# **Тестовое задание**

## **Описание**

В качестве тестового задания хотим вам дать задачу по прогнозированию победителя по картам. Для этого мы выгрузили небольшую часть нашего датасета с начала года до середины лета. В выборке находятся матчи, сыгранные на S-турнирах (турнирах самого высокого уровня) за этот период. Необходимо добиться максимальной метрики ROC-AUC на тестовой выборке.

## **Описание датасета (ссылка на** [**датасет**](https://drive.google.com/drive/folders/1QtGG4gFaKKEmXrlmiOQ-oo_XNuOB4EhU?usp=sharing)**)**

В датасете вы найдете:

1. **train.csv** - история сыгранных карт между двумя командами.
   1. `map\_id`- уникальный идентификатор сыгранной карты
   2. `team1\_id`- уникальный идентификатор команды
   3. `team2\_id`- уникальный идентификатор команды
   4. `map\_name`- название карты
   5. `who\_win` - переменная, которую хотим прогнозировать. Значение 0 означает, что победила команда `team1\_id`, значение 1 - победила команда `team2\_id`.
2. **players\_feats.csv** - признаки игроков, взятых из статистики сайта HLTV за полугодовой период для каждого `map\_id` из файлов `train.csv` и `test.csv`. То есть, одна строка является агрегацией за полгода признаков игроков команды `team\_id` для карты `map\_id` (если дата матча 2022-01-10, то признаки будут сумма/среднее за период с 2021-06-10 до 2022-01-09). Названия колонок состоят из номера игрока и названия метрики. Например, признак `p1\_total\_kills` означает суммарное количество убийств, совершенных игроком `p1` за полгода до даты карты `map\_id`. Вы можете использовать эти признаки, чтобы как то оценить силу команд. Файлы `train.csv` и `players\_feats.csv` можно смерджить по `train.map\_id == players\_feats.map\_id and train.team{1|2}\_id == players\_feats.team\_id`.
3. **test.csv** - история сыгранных карт, для которых не известен победитель

## **Отправка ответов**

Предполагается, что вы построите процесс валидации, используя данные `train.csv`. В качестве решения нужно прислать на почту [demand-sunnier0k@icloud.com](mailto:demand-sunnier0k@icloud.com) свои ФИО и ссылку на github репозиторий

с решением тестового задания. Свои прогнозы для `test.csv` можете загрузить в репозиторий.

**Дедлайн для ответов**: 12:00 мск 5 декабря.

## **Ограничения**

1. Разрешено иметь до 3 файлов с прогнозами для `test.csv`. Если файлов будет больше, из них выберется случайно 3.
2. Важно, чтобы была воспроизводимость результата, то есть чтобы была возможность заново прогнать эксперимент и получить такие же прогнозы (ну или почти такие же).

## **Помощь**

По любым техническим вопросам можете писать на почту [demand-sunnier0k@icloud.com](mailto:demand-sunnier0k@icloud.com).