DOI: https://doi.org/10.26642/ten-2022-2(90)-70-77
УДК 004

Н.О. Кушнір, ст. викладач
Ю.І. Лисогор, ст. викладач
І.Д. Лімінович, магістрант
Т.М. Локтікова, ст. викладач
А.В. Морозов, к.т.н., доц.
Державний університет «Житомирська політехніка»

Програмний комплекс для аналізу статистики футбольних матчів та прогнозування результатів на основі машинного навчання

Досліджується застосування методів машинного навчання для прогнозування результатів футбольних матчів. Розглядається задача аналізу футбольної статистики за допомогою контрольованого машинного навчання. Для цього було сформовано спеціальний набір даних, який подається на вхід системи. Модель машинного навчання аналізує зв'язки між різними статистичними даними та відстежує залежності між ними. Здійснено побудову моделі, її навчання та тестування із використанням бібліотеки «Keras». Для візуального відображення отриманих даних було створено сайт. Для його розробки використовувалися сучасні технології. Клієнтська частина розроблена засобами HTML, CSS та JavaScript. Серверну частину розроблено з використанням мови PHP та фреймворку Yii2. Досліджено якість роботи пропонованої моделі машинного навчання порівняно з реальними результатами футбольних матчів. Найкращі результати роботи досягаються при комбінуванні даних, що подаються на вхід системи, з турнірних таблиць п'яти найбільших чемпіонатів Європи, починаючи з сезону 2014/2015, в один великий масив даних, а не для кожного чемпіонату окремо. Оптимальної точності прогнозів результатів запропонована модель машинного навчання набуває при її навчанні протягом 20 епох.

Ключові слова: *статистика; футбол; аналіз даних; прогнозування; машинне навчання.*

Актуальність теми. В наш час усі сфери людського життя описуються у вигляді статистики. Статистика допомагає зрозуміти навколишній світ. Спорт є однією з найпоширеніших розваг, тому статистика виступів спортсменів, результати змагань та інші дані цікавлять глядачів звідусіль [1]. У машинному навчанні є широко розповсюдженою задача аналізу спортивної статистики. Футбол як найпопулярніший вид спорту обирається для таких досліджень у тому числі. Причому, окрім наукових цілей, також для комерційних, таких як ставки, наприклад. На основі статистики здійснюються передбачення на окремі матчі або навіть футбольні сезони. Такі прогнози можуть виконуватися як людиною, так і автоматизовано, за допомогою комп'ютера.

Аналіз останніх досліджень та публікацій. Суттєвим внеском у розвиток використання машинного навчання для аналізу футбольної статистики стали проєкти британських телерадіокомпаній BT Sport та talkSport [2, 3]. Це були прогнози на футбольні сезони Англійської Прем'єр-ліги 2019/2020 та 2020/2021 років відповідно. Компанії не розповідали про методи та засоби, які вони використовували під час розробки своїх систем, проте наголосили, що прогнози зроблені за допомогою машинного навчання. Також потрібно зазначити, що сучасні підходи до аналізу та прогнозування результатів футбольних матчів зазвичай використовують різні методи й алгоритми, а саме: баєсівська мережа із технікою машинного навчання, враховуючи модуль навчання дерева рішень і K-найближчого сусіда та методи машинного навчання. Під час огляду й аналізу інформаційних джерел було з'ясовано, що метод машинного навчання дає найвищу точність прогнозів порівняно з реальними результатами.

Метою статті є дослідження використання машинного навчання в задачах аналізу футбольної статистики та прогнозування результатів футбольних матчів, зокрема визначення переможця та точного рахунку матча. Для проведення дослідження було розроблено модель машинного навчання для аналізу статистики та прогнозів на результат і сайт для візуалізації даних.

Викладення основного матеріалу. Як уже було зазначено, в наш час усі сфери людського життя описуються у вигляді статистики. Вона широко використовується у повсякденному житті. Завдяки статистичним даним можна завжди проаналізувати ситуацію, порівняти її з іншими, прийняти рішення, провести наукові дослідження. Статистичні дані мають широкий спектр використання. Від допомоги компаніям зрозуміти смаки споживачів до аналізу клінічних досліджень. Знання статистики надає можливість грамотно аналізувати: перелік і параметри товарів та послуг, вплив інфляції на ціни й на доходи, цінові пропозиції тощо. Без статистичних даних неможливі управління державою, розвиток економіки і культури, розробка програм розвитку країни [4].

Ці міркування розповсюджуються також і на спорт, зокрема футбол, як на одну з найпоширеніших розваг. Розглянемо особливості здійснення прогнозування результатів футбольних матчів людиною та автоматизовано, за допомогою комп'ютера.

