

Визуализация внешних данных

На наборе данных «Lichess»

Козловцев Константин

517 группа

Данные



«Lichess» — Набор результатов более чем 20 тысяч игр в шахматы.

Данные включают в себя:

- Ники и рейтинги игроков
- Время игры
- Ходы
- Победителя
- Название разыгранного дебюта
- Количество ходов

Визуализация

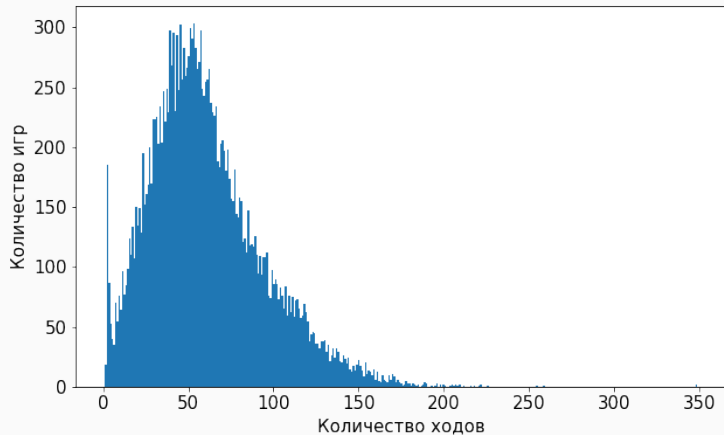


Рис. 1: Гистограмма количества ходов на которых завершается игра

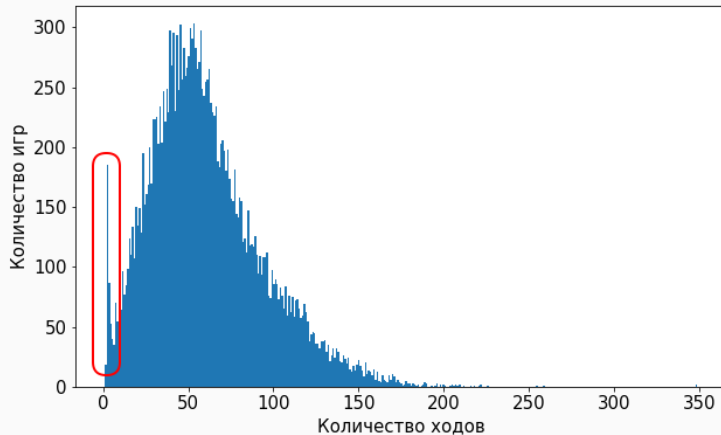


Рис. 2: Гистограмма количества ходов на которых завершается игра

	turns	victory_status
4828	1	resign
1946	1	resign
4115	1	outoftime
13557	1	outoftime
13560	1	outoftime
11742	1	resign
6040	1	outoftime
14117	1	outoftime
5875	1	resign
18583	1	resign
13701	1	resign
5496	1	outoftime
1977	1	outoftime
4473	1	outoftime
15007	1	outoftime
11963	1	outoftime
9229	1	resign
12679	1	outoftime

Правда ли что белые имеют преимущество?

Белые всегда начинают игру. Получают ли они при этом существенное преимущество в игре?

Например если мы представим игру в шахматы как подбрасывание монетки: орел — значит белый выигрывает (для простоты исключим паттовые игры), будет ли монетка чаще падать орлом?

Правда ли что белые имеют преимущество?

