

Задача 9. Определение опытного игрока Dota2

В игре [Dota 2](#) участвуют две стороны: команды сил Света (The Radiant) и сил Тьмы (The Dire). Каждая команда состоит из пяти игроков — персонажей со своими уникальными способностями. Перед началом матча каждый игрок выбирает себе одного героя из 116 возможных. Цель игры — разрушить трон, находящийся на базе противника.

Повышая игровые навыки, игроки повышают свой рейтинг и получают медали опыта. Эксперты по игре Dota 2 утверждают, что легко могут отличить матч новичка от игры опытного игрока. Ваша задача — построить алгоритм, который сможет определять опытность игрока по статистике из одного матча.



Медали новичка

Медали опытного игрока

Для решения задачи участникам предоставляется набор из почти 50'000 примеров матчей, уровень игрока в которых известен. Для каждого из $\approx 21'000$ “тестовых” примеров, участникам необходимо проанализировать данные о матче при помощи своих алгоритмов и дать ответ — опытный ли был игрок. Допускается использование любых языков программирования, библиотек, методов машинного обучения, разрешено использование внешних данных из сети интернет.



Доступные данные

Участникам предоставляются наборы данных с примерами матчей, которые запакованы в файл [dota2_skill_data.zip](#). Каждый пример описывает характеристики матча и статистику одного из игроков на момент окончания матча. Все примеры имеют уникальный идентификатор “id”.

Всего два набора:

- 1) обучающий набор (train), в котором для каждого примера известен тип медали опыта игрока
- 2) тестовый набор (test), для которого участникам неизвестна опытность игрока — ее необходимо вычислить.

Наборы данных предоставляются в двух форматах:

- [CSV](#)-таблицы: skill_train.csv, skill_test.csv;
- [JSON](#)-документы: skill_train.jsonlines, skill_test.jsonlines (расширенный, содержит дополнительную информацию, которой нет в CSV-таблицах).

Описание полей с данными можно найти в руководстве по решению задачи. Про данные достаточно знать, что это некоторый набор чисел, идентификаторов и временных рядов, который потенциально несет в себе полезную информацию. Опыт игры в Dota 2 для использования данных не требуется.

Выходные данные

Для каждого примера из тестового набора необходимо предсказать опытность игрока. В систему необходимо предоставить для проверки CSV-таблицу с предсказаниями, она должна содержать две колонки: id — идентификатор примера, skilled — тип медали (0 — новичок, 1 — опытный).

Пример выходных данных (файл sample_submission.csv):

```
id,skilled
0,0
7,1
9,0
10,0
11,1
15,0
...
```

Оценка качества решения

Доля верно угаданных типов медали опыта по всем тестовым примерам:

$$Accuracy = \frac{1}{N} \sum_{i=1}^N [y_i = p_i],$$

где

- y_i — настоящая опытность игрока,
- p_i — ответ, выданный участником,
- $[x] = 1$ если выражение x верно, 0 — иначе.

Оценка за задачу вычисляется по формуле: $Accuracy * 100$.

Таким образом, максимальная оценка за задачу равна 100 баллам.



Рекомендации по решению задачи

Данная задача не типична для олимпиад по программированию. В отличие от обычных олимпиадных задач, точного решения у организаторов нет, более того — его, скорее всего, не существует, так как данные взяты из реального мира: в компьютерных играх всегда присутствует случайность и человеческий фактор, который не позволяет во всех случаях безошибочно определять целевую переменную. Это открывает огромный простор для творчества и экспериментов. Решая задачу, участникам предлагается ознакомиться с техниками анализа данных и машинного обучения. В частности:

- посмотреть материалы [Академии Искусственного Интеллекта Сбербанка](#);
- обратить внимание на разбор похожих задач на [тренировках по машинному обучению](#).

Для решения удобнее всего использовать язык программирования Python, так как для него есть большое число библиотек для анализа данных: [NumPy](#), [Pandas](#), [SciKit-Learn](#) и другие. В качестве инструмента разработки — интерактивную среду [Jupyter](#).

Базовый пример решения от организаторов доступен в виде Jupyter-тетради:
[Dota2SkillPrediction_Tutorial.ipynb](#)