

## 6.1 Machine Learning and Data Science

In supervised learning, particularly linear regression, the pseudo-inverse provides a closed-form solution for the optimal weights for design matrix and target vector. This is especially useful when the design matrix is not full rank or when the number of features exceeds the number of samples (underdetermined systems). Although gradient-based methods dominate large-scale learning, the pseudo-inverse remains essential for understanding model behavior and for initializing iterative algorithms. In applications like collaborative filtering (e.g., recommendation systems), the pseudo-inverse aids in estimating missing entries by leveraging low-rank structure via SVD.

## 6.2 Signal Processing and Control Theory

In signal reconstruction and system identification, the pseudo-inverse helps recover signals from incomplete or noisy measurements. For example, in compressed sensing, it can be used to approximate solutions when sparsity constraints are relaxed. In control theory, the pseudo-inverse is used to compute the least-squares solution for actuator commands in redundant robotic systems, ensuring minimal energy usage or joint movement.

## 6.3 Quantum Computing and Physics

In quantum mechanics, the pseudo-inverse appears in the context of generalized measurements and state reconstruction. It is used to define the inverse of non-unitary operators and to solve systems arising from quantum tomography. The spectral perspective via SVD aligns well with the decomposition of quantum states and operators, providing a bridge between linear algebra and quantum theory.

## 6.4 Graph-Based Learning and Network Analysis

The algebraic graph theory perspective opens doors to applications in network science. Effective resistance, derived from the pseudo-inverse of the Laplacian, serves as a metric for connectivity and robustness. In semi-supervised learning on graphs, the pseudo-inverse helps propagate labels across nodes by solving Laplacian-based optimization problems.

## Conclusion

The Moore-Penrose pseudo-inverse is more than a computational tool — it is a conceptual bridge linking diverse mathematical domains. Whether viewed through the lens of axioms, optimization, regularization, spectral theory, or graph combinatorics, its versatility and depth make it indispensable in both theory and practice. As data-driven disciplines continue to evolve, the pseudo-inverse remains a cornerstone for solving systems that defy classical assumptions.

### **ГРА З НЕПОВНОЮ ІНФОРМАЦІЄЮ НА ПРИКЛАДІ СПОРТИВНОГО БЕТІНГУ / A GAME WITH INCOMPLETE INFORMATION: THE CASE OF SPORTS BETTING**

*Кущалаба Н.В., Чорней П.К. / Kutsalaba N., Chornei R.* / This work was partly supported by a grant from the Simons Foundation International [SFI-PD-Ukraine-00014577, O.G.]

Національний університет “Києво-Могилянська Академія” / National University of Kyiv-Mohyla Academy

04655, Київ, вул. Григорія Сковороди, 2, факультет інформатики, кафедра математики  
E-mail: n.kutsalaba@ukma.edu.ua, r.chornei@ukma.edu.ua

This study presents a predictive framework for football match outcomes that integrates the Elo rating system, multiple linear regression, and a Bayesian game-theoretic interpretation of sports betting under incomplete information. Match statistics and bookmaker odds were collected via web scraping from Sofascore. The regression model achieved an adjusted  $R^2 = 0.3493$ , with significant predictors including shots on target, total chances, and total shots. Bookmaker accuracy was 22.37% for over/under goals and 28.33% for win/loss outcomes, revealing market inefficiency. Modeling the interaction between the bettor and the bookmaker as a Bayesian game with asymmetric information allows identifying value bets and optimizing decision-making strategies under uncertainty.

У сучасному спортивному аналітичному середовищі прогнозування результатів матчів є однією з ключових задач, що поєднує методи статистики, машинного навчання та теорії ігор.

Спортивний беттинг, зокрема футбольний, може розглядатися як взаємодія раціональних агентів — команд, аналітиків та букмекерів — які приймають рішення в умовах неповної та асиметричної інформації. У даному дослідженні запропоновано поєднати класичні статистичні методи прогнозування з теоретико-ігровою інтерпретацією через концепцію Баєсівських ігор (Bayesian games).

Через обмеженість наявних відкритих датасетів дані були зібрані самостійно з платформи *Sofascore* за допомогою технологій веб-скрапінгу [1]. Отримані масиви містять статистику матчів, коефіцієнти букмекерів і показники гравців, що дозволяє проводити комплексний кількісний аналіз.

Для первинного прогнозування використовувався рейтинг Elo, розроблений Арпадом Ело для оцінки сили гравців у шахах [2]. Цей рейтинг забезпечує можливість не лише ранжувати команди, але й оцінювати ймовірності результатів матчів. Його базова форма задається як

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}},$$

де  $E_A$  — очікуване число очок для команди  $A$ ,  $R_A$ ,  $R_B$  — попередні рейтинги команд. Новий рейтинг визначається за формулою

$$R'_A = R_A + K(S_A - E_A),$$

де  $K$  — коефіцієнт регуляції,  $S_A$  — набрані очки.

Для аналізу факторів, що впливають на кількість забитих голів, застосовано множинну лінійну регресію [3]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon.$$

У нашій моделі прогнозованою змінною ( $Y$ ) була *total\_goals*, а найбільш інформативними предикторами виявилися *total\_on\_target*, *total\_chances* та *total\_shots*. Отримана модель:

$$total_{goals} = 1.40 + 0.336 \cdot total_{\square} + 0.172 \cdot total_{chances} - 0.059 \cdot total_{shots},$$

із скоригованим коефіцієнтом детермінації  $R_{adj}^2 = 0.3493$ , що пояснює близько 35% варіації результату. Незалежність змінних підтверджується низькими значеннями VIF (<2), що свідчить про відсутність мультиколінеарності.

Було також перевірено точність букмекерських прогнозів. Для найпоширеніших тоталів (>2.5, >3.5) середня точність склала 22.37%, а для передбачення результатів *win/loss* — 28.33%. Це свідчить про наявність значного рівня непередбачуваності у футбольних матчах.

У моделі Баєсівської гри букмекер володіє приватною інформацією  $\theta_B \in \Theta$  про точність власної моделі, а гравець спостерігає лише публічний сигнал — коефіцієнти  $k = (k_W, k_D, k_L)$ . Ймовірність, імплікована коефіцієнтами, обчислюється як  $p_{implied} = 1/k_i$ , а value bet визначається умовою  $p_{model} > p_{implied}$ , що сигналізує про потенційну неефективність ринку [4].

Подальший аналіз включатиме перевірку гіпотези про ринкову ефективність  $H_0: E[Value] = 0$ , а також застосування критерію Келлі для оптимального вибору розміру ставки

$$f^{\square} = \frac{p \cdot k - 1}{k - 1},$$

і подальше тестування на історичних даних.

Отримані результати підтверджують можливість формалізації спортивного беттингу як Баєсівської гри та відкривають перспективи для побудови ефективних стратегій прогнозування і керування ризиками на основі ймовірнісних моделей.

### Список літератури

1. Як використовувати web scraper для збору даних з інтернету з Python. URL: <https://dou.ua/forums/topic/43070/> (дата звернення: 23.09.2025).
2. Winston W. L., Pelechrisis K., Nestler S. *Mathletics: How Gamblers, Managers, and Fans Use Mathematics in Sports*. Princeton University Press, 2022. 432 p.
3. *Simple Linear Regression. Linear Models in Statistics*. Hoboken, NJ, USA, pp. 127–136.
4. Zou Q., Song K., Shi J. *A Bayesian In-Play Prediction Model for Association Football Outcomes*. Applied Sciences, 2020, Vol. 10, No. 8, p. 2904.