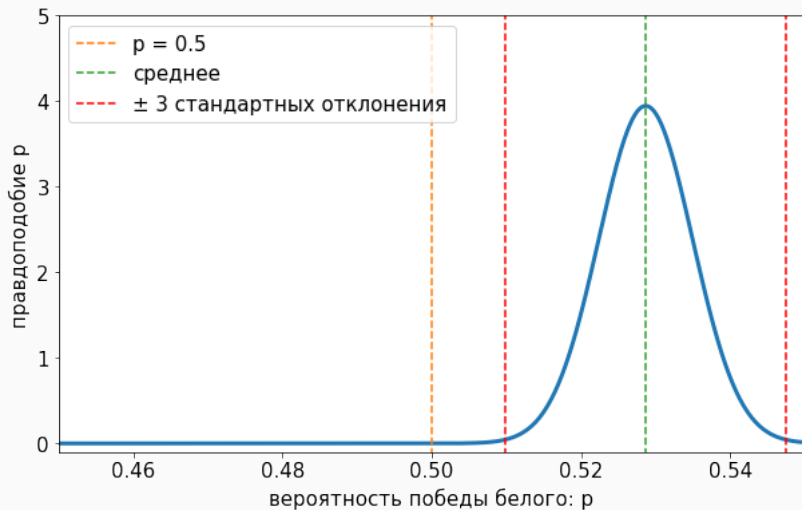


Рис. 3: Правдоподобие вероятности выигрыша белых

Шахматы во многом задаются началом игры — дебютами.

Интересно посмотреть какие дебюты дают большую вероятность выигрыша для каждого из игроков.

Дебюты

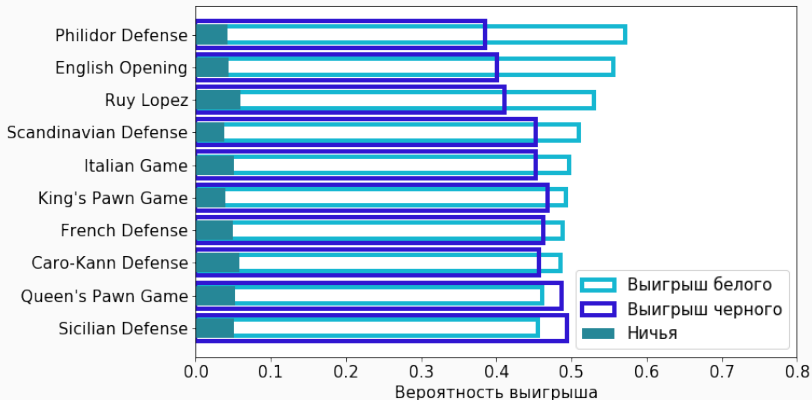


Рис. 4: Вероятность выигрыша при розыгрыше различных дебютов. Каждый дебют, представленный на графике, был разыгран хотя бы 500 раз

*На третьем ходу выяснилось, что гроссмейстер играет восемнадцать испанских партий. В остальных двенадцати черные применили хотя и устаревшую, но довольно верную **защиту Филидора**. Если б Остап узнал, что он играет такие мудреные партии и сталкивается с такой испытанной защитой, он крайне бы удивился. Дело в том, что великий комбинатор играл в шахматы второй раз в жизни.*

Еще немного графиков

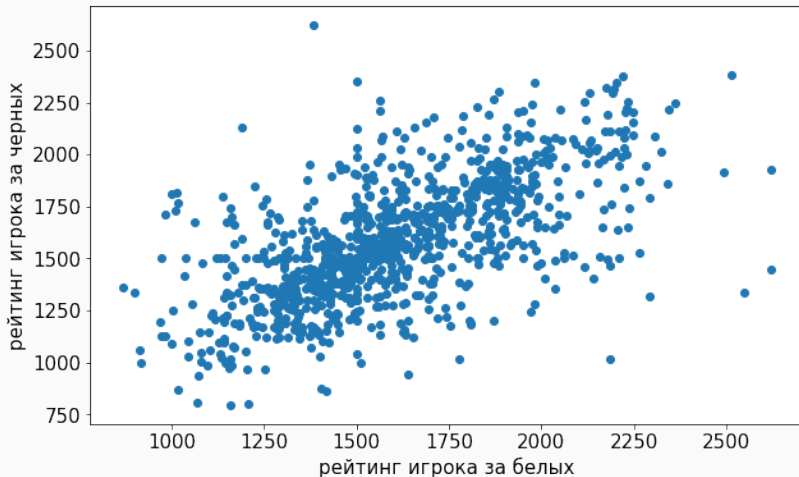


Рис. 5: Рейтинги игроков. Видно что система подбирает игроков с похожим рейтингом, но делает это не слишком тщательно

