відома як автоматична диференціація, дуже корисна під час роботи зі складними моделями.

Крім того, TensorFlow розроблено з урахуванням масштабованості. Він може обробляти великі набори даних і розгортати обчислення на одному або

кількох центральних або графічних процесорах на настільному комп'ютері, сервері чи мобільному пристрої за допомогою єдиного API, що робить його дуже універсальною платформою для розробки та розгортання моделей.

Іншим фактором є активна спільнота та постійна підтримка розробників і учасників TensorFlow. Ця розгалужена мережа підтримки допомагає у вирішенні проблем, обміні найкращими практиками та підтримці бібліотеки в актуальному стані завдяки останнім досягненням машинного навчання.

Конфігурація параметрів і гіперпараметрів моделі в TensorFlow може бути точно налаштована, що дозволяє точно контролювати процес навчання. Це включає встановлення кількості шарів, нейронів, швидкості навчання та функцій активації, які можна оптимізувати за допомогою таких методів, як пошук у сітці або випадковий пошук, щоб досягти найкращої продуктивності моделі.

Нарешті, TensorFlow забезпечує комплексну візуалізацію через TensorBoard, яка допомагає контролювати процес навчання, розуміти структуру моделі та, якщо необхідно, налагодити. Такий рівень прозорості є важливим для розробки моделі, особливо під час пояснення процесу прийняття рішень щодо моделі зацікавленим сторонам, які можуть не мати глибокого технічного досвіду.

Таким чином, універсальність TensorFlow, розширені можливості обробки алгоритмів глибокого навчання, простота використання, масштабованість, потужна підтримка спільноти та комплексні інструменти візуалізації роблять його оптимальним вибором для розробки прогнозної моделі для результатів тенісного матчу.

На рисунку (Рис. 2.5) показано початок роботи з моделлю. Спочатку ми визначаємо кількість унікальних гравців і ліг у своєму наборі даних. Це важливий крок, оскільки він допомагає зрозуміти широту категоріальних даних, які впливатимуть на дизайн шарів вбудовування вашої моделі. Унікальні підрахунки дозволяють розподілити вектор вбудованого розміру фіксованого розміру для кожної унікальної категорії.

```
# Розміри словників для Embeddings
num_players = df_combined['player_1'].nunique()
num_leagues = df_combined['league_name'].nunique()
executed in 5ms, finished 12:36:00 2023-12-15
```

```
# Розмір Embedding
embedding_size = 4
executed in 4ms, finished 12:36:00 2023-12-15
```

```
# Input Layers
input_player_1 = Input(shape=(1,))
input_player_2 = Input(shape=(1,))
input_league = Input(shape=(1,))
input_numerical = Input(shape=(4,)) # Для coef_1, coef_2, answer_1, answer_2
executed in 188ms, finished 12:36:01 2023-12-15
```

Рис. 2.5. Початок роботи з Tensorflow

Розмір Embedding встановлено на 4, що означає, що кожен унікальний гравець і ліга будуть представлені 4-вимірним вектором. Цей гіперпараметр вибрано, щоб збалансувати складність моделі з обчислювальною ефективністю та ризиком перенавчання. Embedding — це потужний спосіб охопити суть категоріальних змінних у просторі нижчої вимірності, на відміну від одноразового кодування, яке може призвести до розрідженості та високої розмірності.

Визначаючи окремі входні рівні для різних типів об'єктів, ми структуруємо свою модель для індивідуальної обробки різнорідних джерел даних. Цей підхід дозволяє нейронній мережі вивчати базові закономірності в кожному типі даних перед тим, як об'єднувати їх для подальшої обробки або прогнозування.

Далі (Рис. 2.6) ми бачимо роботу з шарами та створення моделі:

- Шари **Embedding** використовуються для перетворення ідентифікаторів гравців і ліг у щільні вектори фіксованого розміру (**embedding_size**).
- **input_dim** встановлюється на кількість унікальних гравців або ліг, що визначає розмір словника, а **output_dim** — це розмір щільного вектора. Ці щільні вектори можуть фіксувати більше інформації та взаємозв'язків про гравців і ліги, ніж одноразово закодовані вектори.
- Після створення **Embedding** шари **Flatten** переформовують вихідні дані

вбудовування до єдиного виміру, необхідного для об'єднання цих векторів пізніше з числовими вхідними даними. Ця операція об'єднує зведені вектори **Embedding** з числовими вхідними даними. Об'єднання цих вхідних даних дозволяє моделі паралельно вивчати як категоричні **Embedding**, так і числові дані.

```
# Embeddings
player_1_embedding = Embedding(input_dim=num_players, output_dim=embedding_size, input_length=1)(input_player_1)
player_2_embedding = Embedding(input_dim=num_players, output_dim=embedding_size, input_length=1)(input_player_2)
league_embedding = Embedding(input_dim=num_leagues, output_dim=embedding_size, input_length=1)(input_league)
executed in 85ms, finished 12:36:01 2023-12-15

player_1_embedding = Flatten()(player_1_embedding)
player_2_embedding = Flatten()(player_2_embedding)
league_embedding = Flatten()(league_embedding)
executed in 19ms, finished 12:36:01 2023-12-15

# Конкатенація Embeddings з числовими даними
concatenated = concatenate([player_1_embedding, player_2_embedding, league_embedding, input_numerical])
executed in 16ms, finished 12:36:01 2023-12-15

# Denses
dense_1 = Dense(64, activation='relu')(concatenated)
dense_2 = Dense(32, activation='relu')(dense_1)
output = Dense(1, activation='sigmoid')(dense_2)
executed in 58ms, finished 12:36:01 2023-12-15

# Створення та компіляція моделі
model = Model(inputs=[input_player_1, input_player_2, input_league, input_numerical], outputs=output)
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
executed in 132ms, finished 12:36:01 2023-12-15
```

Рис. 2.6. Шари та створення моделі

- **Dense** шари — це повністю зв'язані шари, які вивчають нелінійні комбінації ознак, витягнутих із даних. Вибір **64** і **32** нейронів у перших двох **Dense** шарах є дещо довільним, але зазвичай використовуються розміри, які забезпечують достатню складність для вивчення шаблонів, не будучи надто дорогим з точки зору обчислень.
- **relu** обрано як функцію активації через її ефективність і ефективність у нелінійних перетвореннях.
- Кінцевий вихідний рівень використовує функцію активації **сигмоїд**, оскільки ця модель призначена для завдання двійкової класифікації (виграти чи програти).
- Оптимізатор **adam** є популярним вибором, оскільки він автоматично регулює швидкість навчання під час навчання та, як правило, стійкий до вибору гіперпараметрів.

- Функція втрат **binary_crossentropy** підходить для задач бінарної класифікації, а **accuracy** вибирається як метрика для моніторингу ефективності моделі.

Цей вибір параметрів і архітектури моделі керується звичайними методами машинного навчання та конкретними вимогами проблеми, що розглядається. Мета полягає в тому, щоб побудувати модель, яка може ефективно обробляти категоріальні та числові дані та вивчати їх, щоб робити точні прогнози щодо результатів тенісного матчу.