Прогнози, виконані людиною, мають свої переваги та недоліки. Перевагою є те, що людина, окрім цифр, бачить показники, які комп'ютер може не враховувати. Наприклад, атмосферу в команді та фізичну форму окремих гравців, які можуть створювати різницю на футбольному полі. Проте людина може здійснювати прогноз, керуючись емоціями й особистими очікуваннями від матчу, що безсумнівно є недоліком.

Автоматизовані комп'ютерні прогнози постійно розвиваються та стають все точнішими, враховуючи більшу кількість факторів із кожною новою системою. Це тісно пов'язано і з розвитком футбольної статистики. З'являються нові показники, які дозволяють підвищити точність оцінки як команди, так і окремих гравців.

За останні роки було створено декілька автоматизованих комп'ютерних систем, які прогнозували результати різних футбольних сезонів. Зокрема, було оглянуто та проаналізовано систему аналізу від BT Sport, «суперкомп'ютер» talkSPORT і «суперкомп'ютер» Mirror Football.

Авторитетний британський телеканал BT Sport є одним із найперспективніших гравців на ринку прогнозування результатів футбольних матчів за допомогою аналізу статистики. Ним виконуються прогнози на цілі футбольні сезони, тобто на матчі, які будуть проходити протягом 9 місяців, не зупиняючись лише на одному матчі. Перший такий прогноз було здійснено в 2019 році на сезон Англійської Прем'єр-ліги (АПЛ) 2019/2020. Компанія зробила заяву, що провідні науковці й експерти зі спортивного аналізу в режимі реального часу Squawka проаналізували футбольні дані від сайту «Opta» та розмістили на Google Cloud Platform, щоби створити сценарій сезону 2019/2020, включаючи переможців Прем'єр-ліги, переможців Ліги чемпіонів УЄФА, кращих бомбардирів, кандидатів на виліт та інші визначні сюжетні лінії. На рисунку 1 представлено прогноз результатів Англійської Прем'єр-ліги та її реальний результат [5]. Проаналізуємо наведені дані.

PREMIER LEAGUE 2019/20: THE FINAL STANDING

CLUB	P	W	D	L	GF	GA	GD	PTS
MAN CITY	38	29	7	2	96	21	75	94
LIVERPOOL	38	27	7	4	80	28	52	88
TOTTENHAM	38	23	6	9	71	39	32	75
CHELSEA	38	19	10	9	70	37	33	67
ARSENAL	38	19	9	10	77	50	27	66
MAN UTD	38	20	6	12	72	51	21	66
EVERTON	38	14	12	12	57	53	4	54
WOLVES	38	14	11	13	45	52	-7	53
LEICESTER CITY	38	14	9	15	47	53	-6	51
CRYSTAL PALACE	38	11	14	13	42	51	-9	47
WEST HAM UTD	38	11	11	16	43	63	-20	44
WATFORD	38	11	10	17	47	58	-11	43
BOURNEMOUTH	38	11	9	18	48	65	-17	42
SOUTHAMPTON	38	10	11	17	50	65	-15	41
ASTON VILLA	38	12	5	21	45	63	-18	41
BRIGHTON	38	12	4	22	33	58	-25	40
BURNLEY	38	8	13	17	31	51	-20	37
NEWCASTLE	38	10	7	21	41	70	-29	37
NORWICH	38	9	7	22	38	71	-33	34
SHEFFIELD UTD	38	8	8	22	34	68	-34	32

1	Ліверпуль	38	32	3	3	85	33	52	99
2	Манчестер Сіті	38	26	3	9	102	35	67	81
3	МЮ	38	18	12	8	66	36	30	66
4	Челсі	38	20	6	12	69	54	15	66
5	Лестер Сіті	38	18	8	12	67	41	26	62
6	Тоттенгем	38	16	11	11	61	47	14	59
7	Вулвергемптон	38	15	14	9	51	40	11	59
8	Арсенал	38	14	14	10	56	48	8	56
9	Шеффілд Юн…	38	14	12	12	39	39	0	54
10	Бернлі	38	15	9	14	43	50	-7	54
11	Саутгемптон	38	15	7	16	51	60	-9	52
12	Евертон	38	13	10	15	44	56	-12	49
13	Ньюкасл Юн…	38	11	11	16	38	58	-20	44
14	Крістал Пелес	38	11	10	17	31	50	-19	43
15	Брайтон	38	9	14	15	39	54	-15	41
16	Вест Гем	38	10	9	19	49	62	-13	39
17	Астон Вілла	38	9	8	21	41	67	-26	35
18	Борнмут	38	9	7	22	40	65	-25	34
19	Вотфорд	38	8	10	20	36	64	-28	34
20	Норвіч	38	5	6	27	26	75	-49	21

