МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра «Прикладная информатика»

Форма обучения: очная

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

по направлению <u>09.03.03 «Прикладная информатика»</u>
на тему <u>«Прогнозирование результатов киберспортивных матчей в игре</u>
<u>«Rocket league» с применением технологий больших данных»</u>

Студент: Алексей Дмитриевич Перехожих

Руководитель работы: доцент, к.э.н. Светлана Викторовна Куликова



Цель и задачи работы

Целью данной выпускной квалификационной работы является прогнозирование результата предстоящих киберспортивных матчей в игре «Rocket league», используя статистические данные прошедших игр.

Актуальность данной ВКР заключается в том, что в настоящий момент, несмотря на развитее сферы киберспорта и широкий интерес к ней, почти все прогнозы основываются на мнении и знаниях людей, которые являются субъективными, зачастую недостаточно точны или неполны. Немногочисленные разработанные автоматизированные решения применяются не во всех играх и как правило являются частными проектами с закрытым или ограниченным доступом.

В соответствии с целью были поставлены следующие задачи:

- Провести анализ предметной области, определив существующие инструменты и подходы к прогнозированию в киберспорте.
- Найти и подготовить исходные данные, на которых основывается исследование.
- Обработать данные, очистив от выбросов и неопределенных событий.
- Выбрать наилучшие модели для прогнозирования исходя из специфики входных данных, а также адаптировать их под особенности игровой дисциплины.
- Разработать программу, технически реализующую выбранные модели.
- Оценить точность моделей и получаемых прогнозов.
- На основе оценки модели, выбрать наилучшую, с ее помощью спрогнозировать результат предстоящих турниров.
- Сравнить прогноз с реальными значениями для турниров, которые завершатся до конца выполнения ВКР



Характеристика киберспорта

Киберспорт - область спортивной деятельности, в которой люди развивают и тренируют свои ментальные и физические возможности, используя информационные и коммуникативные технологии.

- Официально признан видом спорта в России и других странах
- Альтернативный вид соревнований во время пандемии COVID-19
- Участие крупных компаний в качестве спонсоров
- Наличие турниров с большим призовым фондом
- Большая зрительская аудитория
- Доступный вид спорта







Методы сбора статистической информации

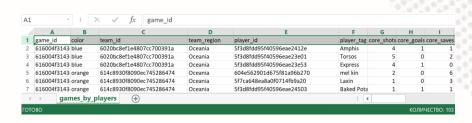


Бумажный протокол

- Рукописные записи
- Бумажные протоколы
- Электронные протоколы
- Видео-запись
- Реплей



Электронный протокол



Обработанные реплеи

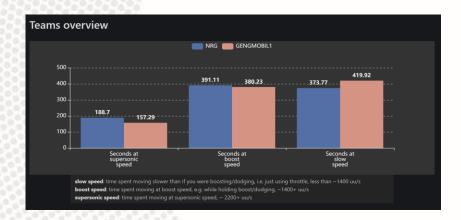


Файлы с реплеями



Big data в киберспорте

Date ↑↓	xGF ↑↓	xGC ↑↓	GF ↑↓	GC ↑↓	HF ↑↓	HC ↑↓	xGD ↑↓	GD ↑↓	HD ↑↓	P(Win) ↑↓	P(Score) ↑↓	Luck % 🗇	Res †
021-07-24T21:47	3.06	2.04	3	2	73	78	1.03	1	-5	70.13	10.27	29.87	W
021-07-24T21:36	3.35	6.11	2	5	65	76	-2.76	-3	-11	12.25	4.1	-12.25	L
021-07-24T21:28	7.25	7.01	3	2	90	113	0.24	1	-23	53.55	0.01	46.45	W*
021-07-24T21:20	2.25	5.7	1	3	64	86	-3.45	-2	-22	6.52	2.23	-6.52	L
021-07-24T21:13	2.55	4.55	3	4	54	80	-2	-1	-26	17.3	7.4	-17.3	L
021-07-24T21:06	3.65	6.18	2	5	58	90	-2.54	-3	-32	14.56	3.38	-14.56	L
021-07-24T20:58	3.88	1.14	3	0	35	47	2.74	3	-12	95.23	8.34	4.77	W
021-07-24T20:53	2.93	4.34	1	5	64	95	-1.41	-4	-31	26.49	3.15	-26.49	L
021-07-24T20:46	3.4	8.37	2	4	57	121	-4.98	-2	-64	3.04	0.49	-3.04	L
021-07-24T20:39	4.38	6.46	3	6	72	80	-2.08	-3	-8	19.84	4.46	-19.84	L



