

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження та розробка системи прогнозування спортивних подій для
букмекерських контор
(тема)

Виконав:

здобувач другого року навчання,

групи ДСМ-23-1

Терехов М.В.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Науки про дані (Data Science)

(повна назва спеціалізації)

Керівник доц. Волощук О.Б.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

О.В. Золотухін

(прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Штучного інтелекту _____
(повна назва)
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-професійна _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Науки про дані (Data Science) _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Терехову Михайлу Владиславовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Дослідження та розробка системи прогнозування спортивних подій для букмекерських контор _____

затверджена наказом університету від 22 листопада 20 24 р. № 1238Ст

2. Термін подання студентом роботи до екзаменаційної комісії 23 січня 20 25 р.

3. Вихідні дані до роботи Наукові статі та роботи про прогнозування спортивних подій, доповіді про використання інформаційних технологій у спорті, технічна документація по процесу формування та функціонування серверних систем букмекерських контор, документація по .NET платформі та Python, документація по побудові багатoshарових, системних архітектур, дослідження по оцінці точності роботи алгоритмів по прогнозуванню

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Теоретичні основи спортивного прогнозування

2) Дослідження алгоритмів аналізу статистичних

3) Проектування та розробка системи прогнозування

РЕФЕРАТ

Пояснювальна записка: 66 с., 43 рис., 1 дод., 21 джерело.

АЛГОРИТМИ ПРОГНОЗУВАННЯ, БУКМЕКЕРСЬКА ДІЯЛЬНІСТЬ, СПОРТИВНА АНАЛІТИКА, DOTNET, PREDICTION SYSTEM, PYTHON.

Об'єкт дослідження – процес прогнозування результатів спортивних подій у букмекерській діяльності.

Предмет дослідження – методи та алгоритми автоматизованого прогнозування спортивних подій на основі статистичних даних та машинного навчання.

Мета роботи – підвищення точності прогнозування результатів спортивних подій шляхом розробки автоматизованої системи, що використовує комплексний аналіз статистичних даних та алгоритми машинного навчання.

Методи дослідження – базуються на використанні системного аналізу для вивчення взаємозв'язків у процесах прогнозування спортивних подій, математичного моделювання та теорії ймовірностей для створення моделей, алгоритмів машинного навчання, включаючи нейронні мережі та методи глибокого навчання для аналізу великих масивів спортивних даних.

У даній роботі розроблено архітектуру системи прогнозування, реалізовано алгоритми обробки статистичних даних спортивних подій та створено систему для автоматизації процесу прогнозування. Проведено комплексне тестування системи, яке підтвердило підвищення точності прогнозів.

Розроблена система може бути використана букмекерськими конторами для підвищення ефективності їхньої діяльності та точності прогнозів спортивних подій.

ABSTRACT

Master's thesis contains: 66 pp., 43 fig., 1 ann., 21 references.

ALGORITHMIC PREDICTION, BETTING ACTIVITIES, DOTNET, PREDICTION SYSTEM, PYTHON, SPORTS ANALYTICS.

The object of research is the process of forecasting the outcomes of sports events in betting.

The subject of research is methods and algorithms for automated forecasting of sports events based on statistical data and machine learning.

The main purpose of the work is to improve the accuracy of predicting the outcomes of sports events by developing an automated system that uses a comprehensive analysis of statistical data and machine learning algorithms.

Research methods – based on the use of system analysis to study the relationships in the processes of predicting sports events, mathematical modelling and probability theory to create models, machine learning algorithms, including neural networks and deep learning methods to analyse large amounts of sports data.

As a result of work, the architecture of the forecasting system was developed, algorithms for processing sports event statistics were implemented, and a system for automating the forecasting process was created. A comprehensive testing of the system was conducted, which confirmed the increase in forecast accuracy.

The developed system can be used by bookmakers to improve the efficiency of their operations and the accuracy of sports event forecasts.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Теоретичні основи спортивного прогнозування.....	10
1.1 Аналіз функціонування букмекерських контор.....	10
1.1.1 Букмекерські контори.....	10
1.1.2 Формування коефіцієнтів.....	12
1.2 Спортивне прогнозування	15
1.2.1 Методи прогнозування спортивних подій	15
1.2.2 Системи обробки великих масивів даних для спортів.....	18
1.2.3 Оцінка достовірності джерел інформації	22
2 Дослідження алгоритмів статистичного аналізу	28
2.1 Сучасні підходи до аналізу даних	28
2.1.1 Машинне навчання	28
2.1.2 Предиктивна аналітика.....	34
2.2 Практичне застосування алгоритмів	40
3 Проектування та розробка системи прогнозування	43
3.1 Проектування системної архітектури застосунку	43
3.2 Проектування алгоритмічної та доменної логіки застосунку	45
3.2.1 Проектування методів отримання даних статистики.....	45
3.2.2 Проектування алгоритмічних скриптів для прогнозування.....	48
3.3 Інтеграційне тестування	54
3.4 Переваги та виклики	58
Висновки	61
Перелік джерел посилання	63
Додаток А Відомість кваліфікаційної роботи	66

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AI – Artificial Intelligence – штучний інтелект;

API – Application Programming Interface – інтерфейс програмування застосунків;

Big Data – великі дані, які обробляються системами аналітики для масштабних прогнозів;

ODDS – коефіцієнти (ймовірності) спортивних подій;

Poisson – Poisson Distribution – пуассонівський розподіл, використовується для прогнозування кількості подій (наприклад, голів у матчі);

RMSE – Root Mean Square Error – середньоквадратична похибка;

RNN – Recurrent Neural Network – рекурентна нейронна мережа.

ВСТУП

У сучасному світі ставки на спорт стали невід'ємною частиною індустрії розваг, привертаючи увагу мільйонів людей по всьому світу. Для букмекерських контор точність прогнозування результатів спортивних подій є найважливішим фактором успіху, що впливає як на формування коефіцієнтів, так і на загальну прибутковість компанії. Однак процес прогнозування залишається складним і супроводжується безліччю проблем, які потребують ефективного вирішення.

Однією з головних проблем є обмеженість традиційних методів аналізу. Статистичні моделі, що ґрунтуються на історичних даних, не завжди можуть врахувати динамічні зміни у складі команд, травми гравців, погодні умови та інші несподівані фактори. Крім того, експертні оцінки часто суб'єктивні й залежать від людського фактора, що може призвести до помилок у прогнозуванні.

Ще одна важлива проблема – обробка величезних обсягів інформації. З кожним роком кількість даних, які можуть вплинути на результат події, зростає: статистика матчів, індивідуальні виступи, соціальні мережі, новини, прогнози погоди тощо. Застарілі аналітичні системи часто не здатні швидко й ефективно обробляти ці дані, що обмежує їхню застосовність у реальних умовах.

Існує також проблема обмеженої адаптивності традиційних моделей. У світі спорту ситуація може змінюватися дуже швидко, і системи, не здатні адаптуватися до цих змін у режимі реального часу, стають неефективними. Наприклад, прихід нового тренера, зміна тактичної стратегії або фізичного стану ключових гравців може кардинально вплинути на результат гри.

Усі ці проблеми вимагають впровадження інноваційних підходів, які дали б змогу підвищити якість прогнозів і дати операторам ставок конкурентну перевагу. Сучасні методи машинного навчання та штучного інтелекту здатні розв'язати ці проблеми, оскільки вони здатні автоматично

розпізнавати складні взаємозв'язки між даними, аналізувати їх у режимі реального часу та швидко адаптуватися до нових умов.

Основною задачею цієї роботи стає розробити систему прогнозування спортивних подій, яка дасть змогу подолати наявні проблеми, підвищити точність і ефективність прогнозів, а також забезпечити гнучкість і адаптивність до спортивного середовища, що швидко змінюється.

1 ТЕОРЕТИЧНІ ОСНОВИ СПОРТИВНОГО ПРОГНОЗУВАННЯ

1.1 Аналіз функціонування букмекерських контор

1.1.1 Букмекерські контори

Букмекери представляються як організації або приватні особи, які приймають ставки на результати різних подій, здебільшого спортивних, з метою отримання прибутку. Основою їхньої діяльності є розрахунок коефіцієнтів, які відображають імовірність певного результату. Ці коефіцієнти створюються шляхом аналізу статистики, врахування зовнішніх факторів і використання спеціальних аналітичних алгоритмів. Виходячи з цього, основною метою їх є забезпечити прибуток незалежно від результату події завдяки вбудованій маржі. Букмекерські контори пропонують різні види ставок: ординарні (на одну подію), експреси (на кілька подій), системи, а також лайв-ставки, що робляться в реальному часі, для прикладу на подію зображену на рисунку 1.1. Крім спортивних подій, доступні також ставки на політику, шоу-бізнес чи навіть погоду.



Рисунок 1.1 – Приклад спортивної події у боксі

Однією з ключових тенденцій останнього десятиліття є інтеграція ставок у режимі реального часу, що дозволяє клієнтам робити ставки під час події. Такі технології, як стрімінг або використання датчиків Інтернету речей на стадіонах, у поєднанні з високошвидкісним обміном інформацією, допомагають оновлювати коефіцієнти в режимі реального часу, виходячи з ситуації на полі.

Інноваційність ідеї полягала у використанні технологій блокчейн, які забезпечують прозорість, анонімність і безпеку фінансових операцій. Це особливо актуально для міжнародних компаній, діяльність яких ведеться на ринках з різними валютами та регуляторними стандартами.

Багато букмекерів впроваджують гейміфікацію для утримання клієнтів. Прикладами є бонуси, які зображені для прикладу на рисунку 1.2, на основі активності користувачів, рівні системи лояльності, персоналізовані акції – аж до рівня гейміфікації між клієнтами. Це перетворює на гру буквально все, включно з беттингом.

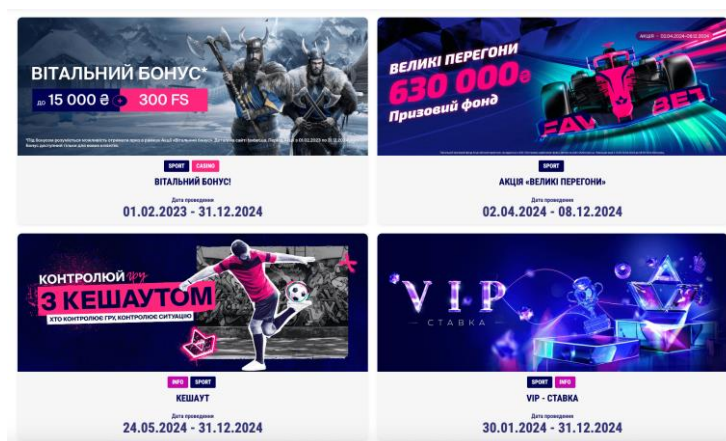


Рисунок 1.2 – Приклад бонусної гейміфікації користувачів

Соціальний аспект діяльності букмекерів також виходить на новий рівень. У відповідь на зростаючу критику, пов'язану з азартними іграми, багато компаній інвестують у програми відповідальної гри (Responsible Gambling). Це включає впровадження систем самовиключення, технологій

розпізнавання проблемної поведінки гравців за допомогою ШІ та підтримку центрів психологічної допомоги.

Вони вже стали важливим партнером у розвитку цієї індустрії. Вони створюють спеціальні пропозиції для ставок на кіберспортивні події, розробляють аналітичні моделі, які враховують унікальні метрики гравців і команд, і навіть спонсорують турніри для популяризації цього виду спорту.

Крім того, особлива увага приділяється регіональним особливостям. Так, у Південній Америці широко поширені ставки на футбол, у США активно розвиваються ставки на американський футбол, бейсбол і баскетбол. Букмекери орієнтують свою маркетингову діяльність відповідно до культурних особливостей і популярності того чи іншого виду спорту в конкретних регіонах.

Таким чином, сьогодні букмекерські контори – це комплексні платформи, які поєднують традиційний гральний бізнес із найсучаснішими технологіями, включаючи соціальну відповідальність і глибокий аналіз для створення конкурентоспроможного та безпечного бізнесу в сфері розваг.

1.1.2 Формування коефіцієнтів

Діяльність букмекерів охоплює кілька ключових аспектів. Основним є прийом ставок, коли клієнти роблять прогнози, а букмекер їх фіксує. Важливою складовою є управління ризиками, що включає балансування ставок, мінімізацію ймовірності значних втрат через несподівані результати, а також обмеження максимальних сум ставок для окремих клієнтів. Для цього використовуються системи аналізу, приклад якої зображено на рисунку 1.3, які допомагають відстежувати динаміку ставок і своєчасно коригувати коефіцієнти.

Manchester City - Arsenal
 Today, 28 Aug 2021, 14:30

1X2 Home/Away AH O/U DNB EH More bets

Full Time 1st Half 2nd Half

Bookmakers	1	X	2	Payout
10x10bet	↓ 1.35	↑ 5.20	↑ 10.00	96.8%
1xBET	↓ 1.38	↓ 5.15	↑ 9.80	98.0%
ASIANODDS	↓ 1.36	↑ 5.31	↑ 10.80	98.4%
bet-at-home	↓ 1.35	↓ 5.00	↑ 8.40	94.4%
bet365	↑ 1.33	↓ 5.25	↓ 9.00	94.9%
BETHARD	↓ 1.34	↓ 5.25	↑ 9.50	96.0%
BETINASIA	↓ 1.35	↑ 5.25	↑ 10.25	97.2%

Рисунок 1.3 – Відстежування зміни коефіцієнтів у різних конторах

Формування коефіцієнтів у букмекерських конторах представляє собою складний процес, який базується на аналітиці, математичних розрахунках та управлінні ризиками. Коефіцієнт відображає ймовірність певного результату події та визначає можливий виграш гравця. Цю ймовірність оцінюють аналітики букмекерських контор. Перш за все, аналітики оцінюють всілякі фактори, які можуть вплинути на результат події. Сюди входить статистика команд або гравців, які беруть участь у грі, їхня поточна форма, погодні умови, місце проведення, історія очних зустрічей між двома командами-учасниками, мотивація, склад на матч і так далі. Всі зібрані дані потім проходять через спеціальні математичні моделі, які оцінюють ймовірність кожного з них [12].

