

Балдинюк Анастасія Олександрівна

студентка

Національний технічний університет України «Київський політехнічний інститут»

Балдынюк Анастасия Александровна

студентка

Национальный технический университет Украины «Киевский политехнический институт»

Baldyniuk A.

student

National Technical University of Ukraine «Kyiv Polytechnic Institute»

**ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ СПОРТИВНИХ ПОДІЙ
НА ОСНОВІ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ
ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ СПОРТИВНЫХ СОБЫТИЙ
НА ОСНОВЕ МЕТОДОВ ИНТЕЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ
FORECASTING THE RESULTS OF SPORTS EVENTS BASED
ON DATA MINING METHODS**

Анотація. Проаналізовано популярні підходи та методи інтелектуального аналізу та ефективність і доцільність їх використання для прогнозування результатів спортивних подій.

Ключові слова: прогнозування, інтелектуальний аналіз даних, глибокі нейронні мережі.

Аннотация. Проанализировано популярные подходы и методы интеллектуального анализа данных, а также эффективность и уместность их использования для прогнозирования результатов спортивных событий.

Ключевые слова: прогнозирование, интеллектуальный анализ данных, глубокие нейронные сети.

Summary. Some of the most commonly used, existing at the moment, today's data mining methods were reviewed and analyzed. Also the relevance of using them in forecasting the results of sports events was considered.

Key words: forecasting, data mining, deep neural networks.

Вступ. На сьогоднішній день сфера беттингу та букмекерства користується великою популярністю у широкого кола прихильників спорту.

З розвитком методів інтелектуального аналізу стало можливим прогнозувати результати спортивних подій за допомогою відповідних математичних алгоритмів, отримуючи таким чином значно більшу точність прогнозу результату, ніж виходячи з суб'єктивних припущень футболістських експертів.

Питанням знаходження універсального алгоритму для побудови точних прогнозів результатів футбольних матчів займаються провідні науковці університетів Сполучених Штатів Америки, Великої Британії, Німеччини, Японії та інших країн. Підходи та методи, що використовуються у цих дослідженнях мають ряд недоліків, через що постає гостра потреба у їх постійній модернізації та пошуку шляхів вирішення ключових проблем.

Дане дослідження є актуальним та корисним у сфері беттингу, а саме у якості інструменту вибору правильної стратегії формування футбольних ставок. Також воно має тісний зв'язок із сектором букмекерства як метод більш доцільного способу розрахунку ігрових коефіцієнтів на спортивні події. Робота, що пропонується, присвячена висвітленню таких можливостей прогнозування, які надають методи Data Mining та глибокі нейронні мережі зокрема у контексті прогнозування результатів спортивних подій.

Постановка задачі. На початку дослідження були поставлені наступні завдання:

- проаналізувати існуючі підходи та методи інтелектуального аналізу;
- оцінити точність прогнозів при застосуванні технологій Data Mining, а саме: методу k-середніх, наївного байєсівського класифікатора та різних типів

нейронних мереж у контексті прогнозування результатів спортивних подій;

- реалізувати чотири типи нейронних мереж із різною топологією з урахуванням особливостей футбольних матчів для прогнозування результатів спортивних подій.

Формування параметрів. Результат футбольного матчу є наслідком взаємодії певних параметрів. Після детального аналізу складових футбольного спорту було обрано 25 вхідних параметрів та побудовано кореляційну матрицю з метою виявлення тих факторів, які найменше впливають на результат. Із 25 параметрів було усунено 2 фактори, які майже не чинили впливу на кінцевий результат.

Таким чином, задля зручності побудови моделей інтелектуального аналізу різної складності було сформовано дві множини параметрів — із 23 та 13 параметрами.

При формуванні вхідної вибірки використано статистичні дані останніх 60 матчів Англійської Прем'єр-ліги сезону 2015–2016.

Застосування методу k-середніх. Метод k-середніх є методом кластерного аналізу, ціллю якого є розділ m спостережень з простору на k кластерів, при чому кожне спостереження відноситься до того кластеру, до центра (центроїда) якого воно є найближчим.

Оскільки кожен матч характеризується низкою параметрів, які є взаємопов'язаними між собою, то

застосування методу k-середніх дає змогу прослідкувати, чи є доцільним розбити схожі між собою матчі на певні угруповання або, іншими словами, кластери. Доцільно постановити кількість кластерів $k = 3$ у відповідності до кількості можливих результатів — «перемога першої команди», «нічия» та «перемога другої команди».