# Виведення структури моделі			
model.summary()			
executed in 54ms, finished 12:36:01 2023-12-15			
Model: "model"			
Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 1)]	0	[]
input_2 (InputLayer)	[(None, 1)]	0	[]
input_3 (InputLayer)	[(None, 1)]	0	[]
embedding (Embedding)	(None, 1, 4)	744	['input_1[0][0]']
embedding_1 (Embedding)	(None, 1, 4)	744	['input_2[0][0]']
embedding_2 (Embedding)	(None, 1, 4)	56	['input_3[0][0]']
flatten (Flatten)	(None, 4)	0	['embedding[0][0]']
flatten_1 (Flatten)	(None, 4)	0	['embedding_1[0][0]']
flatten_2 (Flatten)	(None, 4)	0	['embedding_2[0][0]']
input_4 (InputLayer)	[(None, 4)]	0	[]
concatenate (Concatenate)	(None, 16)	0	['flatten[0][0]', 'flatten_1[0][0]', 'flatten_2[0][0]', 'input_4[0][0]']
dense (Dense)	(None, 64)	1088	['concatenate[0][0]']
dense_1 (Dense)	(None, 32)	2080	['dense[0][0]']
dense_2 (Dense)	(None, 1)	33	['dense_1[0][0]']
=====			
Total params: 4745 (18.54 KB)			
Trainable params: 4745 (18.54 KB)			
Non-trainable params: 0 (0.00 Byte)			

Рис. 2.7. Структура моделі

```

train_inputs = [X_train['player_1'], X_train['player_2'],
                X_train['league_name'],
                X_train[['coef_1', 'coef_2', 'answer_1', 'answer_2']]]
test_inputs = [X_test['player_1'], X_test['player_2'],
               X_test['league_name'],
               X_test[['coef_1', 'coef_2', 'answer_1', 'answer_2']]]

history = model.fit(train_inputs, y_train, epochs=100, validation_split=0.2)

executed in 8.05s, finished 12:36:09 2023-12-15
7/7 [=====] - 0s 9ms/step - loss: 0.2601 - accuracy: 0.9648 - val_loss: 0.4595 - val_accuracy: 0.8000
Epoch 21/100
7/7 [=====] - 0s 9ms/step - loss: 0.2100 - accuracy: 0.9698 - val_loss: 0.4498 - val_accuracy: 0.8400
Epoch 22/100
7/7 [=====] - 0s 9ms/step - loss: 0.1612 - accuracy: 0.9799 - val_loss: 0.4398 - val_accuracy: 0.8200
Epoch 23/100
7/7 [=====] - 0s 8ms/step - loss: 0.1282 - accuracy: 0.9899 - val_loss: 0.4422 - val_accuracy: 0.8000
Epoch 24/100
7/7 [=====] - 0s 11ms/step - loss: 0.0966 - accuracy: 1.0000 - val_loss: 0.4431 - val_accuracy: 0.8200
Epoch 25/100
7/7 [=====] - 0s 11ms/step - loss: 0.0728 - accuracy: 1.0000 - val_loss: 0.4495 - val_accuracy: 0.8200
Epoch 26/100
7/7 [=====] - 0s 11ms/step - loss: 0.0556 - accuracy: 1.0000 - val_loss: 0.4587 - val_accuracy: 0.8200

test_loss, test_accuracy = model.evaluate(test_inputs, y_test)
print(f"Точність на тестових даних: {test_accuracy}")

executed in 114ms, finished 12:36:09 2023-12-15
2/2 [=====] - 0s 6ms/step - loss: 0.4178 - accuracy: 0.8730
Точність на тестових даних: 0.8730158805847168

```

Рис. 2.8. Показники моделі

На Рис.2.7 ми бачимо структуру нашої моделі. Далі (Рис. 2.8) ми можемо бачити показники моделі. Було підібрано 100 епох, так як стандартних 10 було мало і я думаю зі збільшенням даних у майбутньому мало буде і 100:

❖ Метрики навчання та перевірки:

- точність моделі на даних навчання починається з високого рівня та досягає 1 (або 100%) до епохи 24, що вказує на те, що модель правильно прогнозує всі навчальні вибірки з цього моменту.
- точність набору перевірки (підмножини навчальних даних, які використовуються для налаштування гіперпараметрів і запобігання перенавчанню) починається з 0,840 (84%) і незначно коливається, але зазвичай залишається навколо цього значення. Це вказує на те, що модель добре узагальнює, але не покращується суттєво після першої епохи.

❖ Показники втрат:

- втрата даних навчання з часом зменшується, що й очікувалося, оскільки модель вивчає дані.
- втрата перевірки трохи зменшується, але потім починає коливатися.

Примітно, що втрати під час навчання значно нижчі, ніж втрати підтвердження, що може бути показником перенавчання, особливо враховуючи, що точність навчання досягла 100%.

❖ Остаточна оцінка тестових даних:

- Остаточна оцінка даних тесту показує втрату приблизно 0,4178 і точність 0,8730 (87,3%). Ці показники мають вирішальне значення, оскільки вони забезпечують неупереджену оцінку моделі на основі даних, яких вона ніколи раніше не бачила.

Згідно з отриманою інформацією, хоча модель досягає ідеальної точності на навчальному наборі, точність перевірки вказує на те, що може бути розбіжність між характеристиками навчання та перевірки. Це загальна ознака перенавчання, коли модель вивчає закономірності, специфічні для навчальних даних, які не узагальнюються на невидимі дані.

Точність тесту 87,3% є досить високою і свідчить про те, що модель має хорошу прогностичну силу. Однак різниця між точністю навчання (100%) і точністю тесту (87,3%) посилює можливість перенавчання.

Щоб усунути потенційне перенавчання, можна розглянути такі методи, як додавання відсіву, регуляризація або використання різноманітнішого набору навчальних даних. Крім того, можна застосувати подальше налаштування гіперпараметрів або ранню зупинку, щоб запобігти надмірному засвоєнню моделлю навчальних даних.

2.3. Оцінка продуктивності моделі та значущості даних

Головною метою нашого аналізу було зіставити передбачувані результати з урахуванням і без урахування публікацій у соціальних мережах. Ми проаналізували (Рис. 2.9) отримані візуальні представлення даних, щоб розгадати вплив і ефективність включення настроїв соціальних медіа в наші прогнози моделі.