Рис. 1. Прогнозована (зліва) та реальна (справа) таблиці АПЛ 2019/2020

Як видно, переможця ліги було передбачено неправильно. За прогнозом «Ліверпуль» мав фінішувати з меншою кількістю очок, аніж «Манчестер Сіті», проте вони обійшли своїх суперників аж на 18 залікових балів. Третє місце також було перебачено неправильно – його зайняв не «Тоттенгем», а «Манчестер Юнайтед». Четверте місце було передбачено правильно. Звернемо увагу ще на останні три місця в турнірній таблиці. За прогнозом, вищий англійський дивізіон мали покинути «Ньюкасл», «Норвіч» та «Шефілд Юнайтед». Проте з трьох вищевказаних команд покинув турнір лише «Норвіч», а «Ньюкасл» та «Шефілд» опинились у середині турнірної таблиці. Замість них турнір покинули «Борнмут» та «Вотфорд».

В цілому вгадано лише одну з трьох команд, якщо не звертати увагу на турнірне положення. З огляду на точне положення в турнірній таблиці було спрогнозоване точне місце лише однієї з двадцяти команд.

Однією з найкращих можна вважати систему, яку в 2020 році представила британська радіокомпанія talkSPORT. Система, яка отримала назву «суперкомп'ютер», виконала прогноз на сезон Англійської Прем'єр-ліги 2020/2021. Було передбачено, що перше місце в лізі займе «Манчестер Сіті», другим фінішує «Ліверпуль», а четвірку лідерів замкнуть «Челсі» та «Манчестер Юнайтед» відповідно. Порівняємо цей прогноз із реальним результатом турніру (рис. 2).

Із рисунка 2 видно, що чемпіона країни було передбачено правильно. В цілому якщо незважати на точне розташування команд, першу четвірку турнірної таблиці також було передбачено правильно, як і дві з трьох команд, які покидають турнір. Проте мають місце й такі грубі неточності, як передбачення вильоту для «Вест Гему», який в результаті посів досить високе 6 місце.

На сьогодні розроблено велику кількість додатків для прогнозування результатів футбольних матчів на основі аналізу статистики. Найбільш поширеними з них є Bet-Plus, FootBet, FlashScore, Robo-Win. Було здійснено аналіз згаданих вище додатків та визначено їх переваги та недоліки (табл. 1).

Порівнявши існуючі додатки для прогнозування результатів футбольних матчів, можна оцінити точність прогнозів, які вони виконують, та дійти висновку, що жоден із них не використовує машинне навчання в своїй роботі.

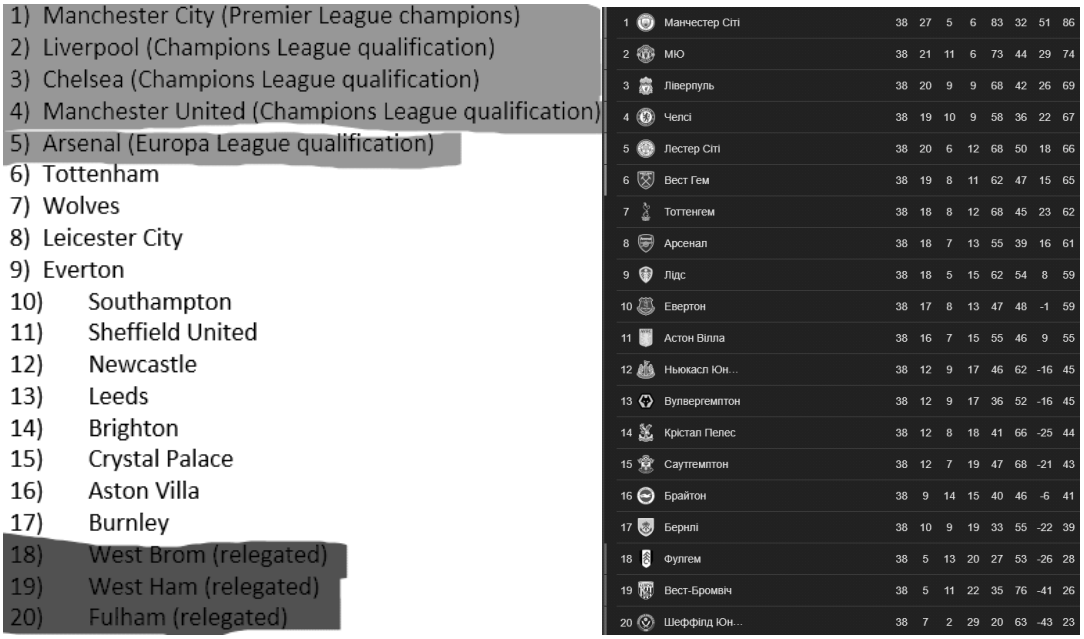

Рис. 2. Прогнозована (зліва) та реальна (справа) таблиці АПЛ 2020/2021

Таблиця 1

Порівняльна характеристика додатків для прогнозування результатів футбольних матчів

