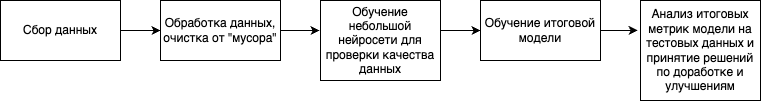
Тестовое задание для соискателя на позицию

# Инженер по созданию и настройке модулей нейросетей и программных компонентов

1. **Проектирование обучения распознавания изображений**

Мы просим схематично спроектировать процесс сбора и подготовки данных для обучения нейросетевой модели распознавания изображений. Например, это могут быть документы, подтверждающие личность отдельно или группой (паспорт, права, загранпаспорт, военный билет). Какие этапы есть, как оценить качество и объем данных?

**Решение**



1. Собираем данные от пользователей: пользователь при заполнении анкеты отправляет фотографию паспорта в определённую графу. Также важно, чтобы он вписывал данные паспорта вручную в соответствующие строки. Это необходимо для последующего обучения модели.
2. После того, как мы набрали достаточное количество данных необходимо оценить их качество, ведь от этого зависит качество самой модели машинного обучения.

Во-первых, человек вряд ли будет вписывать другие данные, отличающиеся от фотографии документа. С другой стороны, человек может ошибиться в 1-2 символах при наборе паспортных данных.

Следующие шаги, это ручной отбор правильных данных, которые безошибочно повторяют данные с фотографии документа. По ним мы обучаем небольшую нейросеть, которая сможет «разметить», а точнее проверить на правильность другие данные. Если модель будет сомневаться, она будет отправлять данные нам, и мы будем лично принимать решение о том правильно ли написаны данные в соответствии с фотографией документа.

1. Теперь мы имеем большое количество качественных данных, что является основным залогом хорошей модели. Обучаем модель на этих данных и далее проверяем ее на тестовых данных, которые модель не видела. Смотрим на метрики и делаем выводы о модели, предпринимаем дальнейшие шаги.
2. **Построить модель для предсказания**

Ниже приведен выдуманный набор данных о клиентах компании по аренде техники.

Где:

* client\_id - уникальный идентификатор клиента
* ФИО клиентов
* gender - пол
* age - возраст
* city - город
* rented\_equipment - арендованное оборудование
* rental\_period - период аренды в днях
* даты начала и конца аренды
* late\_return - опоздание с возвратом (1 - было, 0 - нет)
* registration - наличие регистрации в РФ (0 - нет, 1 - да, но не в Москве, 2 - да, в Москве)
* debt\_history - история долгов перед компанией (0 - не было, 1 - до 5 000 р один раз, 2 - более 5 000 р один раз, 3 - до 5 000 р много раз, 4 - долгов много раз)

Задача - на основе этих данных построить модель для предсказания вероятности опоздания клиента с возвратом оборудования. Среда реализации: Python; опционально - в R.

Порядок данных: client\_id, first\_name, last\_name, gender, age, city, rented\_equipment, rental\_period, rental\_start\_date, rental\_end\_date, late\_return, registration, debt\_history

1, Иван, Петров, M, 35, Moscow, Drill, 5 days, 01/01/2023, 01/05/2023, 0, 2, 0

2, Мария, Сидорова, F, 28, SPb, Ladder, 3 days, 03/01/2023, 01/06/2023, 1, 1, 3

3, Алексей, Орлов, M, 42, Kazan, Saw, 7 days, 02/02/2023, 02/09/2023, 0, 0, 1

4, Ольга, Смирнова, F, 25, Moscow, Sander, 2 days, 05/02/2023, 02/12/2023, 1, 2, 4

5, Сергей, Егоров, M, 37, Rostov, Drill, 4 days, 09/02/2023, 02/13/2023, 0, 1, 2

6, Юлия, Соколова, F, 29, SPb, Perforator, 4 days, 11/02/2023, 02/15/2023, 1, 0, 0

7, Андрей, Кузнецов, M, 36, Yekaterinburg, Jack Hammer, 6 days, 10/02/2023, 02/16/2023, 0, 1, 1

8, Наталья, Козлова, F, 41, Novosibirsk, Tile Cutter, 3 days, 15/02/2023, 02/18/2023, 0, 0, 0

9, Пётр, Волков, M, 38, Moscow, Spray Gun, 1 day, 17/02/2023, 02/18/2023, 1, 2, 4

10, Лидия, Фролова, F, 32, Rostov, Nail Gun, 5 days, 20/02/2023, 02/25/2023, 0, 1, 3

11, Александр, Кузьмин, M, 44, Nizhny Novgorod, Paint Sprayer, 4 days, 22/02/2023, 02/26/2023, 0, 1, 2

12, Елена, Соловьева, F, 37, Kazan, Tile Cutter, 5 days, 25/02/2023, 03/02/2023, 1, 0, 1

13, Антон, Королев, M, 29, Moscow, Nail Gun, 2 days, 28/02/2023, 03/02/2023, 0, 2, 0

14, Анна, Орлова, F, 26, SPb, Jack Hammer, 6 days, 02/03/2023, 03/09/2023, 1, 1, 3

15, Игорь, Савельев, M, 41, Rostov, Paint Sprayer, 3 days, 04/03/2023, 03/07/2023, 0, 1, 0

16, Юлия, Павлова, F, 33, Yekaterinburg, Drill, 1 day, 09/03/2023, 03/10/2023, 1, 1, 4

17, Роман, Соколов, M, 30, Moscow, Saw, 7 days, 06/03/2023, 03/13/2023, 0, 2, 2

18, Олеся, Кузьмина, F, 28, Krasnoyarsk, Ladder, 4 days, 11/03/2023, 03/15/2023, 0, 0, 1

19, Максим, Виноградов, M, 36, Novosibirsk, Wall Sander, 5 days, 10/03/2023, 03/15/2023, 1, 0, 3

20, Валерия, Семёнова, F, 29, Perm, Perforator, 3 days, 16/03/2023, 03/19/2023, 0, 1, 0

21, Фёдор, Смирнов, M, 42, Ufa, Tile Cutter, 2 days, 14/03/2023, 03/16/2023, 1, 1, 4

22, Любовь, Кузьмина, F, 38, Voronezh, Spray Gun, 6 days, 20/03/2023, 03/26/2023, 0, 1, 1

23, Даниил, Орлов, M, 25, Moscow, Paint Sprayer, 5 days, 18/03/2023, 03/23/2023, 1, 2, 2

24, Карина, Ефимова, F, 19, Omsk, Jack Hammer, 4 days, 22/03/2023, 03/26/2023, 0, 0, 0

25, Виктор, Ковалёв, M, 29, Rostov, Saw, 3 days, 21/03/2023, 03/24/2023, 0, 1, 0

26, Анастасия, Суворова, F, 26, Samara, Drill, 7 days, 25/03/2023, 04/01/2023, 1, 1, 3

27, Никита, Волков, M, 34, Moscow, Ladder, 2 days, 27/03/2023, 03/29/2023, 0, 2, 1

28, Алина, Кулагина, F, 24, St. Petersburg, Nail Gun, 5 days, 01/04/2023, 04/06/2023, 0, 1, 0

29, Денис, Соболев, M, 32, Novosibirsk, Wall Sander, 4 days, 30/03/2023, 04/03/2023, 1, 0, 2

30, Алёна, Широкова, F, 27, Yekaterinburg, Perforator, 6 days, 03/04/2023, 04/09/2023, 0, 1, 0

31, Анастасия, Морозова, F, 23, Perm, Wall Sander, 2 days, 17/03/2023, 03/19/2023, 0, 0, 0  
  
**Решение**:  