Розрахунок починається з визначення ймовірності певної події, яка виражається у відсотках. Наприклад, якщо аналітики вважають, що ймовірність перемоги команди «Шахтар» становить 50%, то коефіцієнт для команди Шахтар дорівнює 2.00 без запасу ($100\% / 50\%$) [14].

Наприклад, якщо ймовірність перемоги команди «Динамо» становить 30%, то коефіцієнт буде 3,33 ($100\% / 30\%$). Ці значення, однак, не враховують маржу, яку букмекерські контори закладають у свої ціни, щоб отримати прибуток. Маржа представляє собою різницю між істинною

ймовірністю та закладеною в коефіцієнті ймовірністю. Отже, якщо подія має істинну ймовірність 50%, букмекер може ввести маржу, знизивши коефіцієнт до 1.90 з 2.00. Крім статистичних даних, букмекер також враховує грошові потоки – кількість і обсяг ставок, зроблених на різні результати [15]. Це дозволяє коригувати коефіцієнти залежно від поведінки гравців. Наприклад, якщо більшість ставок зроблено на перемогу команди «Шахтар», букмекер може знизити коефіцієнт, щоб зменшити ризик великих виплат, і одночасно підвищити коефіцієнт на команду «Динамо», щоб заохотити ставки на цей результат, приклад руху коефіцієнтів показано на рисунку 1.4. У лайв ставках коефіцієнти динамічно змінюються під час події. Це необхідно враховувати при прогнозуванні рахунку матчу: забиті голи, червоні картки, травми гравців та інші фактори, які можуть змінити хід гри [7].



Рисунок 1.4 – Відстежування руху лінії по вибраним турнірам

Такі алгоритми і програмне забезпечення автоматизують процес оновлення коефіцієнтів. Величезну роль у цьому відіграє штучний інтелект, який аналізує великі обсяги даних і дуже швидко реагує на зміни. Таким

чином, процес формування коефіцієнтів дуже складний симбіоз аналізу даних, математичних розрахунків, технологій і ризик-менеджменту, спрямований на забезпечення прибутковості букмекерської компанії та збалансованості грошових потоків [4].

1.2 Спортивне прогнозування

1.2.1 Методи прогнозування спортивних подій

Найскладнішим слід вважати спортивне прогнозування, в рамках якого відбувається багатовимірний аналіз даних, залучення сучасних технологій, досвіду аналітика та розуміння культури в спорті. Це набагато більше, ніж арифметика, і охоплює психологію, соціологію, фізіологію та поведінкову економіку. Технологія Big Data та машинне навчання мають великий вплив на сучасні технології прогнозування. Дані надходять з різних джерел: офіційна статистика, для прикладу зображена статистика на рисунку 1.5, результати матчів, результати окремих гравців, метеорологічна інформація, новини і навіть тенденції в соціальних мережах [5].



Дата	Зустріч	Результат	Статус матчу
29.03.1995	Україна – Італія	0:2	Відбір до ЧЄ-1996
11.11.1995	Італія – Україна	3:1	Відбір до ЧЄ-1996
02.06.2006	Італія – Україна	0:0	Товариський матч
30.06.2006	Італія – Україна	3:0	1/4 фіналу ЧЄ-2006
07.10.2006	Італія – Україна	2:0	Відбір до ЧЄ-2008
12.09.2007	Україна – Італія	1:2	Відбір до ЧЄ-2008
29.03.2011	Україна – Італія	0:2	Товариський матч
10.10.2018	Італія – Україна	1:1	Товариський матч

Рисунок 1.5 – Статистика зустрічей збірної України проти Італії

Машинне навчання дозволяє розгадувати приховані закономірності, створювати прогностичні моделі та адаптувати їх залежно від зміни обставин. Найсуттєвішими факторами прогнозування є фізичний стан об'єкта. Також значну роль відіграють історичні дані: статистика результатів попередніх зустрічей, статистика команди в різних умовах або пори року [3].

Також враховуються психологічні фактори, такі як рівень мотивації, конфлікти в команді, публічні заяви як гравців, так і тренерів. Непередбачувані фактори, такі як рішення арбітрів, погодні умови або деякі випадкові події під час матчу, можуть сильно вплинути на результат. Такі моделі враховують ці ризики, оскільки вони імітують «шум» у даних. Що стосується власне прогнозування, то воно також передбачає аналіз поведінки гравців та вболівальників. Наприклад, закономірності в поведінці вболівальників можуть вказувати на те, як вони створюватимуть атмосферу в плей-офф, приклад якої зображено на рисунку 1.6.



Рисунок 1.6 – Вигляд вболівальників української збірної по футболу

Емоційний фон гравців визначається за їхньою мімікою, поведінкою на полі чи реакцією на соціальні виклики. Інша важлива роль відводиться медіа-середовищу в процесі формування громадської думки та впливу на ставки. Це часто може включати навіть новини про потенційну травму

ключового гравця. Тексти в соціальних мережах, наприклад, у Twitter, можуть дати уявлення про настрої вболівальників і психологічну напругу, яка може бути в команді [5].

Прогнозування кіберспортивних змагань все більше пов'язане з аналізом даних про поточну ефективність гравців. Йдеться не лише про статистику попередніх матчів, а й про те, як гравці реагують у стресових ситуаціях та адаптуються до ігрових обставин, приклад реакції гравців зображено на рисунку 1.7. Використання технологій AR і VR дає можливість створювати симуляцію матчів, де можна змоделювати будь-який із сценаріїв розвитку подій. Це особливо корисно для прогнозування рідкісних або нестандартних умов [6].



Рисунок 1.7 – Реакція гравця команди на перемогу в раунді

У спортивних прогнозах слід враховувати етичний аспект, а це означає, що має бути гарантія від того, що результати або інша інформація може бути отримана незаконним шляхом. З цієї причини уряди всього світу впроваджують суворі правила, які контролюють діяльність прогнозистів і букмекерів. Раніше прогнози базувалися на класичних статистичних моделях, таких як регресії та кореляції. Зараз перевага надається саме нейронним мережам, які можуть навчатися на даних і адаптуватися до нових

трендів у спорті. Прогнози поведінки команд. Це дає змогу букмекерській конторі надавати рекомендаційні послуги на основі кожного унікального смаку для конкретного користувача.

Тому можна впевнено сказати, що спортивне прогнозування представляє собою міждисциплінарну сферу, яка об'єднує аналітику, технології, соціологію та психологію. Цей процес постійно ускладнювався завдяки новим відкриттям і технологічним інноваціям, які залишали місце для творчості та інтуїції аналітиків.

1.2.2 Системи обробки великих масивів даних для спортів

Системи обробки великих даних є ключовим інструментом для роботи з великими обсягами спортивної інформації, включаючи статистику, фізіологічні параметри, тактичну інформацію та поведінкові характеристики, для прикладу особливості гри зображені на рисунку 1.8. У спорті такі системи дозволяють не тільки аналізувати історичну інформацію, але й обробляти дані в режимі реального часу, що дає можливість швидко приймати рішення тренерам, аналітикам і спортивним менеджерам.



Рисунок 1.8 – Особливості живої гри у парному тенісі

Однією з найважливіших особливостей систем обробки великих даних є їхня здатність працювати з різноманітними типами даних. Наприклад, тактичні дані часто представлені у вигляді координат гравців і м'яча, фізіологічні – у формі показників, які змінюються щомиті, а поведінкові дані можуть включати текстову інформацію або відео. Системи великих даних дозволяють інтегрувати ці різноманітні формати в єдине середовище, що спрощує аналіз.

Такі системи використовують платформи обробки потоків даних, такі як Apache Kafka або Spark Streaming, принцип роботи якого зображено на рисунку 1.9, для безперервної обробки інформації в реальному часі, що є критично важливим під час матчів і тренувань. Наприклад, носимі пристрої можуть передавати біометричні дані гравця в аналітичні системи, такі як його серцебиття або рівень фізичної активності, з позначенням відповідних випадків.

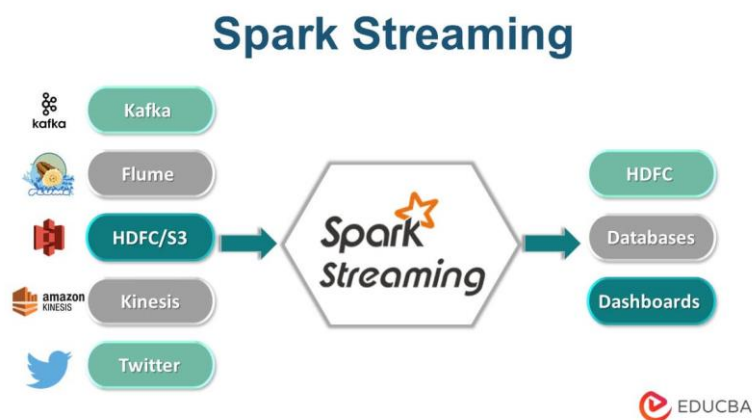


Рисунок 1.9 – Принцип роботи Spark Streaming

Хмарні обчислення відіграють вирішальну роль у спортивній аналітиці, забезпечуючи можливість обробки даних у реальному часі через сервіси, такі як AWS або Google Cloud. Ці платформи дозволяють зберігати історичні дані у розподілених базах, масштабуючи систему відповідно до зростання обсягу даних. Крім цього, аналіз емоційних даних стає новим

трендом. Наприклад, аналіз інформації із соціальних мереж, аудіоаналіз шуму зі стадіонів або оцінка поведінкових реакцій гравців допомагають зрозуміти емоційний стан учасників та їхній вплив на гру. Наприклад, аналіз текстів і постів фанатів дозволяє оцінити рівень підтримки перед важливими матчами або визначити мотивацію команди.

Інтеграція блокчейн-технологій у спортивну аналітику додає прозорості та безпеки. Запис подій і статистики у блокчейні гарантує їхню незмінність, а також забезпечує безпеку фінансових операцій, таких як продаж квитків або ставки на спортивні події, революційність його показана на рисунку 1.10. Крім того, використання блокчейну дозволяє фанатам і аналітикам перевіряти достовірність історичних даних і важливих подій. Таким чином, системи обробки великих даних у спорті забезпечують інтеграцію різних типів інформації, відстеження в реальному часі, прогнозування та створення ефективних рішень для команд і спортсменів, підвищуючи якість гри та взаємодію з фанатами.

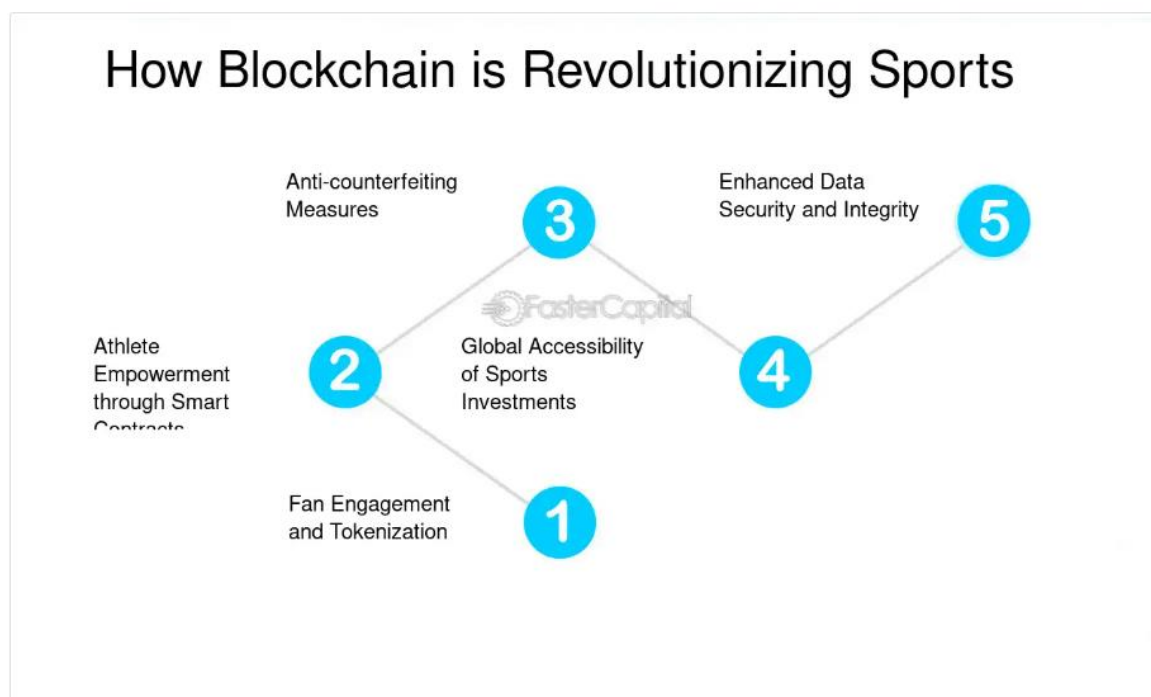


Рисунок 1.10 – Революційний вплив блокчейну на спорт

Фактично це означає розподілене сховище для зберігання даних – для роботи з петабайтами інформації. Важливою буде оптимізація доступу до даних, щоб, наприклад, історичні дані могли зберігатися в «холодному» сховищі, а актуальні дані – в «гарячому» для швидкого доступу.

