

DOI 10.36074/logos-19.12.2025.027

ДОСЛІДЖЕННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ І ПРОГНОЗУВАННЯ ІГРОВОЇ ЕФЕКТИВНОСТІ ФУТБОЛІСТІВ В ІНФОРМАЦІЙНИХ СИСТЕМАХ СПОРТИВНОЇ АНАЛІТИКИ

Малець Єгор Олександрович¹, Міхнова Аліна Володимирівна²

1. здобувач вищої освіти факультету комп'ютерних наук
Харківський національний університет радіоелектроніки, УКРАЇНА

2. доцент кафедри інформаційних управляючих систем
Харківський національний університет радіоелектроніки, УКРАЇНА

Анотація. Ця робота присвячена вдосконаленню моделі машинного навчання XGBoost шляхом інтеграції байєсівської оптимізації для аналізу та прогнозування результатів футболістів в інформаційних системах спортивної аналітики. Розглянуто актуальність підвищення точності прогнозування для мінімізації фінансових ризиків у трансферній діяльності. Система забезпечує адаптивний підхід до регуляризації та аналізу даних. Вдосконалена модель демонструє значне зниження середньої абсолютної похибки (MAE) порівняно з традиційними методами.

У сучасному професійному спорті, зокрема у футболі, ефективність управління тренувальним процесом та трансферною діяльністю критично залежить від якості обробки великих масивів даних. Попри те, що наявні аналітичні платформи, такі як Wyscout або Opta, автоматизують збір статистики, більшість із них обмежується наданням описових звітів, не забезпечуючи необхідної точності прогнозування майбутньої ігрової ефективності [1]. Ця проблема виникає через використання у таких системах лінійних методів оцінювання, які не здатні врахувати складну нелінійну природу взаємодії факторів, таких як накопичена втома, контекст конкретного матчу або позиційний вплив на полі [2]. Ігнорування цих нюансів призводить до суб'єктивності в оцінці гравців, що, своєю чергою, тягне за собою високі фінансові ризики для клубів, пов'язані з неефективними трансферами та

управлінськими рішеннями [3]. Актуальність цього дослідження зумовлена необхідністю підвищення прогностичної точності шляхом інтеграції адаптивних моделей машинного навчання, здатних нівелювати вплив статистичного шуму та об'єктивно оцінювати гравців навіть в умовах високої невизначеності.

Головною метою роботи є вдосконалення моделі градієнтного бустингу XGBoost шляхом інтеграції методів Байєсівської оптимізації для підвищення точності прогнозування ігрової ефективності футболістів та обґрунтування її практичного застосування у складі інформаційних систем спортивної аналітики. Об'єктом дослідження виступає процес аналізу та прогнозування ігрової ефективності на основі статистичних та фізичних даних, а предметом — удосконалена адаптивна модель та методика її застосування.

Як базовий алгоритм для дослідження було обрано модель Extreme Gradient Boosting (XGBoost), яка є галузевим стандартом для роботи з табличними спортивними даними завдяки своїй швидкості та здатності обробляти пропущені значення [4]. Однак використання цієї моделі у стандартній конфігурації має суттєві обмеження, оскільки статичні параметри навчання не враховують динаміку спортивного середовища та специфіку конкретних гравців, що часто призводить до перенавчання на історичних даних. Наукова новизна отриманих результатів полягає у структурно-параметричному вдосконаленні алгоритму XGBoost шляхом інтеграції модуля Байєсівської оптимізації для автоматичного налаштування параметрів регуляризації, а саме коефіцієнтів гамма (γ) та лямбда (λ). На відміну від стандартного підходу, де ці параметри фіксовані, у розробленій моделі вони стають динамічними та контролюються функцією очікуваного покращення (E) [5]. Це забезпечує суттєве підвищення робастності моделі, тобто її стійкості до зашумлених даних, дозволяючи ефективно фільтрувати випадкові статистичні викиди та виявляти реальні ігрові закономірності.