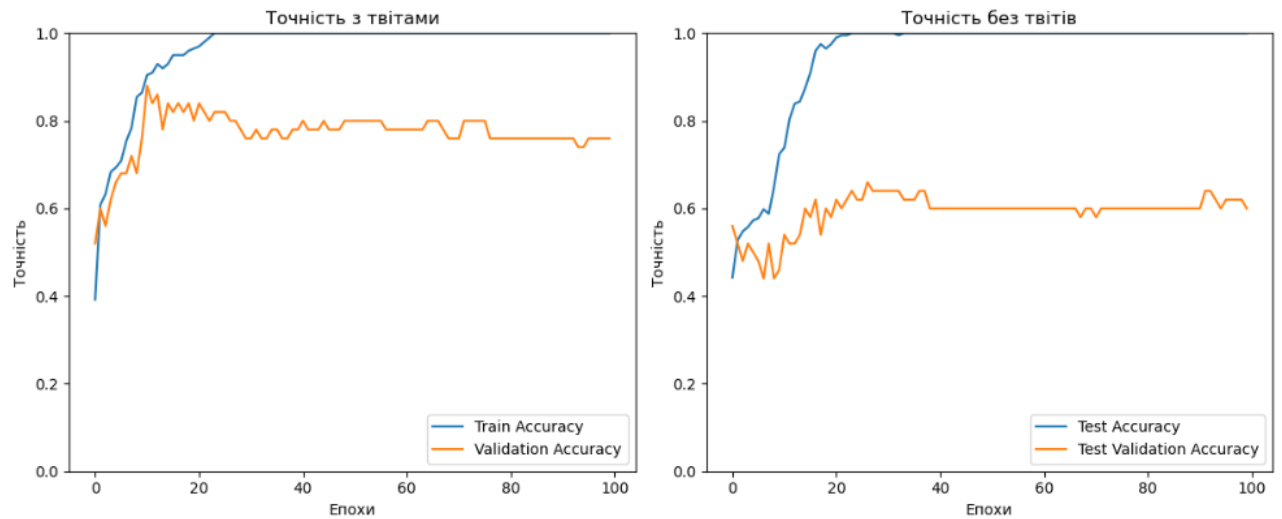


Рис. 2.9. Аналіз точності моделей

Ці два графіки порівнюють точність моделей за 100 епох, один із використанням твітів (ліворуч), а інший виключаючи їх (праворуч). На лівому графіку показано модель, навчену твітам. Спочатку точність як навчання, так і перевірки різко зростає, що свідчить про швидке навчання. Однак із збільшенням епох точність перевірки досягає плато близько 0,8, тоді як точність навчання продовжує дещо покращуватися. Ця розбіжність свідчить про те, що модель починає переповнюватись, шаблони навчання, специфічні для навчальних даних не узагальнюються для набору перевірки. Незважаючи на це, модель підтримує відносно високу точність перевірки, що свідчить про те, що твіти містять цінну інформацію, яка допомагає прогнозувати.

Праворуч модель без твітів демонструє більш хаотичний шаблон. Точність навчання знову зростає швидко, але з більшими коливаннями, а точність перевірки має менш стабільну траєкторію. Після початкового підвищення точність перевірки знижується, а потім знаходиться на плато біля 0,6, що значно нижче, ніж у моделі, яка включає дані твітів. Ця нестабільна поведінка та нижнє плато свідчать про те, що без твітів модель має меншу передбачувану здатність і, можливо, важко охопити основну структуру даних.

Порівнюючи ці два, стає очевидним, що включення твітів веде до вищої загальної точності. Більш плавна крива та вище плато на лівому графіку означають, що модель краще налаштована на нюанси даних, використовуючи додаткову інформацію про настрої в твітах. Модель без твітів, ще навчаючись,

демонструє ознаки недостатнього пристосування, не в змозі захопити якомога більше передбачуваного сигналу.

Ці графіки підкреслюють цінність аналізу настроїв у соціальних мережах у прогностичних моделях. Підвищена точність урахування твітів свідчить про те, що дописи гравців у соціальних мережах мають значну прогностичну цінність і їх слід враховувати при розробці моделей для прогнозування результатів тенісних матчів.

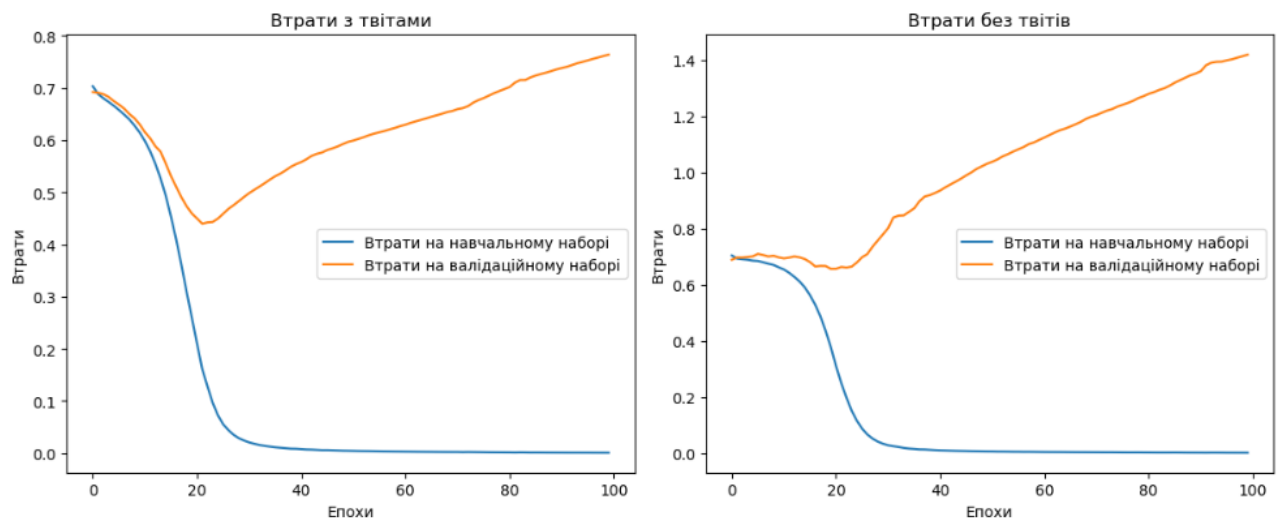


Рис. 2.10. Аналіз втрат моделей

На наступних графіках (Рис. 2.10) представлено порівняння втрат (loss) моделі під час навчання та валідації для двох сценаріїв: з використанням твітів (зліва) та без них (справа). Втрати відображають рівень помилок, які робить модель при прогнозуванні, і є ключовим показником для оптимізації під час тренування.

На лівому графіку втрати на навчальному наборі даних швидко знижуються і стабілізуються на низькому рівні, що свідчить про ефективне навчання моделі. Втрати на валідаційному наборі після певного зниження починають зростати, що може свідчити про початок перенавчання моделі. Тим не менш, загальний рівень втрат залишається порівняно низьким, що підкреслює корисність включення аналізу соціальних медіа у модель.

На противагу, правий графік показує вищий рівень втрат як на навчальному, так і на валідаційному наборах даних. Втрати на навчальному

наборі даних спочатку знижуються, але потім починають зростати, що свідчить про недостатнє навчання та можливу неадекватність моделі для даної задачі без врахування інформації з твітів.

Зіставлення графіків вказує на те, що використання твітів може допомагати у зменшенні втрат та підвищенні загальної точності моделі. Модель із твітами продемонструвала кращу здатність узагальнення та адаптації до нових даних, що підкреслює значення соціальних медіа як важливого джерела інформації для прогнозування результатів тенісних матчів.