Системи великих даних обов'язково включають в себе штучний інтелект і машинне навчання. Зазвичай вони призначені для побудови прогнозних моделей, що включають десятки і сотні факторів. Наприклад, моделі на основі нейронних мереж можуть аналізувати такі параметри, як те, як рухається гравець і що відбувається з м'ячем, де, ймовірно, відбудеться наступна ключова подія в грі. Такі системи також підтримують оптимізацію тренувального процесу шляхом генерації індивідуальних рекомендацій для спортсменів.

Водночас, це дозволяє в режимі реального часу представляти інформацію, яка може бути тепловими картами руху гравців, приклад зображено на рисунку 1.11, графіками фізичних навантажень або аналітикою соціальних мереж, що вказує на емоційний фон вболівальників перед ключовими матчами. Великий обсяг даних стає легше інтерпретувати візуально, і це відкриває перспективу для швидкого прийняття рішень тренерами.

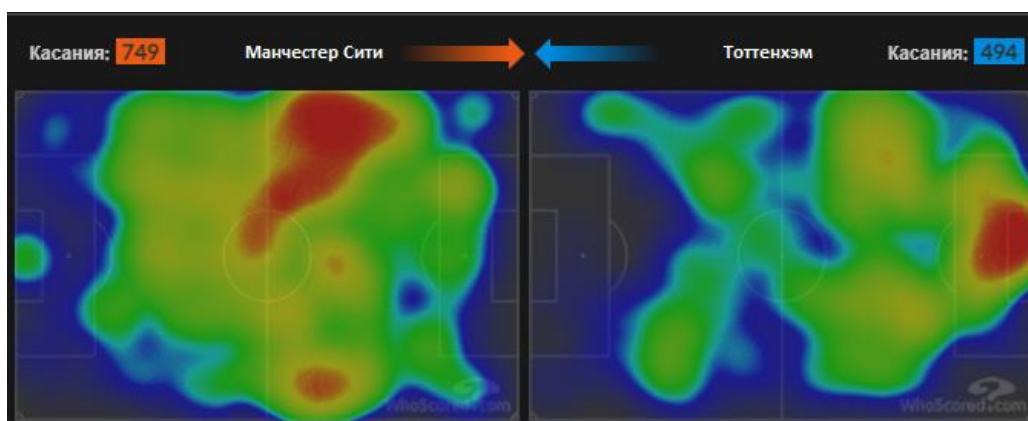


Рисунок 1.11 – Приклад теплової карти руху гравців на полі

Приватність і конфіденційність представляють дві найважливіші характеристики великих даних. Особисті фізіологічні та медичні дані у спорті повинні бути засекречені з обмеженим доступом. Стандартні методи захисту даних включають шифрування, аутентифікацію та контроль доступу.

Системи обробки великих даних у спорті забезпечують точність, швидкість і багатовимірність аналізу, необхідну для підвищення ефективності команд і спортсменів. Ці інструменти дозволяють інтегрувати різні джерела даних, шукати приховані тенденції, будувати прогнозні моделі та надавати візуальну інформацію для прийняття рішень.

1.2.3 Оцінка достовірності джерел інформації

Серед найважливіших питань – вимірювання, оскільки коректність даних є одним з основних факторів, що впливають на якість прогнозів, рішень і стратегій; важливим є питання оцінки достовірності джерел для аналітики у сфері спорту. Серед ключових методів вимірювання в цьому аспекті – технічний, приклад якого зображено на рисунку 1.12, аналітичний та контекстний аналізи.

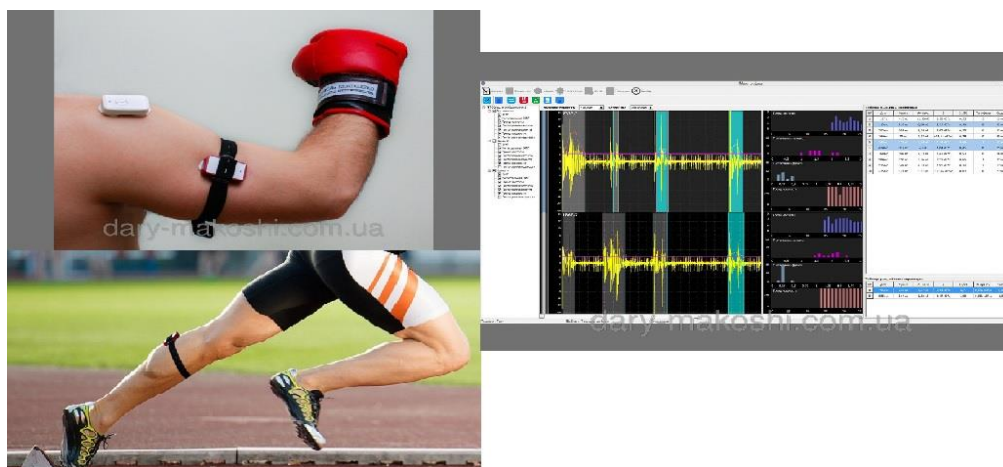


Рисунок 1.12 – Технічний аналіз можливостей гравця

Надійними джерелами можуть бути офіційні спортивні організації, ліги та клуби, а також платформи, які надають перевірену аналітику та мають великий досвід у спортивній спільноті. Наприклад, ФІФА, НБА або АТР, як правило, мають дуже суворі стандарти щодо процесу збору та публікації даних, але підтримують їхню точність. Необхідно враховувати спеціалізацію, методи збору даних та акредитацію платформ, що використовуються.

Іншим важливим моментом є перехресна перевірка джерел іншими незалежними джерелами. Інформація про подію, статистику або фізичний стан спортсмена вважається більш достовірною, коли вона підтверджується інформацією з кількох авторитетних джерел. Наприклад, якщо дані про травму якогось футболіста з'явилися на сайті клубу, до якого він належить, і водночас у провідних спортивних ЗМІ, наприклад, на ESPN або BBC Sport, де логотип першого зображено на рисунку 1.13, і ці дані відповідають один одному, то це означає, що помилки бути не може.



Рисунок 1.13 – Презентація ESPN New York турніру

Специфічна подія може дуже вплинути на достовірність джерела. До параметрів технічних даних відносяться, наприклад, коректність роботи обладнання, сертифікація, відгуки клієнтів, частота оновлення інформації. Наприклад, якщо GPS-трекер має затримку в оновленні позиції або низький

рівень точності, то дані можуть бути перекручені і вплинуть на результати аналізу.

Якщо дані мають пропуски, явні помилки або виглядають надто узагальненими, то це може свідчити. Якщо джерело, ймовірно, має особистий інтерес у даних – скажімо, для маркетингу азартних ігор або маніпулювання громадською думкою – тоді його слід використовувати з ще більшою обережністю, ніж зазвичай.

Нарешті, це процес перевірки отриманих даних на практиці. В аналітиці даних часто проводять контрольні перевірки історичних подій, перевіркою хронометражу, прилад для фіксації якого відображений на рисунку 1.14, тестування даних – для визначення точності та відповідності реальності. Дані також можуть свідчити про помилки в джерелі або методах збору, якщо вони не збігаються з уже відомими результатами.



Рисунок 1.14 – Прилад автоматичного хронометражу для спорту

Оцінка достовірності джерел інформації в спортивному аналізі також залежить від контексту даних і характеру подій, які вони описують. Одним із найважливіших інструментів є оцінка часової узгодженості: важливо перевірити, наскільки швидко джерело оновлює свої дані. Для спортивних

подій у прямому ефірі, таких як матчі або трансфери, своєчасність має вирішальне значення, оскільки застарілі дані можуть призвести до неправильних висновків. Наприклад, джерела, які надають прямі трансляції або статистику в режимі реального часу, приклад якої зображено на рисунку 1.15, зазвичай мають вищу надійність, ніж ті, що публікують інформацію із запізненням. Окрім швидкості оновлення, важливо враховувати точність цих даних. Навіть джерела, що працюють у реальному часі, можуть допускати помилки через технічні несправності, людський фактор або недоліки алгоритмів. Тому необхідно доповнювати дані таких джерел інформацією з інших, незалежних ресурсів, щоб забезпечити крос-перевірку.

Уявімо ситуацію, коли під час прямої трансляції фіксується помилковий гол через технічну затримку системи Hawk-Eye. Якщо цей момент не буде перевірено іншими джерелами, аналітики можуть зробити помилковий висновок щодо результату гри. Саме тому важливо інтегрувати кілька джерел даних, щоб мінімізувати ймовірність помилок.



Рисунок 1.15 – Онлайн скороборд статистики по спортивній події

Іншим фактором є рівень деталізації даних. Джерела, які надають поглиблений аналіз з використанням контекстних даних (наприклад, поєднують статистичні показники з тактичними оглядами), часто

вважаються більш надійними. Просте представлення цифр без пояснень або аналітичного огляду може бути ознакою низькопробного джерела. Крім того, високий рівень деталізації допомагає аналітикам перевірити дані на логічну послідовність. Наприклад, якщо джерело стверджує, що команда досягла найвищої швидкості в історії гри, але це не відповідає загальним фізичним показникам гравців, це може свідчити про помилку або маніпуляцію.

Іншим важливим аспектом є географічна та культурна специфіка джерела. Наприклад, місцеві спортивні ресурси можуть надавати більш точну інформацію про регіональні події, для прикладу на рисунку 1.16 зображено більш типовий для Індії спорт – крикет, ніж глобальні платформи, які не завжди враховують місцевий контекст. Особливо це стосується менш популярних видів спорту або місцевих ліг, де точність даних залежить від знання регіону. Однак такі джерела можуть бути обмеженими через їхню закритість або обмежену аудиторію.



Рисунок 1.16 –Типова гра в крикет для Океанії та Індії

Окремої уваги заслуговують алгоритмічні джерела інформації, засновані на автоматизованих системах збору та аналізу даних. Хоча такі

джерела часто вважаються об'єктивними, їхня достовірність залежить від якості алгоритмів і точності пристроїв. Наприклад, системи визначення місцезнаходження можуть мати помилки в складних умовах, таких як погана видимість або велике скупчення людей. Тому важливо знати, які технології використовує джерело і які обмеження воно може мати.

Нарешті, важливо враховувати історичну точність джерела. Аналізуючи попередні прогнози, аналітичні звіти та публікації, можна визначити, наскільки надійним є джерело в довгостроковій перспективі. Якщо джерело має високий відсоток правильних прогнозів або точних даних у минулому, це підвищує довіру до його поточних публікацій. Такий підхід створює основу для систематичної оцінки якості інформації у спортивній аналітиці.

З огляду на це, джерела інформації про спорт оцінюють на надійність за допомогою комплексного аналізу, незалежної перевірки, специфікацій, контексту та практичного тестування даних. Лише поєднання цих факторів дає змогу забезпечити якісну аналітику та точність прогнозів.

2 ДОСЛІДЖЕННЯ АЛГОРИТМІВ СТАТИСТИЧНОГО АНАЛІЗУ

2.1 Сучасні підходи до аналізу даних

2.1.1 Машинне навчання

Важливість нейронних мереж у світі букмекерства зумовлена їхнім потенціалом обробляти величезні обсяги даних, знаходити неявні закономірності та з високою точністю прогнозувати результати спортивних подій. Найпоширенішими архітектурами нейронних мереж, що застосовуються в цій галузі, є багатошарові перцептрони, RNN, CNN та трансформатори. Багатошаровий перцептрон – це класична архітектура, яка може працювати з табличними даними, такими як статистика матчів, форми команд та числові характеристики. Рекурентні нейронні мережі та їх більш просунуті версії, такі як LSTM і GRU, принцип роботи яких зображено на рисунку 2.1, можна застосовувати для аналізу послідовностей даних, таких як зміни коефіцієнтів у реальному часі або динаміка складу команд за останні кілька матчів. З іншого боку, згорткові нейронні мережі обробляють зображення і відео, аналізуючи протоколи гри, щоб виявити найважливіші моменти або оцінити активність гравця. Крім того, вони застосовуються в таких сферах, як обробка природної мови в спортивних новинах, текстовий аналіз експертних коментарів або навіть постів із соціальних мереж.

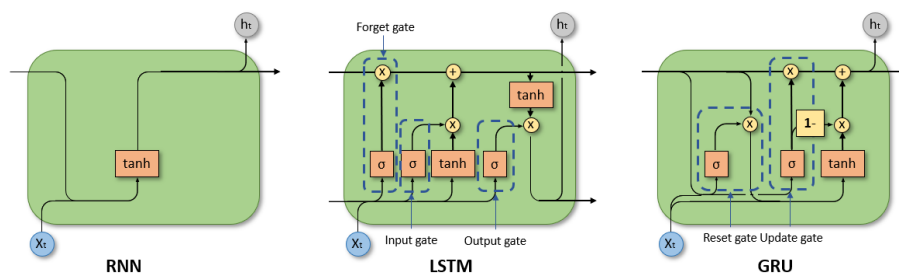


Рисунок 2.1 – Принципи роботи рекурентних нейронних мереж

Основна сила неймереж – це глибоке навчання, яке працює через багатошарові архітектури для аналізу складних і різномірних даних. Все це покликане досягти дуже високого рівня точності прогнозування. Застосовуються алгоритми оптимізації, що мінімізують функцію втрат, такі як градієнтний спуск, екземпляр якого зображено на рисунку 2.2, і його більш прогресивні реалізації, такі як Adam або RMSprop. Для роботи з великими масивами даних, що характеризують діяльність букмекерських контор, застосовуються технології обробки великих даних. Наприклад, Apache Hadoop або Spark можуть ефективно працювати з величезними масивами інформації, що характеризують спортивні події, історію ставок, поведінку клієнтів. Регуляризація моделей, зокрема L1/L2-регуляризація, Dropout або Batch Normalization, запобігає перенастроюванню, а отже, підвищує загальну стабільність в роботі [10].