Метод k-середніх правильно розпізнав у кращому випадку 28 з 60 результатів, що демонструє точність 47%.

Застосування «наївного» байєсовського класифікатора. НБК базується на теоремі Байєса та є однією з найбільш відомих технік класифікації у машинному навчанні, однак одним з головних обмежень є припущення про незалежність ознак.

При застосуванні «наївного» байєсовського класифікатора кожному запису навчальної вибірки необхідно співставити ярлик або, іншими словами, клас, що відповідає опису результату конкретного матчу. У даному випадку ярлики прийматимуть три значення: перемога першої команди, нічия, перемога другої команди.

Після навчання НБК на тестовій вибірці, було проведено тестування на тестовій вибірці. Точність прогнозування склала 33%. Можна зробити висновок, що саме реальна взаємозалежність між параметрами вплинула на точність прогнозу.

Таблиця 1

Точність прогнозів, отриманих за допомогою нейронних мереж

Вид мережі	Кількість вхідних параметрів	Кількість прихованих шарів	Точність прогнозу результату матчів, %	Точність прогнозу тоталів матчів, %
Нейронна мережа прямого поширення	23	2	42,8	–
	13		57,1	71,4
Нейронна мережа Елмана	23	2	42,8	–
	13	2	57,1	71,4
		4	–	42,8
Каскадна нейронна мережа	23	4	28,5	–
	13		57,1	71,4
Нейронна мережа із двома автоенкодерами	23	3	71,4	–
	13		66,6	–

Застосування штучних нейронних мереж.

У роботі було реалізовано та досліджено 4 типи нейронних мереж із різною топологією:

- нейронна мережа прямого поширення. Нейрони організовані в пошарову структуру з прямою передачею сигналу [1].
- нейронна мережа Елмана. Є т. зв. рекурентною мережею, яка створюється наступним чином: у багатшаровий перцептрон вводять зворотні зв'язки, що йдуть від виходів внутрішніх нейронів.
- каскадна нейронна мережа. Каскадна нейронна мережа є подібною до нейронних мереж прямого поширення, але основна відмінність полягає у тому, що її структура передбачає зв'язок між вхідними параметрами та кожним шаром мережі [2].
- глибока нейронна мережа із застосуванням автоенкодерів. Нейронна мережа із модифікованою топологією, що передбачає послідовне використання двох автоенкодерів для попереднього навчання глибокої нейронної мережі без вчителя. Модифікація дозволяє однаково ефективно навчати усі шари мережі, а не лише кінцеві.

Кожна з побудованих нейронних мереж демонструвала мінімальну точність прогнозу 42,8%, окрім каскадної, мінімальна точність якої склала 28,5%. Найвищу точність прогнозування результатів спортивних матчів було досягнуто за допомогою розробленої нейронної мережі з двома енкодерами, а саме – 71,4%.

Точності побудованих моделей чотирьох нейронних мереж занесено у таблицю 1.

Висновки. Отже, було проведено аналіз доцільності та ефективності використання методів ІАД у контексті прогнозування результатів спортивних подій, який показав, що використання методів кластеризації дають значно нижчу точність прогнозування, ніж нейронні мережі, та що застосування методу «наївного» байєсівського класифікатора є найменш ефективним серед усіх розглянутих підходів. Окрім того, було виділено методи, що дають найвищу точність прогнозу, а саме – нейронні мережі Елмана та розроблена модифікована мережа прямого поширення із використанням двох автоенкодерів, які дозволили досягти максимального відсотку правильно спрогнозованих результатів – 71,4%.

Література

1. Штучні нейронні мережі [Електронний ресурс]. – Режим доступу: http://www.victoria.lviv.ua/html/neural_nets/Lecture1.htm.
2. Badde S. Cascade and Feed Forward Back propagation Artificial Neural Network Models for Prediction of Compressive Strength of Ready Mix Concrete [Електронний ресурс] / S. Badde, A. Gupta, K. Patki // Second International Conference on Emerging Trends in Engineering (SICETE). – 2009. – Режим доступу: [http://www.iosrjournals.org/iosr-jmce/papers/sicete\(-civil\)-volume3/26.pdf](http://www.iosrjournals.org/iosr-jmce/papers/sicete(-civil)-volume3/26.pdf)