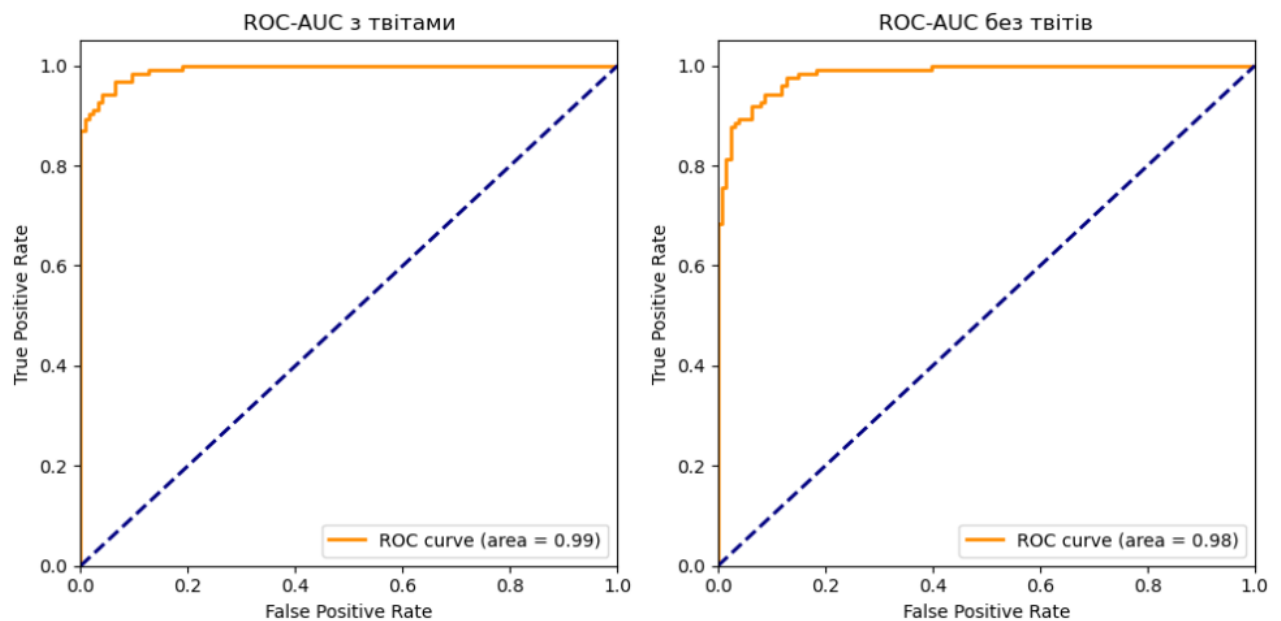


Рис. 2.11. Аналіз AUC-ROC

Далі на графіках AUC-ROC (Рис. 2.11) представлені криві операційних характеристик приймача (ROC) та площі під кривою (AUC) для моделей з використанням твітів (зліва) та без них (справа). Графіки AUC-ROC є мірою здатності моделі класифікувати два класи і є важливим інструментом для оцінки якості бінарних класифікаторів.

Обидва графіки демонструють високі значення AUC, що свідчить про високу здатність моделей розрізняти між класами перемог та поразок. Графік "ROC-AUC з твітами" показує AUC 0.99, що майже ідеально. Це означає, що модель з дуже високою точністю може відрізняти виграші від програшів, використовуючи дані з твітів.

З іншого боку, графік "ROC-AUC без твітів" також показує високу площу

під кривою (AUC 0.98), хоч і трохи нижчу, ніж модель з твітами. Це свідчить про те, що модель без даних з соціальних мереж також ефективно розрізняє результати, хоча можливо з незначно нижчою точністю.

Порівняння двох графіків підтверджує, що інформація з соціальних медіа може додати цінності до моделі прогнозування, покращуючи її здатність до класифікації. Тим не менш, обидві моделі показують високу продуктивність, і різниця в AUC між ними не є значною. Це вказує на те, що хоча твіти можуть бути корисними для збільшення точності прогнозу, самі по собі статистичні дані вже надають сильний прогнозуючий сигнал.

ВИСНОВКИ ДО РОЗДІЛУ 2

Модель машинного навчання, розроблена для прогнозування результатів тенісних матчів, демонструє ефективну інтеграцію статистичних даних і публікацій у соціальних мережах. Доведено, що здатність моделі обробляти й аналізувати різноманітні набори даних, включаючи історичну статистику матчів і настрої в соціальних мережах у реальному часі, значно підвищує точність прогнозування. Використовуючи методи обробки природної мови, модель інтерпретує почуття, виражені в публікаціях спортсменів у соціальних мережах, додаючи рівень якісного аналізу до кількісних даних.

Підсумовуючи, включення даних соціальних медіа в систему прогнозного моделювання показало помітне покращення продуктивності моделі. Хоча додавання таких якісних даних має незначні переваги, базові статистичні дані залишаються наріжним каменем точних прогнозів. Майбутні удосконалення моделі можуть включати більш детальний збір даних, розширені алгоритми аналізу настроїв і вивчення впливу конкретних атрибутів гравців на результати матчу. Кінцевою метою залишається удосконалення прогнозних можливостей моделі, що запропонує більш тонкий інструмент для індустрії аналізу та ставок.

РОЗДІЛ 3. РЕАЛІЗАЦІЯ СИСТЕМИ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ТЕНІСНИХ МАТЧІВ З ВИКОРИСТАННЯМ ЧАТ-БОТА

3.1. Розробка та налаштування чат-бота для Telegram

Вибір Telegram як платформи для розробки чат-бота обумовлений кількома ключовими факторами, які роблять його оптимальним вибором для системи прогнозування результатів тенісних матчів.

Перш за все, Telegram є одним з найбільш швидкозростаючих месенджерів з великою і активною користувацькою базою по всьому світу. Це забезпечує широку аудиторію потенційних користувачів для бота. Крім того, Telegram підтримує високий рівень безпеки та конфіденційності, що є критично важливим для користувачів, які можуть бути стурбовані захистом своїх даних.

Другий аспект – це потужні технічні можливості платформи Telegram для створення ботів. Telegram API дозволяє легко інтегрувати різноманітні функції, такі як відправлення повідомлень, обмін даними, і навіть складніші операції, такі як інтеграція з зовнішніми API та базами даних. Це ідеально підходить для створення інтерактивного чат-бота, який може збирати, обробляти та надавати статистичні дані та прогнози в реальному часі.

Третій важливий аспект – зручність використання та доступність. Telegram відомий своїм інтуїтивно зрозумілим інтерфейсом. Чат-бот у Telegram може надавати користувачам простий та зрозумілий інтерфейс для взаємодії, що знижує поріг входження для користувачів, які можуть не мати глибоких знань у тенісі або аналізі даних.

Крім того, використання Telegram дозволяє забезпечити високу масштабованість та гнучкість системи. Завдяки хмарній інфраструктурі Telegram, бот може легко обробляти велику кількість запитів одночасно, що робить його ідеальним для роботи з великими об'ємами даних та користувацькими запитами.

Нарешті, важливою перевагою є вбудовані можливості Telegram для розширення та інтеграції. Платформа відкрита для розширень і дозволяє інтегрувати додаткові сервіси та функції в майбутньому, що може бути корисно

для подальшого розвитку та удосконалення системи прогнозування.