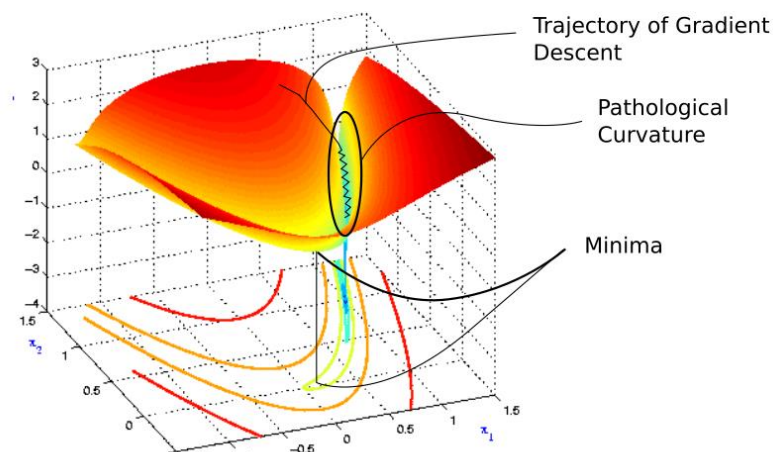


Рисунок 2.2 – Оптимізація мінімізуюча функцію втрат

Крім того, нейронні мережі використовуються для виявлення аномалій у даних, які можуть свідчити про шахрайські дії або аномальні ставки. Серед інших прикладних алгоритмів є ізоляційний ліс і автокодери. Завдяки інтеграції з API спортивних порталів, таких як Opta Sports або

Sportradar, букмекери отримують дані в режимі реального часу, що дозволяє їм оперативно оновлювати коефіцієнти та коригувати стратегії. Наприклад, при прогнозуванні результату матчу нейромережі аналізують історичні дані переглядів користувачів, приклад фільтрації нейромережою зображено на рисунку 2.3. Ще одне застосування нейромереж – створення інтелектуальних чат-ботів для обслуговування клієнтів. Вони здатні відповідати на запити користувачів, давати рекомендації та аналізувати відгуки користувачів, оскільки мають можливість обробки природної мови. Загалом використання нейронних мереж у букмекерських конторах дає змогу підвищити точність прогнозів, автоматизувати процеси, оптимізувати управління ризиками та покращити користувацький досвід. Однак для того, щоб усе це стало можливим, потрібні величезні ресурси, сучасне обладнання та кваліфіковані фахівці, що створює додаткові виклики для цих технологій.

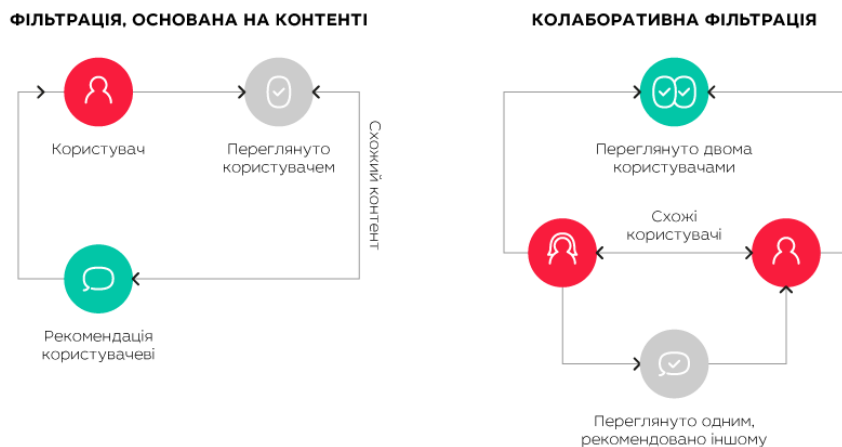


Рисунок 2.3 – Фільтрація нейромережою опираючись на контенті

Градiєнтний бустiнг є одним з найефективнiших методiв машинного навчання, який використовується у сферi беттiнг-прогнозування. Можливість побудови ансамблiв моделей, в яких кожна наступна модель компенсує помилки, допущенi попередньою, дозволяє будувати точнi прогнози.

У результаті їхні моделі краще передбачають ймовірність перемоги команди або гравця, кількість забитих голів, виграних сетів або інші ключові метрики. Одна з головних речей, яка робить градієнтний бустінг особливим.

Наприклад, робота з відсутніми даними може бути ефективно виконана за допомогою моделей на основі XGBoost, примітивний вигляд роботи якого зображено на рисунку 2.4, LightGBM або CatBoost.

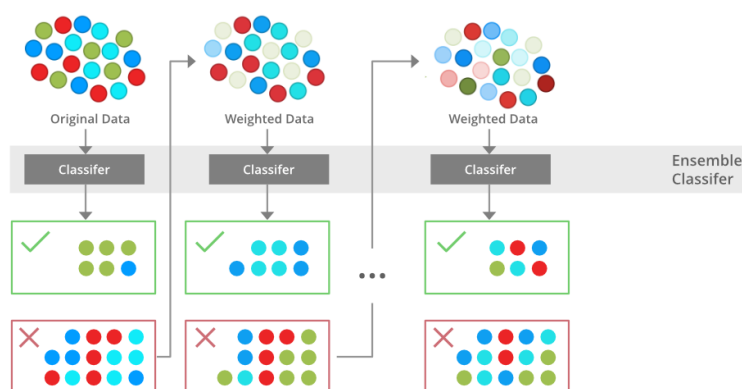


Рисунок 2.4 – Принцип роботи моделі на основі XGBoost

Випадкові ліси, примітивний принцип роботи якого показано на рисунку 2.5 також відіграють важливу роль у сфері беттінг-прогнозування, оскільки забезпечують надійність і стійкість до перенавчання. Цей метод поєднує велику кількість дерев рішень, кожне з яких навчається на різних підмножинах даних і характеристик. Зрештою, результати узагальнюються шляхом голосування або усереднення. Це дозволяє будувати моделі, які добре працюють навіть за наявності шуму в даних або неповної інформації.

У беттінг-індустрії випадкові ліси часто використовують для визначення значущості таких факторів, як перевага свого поля, кількість зіграних матчів поспіль або форма гравців.

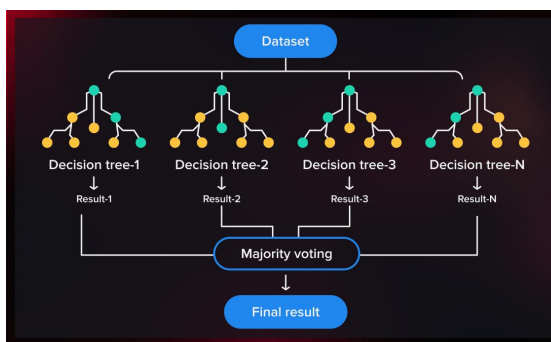


Рисунок 2.5 – Принцип роботи Random Forest

Глибоке навчання, з іншого боку, є ще більш потужним інструментом, який можна використовувати для складних, неоднорідних і багатовимірних даних.

У сучасній букмекерській сфері активно використовуються передові методи машинного навчання для підвищення точності прогнозування, зниження ризиків і оптимізації операційних процесів. Згорткові нейронні мережі (CNN) застосовуються для аналізу відеозаписів матчів з метою визначення ключових моментів гри. Вони здатні автоматично виявляти події, такі як забиті голи, помилки гравців, стратегічні зміни в розташуванні команд на полі та емоції учасників матчу. Завдяки обробці відео CNN забезпечують глибше розуміння динаміки гри, що допомагає створювати більш точні прогнози.

Рекурентні нейронні мережі (RNN), а також їх сучасні модифікації, такі як LSTM (Long Short-Term Memory) і GRU (Gated Recurrent Unit), використовуються для моделювання часових рядів. Ці методи особливо корисні для аналізу динаміки змін у формі команд, прогнозування їхньої продуктивності на основі результатів останніх матчів і моделювання змін коефіцієнтів у реальному часі. Наприклад, вони можуть враховувати вплив безперервних програшів або перемог, а також прогнозувати, як ці тренди зміняться в наступному матчі.

Для обробки текстової інформації, такої як новини, коментарі експертів і дописи в соціальних мережах, застосовуються моделі на основі трансформерів, такі як BERT (Bidirectional Encoder Representations from Transformers), базову архітектуру якого зображено на рисунку 2.6, або GPT (Generative Pre-trained Transformer). Ці моделі здатні витягувати контекстуальну інформацію з текстів, що дозволяє враховувати вплив зовнішніх факторів, наприклад, зміни в складі команди через травми або стратегічні рішення тренера. Завдяки цьому букмекери можуть швидко реагувати на нову інформацію, оновлюючи коефіцієнти ставок у режимі реального часу.

Поєднання згорткових і рекурентних нейронних мереж, а також трансформерів, дозволяє створювати багатошарові системи, що враховують як візуальні, так і текстові дані, а також динамічні зміни у грі. Це забезпечує не лише точніше прогнозування, але й адаптацію до нових даних, що надходять у режимі реального часу. Завдяки цим інноваціям букмекери можуть створювати коефіцієнти, максимально наближені до реальних ймовірностей, що підвищує їхню конкурентоспроможність на ринку та дозволяє оптимізувати управління ризиками. Усе це формує основу для нової ери аналітики в індустрії ставок.

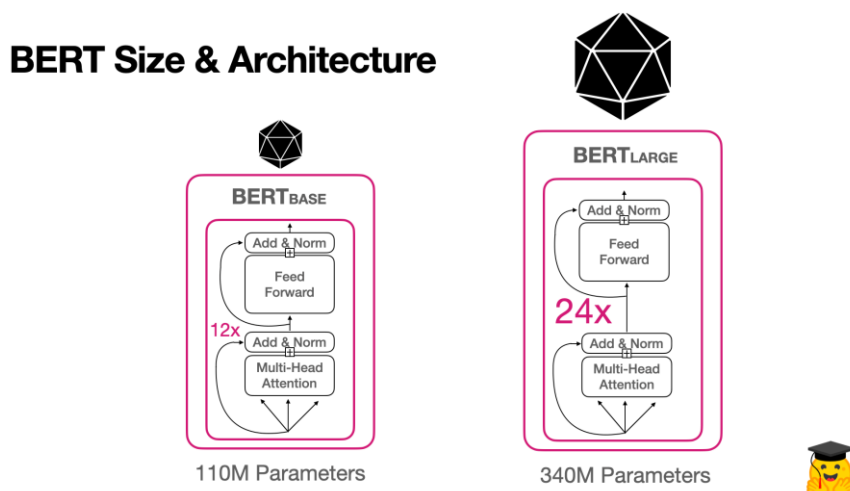


Рисунок 2.6 – Архітектура BERT

2.1.2 Предиктивна аналітика

Предиктивна аналітика в букмекерській сфері – це багатофункціональний підхід, який поєднує аналіз великих даних, алгоритми машинного навчання, математичні моделі та методи прогнозування для визначення ймовірностей результатів спортивних подій, оцінки ризиків і моделювання сценаріїв. Завдяки використанню передових технологій та інструментів вона значно покращує точність прогнозів, оптимізує управління ризиками та автоматизує бізнес-процеси. Прогнозування результатів спортивних подій є центральною функцією діяльності букмекерських контор, що базується на аналізі історичних даних, поточних статистичних показників та умов проведення подій. Це робиться за допомогою алгоритмів, що враховують взаємозв'язки між великою кількістю змінних, таких як історія зустрічей даних команд, результати попередніх ігор, форма гравців, атмосферні умови, розклад ігор, травми або дискваліфікація важливих гравців тощо [5].

Спеціалізовані моделі, такі як XGBoost, LightGBM – принцип роботи якого зображено на рисунку 2.7, та рекурентні нейронні мережі, дозволяють аналізувати динамічну зміну даних з часовими рядами в командній формі, враховуючи тенденції та підлаштовуючись під нові дані автоматично в режимі реального часу.

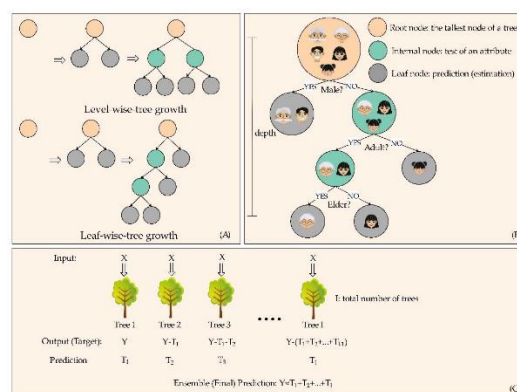


Рисунок 2.7 – Архітектура BERT

Оцінки ймовірностей відіграють важливу роль у роботі букмекера, оскільки точність оцінки ймовірностей впливає на точність встановлення коефіцієнтів. Методами таких оцінок є класифікація, мережі Байєса, приклад яких зображено на рисунку 2.8 та логістична регресія, які допомагають моделювати залежності між ключовими факторами. Наприклад, при визначенні ймовірності перемоги враховуються такі аспекти, як вплив домашнього поля, результати попередніх ігор, психологічний стан команди. Дані в режимі реального часу, наприклад, раптові зміни в складі або інформація про погодні умови, оперативно вносяться в модель, щоб зробити прогнози завжди актуальними [2].