Проекты по машинному обучению

Детализированная статистика



Использование для маркетинга



Теоретическая часть

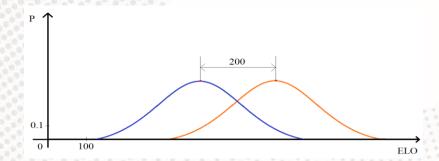
Математическая постановка задачи:

Для решения задачи необходимо для входящих данных, которые представлены в виде n- количества строк, где каждая строка $M: m_1, m_2, ..., m_n$ представляет собой матч с j - набором параметров $x_1, x_2, ..., x_j$, определить принадлежность к одному из двух классов значения Y для каждой из m_n -строк, т.е. результат матча $R: M \to Y$.

Рейтинг ЕІо

$$R' = R + K * (S - E)$$

- R' новый рейтинг команды
- R текущий рейтинг команды
- *К* коэффициент (равен максимально возможному количеству очков, которое может за матч набрать или потерять команда)
- E ожидаемый результат игры



$$E_a = \frac{1}{1+10^{(\frac{R_b - R_a}{400})}}$$

- R_b текущий рейтинг команды В (синей команды)
- R_a текущий рейтинг команды A (оранжевой команды)
- E_a ожидаемая вероятность победы команды А



Рейтинг Gliko

$$RD = \min\left(\sqrt{RD_{old}^2 + c^2t}, 350\right)$$

$$r' = r + \frac{q}{\frac{1}{RD^2} + \frac{1}{d^2}} \sum_{j=1}^{m} g_j (s_j - E_j)$$

$$RD^{\hat{}} = \sqrt{(\frac{1}{RD^2} + \frac{1}{d^2})^{-1}}$$

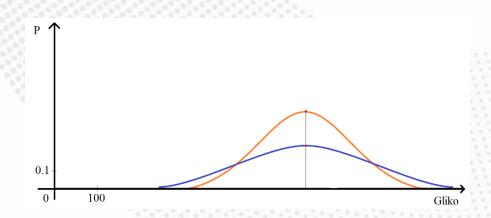
$$E_j = \frac{1}{1+10^{-g_j(r-r_j)/400}}$$

$$q = \frac{\ln 10}{400} = 0.0057565$$

$$g_j = \frac{1}{\sqrt{1 + 3q^2 (RD_j^2)/\pi^2}}$$

$$d^{2} = (q^{2} \sum_{j=1}^{m} g_{j}^{2} E_{j} (1 - E_{j}))^{-1}$$

- RD достоверность рейтинга
- RD` измененная достоверность рейтинга
- RD_i достоверность рейтинга соперника
- r текущий рейтинг команды
- r` измененный рейтинг команды
- r_i рейтинг соперника
- s_i результат игры (1 победа, 0 поражение)
- E_i вероятность победы против j-ой команды



Оптимизация методов под особенности соревнований

Главная особенность состоит в режиме игры 3 на 3, где игроки не имеют своих однозначных ролей, а успешность их игры в одном конкретном матче отражает количество полученных очков, которые учитывают голы, удары, спасения, передачи, касания и другие игровые события.

$$K_a = \frac{P_a}{P_a + P_b + P_c}$$

- K_a коэффициент вклада игрока A $\overline{\mathbf{K}}_a$ коэффициент ожидаемого вклада игрока A
- P_a, P_b, P_c количество набранных очков R_a, R_b, R_c рейтинг игроков a, b, c за матч игроками а, b, с
- $\overline{K}_a = \frac{R_a}{R_a + R_b + R_c}$

Если команда заработала за матч 25 очков рейтинга, количество очков за матч у первого игрока было 400, у двух других по 350, то определяя, что первый игрок получает 25*1.09 = 27.25 очков. Его партнеры по команде получают 23.875 очка каждый. При разных начальных значениях рейтинга используется соотношение K_a / \overline{K}_a таким образом, чтобы сумма коэффициентов была равна 3.



