

Богдан Пришляк, Андрій Нечепуренко*

¹ НЛТУ України, Львів, Україна, 23pryshliak.b@nltu.lviv.ua

² НЛТУ України, Львів, Україна, ORCID: 0000-0003-3955-1849,
andriy.nechepurenko@gmail.com

Інформаційна система аналізу результатів баскетбольних матчів

Анотація. Представлено модель на основі даних для аналізу результатів баскетбольних матчів НБА з використанням методів машинного навчання. Модель ґрунтується на математичній формалізації статистичних показників баскетболу, зосереджуючись на індексі ефективності команди, який враховує як індивідуальні, так і порівняльні показники ефективності. Прогнози робляться шляхом обчислення розширених індексів для команд-суперників, а функція виграшу визначається на основі їхніх відносних значень. Модель використовує навчальні набори даних з п'яти сезонів НБА (2018/2019–2022/2023).

Ключові слова – інформаційна система, машинне навчання, аналіз баскетбольних матчів, індекс ефективності команди, функція виграшу, Python, Pandas, Matplotlib.

Аналіз спортивних результатів за допомогою штучного інтелекту набув популярності серед професіоналів і вболівальників. Це питання цікавить не лише букмекерів, а й тренерів та спортивних експертів. Об'єктивні дані аналізу важливі для оцінки розвитку гравців і команд, запобігання перевантаженням і травмам, а також для прийняття рішень на перспективу. Представлено модель машинного навчання для аналізу результатів ігор НБА (Національної баскетбольної асоціації) [1]. Машинне навчання є підгалуззю штучного інтелекту, що використовує історичні дані для розробки методів аналізу спортивних матчів. Найпоширеніші його методи вимагають вхідних даних, таких як статистика гравців і команд, з відповідними вихідними значеннями – результатами ігор.

Опис моделі. Аналіз спортивних результатів найчастіше розглядають як задачу контрольованої бінарної класифікації. Серед двох основних категорій машинного навчання, класифікації та регресії, перша є більш ефективною для прогнозування результатів командних спортивних ігор. В моделі крім основної

статистики НБА враховано розширену статистику ліги, що стосується конкретних команд, впроваджено індекс ефективності команди та функцію виграшу [2].

Продуктивність гравців і команд об'єктивно оцінюється за допомогою різних показників ефективності. Одним із найпоширеніших є індекс ефективності гравців НБА, який оцінює результативність одного гравця в окремій грі. Цей індекс може бути усереднений за кілька ігор або розрахований для всієї команди. Індекс ефективності гравців НБА враховує тринадцять основних елементів гри, які фіксуються у таблицях баскетбольної статистики [3]. Ці елементи наведені в таблиці 1.

Таблиця 1. Основні елементи гри в баскетбол: позитивні (+), нейтральні (0), негативні (-)

Елемент	Скорочення	Опис
Basic (+)	gmd2F(3F)(FT)	забиті м'ячі з 2-х (3-х) очок (штрафні кидки).
	asts	паси, які безпосередньо призводять до забитого м'яча.
	blcks	блоки - моменти, коли гравець перешкоджає кидку суперника, зупиняючи м'яч у повітрі.
	rbd, rbo	rbd – захисні підбори м'яча після промахів суперника. rbo – нападні підбори м'яча після власних промахів команди.
	stls	перехоплення – моменти, коли гравець успішно забирає м'яч у суперника.
Basic (0)	Gat2F(3F)(FT)	кидки з 2-х (3-х) очок (штрафні кидки) – загальна кількість спроб кидків, які гравець зробив з гри: 2-очкові, 3-очкові та штрафні кидки.
Basic (-)	fls	персональні фоли – кількість порушень, які гравець здійснив проти суперника.
	tos	втрати м'яча – випадки, коли команда втрачає володіння м'ячем.

Індекс ефективності гравців оцінює внесок гравця в гру, це показник загальної роботи, яку виконує гравець під час матчу. Тренер вирішує, скільки часу залишити гравця на полі, залежно від його якості та ситуації в грі. Індекс по-

єднує силу гравця з його ігровим часом, враховуючи також рішення тренера. Це свідчить про те, що внесок гравця визначається не лише його навичками, а й стратегією команди. Індекс ефективності гравців НБА є абсолютним показником їхньої ефективності в грі, його можна обчислити для всієї команди, додавши показники всіх гравців [4].

Результати та обговорення. Було проаналізовано баскетбольні матчі з п'яти сезонів НБА, починаючи з 2018/2019 до 2022/2023. Індекс ефективності гравця є важливим показником, що оцінює внесок кожного баскетболіста в гру на основі різноманітних статистичних даних. Він враховує як позитивні, так і негативні дії гравця, що дає змогу отримати комплексну картину його продуктивності. Індекс ефективності команди, в свою чергу, обчислюється шляхом підсумовування індексів всіх гравців команди, що дає змогу оцінити загальний внесок колективу. Ці індекси допомагають тренерам і аналітикам аналізувати не тільки індивідуальні виступи, а й командну динаміку. Завдяки таким показникам, спортивні фахівці можуть краще планувати стратегії та приймати обґрунтовані рішення щодо складу команди.