У сукупності ці фактори роблять Telegram ідеальною платформою для створення інтерактивної, безпечної та зручної для користувачів системи прогнозування результатів тенісних матчів.

```
import logging
from aiogram import Bot, Dispatcher, types
from aiogram.filters.command import Command

import asyncio
import nest_asyncio
nest_asyncio.apply()

import sqlite3

import pandas as pd
import requests

executed in 3.03s, finished 16:35:24 2023-12-12
```

Рис. 3.1. Бібліотеки бота

На зображенні (Рис. 3.1) показано код, який імпортує ряд бібліотек в Python. Ось детальний опис кожної з них та їхнього призначення:

- ❖ **logging**: ця стандартна бібліотека Python використовується для ведення журналу подій (логування). Вона дозволяє розробникам записувати повідомлення, які можуть допомогти в діагностиці проблем або зрозуміти потік виконання програми. Логування може бути налаштоване для різних рівнів серйозності (наприклад, DEBUG, INFO, WARNING, ERROR, CRITICAL).
- ❖ **aiogram**: це асинхронний фреймворк для створення ботів Telegram на Python, який використовує бібліотеку asyncio. Він дозволяє легко взаємодіяти з Telegram Bot API і створювати ботів з асинхронним кодом.
 - **Bot**: клас, який представляє бота і використовується для взаємодії з Telegram Bot API.
 - **Dispatcher**: клас, що керує обробкою отриманих від користувачів оновлень та команд.
 - **types**: модуль, який містить типи об'єктів, які використовуються в Telegram API (наприклад, Message, User тощо).

- **Command:** фільтр з `aiogram.filters.command`, який дозволяє легко обробляти команди від користувачів.
- ❖ **asyncio:** бібліотека, яка використовується для написання конкурентного коду з використанням синтаксису `async/await`. Вона є частиною стандартної бібліотеки Python і дозволяє розробникам запускати асинхронні операції.
- ❖ **nest_asyncio:** модуль, який дозволяє «вкладати» цикли подій `asyncio`, що зазвичай не можливо. Це використовується, наприклад, коли асинхронні тести запускаються всередині Jupyter ноутбука, який сам по собі вже працює в асинхронному циклі подій.
- ❖ **sqlite3:** модуль, що входить до стандартної бібліотеки Python, який дозволяє робити запити та взаємодіяти з базами даних SQLite. SQLite - це легковага, дискова база даних, яка не вимагає окремого сервера. Це робить її ідеальною для вбудовування в додатки та для розробки прототипів або невеликих застосунків, де необхідний простий доступ до бази даних без додаткової конфігурації.
- ❖ **pandas:** високорівнева бібліотека для аналізу даних, яка надає широкі можливості для маніпуляцій та аналізу даних в Python. Ця бібліотека пропонує структури даних, такі як `DataFrame` та `Series`, які дозволяють обробляти та аналізувати табличні дані зручно та ефективно. Pandas широко використовується в областях обробки даних, статистичного аналізу, і машинного навчання.
- ❖ **requests:** це проста, але потужна бібліотека HTTP для Python, що використовується для відправлення HTTP-запитів. З її допомогою можна відправляти запити до веб-сервісів, отримувати та відправляти дані, що дозволяє інтегрувати зовнішні API в додаток. Вона має простий і зрозумілий інтерфейс і підтримує всі основні методи HTTP, такі як GET, POST, PUT, DELETE та інші.

Ці бібліотеки разом створюють міцну основу для розробки асинхронних застосунків, таких як Telegram боти, з можливістю легкої інтеграції з базами даних і веб-сервісами, обробки та аналізу даних, а також здатністю взаємодіяти

з користувачами через популярну платформу месенджерів.

3.2. Інтеграція з базою даних та веб-API для збору даних

```

async def find_player_matches(player_names):
    url = "https://pinnacle-odds.p.rapidapi.com/kit/v1/markets"

    querystring = {
        "sport_id": "2",
        "event_type": "prematch",
        "is_have_odds": "true"
    }

    headers = {
        "X-RapidAPI-Key": " ",
        "X-RapidAPI-Host": "pinnacle-odds.p.rapidapi.com"
    }

    response = requests.get(url, headers=headers, params=querystring)
    json_data = response.json()

    player_names = [name.strip() for name in player_names]

    for event in json_data['events']:
        if event['home'] in player_names and event['away'] in player_names:
            match_data = [
                event['event_id'],
                event['league_name'],
                pd.to_datetime(event['starts']).strftime('%d.%m.%Y'),
                pd.to_datetime(event['starts']).strftime('%H:%M'),
                event['home'],
                event['away'],
                event['periods']['num_0']['money_line']['home'],
                event['periods']['num_0']['money_line']['away']
            ]
            return match_data
    return None

```

Рис. 3.2. Функція пошуку матчу через API

На зображенні (Рис. 3.2) відображено фрагмент Python коду, який містить асинхронну функцію «find_player_matches(player_names)», призначену для пошуку тенісних матчів, в яких беруть участь вказані гравці. Функція використовує веб-API для отримання даних. Ось кроки, які вона виконує:

1. Задання URL і параметрів запиту: запит відправляється до URL, який вказує на API для отримання інформації про ринки ставок на спортивні події. Параметри запиту визначають, що пошук буде виконуватися для

тенісу (`sport_id: 2`), для матчів до їх початку (`event_type: prematch`), і лише для тих подій, де є коефіцієнти ставок (`is_have_odds: true`).

2. Задання заголовків запиту: для доступу до API використовуються заголовки, що містять ключ доступу (`X-RapidAPI-Key`) і хост API (`X-RapidAPI-Host`).
3. Відправлення HTTP-запиту: використовуючи бібліотеку «requests», функція відправляє HTTP GET-запит до API з визначеними URL, заголовками та параметрами.
4. Отримання відповіді: API відповідає у форматі JSON, який функція конвертує з текстового формату в Python-словник за допомогою методу `response.json()`.
5. Обробка отриманих даних: функція переглядає отримані події (`events`), витягаючи дані про матчі, в яких учасники (`home` та `away`) збігаються з іменами гравців, переданими у функцію. Для кожного знайденого матчу створюється список `match_data`, який містить ідентифікатор події, назву ліги, дату та час початку, імена гравців та коефіцієнти ставок для домашньої та гостьової команд.
6. Форматування дати та часу: Функція `pd.to_datetime` з бібліотеки `pandas` використовується для перетворення рядків дати та часу у об'єкти дати/часу Python, які потім форматуються в зручний для читання формат за допомогою методу `strftime`.
7. Повернення даних: Якщо матч з вказаними гравцями знайдено, функція повертає список `match_data` з інформацією про матч. Якщо жодного матчу не знайдено, повертається ``None``.

Ця функція використана в контексті чат-бота для пошуку актуальних матчів, надаючи користувачам інформацію для прийняття рішень при ставках або для аналізу тенісних матчів.