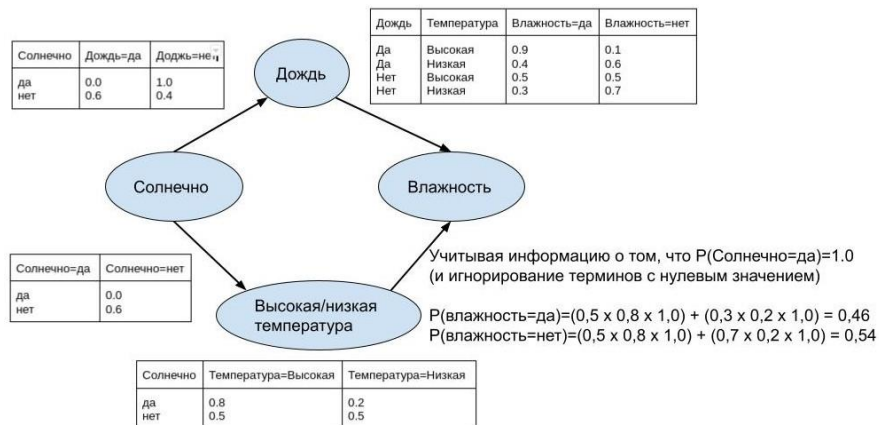


Рисунок 2.8 – Приклад Баєсівської мережі довіри

Безумовно, в основі аналітики лежить розрахунок ризиків, який мінімізує потенційні фінансові втрати букмекерської компанії. Це передбачає використання симуляцій Монте-Карло, приклад симуляцій яких зображено на рисунку 2.9, алгоритмів виявлення аномалій і моделей управління ризиками. Наприклад, система здатна аналізувати великі ставки на несподівані результати або нетипові дії, що свідчать про використання інсайдерської інформації [5].

Автоматизовані системи зміни коефіцієнтів для ставок дають змогу досить оперативно реагувати на зміну умов і ринкових факторів, щоб зберегти баланс між очікуваними виплатами та прибутком компанії.

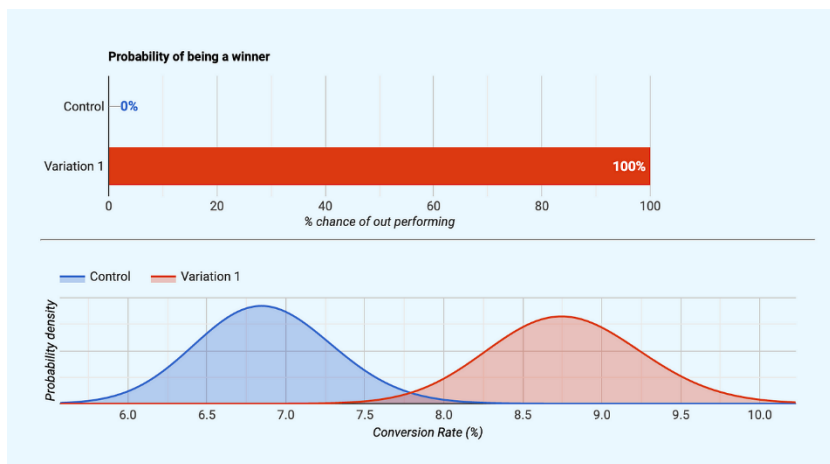


Рисунок 2.9 – Шанси тестової групи на перемоги за Монте-Карло

Моделювання сценаріїв у букмекерській аналітиці дозволяє передбачати, як можуть розвиватися події залежно від різних умов. Для цього застосовуються агентні симуляції, марковські процеси прийняття рішень (MDP), які зображені на рисунку 2.10 і динамічні системи. Наприклад, система може моделювати сценарії, в яких одна команда грає агресивно, а інша зосереджується на обороні, і визначати ймовірності різних результатів гри за таких умов. Це допомагає букмекерам передбачати складні взаємозв'язки між тактикою, стратегією гри та умовами на полі [12].

Букмекер інтегрується в бізнес-процеси і пропонує клієнту автоматизацію управління коефіцієнтами, ризик-менеджмент та персоналізовані пропозиції.

Завдяки хмарним обчислювальним платформам і технологіям обробки великих даних можна аналізувати терабайти інформації в режимі реального часу, динамічно змінюючи прогнози відповідно до нових даних, що надходять.

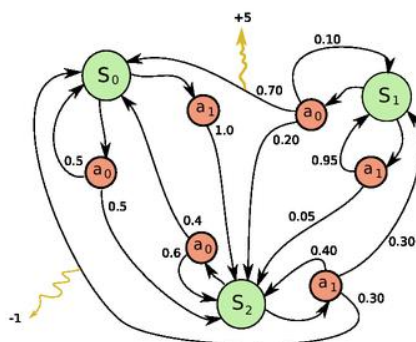


Рисунок 2.10 – Марковський процес вирішування

Завдяки хмарним обчислювальним платформам і технологіям обробки великих даних можна аналізувати терабайти інформації в режимі реального часу, динамічно змінюючи прогнози відповідно до нових даних, що надходять.

Дуже високий рівень автоматизації мінімізує людський фактор, збільшує швидкість прийняття рішень і знижує операційні витрати. Предиктивна аналітика трансформує букмекерську сферу: підвищує точність прогнозів, оптимізує управління ризиками та покращує якість обслуговування клієнтів, надаючи їй конкурентну перевагу на динамічному ринку [1].

У сфері букмекерської аналітики впроваджуються інноваційні методи та технології, які дозволяють досягати ще більшої точності прогнозів та ефективності управління ризиками.

Одним із таких підходів є Hybrid Modeling, який поєднує класичні статистичні моделі (наприклад, ARIMA для аналізу часових рядів, яка зображена на рисунку 2.11) із сучасними глибокими нейронними мережами, такими як LSTM або Transformers.

Це забезпечує високу адаптивність до мінливих ринкових умов і здатність враховувати нелінійні залежності в даних.

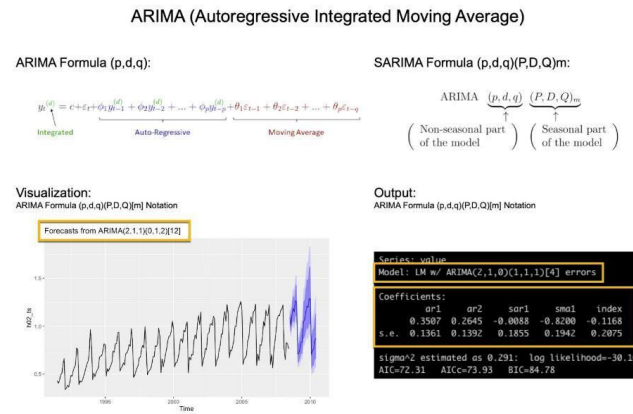


Рисунок 2.11 – Принцип аналізу часових рядів з ARIMA

Ще однією перевагою Federated Learning є можливість навчати модель машинного навчання децентралізовано, що не вимагає локалізації для того, щоб централізувати передачу локальних даних на центральному сервері. Воно також може зберігати дані клієнта приватними, дозволяючи при цьому агрегувати глобальні знання для глобального підвищення точності прогнозів [11].

Все більшої популярності набуває пояснювальний ШІ (Explainable AI, XAI). До таких методів, які застосовуються для моделювання результатів у букмекерській сфері, належать SHAP (Shapley Additive Explanations), візуалізацію якого зображено на рисунку 2.12 і LIME (Local Interpretable Model-agnostic Explanations), які допомагають оцінити внесок кожного з факторів у підсумковий прогноз і сприяють побудові прозорих і зрозумілих коефіцієнтів.

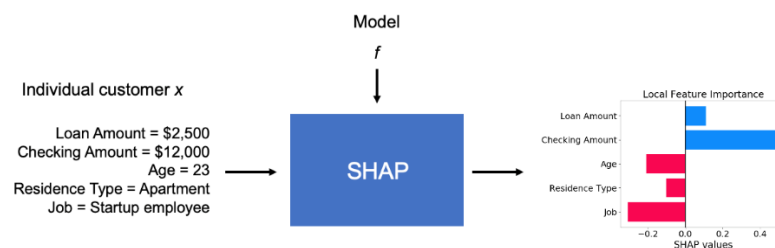


Рисунок 2.12 – Принцип роботи SHAP

Цей метод, стохастична градієнтна динаміка Ланжевена (SGLD), враховує аспекти невизначеності в даних і оцінює можливі відхилення в результатах, тому дає кращу оцінку ймовірності для екстремальних ризиків, таких як великі неочікувані ставки [11].

Виходячи з цього, у питаннях обчислювальної продуктивності значною мірою вирішують проблеми квантування моделі – зменшення розміру моделі без втрати її точності – та периферійних обчислень, які здійснюють обробку даних ближче до джерел, наприклад, на серверах локальної букмекерської платформи.

Ще один тип моделей прогнозування – нейронні мережі з розширенням пам'яті (MANN) [13], [16]. Вони можуть запам'ятовувати необхідну інформацію з минулих матчів і використовувати її для складання прогнозів, навіть якщо вона відсутня в поточному наборі даних.

У сфері обробки даних з'являються нові техніки, такі як Data Augmentation, простий приклад якого зображено на рисунку 2.13 для штучного збільшення обсягів навчальних даних, наприклад, шляхом створення. Це відмінні риси побудови ефективної, конкурентоспроможної букмекерської платформи: висока точність, адаптивність і масштабованість рішень.

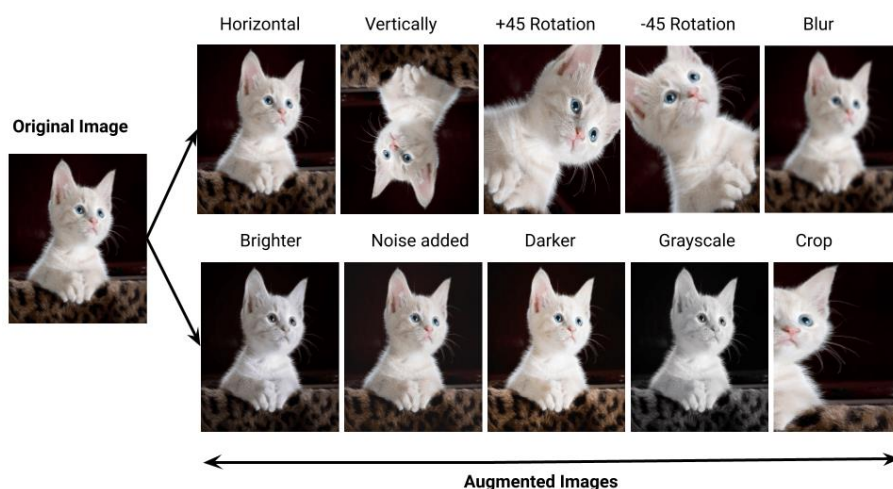


Рисунок 2.13 – Принцип Data Augmentation

2.2 Практичне застосування алгоритмів

Рішення про встановлення фундаментальних коефіцієнтів у букмекерській діяльності є досить багатокомпонентним і заплутаним питанням [17]. Сама концепція узагальнює аналіз історичних даних, оцінку поточних факторів, математичне моделювання, використання новітніх технологій і постійну перевірку результатів. Ця процедура, по суті, робить точні коефіцієнти для букмекерських контор, що відображають ймовірність настання подій і збалансований розподіл ризиків.

Першим етапом є збір та аналіз історичної статистики: дані про попередні ігри досліджуваних команд; результати ігор за участю конкретних гравців; дані про суддівство, а також статистика поведінки на стадіонах. Ці дані використовуються для розрахунку рахунку гри. При цьому враховується форма провідних гравців, психологічний стан команди, наявність мікротравм, тактичні схеми, ротація складів і навіть активність трансферів. Всі ці фактори враховують особливості кожної конкретної гри.

Для розрахунку ймовірностей використовуються складні математичні моделі. Розподіл Пуассона, вигляд якого можна побачити схематично на рисунку 2.14, використовується для прогнозування кількості голів, ланцюги Маркова дозволяють моделювати хід гри, регресійні моделі оцінюють загальну результативність, а байєсівські мережі забезпечують комплексний аналіз залежностей між змінними. Цей процес доповнюється алгоритмами машинного навчання та глибокими нейронними мережами, які враховують складні взаємозв'язки та тенденції, які важко виявити за допомогою традиційних підходів.

Розглянемо Пуассонівський розподіл, який використовується для прогнозування кількості голів у футбольних матчах, базуючись на середній кількості голів, які команда забиває за матч. На основі цього розподілу оцінюється ймовірність того, що команда заб'є задану кількість голів,

припускаючи, що події є незалежними і відбуваються з постійною середньою частотою.

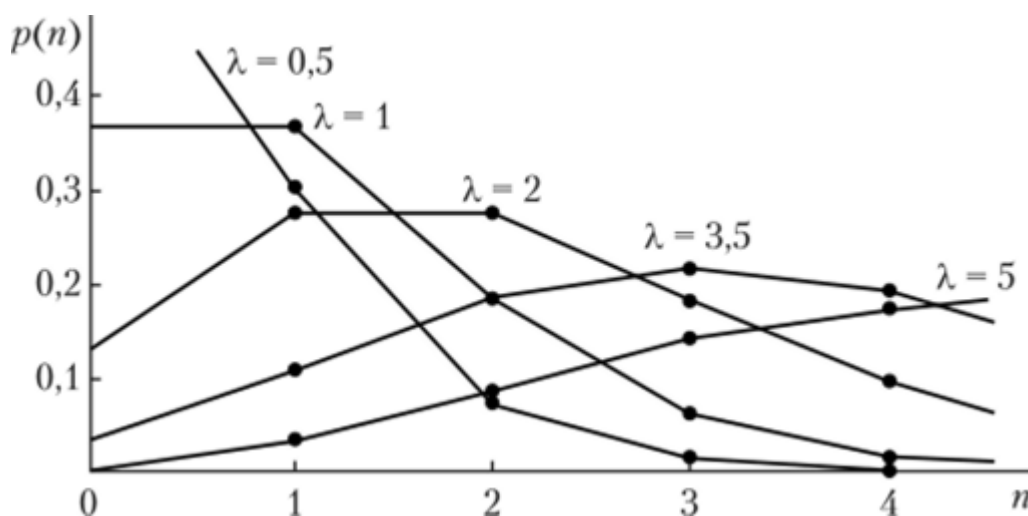


Рисунок 2.14 – Приклад розподілу Пуассона

Це обчислюється. Потім ці оцінки слід скоригувати на такі фактори, як домашня чи виїзна гра. Для прогнозування результату гри застосовується розподіл Пуассона для кожної команди, де обчислюється ймовірність того, що кожна команда заб'є k голів, де k може дорівнювати 0, 1, 2 і більше. Потім вони об'єднуються, щоб отримати шанси на всі можливі результати матчу, наприклад, перемогу, нічию або поразку.