Инструменты и технологии

Инструменты:

- Python
- JSON и CSV модули
- Time и Requests модули
- Openpyxl модуль
- MS Excel

Технологии:

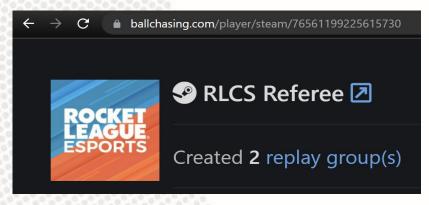
- Веб-скрейпинг
- Парсинг
- API (Application Programming Interface)
- Нереляционные базы данных (No SQL)
- Ограниченный искусственный интеллект (Narrow AI)



Получение и обработка данных

```
map name : Mannfield (Night)
  playlist_id: private
  playlist name: Private
  duration: 351
  overtime: false
  season: 10
  season_type: free2play
  date: 2023-04-30T13:16:17+02:00
  date has tz: ✓ true
  visibility : public
  created: 2023-04-30T11:16:18.009005Z
▶ uploader {4}
▶ groups [1]
▼ blue {2}
     name: INVISBLE NINJA
   ▼ players [3]
     ▼ 0 {6}
           start_time: 0
           end time: 343.46704
           name: scruffy
```

Информация о матче в формате JSON



Источник данных



Пример обработанных данных



Создание программы

Общие составляющие:

- Словарь Teams
- Словарь PlayersTeams

Функции:

- teamCheck()
- playerCheck()
- rosterMove()
- newTeam()
- playerSort()
- playerChangePlusRating()
- playerImpact()
- playerExpectedImpact()
- predict()

Для Elo:

- gamePlayed()
- chanceWin()

Для Gliko:

- nowRD()
- daysFromStart()
- glikoGamePlayed()
- glikoChanceWin()
- teamRD()
- glikoNewRD()
- glikoRateChange()



Оценка результатов

Сравнение метрик оценки результата наилучших моделей

	10101010101010101	01010101010101010101		
Показатель	ELO, K_CONST = 47, Gamma = 2	Gliko, Epoch = 11, C_RD = 59, Gamma = 2	Gliko, Epoch = 1, C_RD = 36, Gamma = 2	
accuracy	0.629	0.630	0.641	
precision	0.637	0.645	0.651	
recall	0.635	0.634	0.646	
L	0.300	0.300	0.297	
log-loss	0.644	0.642	0.650	

Для прогнозирования выбрана модель Gliko с параметрами Epoch = 1, C_RD = 36, Gamma = 2, так как она отличается наилучшей точностью и низким значением коэффициента Тейла.



Прогнозирование кубка Америки и его результаты

```
Первый раунд верхней сетки---
Шанс победы команды COMPLEXITY VS TEAM AXLE = 0.8661299840655232
Шанс победы команды REBELLION VS VERSION1 = 0.05724224732683701
Шанс победы команды G2 VS KOI = 0.657655336129398
Шанс победы команды OPTIC VS M80 = 0.6558048014780348
Шанс победы команды FAZE CLAN VS ZERO2ONE = 0.9407493078550504
Шанс победы команды DIGNITAS VS FURIA = 0.27007398860003634
Шанс победы команды GENGMOBIL1 VS HEY BRO = 0.8064699353021493
Шанс победы команды SPACESTATION VS NRG = 0.55237299755865
Первый раунд нижней сетки-----
 Шанс победы команды TEAM AXLE VS REBELLION = 0.3219629188825459
Шанс побелы команлы KOI VS M80 = 0.5409770471496328
 Шанс победы команды ZERO2ONE VS DIGNITAS = 0.38785575120040217
Шанс победы команды HEY BRO VS NRG = 0.2226652393751229
Второй раунд верхней сетки-----
Шанс победы команды COMPLEXITY VS VERSION1 = 0.15721242184072656
Шанс победы команды GENGMOBIL1 VS REBELLION = 0.6962272309547655
Шанс победы команды FURIA VS KOI = 0.6736833360913197
Шанс победы команды OPTIC VS DIGNITAS = 0.6791205044828137
Шанс победы команды COMPLEXITY VS NRG = 0.6153922974254384
Третий раунд нижней сетки---
Шанс победы команды GENGMOBIL1 VS FURIA = 0.6039797964115982
Шанс победы команды OPTIC VS COMPLEXITY = 0.27695953730957024
Полуфинальный раунд верхней сетки----
Шанс победы команды VERSION1 VS G2 = 0.9217399196290075
Шанс победы команды FAZE CLAN VS SPACESTATION = 0.7024539658949336
Шанс победы команды G2 VS GENGMOBIL1 = 0.37892714198764366
Шанс побелы команлы SPACESTATION VS COMPLEXITY = 0.43541954764520246
Шанс победы команды GENGMOBIL1 VS COMPLEXITY = 0.42726860732141847
Финальный рауннд нижней сетки-----
Шанс победы команды FAZE CLAN VS COMPLEXITY = 0.6454819695618651
 Шанс победы команды VERSION1 VS FAZE CLAN = 0.7464710799345582
```

Прогнозирование сетки целиком

Такой подход набрал целых 112 баллов, угадав победителя турнира, топ-2 и 3 команды из топ-4, уступив всего одному прогнозу из всех участвующих пользователей. сравнения, прогноз без использования программы набрал лишь 84 балла.