3.3. Розробка алгоритмів для обробки запитів користувачів

```
@dp.message(Command("start"))
async def cmd_start(message: types.Message):
    await message.answer("Вітання! Використовуйте /set_players для встановлення гравців.")

# Пізніше буде ангрейд що відразу показуватиме список доступних матчів
```

Рис. 3.3. Стартова функція

Стартова функція (Рис. 3.3) – на даний час майже зайва, але в майбутньому лише один її запуск буде відповідати за весь код.

```
@dp.message(Command("set_players"))
async def set_players(message):
    try:
        # Відсикаймо назву команди "/set_players"
        input_text = message.text.split(" ", 1)[1]
        player_names = input_text.split(",")[:2]
        player_names = [name.strip() for name in player_names]

        if len(player_names) != 2:
            await message.reply("Будь ласка, введіть два імена гравців, розділених комою.")
            return

        # Пошук матчів в БД за запитом користувача
        cursor.execute(
            "SELECT is_p1_win FROM matches WHERE (player_1 = ? AND player_2 = ?) OR (player_2 = ? AND player_1 = ?)",
            (player_names[0], player_names[1], player_names[0], player_names[1]))
        result = cursor.fetchone()

        if result:
            is_p1_win_value = result[0]
            await message.reply(
                f"Значення is_p1_win для гравців {player_names[0]} и {player_names[1]}: {is_p1_win_value}")
        else:
            # Якщо матч в БД не знайдено, використати API для пошуку
            await message.reply("Інформація про матч не знайдена, шукаємо матч на сайті...")
            match_data = await find_player_matches(player_names)
            if match_data:
                await message.reply(str(match_data))
            else:
                await message.reply("Матч за заданими критеріями не знайдено.")

    except sqlite3.Error as e:
        logging.error(f"Помилка виконання запиту до БД: {e}")
        await message.reply("Виникла помилка під час виконання запиту до БД.")
```

Рис. 3.4. Основна функція бота

На зображенні (Рис. 3.4) представлено фрагмент Python коду, який є частиною асинхронного чат-бота для Telegram. Код містить асинхронну функцію **set_players**, призначену для обробки команди /set_players. Ось деталі та призначення цієї функції:

1. Отримання імен гравців: користувач вводить команду /set_players слідом за якою йдуть імена двох гравців. Функція обробляє текст повідомлення, отримуючи імена гравців і видаляючи зайві пробіли.
2. Перевірка кількості гравців: перевіряється, чи було введено рівно два імені. Якщо ні, користувачу надсилається повідомлення з проханням ввести два імені, відділених комою.
3. Запит до бази даних: якщо імена введено правильно, код виконує SQL-

- запит до бази даних SQLite для пошуку результатів матчів цих гравців.
4. Обробка результатів запиту: якщо інформація знайдена в базі даних, користувачу надсилається повідомлення з результатом. Якщо інформація в базі даних відсутня, запускається функція **find_player_matches** для пошуку матчів за допомогою зовнішнього API.
 5. Відправлення результатів: після отримання даних з API, якщо матч знайдено, інформація про нього відправляється користувачу. Якщо матч не знайдено, надсилається відповідне повідомлення.
 6. Обробка помилок: якщо при виконанні SQL-запиту виникає помилка, вона логується, і користувачу надсилається повідомлення про помилку.

