



Тестовое задание для специалиста на должность  
**Machine-learning разработчик**

Ключевой задачей ml-разработчика является генерация и реализация идей по предсказанию поведенческих паттернов, нахождения закономерностей в данных и извлечение практической пользы из этого.

По нашим оценкам опытный специалист справится с этим тестовым заданием за 1 рабочий день. Просьба заранее оценить время, которое Вы планируете потратить на выполнение и сообщить нам. Также, после выполнения задания, напишите фактическое время, которое было потрачено на каждое из заданий в часах.

Ссылка на игру - <https://apps.facebook.com/solitairesocial>

**Задание I. «Space dataset»**

Необходимо собрать датасет, который бы описывал поведение пользователя за определенный промежуток времени, в нашем случае возьмем 24 часа (нужно описание первые 24 часов жизни игрока в игре)

Вместе с этим документом Вы получили данные, они располагаются в файлах **users.csv**, **game\_results.csv**. Данные структурно и по своему качеству похожи на реальные данные игры.

Сохраните собранный в результате выполнения задания датасет как **user\_features.csv**.

### Структура данных

Таблица **users.csv** представляет собой описание пользователя

| Столбец            | Тип       | Описание                       |
|--------------------|-----------|--------------------------------|
| <b>id</b>          | int       | Уникальный id игрока           |
| <b>create_time</b> | timestamp | Время регистрации пользователя |

Таблица **game\_results.csv** представляет собой описание логов игр пользователей.

| Столбец             | Тип       | Описание                                                                                                                                          |
|---------------------|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>id</b>           | bigint    | Уникальный id всех записей. Автоинкремент                                                                                                         |
| <b>user_id</b>      | int       | Уникальный id игрока                                                                                                                              |
| <b>timestamp</b>    | timestamp | Время начала игры                                                                                                                                 |
| <b>winner</b>       | tinyint   | <ul style="list-style-type: none"><li>• 1 - победа</li><li>• 0 поражение</li></ul>                                                                |
| <b>length</b>       | int       | Длительность раунда в секундах                                                                                                                    |
| <b>magic_used</b>   | tinyint   | Кол-во магии, которое игрок использовал в раунде                                                                                                  |
| <b>player_cards</b> | tinyint   | Число карт, которые игрок успел разложить к концу раунда                                                                                          |
| <b>round</b>        | tinyint   | Номер раунда: <ul style="list-style-type: none"><li>• 1</li><li>• 2</li><li>• 3</li></ul>                                                         |
| <b>type</b>         | tinyint   | Тип игры: <ul style="list-style-type: none"><li>• 1 - Tournament</li><li>• 2 - World Tour</li><li>• 3 - Training</li><li>• 4 - Practice</li></ul> |

### Цель формирования таблицы:

- Подготовить данные для создания/обучения предсказательной модели
- Произвести EDA (exploratory data analysis)

### Результаты формирования датасета:

В рамках задачи нужно собрать датасет **user\_features.csv**, который будет дальше использоваться для обучения Вашей модели.

Период сбора данных по конкретному пользователю составляет 24 часа от момента регистрации (create\_time)

Сколько и каких фичей Вы должны собрать по каждому пользователю - решать Вам. Вы можете использовать любой инструмент для формирования датасета (SQL, NoSQL, Python, R, etc )

Если Вы использовали SQL, NoSQL, Python, R или иные инструменты программирования, необходимо приложить программный код. Его можно приложить как исходный файл (в удобном формате) или прикрепить ссылку на github.

### Задание II. «Space predictions»

После выполнения первого задания, Вы уже сформировали свое представление о данных, а также составили свой датасет.

Дальше Вам предстоит использовать данные для построения модели.

Вместе с этим документом Вы получили данные, что описывают некоторую Космическую характеристику пользователя, что располагается в файле **predictions.csv**.

## Структура данных

Таблица `predictions.csv` представляет собой описание пользователя

| Столбец                 | Тип                  | Описание                                                                                   |
|-------------------------|----------------------|--------------------------------------------------------------------------------------------|
| <code>user_id</code>    | <code>int</code>     | Уникальный id игрока                                                                       |
| <code>prediction</code> | <code>tinyint</code> | <ul style="list-style-type: none"><li>• 0 - не космонавт</li><li>• 1 - космонавт</li></ul> |

## Результаты построения модели:

После предыдущей задачи у Вас должен был получиться датасет `user_features.csv`, где каждый пользователь описан некоторыми игровыми метриками.

Теперь **только** для **части** пользователей из изначального датасета Вам известны их Космические характеристики(из файла `predictions.csv`).

Вам необходимо построить модель, чтобы предсказать Космическую характеристику для пользователей, которые были в `users.csv` датасете, но не попали в `predictions.csv` датасет.

Мы оставляем выбор за Вами, какую модель строить и как её обучать. Можем посоветовать использовать любые готовые инструменты языка программирования Python.

Код на построение модели, датасет с полученными предсказаниями необходимо также приложить как исходный файл(в удобном формате) или прикрепить ссылку на github.

Результаты EDA из предыдущего задания и процесс обучения, оценки модели важно описать в отдельном PDF файле и приложите к остальным файлам.

## **Важно**

Перед тем, как приступить к выполнению тестового задания, пожалуйста, обратите ваше внимание на то, что оно не является оплачиваемым и не будет использовано в коммерческих целях. Также, предоставляя результат вашего тестового задания, вы принимаете тот факт, что Космос может в настоящее время и/или в будущем разрабатывать внутри компании или получать от третьих лиц идеи и другие материалы, похожие по содержанию на присланные вами, но ни в коем случае не основанные на них. Мы еще раз благодарим вас за интерес к нашей компании и желаем удачи с выполнением тестового задания!

Очень надеемся, что задание покажется вам интересным :)

Вдохновения, удачи и хорошего настроения.

**Поехали!**