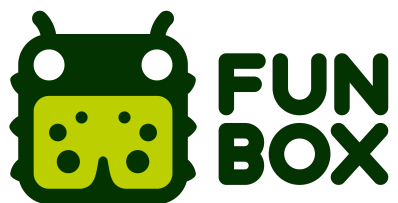


Квалификационное задание для
Data Mining Specialist / Data Scientist



1 О задании

Результат присылайте нам вместе с рассказом о себе на ящик wanted@fun-box.ru с пометкой «Вакансия Data Mining Specialist / Data Scientist». После проверки задания мы обязательно сообщим вам о нашем решении.

2 Вопросы

Здесь несколько теоретических вопросов, на которые нет однозначного ответа. С их помощью вы сможете в свободной форме рассказать нам о своем личном опыте в том или ином аспекте и проиллюстрировать рассказ примерами.

- С какими инструментами обработки данных вы знакомы?
- Есть ли у вас опыт разработки? Какой?
- Что вы понимаете под понятием Data Mining?
- Какие инструменты и методы помогают вам экономить время и избегать ошибок при работе?
- На каких ресурсах вы получаете информацию и новости по используемым технологиям?
- Расскажите нам о себе и о ваших последних работах (в рамках допустимого).

3 Задачи

3.1 Задача о выпускнике ML-курсов

Выпускник популярной онлайн школы по машинному обучению отдыхал с приятелем в баре. Решив похвастаться своими знаниями, он предложил спор:

— Достоверно известно, что в нашем провинциальном городе находится столько же женщин, сколько и мужчин. Поэтому, выйдя на улицу, я с вероятностью $1/2$ встречу мужчину и с вероятностью $1/2$ — женщину. Спорим на 1000 рублей к одному, что я сейчас выйду на улицу, и среди первых 30 встречных будет хоть одна женщина (я прав — ты мне один рубль, я неправ — я тебе 1000).

Приятель согласился. Они вышли из бара и столкнулись с идущей по улице ротой солдат, состоящей из мужчин и численностью явно больше 30.

- Как вы думаете, с каким процессом и с каким распределением выпускник надеялся иметь дело?
- Какую вероятность неблагоприятного для себя исхода он мог предполагать?
- Как он мог оценить математическое ожидание своего выигрыша?
- Что же произошло? Ужасная маловероятная случайность, или выпускник все-таки что-то не учел?

3.2 Задача о штате IT-компании

У одной компании ведется база сотрудников. В ней указано имя сотрудника, его `id` и `id` его руководителя (файл `employee.db` в [папке с данными](#)). Сотрудники без `id` руководителя – руководители отделов.

С некоторого момента в компании стал вестись учет найма, увольнений и переводов. Записи об этих событиях находятся в файле `events.txt` в [папке с данными](#).

Необходимо:

- Разобраться в структуре бд и файла с событиями.
- Для каждого отдела построить график численности с интервалом в неделю.

3.3 Задача о лишних транзакциях

Мы разрабатываем систему обработки транзакций для компании SuperLabs. Её клиенты (с `id` от 0 до 29) проводят через нашу систему транзакции.

Неожиданно заказчик испуганно сообщил, что клиенты жалуются на то, что они обнаружили лишние транзакции, проведенные от их имени. Никаких подробностей заказчик не добавил, но попросил разобраться как можно скорее.

Мы выгрузили лог транзакций за всё время (файл `txs.tzr.bz2` в [папке с данными](#)).

- Можем ли помочь заказчику и предположить `id`-ы пострадавших клиентов?
- Можем ли мы предположить время возможного сбоя (или атаки)?

Спасибо!