

## ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ФУТБОЛЬНИХ МАТЧІВ

Більшість подій можливо спрогнозувати і футбольний матч не є виключенням з правил. Прогнозування – це процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного.

У футболі для аналізу існує дуже велика кількість факторів, які у різній мірі впливають на перебіг матчу. Більшість команд зараз надають доступ лише до двох координат. Тобто футбольне поле представляється двовимірним об'єктом, яким пересуваються точки. Звичайно, було б цікаво здобути третю координату – як високо летить м'яч, яким було положення тіла гравця у просторі під час дії.

Одним із підходів, який ми використовуємо у проєкті, є data-blending: ми об'єднуємо дані з різних джерел для того, щоб зібрати максимальний набір метрик та вимірювань для побудови наскрізної аналітики. Тобто збираємо дані з різних відкритих джерел та агрегуємо їх у єдину базу даних, яка є ядром нашого проєкту. Після збору ми обробляємо та фільтруємо дані. Далі вони завантажуються на сервер і ми візуалізуємо ці дані в інтуїтивно зрозумілій і зручній оболонці Tableau. Тільки після цього відбувається навчання математичних моделей для пошуку відповідей на різні питання та перевірки гіпотез.

На етапі збору та парсингу даних ми використовуємо Python, зокрема бібліотеки BeautifulSoup та Selenium. Зібравши усі дані в один файл, буде створено датасет. Згодом, цей датасет буде використано при створенні моделі прогнозування переможця матчу. Наступним кроком буде створення вектору кожної команди за сезон та почергове порівняння векторів з урахуванням важливості факторів. Порівнявши вектори між собою, буде виведено 2 числа, їх різниця буде результатом прогнозу.

Розглянемо відомий аналог kickoff.ai, який використовує Баєсовський вивід для прогнозування переможця матчу. Суть баєсовського висновку в тому, щоб об'єднати два різні розподіли (ймовірності та апіорний) у більш «розумний» (апостеріорний). Ми використовуємо апостеріорну можливість для пошуку «найкращих» властивостей; «найкращих» у термінах оцінки апостеріорного максимуму з урахуванням даних. Цей процес називається оцінка апостеріорного максимуму. Оптимізація, що використовується в оцінці апостеріорного максимуму, та сама, що і в типовому машинному навчанні (градієнтний спуск, метод Ньютона і так далі). Формула баєсовського виводу:

$$P\left(\frac{A}{B}\right) = \frac{P(B/A)P(A)}{P(B)} \quad (1)$$

Також, при створенні прогнозу була використана лінійна регресія у статистиці, лінійна регресія – це метод моделювання залежності між скалярною змінною Y та векторною (у загальному випадку) змінною X. У разі, якщо змінна X також є скаляром, регресію називають простою. При використанні лінійної регресії взаємозв'язок між даними моделюється за допомогою лінійних функцій, а невідомі параметри моделі оцінюються за вхідними даними. Подібно до інших методів регресійного аналізу, лінійна регресія повертає розподіл умовної імовірності Y в залежності від X, а не розподіл спільної імовірності Y та X, що стосується області мультиваріативного аналізу.

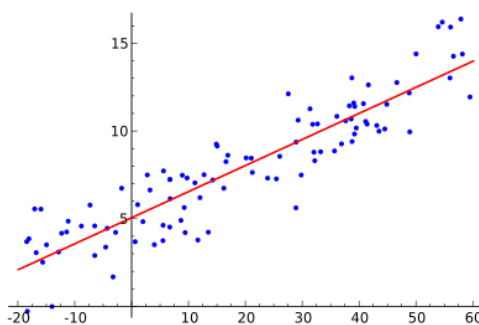


Рис. 1. Лінійна регресія

Таким чином, використання машинного навчання для прогнозування перебігу футбольних матчів може бути ефективним, при наявності великої кількості вхідних даних та правильному підборі алго-ритму.

### Список використаних джерел

1. Лінійна регресія у машинному навчанні – Режим доступу: <https://machinelearningmastery.com/linear-regression-for-machine-learning/>.
2. Основи машинного навчання – Режим доступу: <https://travelscode.com/osnovi-mashinnogo-navchannya>