Параметр	Bet-Plus	FootBet	FlashScore	Robo-Win
Наявність власних алгоритмів аналізу даних	+	+	-	+
Кросплатформенність	–	–	+	–
Зручність інтерфейсу	–	+	+	+
Заявлена точність прогнозів	85 %	70 %	–	83 %
Доступність	–	+	+	–
Мультимовність	–	–	+	+
Можливість налаштування даних для аналізу	+	–	–	–
Наявність детальної статистики команд, гравців, турнірів, доступної для перегляду	–	–	+	–
Використання машинного навчання для аналізу даних та прогнозування результатів матчів	–	–	–	–
Наявність української мови інтерфейсу	–	–	+	–

Для аналізу футбольної статистики можна використовувати різні методи, а саме: систему рейтингів, метод експертних оцінок, метод послідовних порівнянь, метод парних порівнянь, залежний від часу метод Монте-Карло за ланцюгами Маркова [6] та метод VAEP [7]. Розглянемо детальніше деякі з них.

Система рейтингів. Системи футбольних рейтингів надають звання кожній команді на основі результатів їхніх минулих ігор, тому найвищий ранг присвоюється найсильнішій команді. Результат матчу можна спрогнозувати, порівнявши ранги суперників. Існує декілька різних футбольних рейтингових систем, наприклад, Світовий рейтинг ФІФА.

Метод експертних оцінок. Методи експертних оцінок застосовуються в системі підготовки спортсменів як інструмент прогнозування їхніх спортивних результатів. До проведення експертизи з метою прогнозування спортивних результатів залучаються провідні спеціалісти з певного виду спорту, науковці, тренери. До таких методів належать: метод послідовних порівнянь, метод парних порівнянь та інші.

VAEP (Valuing Actions by Estimating Probabilities) метрика. Ця метрика побудована на моделі машинного навчання, що оцінює кожну дію футболіста на полі, обчислюючи те, яким чином змінилася би можливість забити і пропустити гол у результаті певної дії. Наприклад, оцінка дії гравця +0,05 свідчить про те, що в результаті тактико-технічної дії гравця ймовірність забити гол збільшиться на 0,05. Відповідно, оцінка мінус -0,05 означає, що відповідна дія збільшує шанси суперника забити гол на таке саме значення. На підставі таких оцінок визначається рейтинг футболістів VAEP. Залежно від середнього рейтингу футболістів можна розрахувати рейтинг команди та на основі цих даних виконувати прогноз на результат матчу. VAEP для кожної конкретної дії визначається як сума атакуючих та оборонних дій. Нижче наводяться формули визначення VAEP дії.

$$V(a_i, x) = \Delta P_{scores}(a_i, x) + \left(-\Delta P_{concedes}(a_i, x)\right), \quad (1)$$

$$\Delta P_{scores}(a_i, x) = P_{scores}(S_i, x) - P_{scores}(S_{i-1}, x), \quad (2)$$

$$\Delta P_{concedes}(a_i, x) = P_{concedes}(S_i, x) - P_{concedes}(S_{i-1}, x), \quad (3)$$

де $V(a_i, x)$ – VAEP для дії a гравця x;

$\Delta P_{scores}(a_i, x)$ – зміна ймовірності забити гол у результаті дії a гравця команди x;

$\Delta P_{concedes}(a_i, x)$ – зміна ймовірності пропустити гол у результаті дії a гравця команди x;

$P_{scores}(S_i, x)$ – ймовірність забити гол командою x за 10 наступних дій, залежно від поточного стану гри (СГ) S;

$P_{scores}(S_{i-1}, x)$ – ймовірність забити гол командою x за 10 наступних дій відносно попереднього СГ;

$P_{concedes}(S_i, x)$ – ймовірність пропустити гол командою x за 10 наступних дій залежно від поточного СГ;

$P_{concedes}(S_{i-1}, x)$ – ймовірність пропустити гол командою x за 10 наступних дій відносно попереднього СГ.

Для розробки пропонованої системи аналізу статистики та прогнозування результатів футбольних матчів було обрано метод контрольованого навчання. Контрольоване навчання – це задання вивчення функції, яка перетворює вхідні дані у вихідні на основі прикладів пар «вхід-вихід». У нашому випадку потрібно передбачити категорію результатів (перемога вдома / нічия / перемога на виїзді) або кількість голів, забитих командою (постійну кількість). Це й зумовило вибір контрольованого методу машинного навчання [8].