Хоча пуассонівський розподіл є потужним інструментом, він все ще має деякі недоліки, включаючи припущення, що події є незалежними, а ймовірність постійною, що може бути порушено в реальних умовах – наприклад, коли команда змінює свою стратегію гри в ході змагання або через деякі зовнішні фактори. Однак саме ця модель є досить популярною завдяки своїй простоті та ефективності у вирішенні завдань прогнозування забитих голів.

У процесі розрахунку враховуються різні фактори, що впливають на коефіцієнти: турнірна мотивація, щільність календаря матчів, географія поїздок команд, погодні умови, тип покриття поля тощо.

Перед публікацією коефіцієнти проходять ретельну валідацію. Вони перевіряються на відповідність історичним трендам, порівнюються з ринковими показниками, оцінюються експертами-аналітиками, тестуються на історичних даних, піддаються стрес-тестуванню в екстремальних сценаріях. Цей етап забезпечує їхню точність і надійність.

Технологічна реалізація процесу базується на використанні новітніх систем обробки даних, які включають Big Data, алгоритми машинного навчання, нейронні мережі, хмарні обчислення та системи оновлення даних у режимі реального часу. Також застосовуються інші інноваційні методи, такі як комп'ютерний зір при аналізі відеозаписів, аналіз настроїв у соціальних мережах, даних з натільних пристроїв, біометрії та психологічних факторів.

Це є безперервним процесом, який потребує постійного методологічного вдосконалення. Неприпустимість неточностей, необхідність використовувати новинки, проводити глибокий аналіз і адаптуватися до мінливих ринкових умов показує те, що забезпечує точність прогнозів і ефективність букмекерської діяльності, надає компаніям конкурентні переваги.

3 ПРОЕКТУВАННЯ ТА РОЗРОБКА СИСТЕМИ ПРОГНОЗУВАННЯ

3.1 Проектування системної архітектури застосунку

Проектування системної архітектури застосунку для прогнозування з використанням даних від постачальників включає створення багатошарової, масштабованої та модульної системи, яка забезпечує збір, уніфікацію, обробку даних та інтеграцію з фронтенд-запитами через бекенд.

На вході системи працює Актор, який представляє кінцевого користувача, наприклад, адміністратора або аналітика, що взаємодіє із системою через зовнішній інтерфейс. Основним завданням цього компонента є надсилання запитів для отримання прогнозів або аналізу статистичних даних.

Далі запит передається до Gambling Control System, як показано на рисунку 3.1, яка виконує функцію централізованого управління всіма операціями в системі. Ця підсистема відповідає за контроль активності користувачів, управління ризиками, аналіз трендів і забезпечення відповідності запитів встановленим правилам. Activity Module, який є частиною цієї системи, відстежує дії користувачів і взаємодіє з алгоритмами прогнозування для виконання задач у реальному часі.

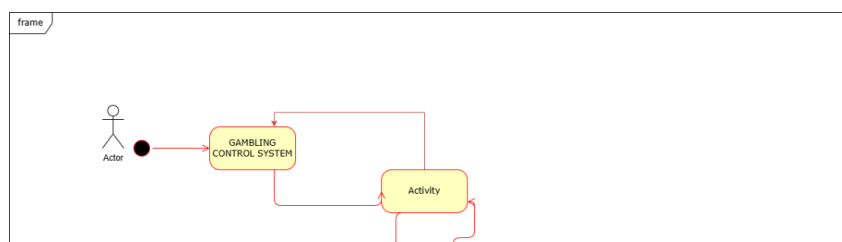


Рисунок 3.1 – Вигляд частини архітектури взаємодії з користувачем

На наступному етапі запит обробляється через Request Kernel, яке виконує функцію маршрутизації між компонентами системи. Воно передає

запити до Prediction Algo, центрального модуля для прогнозування, і взаємодіє із System Kernel, яке забезпечує доступ до необхідних системних ресурсів та історичних даних. Prediction Algo використовує алгоритми, такі як LSTM, розподіл Пуассона, градієнтний бустинг та Монте-Карло, щоб здійснити прогнозування на основі отриманих даних. Результати обробки повертаються в системне ядро для подальшого використання.

Основна обробка даних виконується в межах Sport Processing Module, який представлено на рисунку 3.2, який складається з кількох підсистем. Спочатку дані отримуються від зовнішніх API постачальників і передаються до Data Preprocessing, де вони нормалізуються, перевіряються на повноту, коректність і стандартизуються за допомогою системи повідомлень, наприклад, Kafka. Уніфіковані дані зберігаються у History Saving, який забезпечує їхнє довготривале збереження у базі даних (DB). Цей модуль також обробляє нові дані в реальному часі, гарантуючи їхню актуальність для прогнозування.

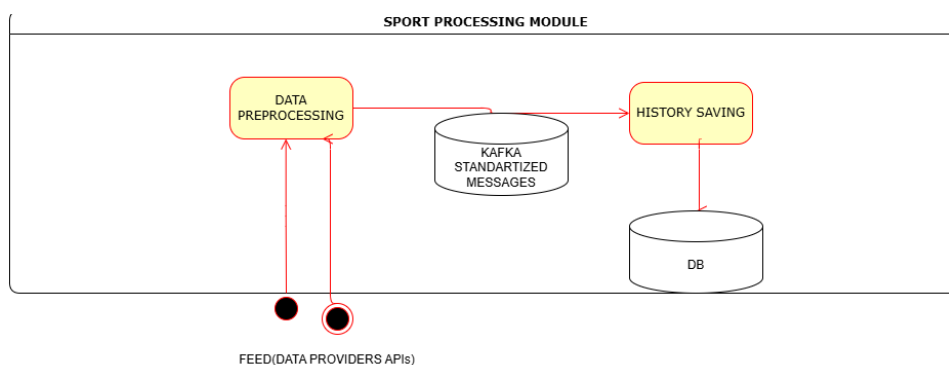


Рисунок 3.2 – Вигляд модуля системи обробки спортивних подій

Системне ядро, або System Kernel, забезпечує інтеграцію між модулями обробки даних, алгоритмами прогнозування та компонентами, які обслуговують запити користувачів. Цей підхід дозволяє створити високоінтегровану, масштабовану та адаптивну систему для прогнозування. Уся архітектура, загальний вигляд якої зображено на рисунку 3.3

побудована так, щоб забезпечити швидке виконання запитів, точність прогнозів і інтерактивність у реальному часі через взаємодію фронтенда та бекенда. Завдяки такій модульній структурі система може легко адаптуватися до нових вимог і змін у джерелах даних, що забезпечує її ефективність і гнучкість.

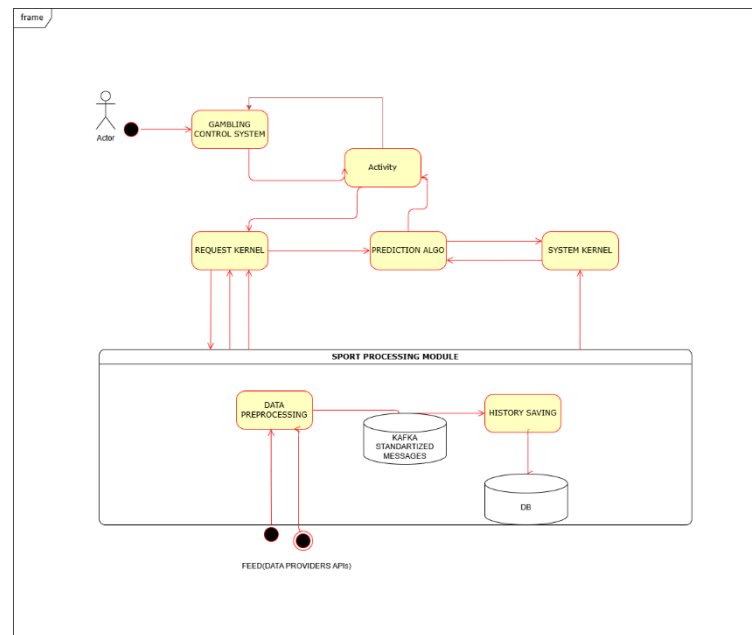


Рисунок 3.3 – Загальний вигляд архітектури

3.2 Проектування алгоритмічної та доменної логіки застосунку

3.2.1 Проектування методів отримання даних статистики

Створення Quantum Adapter як частини Sport Processing Module є ключовим етапом у забезпеченні ефективного підключення системи. Такий адаптер виконує функцію посередника між системою і API, стандартизуючи всі запити та відповіді. Завдяки цьому забезпечується узгодженість і адаптивність у роботі системи навіть за умов змін у структурі API або додавання нових джерел даних. Quantum Adapter перетворює вхідні та вихідні дані в уніфікований формат, що відповідає вимогам системи, і

дозволяє легко масштабуватися та інтегрувати нові компоненти. Це особливо важливо для забезпечення надійності роботи, адже адаптер мінімізує потребу у зміні основного коду під час оновлення API або впровадження нових джерел.

Основний функціонал Quantum Adapter включає методи для отримання та обробки даних. Наприклад, метод `GetQuantumEventsList` реалізує HTTP GET-запит до API провайдера, код якого зображено у лістингу 3.1, що дозволяє завантажувати список подій для конкретної дати. Отримані дані обробляються, десеріалізуючись у JSON-формат у стандартний об'єкт – `DecimalEvent`, який відповідає внутрішнім вимогам системи. Завдяки цьому забезпечується високий рівень узгодженості даних і легкість інтеграції з іншими модулями, зберігаючи при цьому структурну гнучкість та готовність до змін.

Лістинг 3.1 – Програмний код методу `GetQuantumEventsList`

```
public async Task<DecimalEvent[]> GetQuantumEvents()
{
    _logger.LogInformation("GetQuantumEvents");
    var response = await _client.GetAsync("match/list");
    response.EnsureSuccessStatusCode();
    var result = await response.Content.ReadAsStringAsync();
    var root = JObject.Parse(result);
    var feedEvents = root["matches"]?.Children()
        .Select(Convert).Where(m => m != null).ToArray();
    var res = feedEvents ?? Array.Empty<DecimalEvent>();
    if (!res.Any())
    {
        _logger.LogInformation("Feed return empty events
list");
    }
    return res;
}

private DecimalEvent Convert(JToken value)
```

Продовження лістингу 3.1

```

{
    try
    {
        return value.ToObject<DecimalEvent>();
    }
    catch (Exception e)
    {
        _logger.LogError(e, e.Message);
    }
    return null;
}

```

Також розроблено умовний kernel, основін сигнатури методів яких зображено на рисунку 3.4 для препроцесінгу даних отриманих від фіда, частину формату даних зображено на рисунку 3.5.

```

public class DataPreprocessing
{
    public string StandardizeData(string rawData){...}

    public DataTable AggregateToDataTable(List<Dictionary<string, object>> rawData){...}

    public void SaveToExcel(DataTable dataTable, string filePath){...}
}

public class HistorySaving
{
    private readonly List<string> _history = new();

    public void Save(string data){...}

    public void DisplayHistory(){...}
}

public class DataFetchingModule{...}

```

Рисунок 3.4 – Основні сигнатури методів Sport Processing Module

Кожна подія включає розділ Statistics, який містить детальні параметри продуктивності учасників. Для футбольних матчів, наприклад, статистика включає відсоток перемог WinRate, середню кількість забитих

голів за гру `AverageGoals` і кількість ударів у ворота `ShotsOnTarget`. У тенісних подіях статистичні параметри для кожного гравця включають відсоток виграних матчів `WinRate`, середню кількість ейсів за гру `AverageAces` і кількість подвійних помилок `DoubleFaults`. У баскетбольних матчах статистика кожної команди включає відсоток перемог `WinRate`, середню кількість очок за гру `AveragePoints` і кількість підбирань `Rebounds`.

Цей JSON-формат структурований таким чином, щоб забезпечити гнучкість і зручність обробки даних. Він включає як базову інформацію про подію, так і ключові статистичні показники, що дозволяють проводити аналіз ефективності команд або гравців.

```
    "Statistics": {  
      }  
    },  
    {  
      "EventID": "23242321",  
      "EventName": "Borysov D. vs Voitenko M.",  
      "StartTime": "2025-01-16T18:30:00Z",  
      "Sport": "Tennis",  
      "League": "ATP",  
      "Venue": "Sydney ar",  
      "Statistics": {  
        "TeamC": {  
          "WinRate": 58.7,  
          "AveragePoints": 102.5,  
          "Rebounds": 45  
        },  
        "TeamD": {  
          "WinRate": 49.3,  
          "AveragePoints": 98.2,  
          "Rebounds": 42  
        }  
      }  
    }  
  }  
}
```

Рисунок 3.5 – Частина сирих даних від постачальника статистики

3.2.2 Проектування алгоритмічних скриптів для прогнозування

Планування алгоритмів скриптів для прогнозування тенісних матчів з використанням розподілу Пуассона є досить складним завданням, і

необхідне детальне дослідження характеру гри. У центрі уваги – індивідуальна гра кожного з тенісистів, покриття корту, поточна форма гравця, історія [1].