Індекс ефективності гравця обчислюється за формулою, що враховує кілька основних статистичних показників. Він дає змогу оцінити ефективність гравця у контексті його ігрового часу, порівнюючи його продуктивність з іншими гравцями.

Функція виграшу – модель, яка використовується для прогнозування ймовірності виграшу однієї команди над іншою в спортивному матчі. Вона обчислюється на основі різних показників ефективності, які враховують продуктивність команд у минулих іграх. Функція виграшу надає цінну інформацію для тренерів, аналітиків та букмекерів, допомагаючи оцінити ймовірність результату гри.

Датасет містить дані за останні 10 ігор для кожної команди. Ці фактори розділяються на домашні та гостьові значення для кожної з команд. Це допомагає відстежувати, як зміни у показниках чотирьох факторів впливають на результат матчу. Для навчання моделей використовувалися різні алгоритми:

LogisticRegression, KNeighborsClassifier, RandomForestClassifier, Gaussian Naive Bayes, класифікатор опорних векторів для навчання моделей аналізу баскетбольних матчів і порівняння їхньої точності. Для моделей із гіперпараметрами використовується GridSearchCV для пошуку оптимальних значень параметрів. Реалізовано пошук оптимальної кількості ознак для моделі Gaussian Naive Bayes. Для відбору ознак було об'єднано статистики домашньої та гостьової команд. Отже, щоб для кожної ознаки був один стовпець.

Виконано розбиття набору даних на навчальні та тестові підвибірки для подальшого навчання моделі та оцінки якості її роботи відповідно.

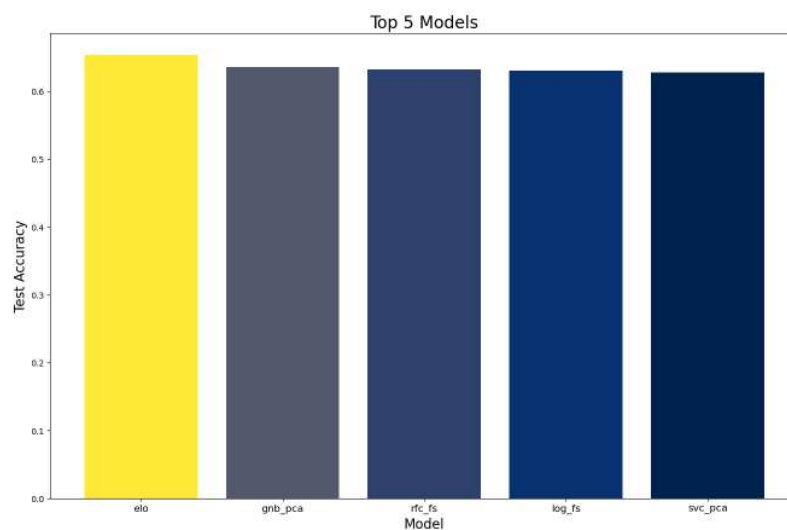


Рисунок 1. Оцінка якості моделей

На рис. 1 представлено графік, який показує точність (частку правильних відповідей) на тестових даних для п'яти досліджених моделей, з кольорами, що відображають різні рівні точності. Кожен стовпчик представляє одну модель, його висота показує, наскільки добре модель працює на тестових даних. По горизонтальній осі представлено назви моделей, по вертикальній – точність моделей на тестових даних. Колір відображає рівень точності (чим він темніший, тим вища точність).

Висновок. При застосуванні машинного навчання для аналізу результатів командних видів спорту важливо зрозуміти складність системи та виявити правильні індикатори, які корелюють з майбутньою продуктивністю команд. Це ускладнюється тим, що багато баскетбольних матчів закінчуються з невеликою

різницею в очках, що робить прогнози менш точними. Детальна баскетбольна статистика, яка базується на вимірюваних факторах, забезпечує надійну основу для оцінки продуктивності команди в минулому та прогнозу її майбутньої діяльності. Введено загальний показник ефективності команди, який базується на індексі ефективності гравців НБА та функцію виграшу. Набір основних елементів гри, що використовуються для його розрахунку, складає базу моделі машинного навчання.

Список використаних літературних джерел

- [1] Wilkens, S. (2021). Sports Prediction and Betting Models in the Machine Learning Age. *Journal of Sports Analytics*, 7, 99–117.
- [2] Lam, M.W. (2018). One-match-ahead forecasting in two-team sports with stacked Bayesian regressions. *Journal of Artificial Intelligence and Soft Computing Research*, 8, 159–171.
- [3] Morgulev, E., Azar, O.H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analysis*, 5, 213–222.
- [4] Pai, P.F.; ChangLiao, L.H.; & Lin, K.P. (2017). Analyzing basketball games by a support vector machines with decision tree model. *Neural Computing and Applications*, 28, 4159–4167.

Information system for analysis of basketball match results

Bohdan Pryshlyak, Andriy Nechepurenko

The paper presents a data-driven model for predicting the outcomes of NBA basketball games using machine learning methods. The model is based on the mathematical formalization of basketball statistics, focusing on the team performance index, which takes into account both individual and comparative performance indicators. Predictions are made by calculating augmented indices for opposing teams, and the win function is determined based on their relative values. The model uses training data sets from five NBA seasons (2018/2019-2022/2023).