Для контрольованого машинного навчання потрібні дані, які будуть подаватися на вхід системи. Це може бути готовий набір даних, завантажений зі спеціальних інтернет-джерел або створений власноруч. Для розробленої системи було взято дані з сайту «Understat» [9] та сформовано з них відповідний набір даних. «Understat» – сайт, який збирає просунуті метрики на основі показників ударів по воротах. Сайт пропонує статистичні дані з п’яти найсильніших європейських футбольних чемпіонатів. Він є безкоштовним та на ньому пораховується багато додаткових статистичних параметрів і зберігаються дані, починаючи з футбольного сезону 2014/2015.

У таблиці даних з сайту «Understat», окрім основних показників, а саме: кількість матчів, перемог, поразок, нічиїх та очок, можна побачити такі статистичні показники, як «xG», «xGA» та інші. «xG» – це модель очікуваних голів. В основі такої моделі лежить показник, який допомагає оцінити, скільки голів за інших рівних мала забити команда з ударами такої гостроти. «xGA» («expected goal attempts») – «допущені моменти голу», коефіцієнт небезпеки біля воріт команди. Він завжди дорівнюватиме «xG» суперника. «PPDA» – футбольний статистичний показник, який дозволяє визначити інтенсивність пресингу команди. Чим нижчий цей показник, тим більшу інтенсивність пресингу показує команда. В цілому цей показник дозволяє частково оцінити стиль гри команди. «OPPDA» – це статистичний показник у футболі, що дозволяє оцінити надійність оборонних дій команди. «DC» – показник, який показує, скільки часу проводить команда на третині поля суперника, в безпосередній близькості до його воріт. «ODC» – статистичний показник, який показує, скільки часу проводить команда на своїй третині поля, в захисті. Префікси «a_» та «h_» вказують на приналежність даних виїзній та домашній командам відповідно.

Для кращого розуміння представлених даних та залежностей між ними було побудовано матрицю кореляції, яка зображена на рисунку 3.

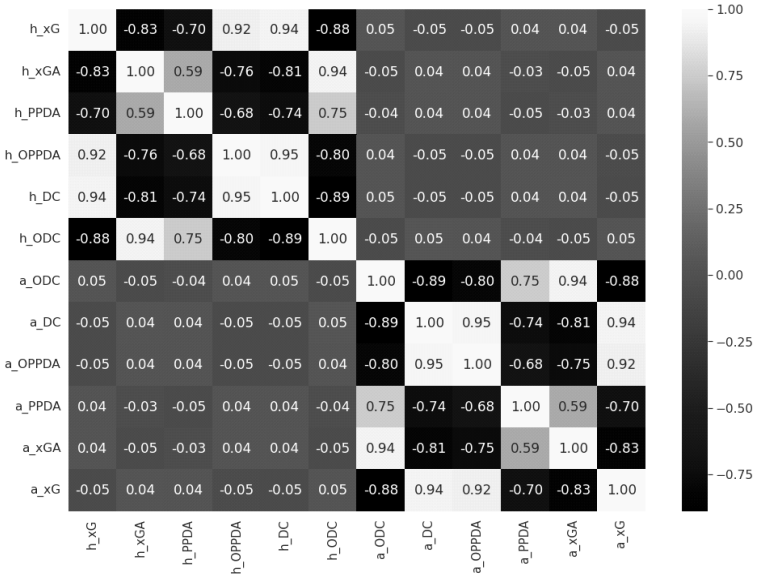

Рис. 3. Результат кореляції

Матриця кореляції наочно відображає зв'язки між статистичними показниками. Дані виїзної та домашньої команд не корелюють між собою, що є логічним. Високо корелюють показники OPPDA та DC. Це є логічним, тому що коли тримаєш м'яч біля воріт суперника, то він виконує більше оборонних дій. Така ж ситуація з PPDA та ODC, хоча й кореляція менша. Деякі показники мають високу від'ємну кореляцію, яка означає, що і збільшенням однієї змінної інша зменшується або навпаки. Наприклад, ODC та xG – чим більше команда проводить часу в обороні, тим менше в неї буде очікуваних голів, тому що вона не буде створювати гольові моменти біля воріт суперника.

Для розробки моделі машинного навчання використовувалися мова програмування Python та бібліотека «Keras». Python – це потужна багатопарадигмова мова програмування, оптимізована для продуктивності програміста, легкої читаності коду та якості програмного забезпечення. «Keras» – це бібліотека для мови програмування Python, призначена для глибокого машинного навчання. Вона дозволяє швидко створювати та налаштовувати моделі – схеми, якими поширюються та підраховуються дані під час навчання. Але складних математичних обчислень «Keras» не виконує та використовується як надбудова над іншими бібліотеками.