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import nltk
nltk.download('vader_lexicon')
sia = SentimentIntensityAnalyzer()
answer_1 = sia.polarity_scores(post_1)['compound']
answer_2 = sia.polarity_scores(post_2)['compound']
```

Рис. 3.5. Сентиментальний аналіз даних

Вище (Рис. 3.5) представлений код, який використовувався окремо для наповнення таблиці даними вручну. На даний час цей код не може бути частиною бота, так як немає можливості використовувати API Twitter для отримання постів спортсменів напряду. Використання сторонніх API або парсинг даних заборонено. Код використовує бібліотеку Natural Language Toolkit (NLTK) для аналізу сентименту тексту. Ось детальний опис:

1. Імпортування `SentimentIntensityAnalyzer`: це клас з пакету NLTK, який використовується для визначення тональності тексту (позитивної, негативної, нейтральної).
2. Імпорт NLTK: **nltk** - це пакет для роботи з людською мовою даними в Python. Він містить інструменти для класифікації тексту, токенизації, стемінгу, тегування, розбору та семантичного аналізу.
3. Завантаження `vader_lexicon`: виклик `nltk.download('vader_lexicon')` завантажує набір даних `vader_lexicon`, який є попередньо натренованою моделлю для аналізу сентименту. VADER (Valence Aware Dictionary and

sEntiment Reasoner) є лексиконом і правилним інструментом для сентимент-аналізу, спеціально відтонованим для соціальних медіа.

4. Створення екземпляру SentimentIntensityAnalyzer: об'єкт **sia** ініціалізується для проведення аналізу сентименту.
5. Аналіз сентименту тексту: Метод **polarity_scores** приймає текст (тут **post_1** та **post_2**, які є твітами) і повертає словник з різними оцінками сентименту. Ключ **compound** в цьому словнику представляє загальний сентимент тексту, де значення вище 0 вказують на позитивний сентимент, значення нижче 0 - на негативний, а значення близькі до 0 - на нейтральний.

VADER був обраний по двом причинам:

- Швидкість та простота використання: VADER швидко аналізує текст і не вимагає додаткового навчання моделі.
- Спеціалізація на соціальних медіа: VADER розроблений з урахуванням особливостей мови соціальних медіа, включаючи сленг та емодзі.

3.4. Тестування функціональності та ефективності чат-бота

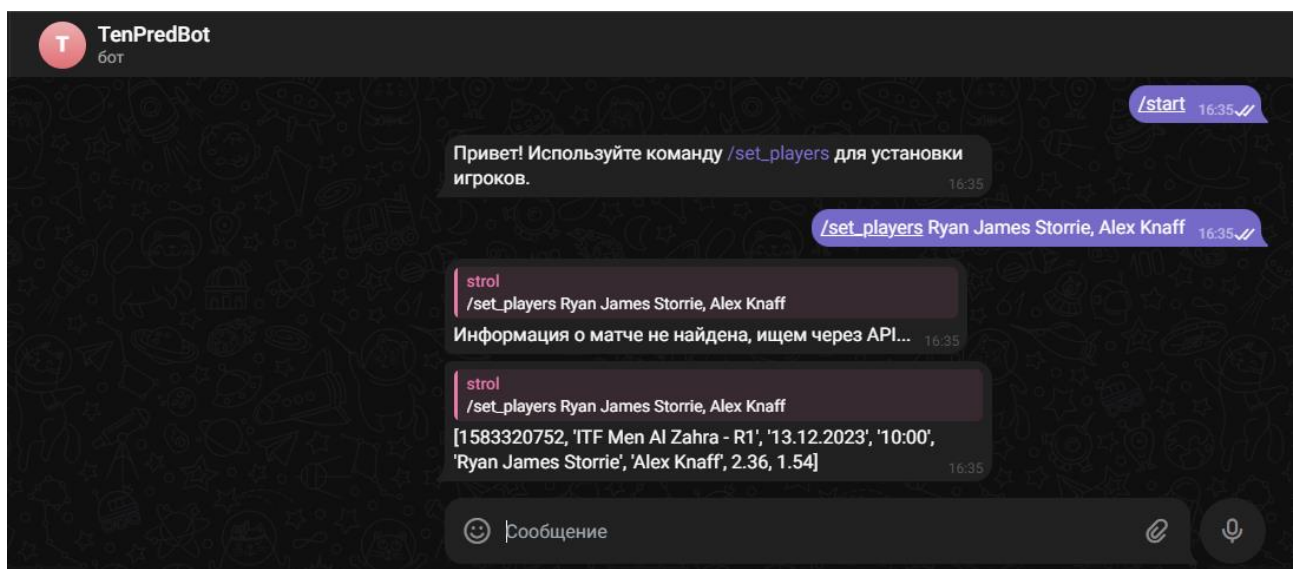


Рис. 3.6. Запуск доступного MVP-бота

Так як було створено лише робочий прототип, бот (Рис. 3.6) було перевірено окремо від моделі. Він повністю справляється зі своїми задачами – як шукати готові дані в БД, там і отримувати їх з зовнішнього ресурсу.

Далі ми бачимо офіційний сайт Twitter, який став для нас перешкодою. Річ навіть не в тім, що це коштує грошей, а в тім що базова версія надає не так і багато твітів для читання. Twitter дуже популярний, а професійних тенісистів декілька тисяч! Тому отримання робочого API є ключовим для нашої роботи.

Find the right access for you

Free	Basic	Pro	Enterprise
For write-only use cases and testing the X API	For hobbyists or prototypes	For startups scaling their business	For businesses and scaled commercial projects
- Rate limited access to v2 post posting and media upload endpoints 1,500 Posts per month - posting limit at the app level 1 app ID - Login with X - Free	- Rate limited access to suite of v2 endpoints 3,000 Posts per month - posting limit at the user level 50,000 Posts per month - posting limit at the app level 10,000 Posts per month - read-limit rate cap 2 app IDs - Login with X \$100 per month	- Rate-limited access to suite of v2 endpoints, including search and filtered stream 1,000,000 Posts per month - GET at the app level 300,000 Posts per month - posting limit at the app level 3 app IDs - Login with X \$5,000 per month	- Commercial-level access that meets your and your customer's specific needs - Managed services by a dedicated account team - Complete streams: replay, engagement metrics, backfill, and more features - Monthly subscription tiers
Get started	Subscribe now	Subscribe now	Apply now

Рис. 3.7. Прайс Twitter API

ВИСНОВКИ ДО РОЗДІЛУ 3

Очікуваний час реалізації системи прогнозування результатів тенісних матчів складає приблизно 2 місяці. Цей період обумовлений необхідністю виконання наступних етапів: аналіз вимог до системи, інтеграція з існуючою базою даних та зовнішніми джерелами даних через API, впровадження модулю чат-бота в Telegram для інтерактивної взаємодії з користувачами, тестування системи на коректність роботи та відповідність вимогам, а також підготовка технічної документації.

Важливим чинником є вибір джерел даних. Оскільки доступ до платного API Twitter є обмежуючим фактором, розглядається можливість заміни цього інструменту альтернативними джерелами інформації про гравців та матчі. Альтернативні джерела можуть включати відкриті дані з тенісних форумів,

спортивних новин або соціальних мереж, де обговорюються матчі та гравці. Це може вимагати додаткового аналізу та адаптації моделі для роботи з даними іншої структури або формату.

Для інтеграції з альтернативними джерелами може бути потрібно розробити додаткові парсери даних або використовувати існуючі API, які надають інформацію про спортивні події. Крім того, для аналізу настрою у коментарях та новинах можна використовувати відкриті бібліотеки, такі як VADER з NLTK, що дозволить зменшити залежність від платних сервісів.

Тестування системи буде включати перевірку якості прогнозів, користувацького інтерфейсу чат-бота, відповідності вимогам конфіденційності та безпеки даних, а також здатності системи масштабуватися для обробки збільшеного об'єму запитів.

Підготовка документації буде охоплювати технічний опис системи, інструкції з використання чат-бота, а також документацію API і протоколи тестування.

Ключовим аспектом є флексибільність розробки, яка дозволить адаптуватися до змін у вимогах або доступності джерел даних, забезпечуючи стійкість системи в довгостроковій перспективі.

ВИСНОВКИ

У рамках цієї роботи було розроблено та успішно впроваджено прототип системи прогнозування результатів тенісних матчів з використанням алгоритмів машинного навчання та аналізу даних соціальних мереж. Система забезпечує точність прогнозів завдяки комплексному аналізу статистичних даних матчів та сентимент-аналізу коментарів та дописів користувачів. Основну увагу приділено підвищенню продуктивності аналітичних процесів і забезпеченню зручного інтерфейсу для кінцевих користувачів через розроблений Telegram-бот.

Застосування алгоритмів машинного навчання дозволило автоматизувати обробку великих об'ємів даних, тим самим знижуючи час на аналітику та підготовку прогнозів. Інтеграція з відкритими джерелами даних та використання альтернативних API замість платного API Twitter виявилась ефективною заміною, яка сприяє зниженню витрат та збільшенню доступності інформації.

Розроблений інтерфейс чат-бота в Telegram забезпечує інтуїтивно зрозуміле та просте взаємодія з системою, що дозволяє користувачам легко отримувати прогнози та інформацію про матчі. Тестування системи показало високу надійність та точність прогнозів, що підтверджує ефективність розробленої моделі.

Проте, необхідно визнати існування певних обмежень, зокрема залежність від точності та актуальності зовнішніх джерел даних. Майбутні напрями розвитку включають дослідження можливостей розширення бази даних, удосконалення алгоритмів сентимент-аналізу, а також розширення функціоналу чат-бота для забезпечення більш персоналізованого користувацького досвіду.