Еще немного графиков

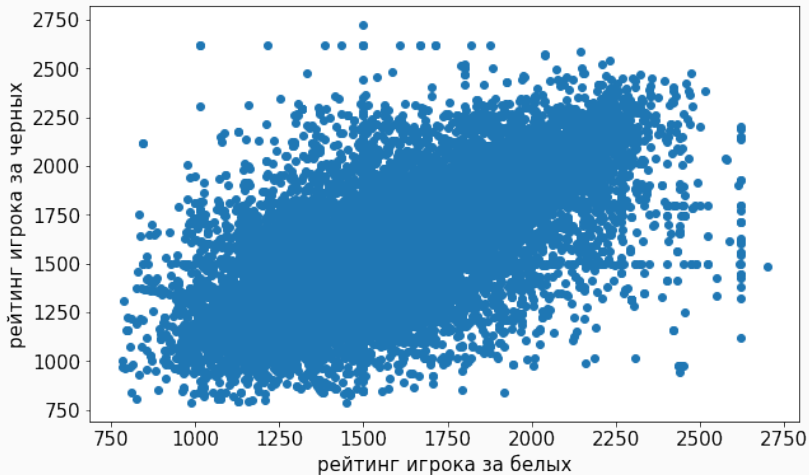


Рис. 6: То же самое, но для большего количества игроков. Сразу видно
завсегдаев

Списки постоянных посетителей

```
In [344]: data[data['white_rating'] > 2500][['white_id', 'black_id']]
```

```
Out[344]:
```

	white_id	black_id
904	lance5500	cookieblake
4426	lexisvar	kerboodle
4467	lance5500	stevenemily
5424	lance5500	artem555
5935	lance5500	ijh
6040	lance5500	justindong
8308	shahoviy_komentator	iordor
10147	blitzbullet	milancuruvija
11584	lance5500	thechessdetective
11600	king-bullet	anonymous_enemy
11602	lance5500	anonymous_enemy
11742	lance5500	antons
11808	lance5500	oguzkagangurer
11819	lance5500	oguzkagangurer
11949	lance5500	danny77

Рис. 7: причиной странной вертикальной прямой на графике выше является игрок lance5500

Рейтинг против winrate

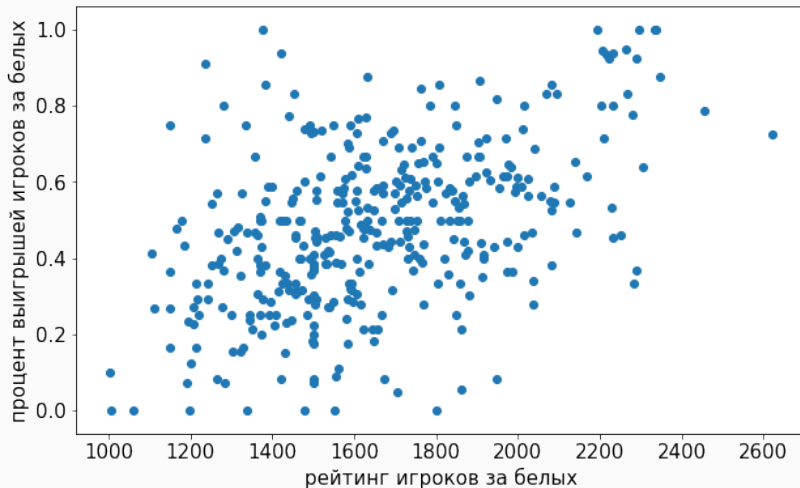


Рис. 8: Сравнение максимального внутриигрового рейтинга игроков и процента побед. Есть ощущение что зависимость минимальная.

О пользе рокировки

	рокировка черных	черные без рокировки
победа белых	0.463	0.564
победа черных	0.485	0.398

	рокировка белых	белые без рокировки
победа белых	0.511	0.469
победа черных	0.438	0.492

Таблица 1: Процент выигрыша в зависимости от того проводили ли игроки рокировку. Ясно видно что рокировка дает дополнительную защиту в игре и своевременное ее выполнение дает дополнительный шанс на успех

Заключение

Я поигрался с данными, нашел пару закономерностей, узнал как можно повысить свой шанс на успех, если не особо играешь в шахматы.

Также я постарался выжать по максимуму из данного набора данных, дальше только уже начинать обучать алгоритм играть, что выходит за пределы визуализации.

Спасибо за внимание!