На рисунку 4 представлено фрагмент програмного коду, що відповідає за підключення бібліотек, які використовуються в системі, та функції отримання окремої турнірної таблиці.

```
import asyncio
import json
import sys
from matplotlib import pyplot
from keras.models import Sequential
from keras.layers import Dense
import matplotlib.pyplot as plt
import seaborn as sns
import aiohttp
import pandas as pd
import numpy as np
from understat import Understat
import nest_asyncio
nest_asyncio.apply()
def get_table():
    async def _table():
        async with aiohttp.ClientSession() as session:
            understat = Understat(session)
            table = await understat.get_league_table("EPL", "2021")
            return table
    loop = asyncio.get_event_loop()
    understat = loop.run_until_complete(_table())
    temp_table = pd.DataFrame(data=understat)
    league_table = pd.DataFrame(temp_table.values[1:], columns=understat[0])

    return league_table
```

Рис. 4. Фрагмент програмного коду

Для зручної візуалізації даних також було розроблено сайт, який враховує всі сучасні вимоги до інтернет-ресурсів, а саме: мультимовність, адаптивність тощо. Клієнтську частину розроблено за допомогою HTML, CSS та JavaScript. Серверна частина сайту була розроблена на мові PHP з використанням фреймворку Yii2 [10].

Розроблений вебсайт має такі сторінки:

- Головна – сторінка, на якій відображаються прогнози на найважливіші футбольні матчі наступного тижня, а також додаткова інформація про проєкт. Зовнішній вигляд головної сторінки показаний на рисунку 5;
- П’ять ідентичних сторінок – для кожної ліги, на матчі якої виконуються прогнози, а саме: Ла Ліга, Англійська Прем’єр-ліга, Бундесліга, Серія А, Ліга 1. На цих сторінках є по три розділи. В першому знаходяться прогнози на найближчі матчі. В другому – матчі, які вже відбулись, їхні результати та прогнози, які на них було виконано, щоби користувач міг зайти та порівняти реальний результат із передбачуваним системою. В третьому розділі знаходиться поточна турнірна таблиця футбольного чемпіонату;
- Сторінка команди – в кожної команди є окрема сторінка, на якій можна переглянути результати останніх матчів, інформацію про гравців команди, їхню статистику та положення команди в турнірній таблиці чемпіонату;
- Контактна інформація – це сторінка з інформацією про проєкт та розробника. Також на ній знаходиться контактна форма для того, щоби користувачі могли написати свої запитання або зауваження.

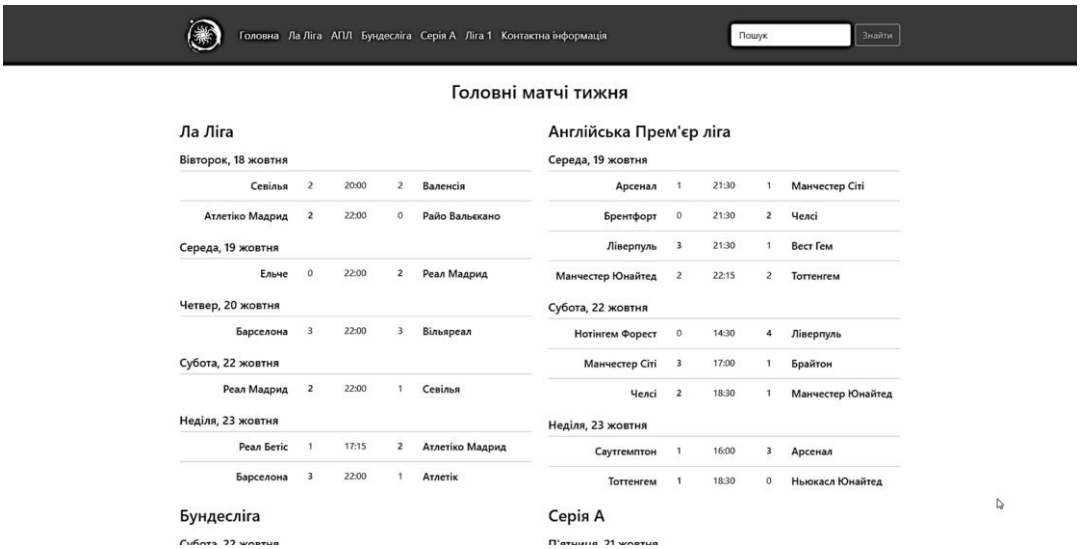

Рис. 5. Головна сторінка сайту

Розроблена база даних, яка складається з чотирьох таблиць. Схема бази даних зображена на рисунку 6.