Для забезпечення коректної роботи вдосконаленого алгоритму було розроблено комплексну методику збору та попередньої обробки вхідних даних. Організація цього процесу базується на інтеграції гетерогенних, тобто різнорідних, даних: тактико-технічних дій гравців зі звітів провайдерів та показників фізичної активності, отриманих з GPS-трекерів. Ключовим елементом підготовки даних є процедура комбінованої нормалізації векторів ознак методом Min-Max Scaling, яка приводить усі числові показники до єдиного діапазону від 0 до 1, що критично важливо для стабільної роботи градієнтних методів оптимізації. Крім того, у процесі інженерії ознак вводяться відносні метрики ефективності, розраховані на 90 хвилин ігрового часу, що дозволяє коректно порівнювати гравців, які провели на полі різну кількість часу.

SEZIONE 14.

TECNOLOGIE E SISTEMI DELL'INFORMAZIONE

Процес навчання та верифікації моделі проводився з використанням специфічної методики розбиття даних, що називається послідовною валідацією або Time Series Split [6]. На відміну від випадкового перемішування, цей метод зберігає хронологічний порядок подій: дані за перші 70% сезону використовуються для навчання, наступні 15% — для налаштування гіперпараметрів, і останні 15% – для фінального тестування. Це дозволяє змодельовати реальну ситуацію, коли системі необхідно робити прогноз на майбутнє, базуючись виключно на минулому досвіді, що значно підвищує достовірність оцінки якості моделі.

Практичне значення роботи підтверджено результатами експериментальної апробації, яка показала значну перевагу вдосконаленої моделі над традиційними методами. Порівняльний аналіз точності прогнозування продемонстрував зниження середньої абсолютної похибки (MAE) оцінювання до 0.18 бала, що становить зменшення помилки на 85.9% порівняно з традиційною статистичною оцінкою, де похибка становила 1.28 бала. Така висока точність була досягнута завдяки успішній роботі механізмів адаптивної регуляризації, які дозволили коректно ідентифікувати складні ігрові ситуації. Зокрема, модель успішно виявила випадки «пасивної гри», коли гравець демонстрував високу статистичну активність, але низьку користь для команди, і скоригувала оцінку в бік зниження, тоді як базова статистика помилково оцінювала такого гравця високо. Також модель довела свою ефективність у прогнозуванні функціонального стану гравців в умовах щільного календаря, виявивши приховані аномалії у співвідношенні механічного навантаження та пульсової вартості, що дозволило своєчасно попередити про високий ризик м'язової травми [7].

Розроблене рішення має виражену практичну цінність у трьох основних аспектах: спортивному, медичному та економічному. У спортивному плані оптимізація тактичного плану під поточний стан гравців підвищує ймовірність позитивного результату матчу. У медичному аспекті раннє виявлення ознак перетренованості дозволяє запобігати безконтактним травмам м'яких тканин. Економічний ефект полягає у зниженні витрат клубу на реабілітацію травмованих гравців та мінімізації фінансових ризиків від помилкових трансферів завдяки точнішій оцінці активів. Для забезпечення довіри до системи з боку тренерського штабу впроваджено інструменти пояснюваного штучного інтелекту (Explainable AI), які трансформують складні математичні прогнози у зрозумілі рекомендації з використанням колірною кодування за принципом «світлофора», роблячи логіку прийняття рішень прозорою та доступною для кінцевого користувача [8]. Таким чином, розроблена адаптивна модель є ефективним інструментом для об'єктивізації управління футбольною командою.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

- [1] Dobson S., Goddard J. The Economics of Football. 2nd ed. Cambridge: Cambridge University Press, 2011. 468 p.
- [2] Sarmiento H. et al. Match analysis in football: a systematic review. Journal of Sports Sciences. 2014. Vol. 32. No. 20. P. 1831–1843.
- [3] Rathke A. An examination of expected goals and shot efficiency in soccer. Journal of Human Sport and Exercise. 2017. Vol. 12. No. 2. P. 451–467.
- [4] Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785–794.
- [5] Snoek J., Larochelle H., Adams R. P. Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems. 2012. Vol. 25. P. 2951–2959.
- [6] Tashman L. J. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting. 2000. Vol. 16(4). P. 437–450.
- [7] Ekstrand J., Waldén M., Hägglund M. Hamstring injury rates increase during the FIFA World Cup. British Journal of Sports Medicine. 2021. Vol. 55(15). P. 869–873.
- [8] Lundberg S. M., Lee S.-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 2017. P. 4765–4774