Наприклад, скільки в середньому вигравав гравець А за останні сезони на ґрунті, і аналогічно для гравця Б? Причина: історичні дані про кількість виграних і програних ігор гравцями на певному типі покриття – глиняному, трав'яному чи хардовому. Ці дані отримуються після агрегації в рамках Sport Processing Module та дають змогу коректно відпрацювати алгоритму та показати всю статистичну інформацію потрібну для прогнозування матчів для букмекерських контор. Перейдемо до проектування алгоритмічного модуля який спочатку буде вираховувати ймовірності по Пуассону, відобразимо це в лістингу 3.2.

Лістинг 3.2 – Реалізація визначення ймовірностей по Пуассону

```

columns_to_process = ['Winner', 'Loser', 'AvgW', 'AvgL']
processed_data = data[columns_to_process].copy()
processed_data.rename(columns={'Winner': 'Player', 'AvgW':
'AvgWin', 'Loser': 'Opponent', 'AvgL': 'AvgLoss'},
inplace=True)

player_statistics = processed_data.groupby('Player').agg({
'AvgWin': 'mean', 'AvgLoss': 'mean' }).reset_index()
player_statistics.head()

def calculate_poisson_probability(average_rate, events):
    return (np.exp(-average_rate) * average_rate**events)
/ np.math.factorial(events)

def simulate_match(player_one, player_two,
player_statistics, max_goals=5):
    stats_player_one =
player_statistics[player_statistics['Player'] == player_one]
    stats_player_two =
player_statistics[player_statistics['Player'] == player_two]
```

Продовження лістингу 3.2

```

    if stats_player_one.empty or stats_player_two.empty:
        return f"Statistics not available for {player_one}
or {player_two}."
    average_goals_player_one =
stats_player_one['AvgWin'].values[0]
    average_goals_player_two =
stats_player_two['AvgLoss'].values[0]
    goal_probabilities = {}
    for goals_one in range(max_goals + 1):
        for goals_two in range(max_goals + 1):
            probability =
(calculate_poisson_probability(average_goals_player_one,
goals_one) *
calculate_poisson_probability(average_goals_player_two,
goals_two))
            goal_probabilities[(goals_one, goals_two)] =
probability
    return {k: v for k, v in
sorted(goal_probabilities.items(), key=lambda x: -x[1])}

```

Перейдемо до проектування метрик прогнозування та відображення статистичної та прогностичної інформації для користувача по його вподобанням.

Окремо від алгоритмів обчислення, у системі буде реалізована функція відображення ймовірності кожного можливого результату матчу між гравцями у вигляді відсотків, результат зображено на рисунку 3.6. Такий підхід забезпечує наочне уявлення про розподіл ймовірностей для різних сценаріїв розвитку подій. Візуалізація буде виконана у зручному графічному форматі, що дозволить користувачеві швидко оцінити найбільш імовірний результат, а також зіставити різні варіанти між собою.

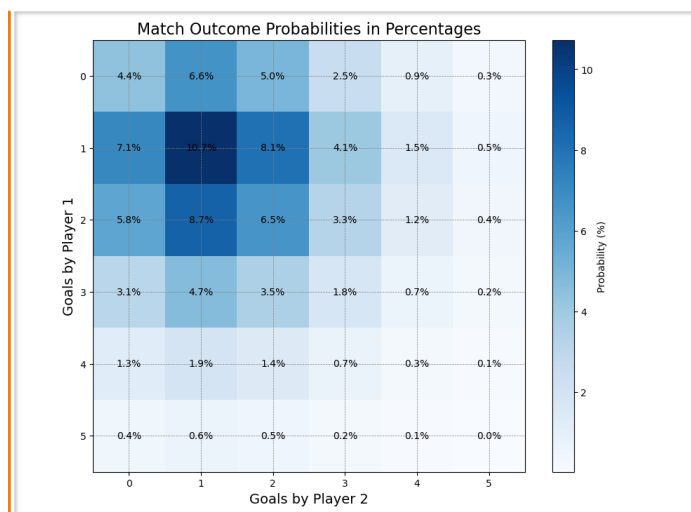


Рисунок 3.6 – Ймовірності результатів між гравцями у розрізі матчів

Додамо метрику по топ гравцям, результат зобразимо на рисунку 3.7, за середнім коефіцієнтом виграшу (AvgWin), що покаже нам, які гравці мають найвищу середню ймовірність виграшу, що корисно для оцінки сильних гравців.

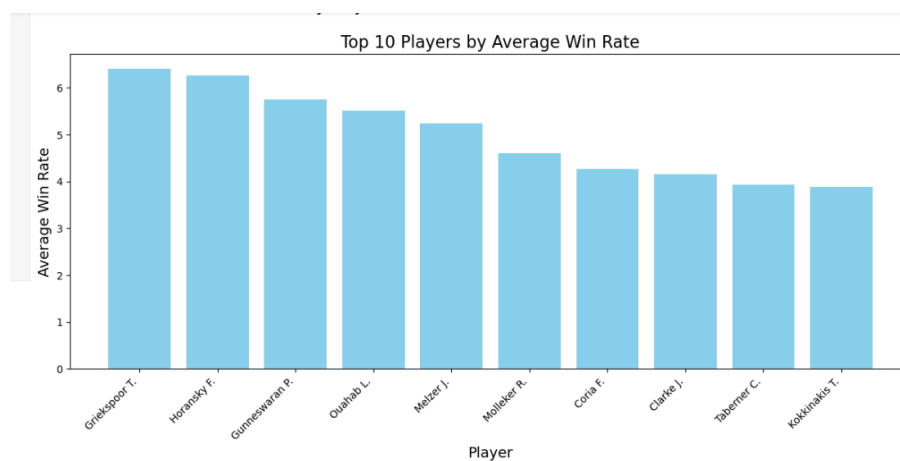


Рисунок 3.7 – Відображення топ-10 гравців по Win Rate

Додамо також дані про залежність між середніми коефіцієнтами виграшу та програшу для всіх гравців, що дозволяє виявляти тренди чи аномалії. Також додамо дві нові метрики, які будуть відображати

відношення вигравів до програшів гравців та метрику консистентності. Частина коду представлено у лістингу 3.3 та відобразимо результати даних метрик, що видно на рисунку 3.8.

Лістинг 3.3 – Додавання метрик Consistency and Win/Loss

```
player_stats['Win/Loss Ratio'] = player_stats['AvgWin'] /
player_stats['AvgLoss']
player_stats['Consistency'] = 1 / player_stats['AvgLoss']
```

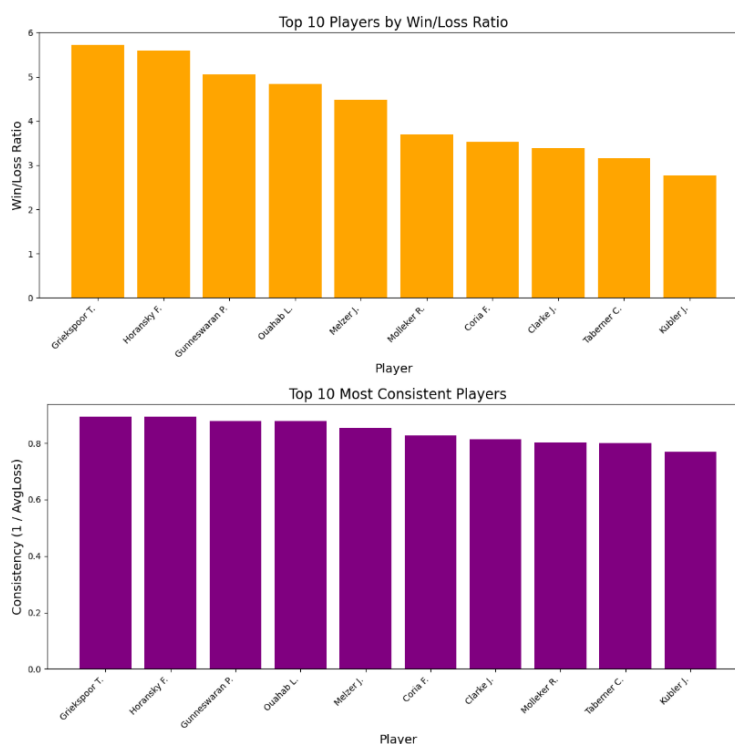


Рисунок 3.8 – Відображення метрик Consistency and Win/Loss

Цільовий стовпець визначає, чи виграв гравець з ім'ям наприклад – Андерсон певний матч. Він перетворюється на двійкову змінну, де 1 – перемога, а 0 – поразка.

Потім клас SimpleImputer буде використано для заміни відсутніх значень у вибраних стовпчиках відповідними значеннями; це гарантує, що модель працюватиме належним чином, навіть якщо в даних є прогалини.

Потім за допомогою функції `train_test_split` вибирається випадкова 70% частина даних, яких було попередньо збережено в модуль історії спортивних подій в системі як навчальна вибірка. Навчальна вибірка тренує модель, а тестова вибірка використовується для оцінки продуктивності.

Для моделювання використовується алгоритм Gradient Boosting, частина коду якого показана у лістингу 3.4, реалізований у класі `GradientBoostingClassifier`. Цей алгоритм комбінує кілька слабких моделей (дерев рішень) для побудови сильної моделі. Параметри моделі включають кількість дерев (`n_estimators=100`), швидкість навчання (`learning_rate=0.1`) та максимальну глибину дерев (`max_depth=3`), що регулює складність моделі.

Лістинг 3.4 – Частина коду з `GradientBoostingClassifier`

```
features = data[['AvgW', 'AvgL']]
labels = (data['Winner'] == 'Anderson K.').astype(int)

imputer = SimpleImputer(strategy='mean')
features_imputed = imputer.fit_transform(features)

X_train, X_test, y_train, y_test =
train_test_split(features_imputed, labels, test_size=0.3,
random_state=42)

gb_model = GradientBoostingClassifier(n_estimators=100,
learning_rate=0.1, max_depth=3, random_state=42)
gb_model.fit(X_train, y_train)

predictions = gb_model.predict(X_test)
```

Після того, як модель навчена на навчальних даних, тестові дані використовуються для прогнозування; результат зберігається у змінній `predicted_labels`.

Точність моделі обчислюється як частка правильних прогнозів для всіх тестових вибірок за метрикою `accuracy_score`, результат якої показано на рисунку 3.9. Результат у цьому випадку зберігається у змінній `model_accuracy`, щоб показати ефективність побудованої моделі. Процедура дозволяє отримати або апроксимувати модель для класифікації результатів матчів на основі історичних даних для коефіцієнтів, пов'язаних з переможцями та переможеними.

```
Classification Report:
              precision    recall  f1-score   support

Not Anderson K.      0.98      0.99      0.99         775
 Anderson K.         0.00      0.00      0.00          17

   accuracy              0.97         792
  macro avg              0.49         792
 weighted avg              0.96         792
```

Рисунок 3.9 – Classification report по скрипту з градієнтним бустінгом

3.3 Інтеграційне тестування

Інтеграційне тестування в букмекерських конторах для перевірки автоматичних коефіцієнтів є важливим етапом, який забезпечує їх точність і відповідність очікуванням перед публікацією. Уявімо ситуацію, коли трейдер, відповідальний за управління коефіцієнтами, хоче протестувати автоматизовану систему прогнозування перед тим, як виставити коефіцієнти для клієнтів. Для цього він використовує внутрішню систему запитів, яка надсилає результати до Telegram-бота, частину програмної реалізації якої показано у лістингу 3.5. Такий підхід дозволяє трейдеру

швидко отримати необхідну інформацію у зручному форматі, оцінити її та прийняти рішення про подальші дії.

Лістинг 3.5 – Частина коду надсилання результатів до Telegram TennisPredictionBot

```
[HttpPost("send-to-telegram")]
public async Task<ActionResult> SendToTelegram([FromBody]
    TelegramRequest telegramRequest)
{
    string telegramApiUrl =
    $"https://api.telegram.org/bot{telegramRequest.BotToken}/sendM
    essage";
    var payload = new
    {
        chat_id = telegramRequest.ChatId,
        text = telegramRequest.Message
    };

    var response = await _httpClient.PostAsync(telegramApiUrl,
    content);

    if (response.IsSuccessStatusCode)
    {
        return Ok(new { success = true, message = "Message
    sent to Telegram successfully." });
    }

    return BadRequest(new { success = false, message = "Failed
    to send message to Telegram." });
}
```

Процес починається із формування запиту трейдером через внутрішній інтерфейс, наприклад, веб-додаток або десктопний інструмент. У цьому запиті трейдер зазначає параметри події, такі як назви команд, дата

та час матчу, історія попередніх зустрічей, результат виконання показано на рисунку 3.10, або додаткові статистичні метрики, наприклад, середню результативність або форму команд. Запит надсилається до автоматизованої системи прогнозування, яка проводить розрахунки на основі заданих параметрів.

