

РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ПОДІЙ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Вінницький національний технічний університет

Анотація

У роботі розроблено інформаційну технологію прогнозування результатів спортивних подій з використанням методів машинного навчання. Основна увага приділена створенню та налаштуванню моделей класифікації на основі алгоритмів Random Forest, XGBoost, LightGBM та логістичної регресії. Розроблена система забезпечує обробку великих обсягів статистичних даних про команди, гравців та історію матчів, що дозволяє підвищити точність прогнозів. Для підвищення стабільності результатів застосовано метод ансамблювання та стекування моделей. Проведено аналіз впливу ключових факторів — рейтингу команд, поточної форми, очних зустрічей та коефіцієнтів букмекерів — на кінцевий результат. Розроблений програмний продукт реалізовано мовою Python із використанням бібліотек Scikit-learn, XGBoost та LightGBM. Отримані результати показали підвищення точності прогнозування порівняно з базовими моделями, що підтверджує ефективність запропонованої технології.

Ключові слова: прогнозування, спортивні події, машинне навчання, класифікація, ансамблювання, XGBoost, LightGBM.

Abstract

The paper presents the development of an information technology for predicting sports event outcomes using machine learning methods. The main focus is on creating and tuning classification models based on Random Forest, XGBoost, LightGBM, and Logistic Regression algorithms. The developed system processes large volumes of statistical data about teams, players, and match history, which improves prediction accuracy. To enhance result stability, ensemble and stacking techniques were applied. The influence of key factors such as team rating, current form, head-to-head results, and bookmaker odds on the final outcome was analyzed. The software was implemented in Python using Scikit-learn, XGBoost, and LightGBM libraries. Experimental results demonstrated higher prediction accuracy compared to baseline models, confirming the efficiency of the proposed technology.

Keywords: matrix multiplication, MPI, Mesh Network, performance.

Keywords: prediction, sports events, machine learning, classification, ensemble learning, XGBoost, LightGBM.

Вступ

Прогнозування результатів спортивних подій є складним завданням, яке вимагає врахування великої кількості факторів, таких як статистика команд, поточна форма, травми гравців, очні зустрічі та інші параметри. З розвитком машинного навчання стало можливим створення систем, здатних аналізувати великі обсяги даних і формувати точні прогнози на основі виявлених закономірностей [1]. Використання сучасних алгоритмів, таких як градієнтний бустинг (XGBoost, LightGBM) та ансамблювання моделей, дозволяє підвищити точність прогнозування у порівнянні з традиційними статистичними методами [2]. Такі технології мають практичне застосування у спортивній аналітиці, букмекерській діяльності та стратегічному плануванні змагань [3].

Постановка задачі дослідження

Для досягнення мети дослідження необхідно: провести аналіз сучасних підходів до прогнозування спортивних подій; дослідити вплив основних статистичних показників на результати матчів; розробити набір моделей машинного навчання для класифікації результатів; реалізувати систему попередньої обробки та стандартизації даних; провести навчання та тестування моделей Random

Forest, XGBoost, LightGBM і Logistic Regression; застосувати ансамблювання та стекування моделей для підвищення точності прогнозів; оцінити ефективність розробленої системи за метриками Accuracy, F1-macro та ROC- AUC.

Виклад основного матеріалу

Розроблена система прогнозування результатів спортивних подій базується на принципах машинного навчання. Для навчання моделей використано історичні дані матчів, що містять показники рейтингу команд, форму, очні зустрічі, кількість забитих і пропущених м'ячів, коефіцієнти букмекерів тощо. Попередня обробка даних включала нормалізацію, імпутацію пропущених значень і балансування класів за допомогою методу SMOTE. Для зменшення впливу випадкових флуктуацій проведено стратифіковану крос-валідацію. Кожна модель (Logistic Regression, Random Forest, XGBoost, LightGBM) була оптимізована за допомогою підбору гіперпараметрів. Остаточне рішення формувалося методом стекування, де метамоделлю виступала логістична регресія. Розробка виконана мовою Python із використанням бібліотек Pandas, NumPy, Scikit-learn, Imbalanced-learn, XGBoost і LightGBM.

Результати дослідження

Експериментальні результати показали, що ансамбль моделей перевищив точність окремих алгоритмів: **Baseline (Logistic Regression)** - 0.64 Accuracy; **Random Forest** - 0.71; **XGBoost** - 0.74; **LightGBM** - 0.76; **Stacking Ensemble** - 0.79 Accuracy, 0.78 F1-macro.

Отримані результати підтверджують ефективність комбінування моделей для підвищення точності прогнозування спортивних результатів.

Висновки

У ході дослідження розроблено інформаційну технологію прогнозування результатів спортивних подій на основі машинного навчання. Запропонований підхід дозволяє об'єднати статистичний аналіз із сучасними алгоритмами класифікації, що забезпечує високу точність прогнозів.

Розроблена система може бути використана для спортивної аналітики, автоматизованого прийняття рішень та підтримки ставок. Подальші дослідження можуть бути спрямовані на впровадження нейронних мереж (LSTM) для обліку часових залежностей у спортивних даних та розширення системи на різні види спорту.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly Media, 2023.
2. Brownlee J. *Ensemble Learning Algorithms with Python*. Machine Learning Mastery, 2022.
3. Liu Y., Wang J. *Predicting Football Match Outcomes Using Machine Learning Techniques*. *Applied Computing and Informatics*, 2021.

Суботін Ілля Андрійович – студент кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: Subotin.illya@gmail.com ;

Денисюк Валерій Олександрович – канд. техн. наук, доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м.Вінниця, e-mail: vad64@i.ua .

Subotin Illia Andriiovych – student of Computer Science Department, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, Ukraine, e-mail: Subotin.illya@gmail.com ;

Denysiuk Valerii Olexandrovich – Ph.D., Assistant Professor, Assistant Professor of the Chair of Computer Science, Vinnytsia National Technical University, Vinnytsia, e-mail: vad64@i.ua .