



Конкурс аналитиков
данных

«Волга-AI»

Задание финального этапа конкурса аналитиков данных «Волга-AI» состоят из двух задач. В ходе работы тебе будут доступны массивы данных от одной из ведущих компаний России в сфере ЖКХ «АИС Город». Наши задачи основаны на реальных потребностях отрасли и представляют собой испытания, требующие нестандартного мышления, затрагивающие знания о моделях ИИ и умение проводить анализ данных.

Для успешного прохождения этапа необходимо выполнить обе задачи. Но даже если получилось сделать только одну, отправляй так. Комиссия рассмотрит все ответы.

Желаем успехов!

Задача 1. Классификация заявок в диспетчерскую

В диспетчерскую службу уездного городка Д поступает множество заявок: то трубу прорвёт, то с крыши капает, то крышку люка кто-то позаимствовал.

Администрация городка просит делить все полученные заявки по специальному классификатору, который состоит из пар «Группа работ»-«Категория работ в группе».

Городок продвинутый, поэтому диспетчерская служба может получать заявки как по телефону, так и через Интернет, и даже

через мессенджеры. В любом случае, это получается набор коротких текстов, которые надо каким-то образом оперативно классифицировать.

Данные для построения модели:

1. Сведения о ранее обработанных заявках (файл «*L-short.csv*») со следующими параметрами:
 - *Number* – Идентификатор заявки;
 - *Date* – дата и время регистрации;
 - *UID* – Уникальный идентификатор адреса дома;
 - *Desc* – Текст заявки, который необходимо классифицировать;
 - *Group* – Группа работ;
 - *Cat* – Категория работ в группе.

Исходные данные:

1. Файл с новыми ещё не классифицированными заявками *C-short.csv* (*предоставляется заранее*) с параметрами, аналогичными файлу «*L-short.csv*» из данных для построения модели, но с пустыми *Group* и *Cat*.

Выходные данные:

1. Файл *A.csv*, который получился заполнением параметров *Group* и *Cat* (в исходном файле они пустые).
2. (не обязательно) Файл *A-2.csv* аналогичного формата, где классификация проведена иначе.
3. (не обязательно) Файл *A-3.csv* аналогичного формата, где классификация проведена иначе.

Примечания:

- Формат всех файлов CSV: с заголовком, разделители - «;».
- Если будет реализовано несколько различных моделей классификаторов, засчитываться будет лучший результат.

На проверку отправляется архив, в котором присутствует:

- Файл *README.pdf* с кратким описанием применённого решения и инструкцией по сборке программ.
- Исходные тексты программы, формирующей модель, в папке */src-train*.

- Исходные тексты программы, преобразующей исходные данные в выходные на основании заранее сформированной (обученной) модели в папке */src-predict*.
- Собственно модель в папке */model*.
- Выходные файлы *A.csv* (*A-2.csv* и *A-3.csv* при наличии) в папке */result*.

Оргкомитет оставляет за собой право самостоятельно сформировать модель по предоставленным скрипам обучения и произвести классификацию на основе предоставленных исходных текстов и инструкции.

Задача 2.

Данное задание выполняется на сайте: <https://volga-ai.aisgorod.ru/>

На проверку отправляется архив, содержащий:

- Исходные файлы решения
- Файл .txt, где должен быть логин/почта, с которого вы регистрировались на сайте для решения задачи 2

ВАЖНО!

После выполнения заданий добавь ссылку на свои решения в личном кабинете конкурса. Решение должно содержать в себе два архива по задачам. Не забудь предоставить доступ к репозиторию или облачному хранилищу, чтобы Оргкомитет мог проверить твои работы.