

Задание кандидату

1. Python

Все задачи необходимо выполнить используя нативный python третьей версии. Количество и качество реализованных задач будет влиять на окончательную оценку. Результат выполнения разместить в репозиторий (github, bitbucket, gitlab...) и в качестве результата прикрепить ссылку на результат.

Task #1.

Описание Реализуйте функцию, определяющую, является ли одна строка перестановкой другой (полиндrom). Под перестановкой понимаем любое изменение порядка символов. Регистр учитывается, пробелы являются существенными.

Task #2.

Описание Реализуйте функцию сжатия строки на основе счетчика повторяющихся символов. Например, строка aabcccccaaa должна превратиться в a2b1c5a3. Если «сжатая» строка оказывается длиннее исходной, метод должен вернуть исходную строку.

Task #3.

Описание Написать отдельный скрипт, принимающий из командной строки число, количество последних строчек, которые нужно прочесть из файла и вывести в консоль. Напишите

предварительно отдельный скрипт, который генерирует текстовый файл и наполняет его случайными числами. Количество строк передается из командной строки.

2. Machine Learning

1. Построить модель (при помощи Python) прогнозирования дефолта клиента на аппликационных и поведенческих данных из сайта с помощью логистической регрессии, спрогнозировать вероятность не выполнения обязательств перед компанией. Финальную модель логистической регрессии необходимо отобразить в Excel (параметры, группы и баллы).

2. То же, с помощью модели на выбор (без отображения в Excel).

3. Предоставить код на Jupiter Notebook.

Для исследования предоставлены три файла:

appl_data.csv, аппликационные данные клиента. Данные закодированы. Расшифровка колонок:

appl_id – уникальный номер заявки

app_crtime – время и дата создания заявки

client_id – уникальный номер клиента

birth – дата рождения клиента

gender – пол клиента

pass_bdate – дата выдачи паспорта

fam_status – семейный статус

quantity_child – количество детей

max_age_child – возраст старшего ребенка

property – имущество в собственности

lived_since – с какой даты клиент проживает по месту жительства

is_same_reg_lived_since – дата регистрации по месту регистрации

region – область проживания

region_reg – область регистрации

jobworksince – с какой даты работает на последнем месте

work_experience – опыт работы

empl_state – тип занятости

empl_type – отрасль труда

empl_worker_count – количество работников на месте труда (сгруппировано)

education_area – отрасль образования

education – тип образования

monthlyincome – месячный доход

monthlycost – месячные расходы

behave_on_site.csv, поведение клиента на сайте. Расшифровка колонок:

device_id – уникальный номер устройства, из которого клиент посещает вебсайт

client_id – уникальный номер клиента

browser – браузер, который использует клиент

platform – операционная система, которую использует клиент

create_time – время и дата посещения конкретной страницы на сайте

id_ref – закодированное название страницы на сайте (ссылка), которую посетил клиент во время create_time

is_default.csv, целевая функция, или выполнил клиент обязательства перед компанией, или нет. Расшифровка колонок:

appl_id – уникальный номер заявки

df – признак дефолта, bad – плохой клиент, good – хороший клиент. Признак проставлен лишь для части клиентов, для остальных нужно спрогнозировать вероятность невыполнения обязательств.

3. SQL

Решить задачи № 45, 62, 64, 93, 96, 130 на сайте <http://www.sql-ex.ru>. Для проверки прислать логин и пароль.