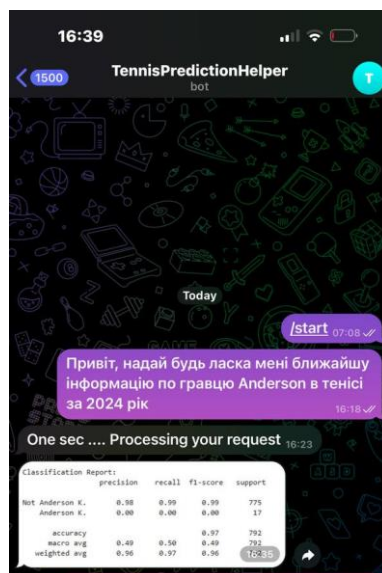


Рисунок 3.10 – Подання результатів зустрічей гравця Anderson

Система обробляє запит, використовуючи аналітичні моделі, статистичні методи або алгоритми машинного навчання. Дані для розрахунку можуть бути отримані з внутрішньої бази або через API зовнішніх постачальників. Система аналізує отримані дані, розраховує ймовірності результатів події (наприклад, перемога команди А, нічия або перемога команди В) і формує автоматичні коефіцієнти, результат виконання якого показано на рисунку 3.11. Крім того, система може надавати трейдеру додаткову інформацію, наприклад, попередження про можливі аномалії в даних або нестандартні результати.

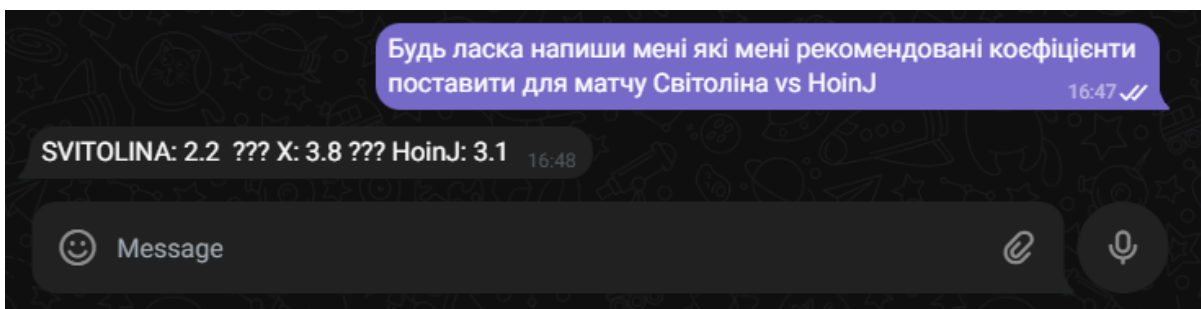


Рисунок 3.11 – Подання базових коефіцієнтів по запиту

Після завершення обчислень система надсилає результати трейдеру через Telegram-бота. У повідомленні міститься вся ключова інформація, необхідна для оцінки події: ймовірності результатів, розраховані коефіцієнти, статистика команд (середня кількість голів, форма за останні матчі тощо) та інші релевантні дані. Також повідомлення може містити попередження, якщо система виявила підозрілу активність або аномалії.

Трейдер, отримавши повідомлення, аналізує дані та приймає рішення. Якщо результати відповідають очікуванням, він підтверджує їх публікацію. У разі необхідності внесення коригувань трейдер може надати системі зворотний зв'язок, після чого коефіцієнти будуть перераховані, також на рисунку 3.12 можна побачити результат виконання алгоритму по статистиці по гравцям по найкращим Win/Loss графіку.



Рисунок 3.12 – Подання базових коефіцієнтів по запиту

Такий підхід забезпечує високу швидкість і зручність тестування автоматизованої системи. Telegram-бот дозволяє трейдеру отримувати дані в реальному часі, не потребуючи доступу до складних інтерфейсів чи додаткових програм. Це мінімізує ризики публікації некоректних коефіцієнтів, дозволяючи трейдеру приймати обґрунтовані рішення. Інтеграція Telegram-бота з системою також надає гнучкість, оскільки трейдер може швидко задавати параметри запитів і отримувати персоналізовану аналітику.

3.4 Переваги та виклики

Комбінування методів у прогнозуванні спортивних подій дозволяє досягати нових рівнів точності й адаптивності в аналітичних системах, проте воно також ставить перед розробниками складні технічні завдання. Поєднання статистичних моделей, таких як пуассонівський розподіл, з сучасними методами машинного навчання, включаючи нейронні мережі та Gradient Boosting, створює синергію, що відкриває нові можливості для аналізу. Зокрема, статистичні моделі забезпечують інтуїтивно зрозумілі основи прогнозування, тоді як нейронні мережі дозволяють опрацьовувати великі обсяги даних та виявляти складні взаємозв'язки, які важко помітити традиційними методами.

Одним із ключових нововведень, які стають доступними завдяки такому підходу, є можливість динамічної верифікації результатів, уявний чек-лист якої зображено на рисунку 3.13. Це означає, що система може самостійно перевіряти свої прогнози, аналізуючи їх узгодженість між різними моделями [18], [19].

Наприклад, якщо статистична модель прогнозує високі шанси на перемогу однієї команди, але нейронна мережа вказує на протилежний результат через виявлені аномалії у формі гравців або зміну їхньої продуктивності, система може автоматично сигналізувати про необхідність

додаткового аналізу. Це може включати перевірку актуальності даних, аналіз позаштатних факторів, таких як травми, погодні умови або внутрішні зміни в команді, що могли вплинути на прогноз [21].

Такий підхід базується на поєднанні різних методів аналізу даних, що дозволяє не лише забезпечувати більш точні прогнози, але й виявляти потенційно критичні ситуації, які могли залишитися поза увагою стандартних алгоритмів. Виявлення розбіжностей між різними підходами прогнозування стимулює систему до додаткової перевірки та корекції.



Рисунок 3.13 – Чек-лист для перевірки якості даних

Унікальним аспектом комбінування є можливість створення спеціалізованих модулів для кожного етапу аналізу. Наприклад, один модуль може бути призначений для опрацювання часових рядів, що враховують поточну форму команд чи гравців, тоді як інший обробляє історичні дані матчів із використанням багатовимірної регресії. Об'єднання таких модулів у загальній архітектурі дозволяє створити гнучку і масштабовану систему, яка може бути адаптована до будь-якого виду спорту чи навіть до конкретних потреб аналітичного відділу [21].

Важливим технічним викликом у таких системах є складність інтерпретації результатів складних моделей. Наприклад, у нейронних мережах важко зрозуміти, які саме фактори вплинули на прогноз, особливо у випадках з багатошаровими архітектурами, такими як LSTM або трансформери. Для вирішення цього питання розробляються спеціальні алгоритми інтерпретації, які дозволяють візуалізувати важливість окремих вхідних даних для кінцевого результату. Це забезпечує більшу довіру до системи з боку трейдерів та аналітиків, оскільки вони отримують змогу оцінити логіку прогнозу.

Значним викликом залишається потреба в обчислювальних ресурсах для підтримки таких систем у реальному часі. Комбінування моделей потребує паралельної обробки великих масивів даних, що вимагає використання розподілених обчислень у хмарних середовищах. Це особливо актуально для обробки багатопотокових запитів у режимі реального часу, коли система одночасно аналізує кілька матчів або подій [20].

Таким чином, комбінування методів у прогнозуванні спортивних подій є перспективним напрямом, який відкриває нові можливості для аналітичних систем. Проте цей підхід вимагає вирішення низки технічних викликів, таких як забезпечення якісних даних, оптимізація ресурсів і підвищення зрозумілості моделей. Успішне подолання цих бар'єрів дозволить створювати більш надійні, адаптивні та точні системи прогнозування.

ВИСНОВКИ

У результаті виконання дослідження та розробки системи принесли важливі результати, пов'язані з використанням сучасних алгоритмів машинного навчання та проектуванням оптимізованої архітектури системи для прогнозування спортивних подій. Це дозволило додатково перевірити ефективність цих алгоритмів у моделюванні ймовірностей подій шляхом застосування розподілу Пуассона, градієнтного бустінгу, рекурентних нейронних мереж, LSTM, GRU та згорткових нейронних мереж. Ці алгоритми допоможуть врахувати численні статистичні показники та фактори, що залежать від історії даних, які впливають на результат, щоб забезпечити певний рівень точності прогнозів.

Особливе місце в даному дослідженні займає розробка багаторівневої системної архітектури, що забезпечує збір, уніфікацію, обробку та зберігання великих обсягів даних. Модульний підхід в архітектурі дозволяє реалізувати ефективне розділення функціональних елементів системи. Наприклад, модуль збору даних поєднується з API постачальників інформації, який автоматично отримує інформацію про події, команди, гравців та їхню статистику.

Окрему увагу було приділено модулю обчислень, в якому реалізовано основні алгоритми машинного навчання для ймовірнісного аналізу результатів подій. Наприклад, за допомогою розподілу Пуассона можна змоделювати кількість забитих голів під час футбольного матчу, а ефективність градієнтного бустінгу – результат аналізу зв'язків між статистичними показниками, за допомогою якого можна встановити перевагу однієї команди над іншою. Для аналізу часових рядів та прогнозування змін показників у реальному часі використовувалися рекурентні нейронні мережі у вигляді LSTM та GRU.

Це означає взаємодію модулів та організацію обчислень у реальному часі; використані технології обробки Big Data дозволять масштабуватися.

Розроблена архітектура дозволяє оперативно проводити аналіз і прогнозування величезних обсягів даних, що дуже важливо для динамічно зростаючої сфери букмекерського бізнесу.

Результати дослідження підтвердили доцільність впровадження таких систем у сферу прогнозування спортивних подій. Завдяки таким системам букмекерські контори отримують можливість не тільки автоматизувати свої процеси, але й покращити якість обслуговування клієнтів через більш обґрунтовані та конкурентоспроможні коефіцієнти.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Бойко Н., Левицький Б. Алгоритми тренування та оцінки моделей машинного навчання для структурованого набору даних. *Information technology: computer science, software engineering and cyber security*. 2023. № 3. С. 3–12. URL: <https://doi.org/10.32782/it/2023-3-1> (дата звернення: 13.01.2025).
2. Васильєв І. П. Алгоритми та технології управління ризиками у букмекерських системах. *Вісник Національного технічного університету "ХПИ"*. 2023. Т. 12. С. 78–85.
3. Гриневич О. П. Інтеграція алгоритмів класифікації для аналізу спортивних даних. *Український журнал прикладної статистики*. Т. 3, № 2022. С. 47–55.
4. Гупта А. Машинне навчання для великих даних. Алгоритми і приклади. Харків : Фактор, 2018. 480 с.
5. Іванов М. П. Інтеграція машинного навчання у прогнозування спортивних результатів: сучасні підходи. *Український журнал прикладної інформатики*. 2023. Т. 4, № 28. С. 45–51.
6. Міллер Дж. Спортивна аналітика: як статистика змінює гру / пер. з англ Т. Картер. Київ : Наш Формат, 2020. 256 с.
7. Петрина В. В., Дорошенко А. В. Ефективність застосування методів класифікації для задач інтелектуального аналізу великих даних. *Scientific Bulletin of UNFU*. 2024. Т. 34, № 5. С. 119–128. URL: <https://doi.org/10.36930/40340516> (дата звернення: 08.01.2025).
8. Сидоров К. В. Побудова математичних моделей для автоматизації розрахунку коефіцієнтів у букмекерських системах. *Вісник Харківського національного університету радіоелектроніки*. 2023. Т. 2. С. 122–129.
9. Ушакова І. О. Підходи до створення інтелектуальних чатботів. *Системи обробки інформації*. 2019. № 2. С. 76–83.

10. Ходаков М. Особливості впровадження алгоритмів машинного навчання для прогнозування у букмекерській діяльності. *Інформаційні системи та технології*. 5-те вид. 2022. С. 101–108.

11. Щукін Б. М. Інвестиційна діяльність : Метод. посіб. Київ : МАУП, 1998. 63 с.

12. A compound framework for sports results prediction: a football case study / M. Byungho et al. *Knowledge based systems*. 21st ed. 2017. P. 551–562.

13. Artificial intelligence : multidisciplinary perspectives on emerging challenges, opportunities: artificial intelligence AI cognitive computing expert systems machine learning research agenda. Independently Published, 2021.

14. Betradar – leading supplier of sports betting data. *Betradar*. URL: <https://www.betradar.com> (дата звернення: 16.01.2025).

15. Challenges and opportunities of artificial intelligence implementation within sports science and sports medicine teams / M. Naughton та ін. *Frontiers in sports and active living*. 2024. Т. 6. URL: <https://doi.org/10.3389/fspor.2024.1332427> (дата звернення: 13.01.2025).

16. Machine learning and data mining for sports analytics / ed. by U. Brefeld et al. Cham : Springer International Publishing, 2022. URL: <https://doi.org/10.1007/978-3-031-02044-5> (date of access: 16.01.2025).

17. Newton P. K., Aslam K. Monte carlo tennis: a stochastic markov chain model. *Journal of quantitative analysis in sports*. 2009. Т. 5, № 3. URL: <https://doi.org/10.2202/1559-0410.1169> (дата звернення: 13.01.2025).

18. Paul R. J., Weinbach A. P. Market efficiency and a profitable betting rule. *Journal of sports economics*. 2002. Vol. 3, no. 3. P. 256–263. URL: <https://doi.org/10.1177/1527002502003003003> (дата звернення: 04.01.2025).

19. Poisson distribution in sports analytics | restackio. *The framework for autonomous intelligence – Restack*. URL: <https://www.restack.io/p/ai-in-sports-analytics-answer-poisson-distribution-cat-ai> (дата звернення: 02.01.2025).

20. Vishwakarma A., Paranjape V., Singh A. "Advancements in natural language processing: integrating machine learning for enhanced accuracy and precision". *International journal of innovative research in computer and communication engineering*. 2023. Т. 11, № 10. С. 11366–11373. URL: <https://doi.org/10.15680/ijircce.2023.1110053> (дата звернення: 16.01.2025).

21. Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008 / S.-M. Ma et al. *Journal of Sports Sciences*. 2013. Vol. 31, no. 11. P. 1147–1155. URL: <https://doi.org/10.1080/02640414.2013.775472> (дата звернення: 12.01.2025).