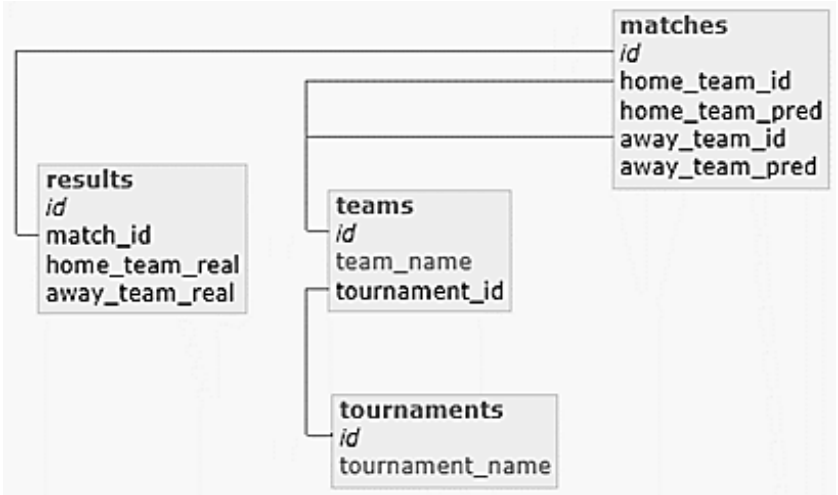

Рис. 6. Схема бази даних

Розглянемо та опишемо кожну з таблиць бази даних:

- «Чемпіонати» – таблиця, в якій знаходяться турніри, на які виконуються прогнози. Має поля «id» – унікальний ідентифікатор та «tournament_name» – назва чемпіонату;

- «Команди» – таблиця, в якій знаходяться команди, що беруть участь у цих чемпіонатах. Ця таблиця має такі поля: «id» – унікальний ідентифікатор команди, «team_name» – назва команди, «tournament_id» – ідентифікатор чемпіонату, в якому бере участь команда. По полю «tournament_id» відбувається зв'язок між таблицями;

- «Матчі» – таблиця, в яку заносяться матчі, що відбуватимуться. Має поля: «id» – унікальний ідентифікатор матча, «home_team_id» – ідентифікатор команди, що грає вдома, «home_team_pred» – прогнозована кількість забитих домашньою командою голів, «away_team_id» – ідентифікатор команди, що грає в гостях, «away_team_pred» – прогнозована кількість голів, що заб'є виїзна команда, «date» – дата проведення матча. Поля «home_team_id» та «away_team_id» пов'язані з таблицею «Команди»;

- «Результати» – таблиця з результатами матчів, що відбулися. Має всього чотири поля, а саме: «id» – унікальний ідентифікатор матча, що відбувся, «match_id» – ідентифікатор матча з таблиці «Матчі», «home_team_real» – реальна кількість голів, які забила домашня команда, «away_team_real» – реальна кількість голів, забитих виїзною командою.

Висновки та перспективи подальших досліджень. Було проведено тестування розробленого програмного комплексу для аналізу статистики футбольних матчів та прогнозування результатів на основі машинного навчання, яке підтвердило коректність роботи програмного додатка та працездатність сайту. При порівнянні прогнозів, здійснених системою, та реальних результатів футбольних матчів була досягнута 74 % точність прогнозування. Об'єктом подальших досліджень буде підвищення точності прогнозів за рахунок удосконалення моделі машинного навчання.

Список використаної літератури:

1. *Kimberly E.* Production Planning and Inventory Control / *E.Kimberly*. – McGraw-Hill, 2010. – 550 с.
2. *Hoskin R.* Revisiting BT Sport's 'The Script' after they attempted to predict the entire 2019/20 season / *R.Hoskin*. – 2020 [Electronic resource]. – Access mode : https://www.givemesport.com/1587003-revisiting-bt-sports-the-script-after-they-attempted-to-predict-the-entire-201920-season.
3. *Riaz A.* Supercomputer Predicts The Full Premier League Table For The 2020-21 Season / *A.Riaz*. – 2020 [Electronic resource]. – Access mode : https://cutt.ly/QMnkd9W.
4. *Kress G.* Forecasting and Market Analysis Techniques: A Practical Approach / *G.Kress, J.Snyder*. – Quorum Books, 1994. – 304 с.
5. FlashScore [Electronic resource]. – Access mode : https://www.flashscore.com.
6. *Yam D.* Attacking Contributions: Markov Models for Football / *D.Yam* // StatsBomb. – 2019 [Electronic resource]. – Access mode : https://statsbomb.com/articles/soccer/attacking-contributions-markov-models-for-football.
7. VAEP (Valuing Actions by Estimating Probabilities) is a framework for valuing player actions in soccer [Electronic resource]. – Access mode : https://dtai.cs.kuleuven.be/sports/vaep.
8. *Ротштейн О.П.* Інтелектуальні технології ідентифікації: нечіткі множини, генетичні алгоритми, нейронні мережі / *О.П. Ротштейн*. – Вінниця : УНІВЕРСУМ-Вінниця, 1999. – 320 с.
9. Understat [Electronic resource]. – Access mode : https://understat.com.
10. Про Yii – Українська спільнота / Yii Framework [Електронний ресурс] – Режим доступу : https://yiiframework.com.ua/uk/doc/guide/2.