Прогнозирование по раундам

В таком варианте модель правильно предсказала исход 26 из 30 игр, ошибившись только во втором раунде верхней и нижней сетки.





Top 12

Top 4

Прогноз на ближайший Европейский турнир

```
Шанс победы команды KARMINE CORP VS SOLARY = 0.99
Шанс победы команды PSG TUNDRA VS GUILD = 0.56178024774475
Шанс победы команды BDS VS WILLIAMS RESOLVE = 0.99
Шанс победы команды ОХУGEN VS LUNA GALAXY = 0.5756638074574217
Шанс победы команды VITALITY VS MONACO = 0.99
Шанс победы команды G1 VS SUHHH = 0.5056687347727333
Шанс победы команды LIQUID VS FUFAXDOP = 0.718516520660638
Шанс победы команды MOIST VS HOGAN MODE = 0.637589114129297
Первый раунд нижней сетки-----
Шанс победы команды SOLARY VS GUILD = 0.34389702617432494
Шанс победы команды WILLIAMS RESOLVE VS LUNA GALAXY = 0.47021321721437725
Шанс победы команды MONACO VS SUHHH = 0.01
Шанс победы команды FUFAXDOP VS HOGAN MODE = 0.42984379459136896
Второй раунд верхней сетки-----
Шанс победы команды KARMINE CORP VS PSG TUNDRA = 0.650012611961802
Шанс победы команды BDS VS OXYGEN = 0.6489681658044487
Шанс победы команды VITALITY VS G1 = 0.99
Шанс победы команды LIQUID VS MOIST = 0.4899361234873885
Второй раунд нижней сетки-----
Шанс победы команды LIOUID VS GUILD = 0.6273640757119854
Шанс победы команды G1 VS LUNA GALAXY = 0.522264676943076
Шанс победы команды ОХУGEN VS SUHHH = 0.5575403945056717
Шанс победы команды PSG TUNDRA VS HOGAN MODE = 0.5599035332010953
Третий раунд нижней сетки-----
Шанс победы команды LIQUID VS G1 = 0.6356959431391769
Шанс победы команды ОХҮGEN VS PSG TUNDRA = 0.48391857400877475
Полуфинальный раунд верхней сетки-----
Шанс победы команды KARMINE CORP VS BDS = 0.5175354355471857
Шанс победы команды VITALITY VS MOIST = 0.5695842416038271
Четвертый раунд нижней сетки-----
Шанс победы команды BDS VS LIQUID = 0.5634494776714338
Шанс победы команды MOIST VS PSG TUNDRA = 0.5712091222907669
Полуфинальный раунд нижней сетки-----
Шанс победы команды BDS VS MOIST = 0.552139724573913
Финальный раунд верхней сетки-----
Waнc победы команды KARMINE CORP VS VITALITY = 0.5015925836988275
Финальный рауннд нижней сетки-----
Шанс победы команды VITALITY VS BDS = 0.5158893985358626
Гранд Финал-----
Шанс победы команды KARMINE CORP VS VITALITY = 0.5015925836988275
```

Прогноз с использованием рейтинга Gliko:

- 1 место: Karmine Corp
- 2 место: Vitality
- 3 место: BDS

Результаты моделирования турнира 100000 раз:

Шансы на победу в турнире

- Vitality 32.4%
- Karmine corp 27,3%
- BDS 21,2%

{'KARMINE CORP': 0.27327, 'VITALITY': 0.32396, 'LIQUID': 0.03978, 'BDS': 0.21183, 'OXYGEN': 0.02943, 'MOIST': 0.0793, 'G1': 0.00039, 'PSG TUNDRA': 0.03286, 'GUILD': 0.00145, 'SUHHH': 0.00042, 'HOGAN MODE': 0.00713, 'LUNA GALAXY': 5e-05, 'WILLIAMS RESOLVE': 3e-05, 'FUFAXDOP': 0.0001, 'MONACO': 0.0, 'SOLARY': 0.0}



Спасибо за внимание!

Прогнозирование результатов киберспортивных матчей в игре «Rocket league» с применением технологий больших данных»

Выпускная квалификационная работа
По направлению подготовки
09.03.03 Прикладная информатика
Образовательная программа «Большие и открытые данные»

Студент:

Алексей Дмитриевич Перехожих, 191-363

Руководитель ВКР:

Светлана Викторовна Куликова, к.э.н., доцент

