

Использование машинного обучения для анализа и прогнозирования результатов футбольных матчей

Ван Цюаньюй (Университет ИТМО),

Научный руководитель-доцент факультет ИКТ, к.т.н. Добреко Н.В.

(Университет ИТМО)

Введение. Используя популярную тему «Чемпионата мира» в качестве отправной точки, идея данной работы — это предоставлять любителям футбола по всему миру информацию и предложение по прогнозированию результатов матчей. В данной работе использовались полные результаты последнего крупного международного футбольного турнира - Чемпионата мира в Катаре в качестве ориентира для результатов прогнозирования классификатора машинного обучения.

Основная часть. Все данные, необходимые для этой работы, были собраны и интегрированы с официального сайта FIFA [1], kaggle [2] и Opta Analyze [3]. Для прогноза результатов матчей используются 5 различных классификаторов:

- 1) Логистическая регрессия (Logistic Regression);
- 2) kNN (KNeighborsClassifier);
- 3) Метод опорных векторов SVM (SVC);
- 4) Метод случайного леса (Random Forest Classifier);
- 5) Ансамбль моделей классификации (VotingClassifier).

Проведен анализ и сравнение полученных результаты с фактическими результатами. В настоящее время, общая точность прогнозов пяти классификаторов для группового этапа чемпионата мира в Катаре составляет 72,91%, общая точность прогнозов на 1/8 финала составила 68,75%, общая точность прогнозов на 1/4 финала составила 75%, общая точность прогнозов на полуфинала составила 50%, и классификатор логической регрессии правильно прогнозировал победителя и проигравшего в финале.

Далее в работе будет добавлено больше функций данных, а также скорректированы соответствующие параметры каждого классификатора и т.д. Параметры некоторых классификаторов будут заимствованы из литературы в Scopus [4]. В результате планируется обучить модель прогнозирования, наиболее близкую к фактическим результатам матча.

Выводы. Благодаря добавлению большего количества исторических данных о матчах и функций данных, а также постоянной корректировке параметров классификатора, уровень точности прогнозирования результатов матчей улучшился по сравнению с предыдущим.

Список использованных источников:

1. «Рейтинг мужских национальных сборных по футболу на 6 октября 2022 года» // ФИФА [Электронный ресурс]: URL: <https://www.fifa.com/fifa-world-ranking/men?dateId=id13792> (дата обращения 17.02.2023).

2. «Международные футбольные результаты с 1872 по 2022 год» // Kaggle [Электронный ресурс]: URL: <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017?select=results.csv> (дата обращения 17.02.2023).
3. «Статистика чемпионата мира: Катар 2022» // Opta Analyze [Электронный ресурс]: URL: https://theanalyst.com/eu/2022/11/world-cup-stats-qatar-2022/?tab=fixtures&_gl=1*esn0ei*_ga*MTk2NjM3ODY2Mi4xNjcwMDc2NTI4*_ga_BGFPTYQE1X*MTY3NDA2MTg4My41Ny4xLjE2NzQwNjM1MzQuNjAuMC4w (дата обращения 17.02.2023).
4. Arntzen H., Hvattum L.M. Predicting match outcomes in association football using team ratings and player ratings [Текст] // Statistical Modelling 21(5), pp.449-470.