References:

1. Kimberly, E. (2010), *Production Planning and Inventory Control*, McGraw-Hill, 550 p.
2. Hoskin, R. (2020), *Revisiting BT Sport's 'The Script' after they attempted to predict the entire 2019/20 Season*, [Online], available at: https://www.givemesport.com/1587003-revisiting-bt-sports-the-script-after-they-attempted-to-predict-the-entire-201920-season
3. Riaz, A. (2020), *Supercomputer Predicts The Full Premier League Table For The 2020-21 Season*, [Online], available at: https://cutt.ly/QMnkd9W
4. Kress, G. and Snyder, J. (1994), *Forecasting and Market Analysis Techniques: A Practical Approach*, Quorum Books, 304 p.
5. FlashScore, [Online], available at: https://www.flashscore.com
6. Yam, D. (2019), *Attacking Contributions: Markov Models for Football*, StatsBomb, [Online], available at: https://statsbomb.com/articles/soccer/attacking-contributions-markov-models-for-football
7. VAEP (Valuing Actions by Estimating Probabilities) is a framework for valuing player actions in soccer, [Online], available at: https://dtai.cs.kuleuven.be/sports/vaep
8. Rotshtejn, O.P. (1999), *Intelektual'ni tehnologii' identyfikacii': nechitki mnozhyny, genetychni algorytmy, nejronni merezhi*, UNIVERSUM-Vinnycja, Vinnycja, 320 p.
9. Understat, [Online], available at: https://understat.com
10. Yii Framework, *Pro Yii – Ukrai'ns'ka spil'nota*, [Online], available at: https://yiiframework.com.ua/uk/doc/guide/2

Кушнір Надія Олександрівна – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка».
https://orcid.org/0000-0002-0797-3687.
Наукові інтереси:
- комбінаторна оптимізація;
- інформаційні технології.

E-mail: kipz_kno@ztu.edu.ua.

Лисогор Юрій Іванович – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка».
https://orcid.org/0000-0003-1194-2813.
Наукові інтереси:
- комп'ютерна графіка та дизайн;
- цифрова обробка сигналів;
- інформаційні системи та технології.

E-mail: lysogor@ztu.edu.ua.

Лімінович Іван Дмитрович – магістрант Державного університету «Житомирська політехніка».
https://orcid.org/0000-0003-2196-4186.
Наукові інтереси:
- інформаційні системи та технології.

E-mail: ivanliminovych@gmail.com.

Локтікова Тамара Миколаївна – старший викладач кафедри інженерії програмного забезпечення Державного університету «Житомирська політехніка».
https://orcid.org/0000-0002-3525-0179.
Наукові інтереси:
- цифрова обробка зображень;
- інформаційні системи та технології.

E-mail: tamlokt@ukr.net.

Морозов Андрій Васильович – кандидат технічних наук, доцент, проректор з науково-педагогічної роботи Державного університету «Житомирська політехніка».
https://orcid.org/0000-0003-3167-0683.
Наукові інтереси:
- комбінаторна оптимізація;
- інформаційні технології.

E-mail: morozov@ztu.edu.ua.

Kushnir N.O., Lysohor Y.I., Liminovych I.D., Loktikova T.M., Morozov A.V.

A software system for analyzing the statistics of football matches and predicting the results based on machine learning

The application of machine learning methods for predicting the results of football matches is under the investigation. The task of analyzing football statistics using supervised machine learning is considered. For this purpose, a special data set was formed, which is input into the system. The machine learning model analyzes the relations between different statistics and tracks the dependencies between them. The model was built, trained and tested using the «Keras» library. A website was created to visually display the data obtained. During its development modern technologies were used. The client part was developed using HTML, CSS and JavaScript. The server part was developed using the PHP language and the Yii2 framework. The quality of the proposed machine learning model was investigated by using the comparison with the real results of football matches. The best results are achieved by combining the input data from the standings of the five major European championships, starting from the 2014/2015 season, into one large data set, and not for each championship separately. The optimal accuracy of predictions of the results of the proposed machine learning model is obtained by training for 20 epochs.

Keywords: statistics; football; data analysis; prognostication; machine learning.

Стаття надійшла до редакції 18.08.2022.