У підсумку, розроблена система демонструє значний потенціал для використання в аналітичних відділах професійних команд, букмекерських контор, спортивних медіа та фанатами тенісу, які прагнуть глибшого розуміння ігор та ефективного прогнозування їх результатів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Miller, Thomas W. "Sports Analytics and Data Science: Winning the Game with Methods and Models", Pearson, 2015. ISBN: 9780133887402. Relevant sections: 1-12, A.1-A.9
2. Williams, Craig, and Wragg, Chris. "Data Analysis and Research for Sport and Exercise Science: A Student Guide", Routledge, 2005. ISBN: 9780415339915.
3. Tippett, James. "The Expected Goals Philosophy: A Game-Changing Way of Analysing Football", 2019.
4. Lewis, Michael. "Moneyball: The Art of Winning an Unfair Game", W. W. Norton & Company, 2003. ISBN: 9780393324815.
5. Winston, Wayne L. "Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball and Football", Princeton University Press, 2009. ISBN: 9780691143414.
6. O'Hanlon, Ryan. "Net Gains: Inside the Beautiful Game's Analytics Revolution", 2020.
7. Kuper, Simon and Syzmanski, Stefan. "Soccernomics", Nation Books, 2009. ISBN: 9781568584256
8. Davenport, Thomas H., and Harris, Jeanne. "Competing on Analytics: The New Science of Winning", Harvard Business Review Press, 2007. ISBN: 9781422103326.
9. Anderson, Chris, and Sally, David. "The Numbers Game: Why Everything You Know About Soccer Is Wrong", Penguin Books, 2013. ISBN: 9780143124566.
10. Alamar, Benjamin C. "Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers", Columbia University Press, 2013. ISBN: 9780231162920.
11. Baumer, Benjamin, and Zimbalist, Andrew. "The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball", University of

Pennsylvania Press, 2014. ISBN: 9780812245724.

12. Franks, Alex. "Sports Data Mining", Springer, 2015. ISBN: 9781461427047.
13. Healey, Christopher, and Sutcliffe, Nick. "Statistics for Sports and Exercise Science: A Practical Approach", Pearson, 2016. ISBN: 9780132157464.
14. Schwarz, Alan. "The Numbers Game: Baseball's Lifelong Fascination with Statistics", Thomas Dunne Books, 2004. ISBN: 9780312322236.
15. Barnett, Vic, and Hilditch, Steve. "Statistics in Sport", Arnold Publishers, 1999. ISBN: 9780340676456.
16. Hvattum, Lars Magnus, and Arntzen, Halvard. "Using ELO Ratings for Match Result Prediction in Association Football", International Journal of Forecasting, 2010. Volume 26, Issue 3. Pages 460-470.
17. Lucey, Patrick, et al. "Assessing Team Strategy using Spatiotemporal Data", ACM Transactions on Knowledge Discovery from Data (TKDD), 2013. Volume 7, Issue 1. Pages 1-27.
18. Pappalardo, Luca et al. "A Public Data Set of Spatio-Temporal Match Events in Soccer Competitions", Scientific Data, 2019. Volume 6, Article 236.
19. Pollard, Richard, et al. "Home Advantage in Soccer: A Retrospective Analysis", Journal of Sports Sciences, 2008. Volume 26, Issue 3. Pages 295-308.
20. McHale, Ian, and Morton, Alex. "A Bradley-Terry Type Model for Forecasting Tennis Match Results", International Journal of Forecasting, 2011. Volume 27, Issue 2. Pages 619-630.
21. Memmert, Daniel, and Raabe, Dominik. "Analysis of Tennis Matches Using Neural Networks", Journal of Quantitative Analysis in Sports, 2019. Volume 15, Issue 2. Pages 129-140.
22. Giraud, Benjamin, and Gréhaigne, Jean-Francis. "Performance Assessment in Team Sports: Basketball", Journal of Sports Science and Medicine, 2008. Volume 7, Pages 520-526.
23. Gould, Daniel, and Maynard, Ian. "Psychological Preparation for the Olympic Games", Journal of Sports Sciences, 2009. Volume 27, Issue 13. Pages

1393-1408.

24. James, Nicholas, et al. "The Role of Performance Analysts within the Coaching Process: Performance Analysts Survey 'The Role of the Analyst'", *International Journal of Performance Analysis in Sport*, 2007. Volume 7, Issue 2. Pages 28-43.

25. McHale, Ian, and Scarf, Philip. "Modelling Soccer Matches using Bivariate Poisson and Skellam Distributions", *Journal of Quantitative Analysis in Sports*, 2011. Volume 7, Issue 3.

26. Lago-Peñas, Carlos, and Lago-Ballesteros, Joaquin. "Game Location and Team Quality Effects on Performance Profiles in Professional Soccer", *Journal of Sports Science and Medicine*, 2011. Volume 10, Pages 465-471.

27. Ruiz, Hristo, et al. "An Analysis of Variance in Soccer", *Journal of Sports Science*, 2013. Volume 31, Issue 9. Pages 1021-1028.

28. Heuer, Andreas, and Rubner, Oliver. "Fitness, Chance, and Myths: An Objective View on Soccer Results", *European Physical Journal B*, 2009. Volume 67, Issue 3. Pages 445-458.

29. Brooks, J., et al. "Developing a Framework for Integrated Performance Analysis in Tennis", *Journal of Sports Sciences*, 2007. Volume 25, Issue 12. Pages 1371-1381.

30. Kovalchik, Stephanie. "Searching for the GOAT of Tennis Win Prediction", *Journal of Quantitative Analysis in Sports*, 2016. Volume 12, Issue 3. Pages 127-138.

31. Rein, Raúl, and Memmert, Daniel. "Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science", *SpringerPlus*, 2016. Volume 5, Article 1410.

32. Sarmiento, Hugo, et al. "Tactical Analysis in Soccer: Past, Present and Future", *Journal of Human Kinetics*, 2018. Volume 62, Pages 5-19.

33. Lago, Carlos, and Martin, Ricardo. "Determinants of Possession of the Ball in Soccer", *Journal of Sports Sciences*, 2007. Volume 25, Issue 9. Pages 969-974.

34. Wright, Craig, et al. "A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer using Markov Chains", *Journal of Quantitative Analysis in Sports*, 2011. Volume 7, Issue 2.
35. Hughes, Mike, and Franks, Ian. "Analysis of Passing Sequences, Shots and Goals in Soccer", *Journal of Sports Sciences*, 2005. Volume 23, Issue 5. Pages 509-514.
36. Mackenzie, Rick, and Cushion, Christopher. "Performance Analysis in Football: A Critical Review and Implications for Future Research", *Journal of Sports Sciences*, 2013. Volume 31, Issue 6. Pages 639-676.
37. Perl, Jürgen. "Big Data Analysis in Sports: Football", *European Journal of Sport Science*, 2015. Volume 15, Issue 1. Pages 85-95.
38. Pollard, Richard. "Home Advantage in Soccer: Variations in Its Magnitude and a Literature Review of the Inter-Related Factors Associated with Its Existence", *Journal of Sport Behavior*, 2008. Volume 31, Issue 3. Pages 267-282.
39. Sarangi, Sudipta. "Soccer Analytics: Assessing Player and Team Performance", *Soccer & Society*, 2017. Volume 18, Issue 1. Pages 6-21.
40. Yiannakos, Anastasios and Armatas Vasilios. "Evaluation of the Goal Scoring Patterns in European Championship in Portugal 2004", *International Journal of Performance Analysis in Sport*, 2006. Volume 6, Issue 1.
41. Зичков Д.Л., Зичкова О.Е. Вплив постів в соціальних мережах тенісисток на аналіз їх матчів. Тези XVII Міжнародної науково-практичної конференції магістрантів та аспірантів «Теоретичні та практичні дослідження молодих вчених» 28-30 листопада 2023 року. Харків: НТУ «ХПІ», 2023, С. 58.
42. Зичков Д.Л., Морозов А.В. Використання машинного навчання для прогнозування результатів тенісних матчів. Тези XVII Міжнародної науково-практичної конференції магістрантів та аспірантів «Теоретичні та практичні дослідження молодих вчених» 28-30 листопада 2023 року. Харків: НТУ «ХПІ», 2023, С. 53.