

Решение IT кейса отборочного этапа Opportunity cup

Команда "Чемпионы"



Задача и методы решения

Задача нашей команды в рамках IT кейса заключалась в разработке алгоритма для определения мошеннических банковских операций. Для этого мы использовали:

Pandas

Библиотека Pandas позволила нам обрабатывать и анализировать данные, полученные из json файла.

Python

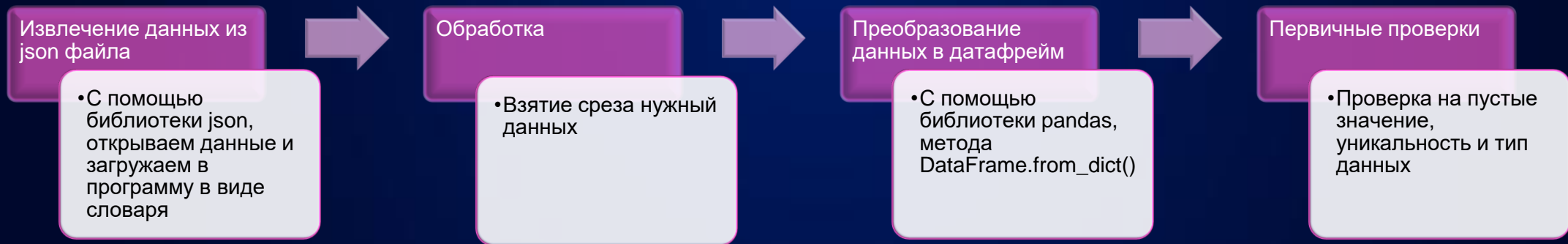
Наш код написан на одном из самых популярных языков программирования в наше время, что делает его более адаптивным.

PosgreSQL

Мы использовали базу данных PostgreSQL для лучшей структуры и переноса данных. Это улучшило качество и понятность нашей работы.

Решение кейса

Наша задача была придумать и реализовать механизм по обнаружению фродов и выявить три основных fraud-паттернов. Мы тщательно проанализировали задание и пришли к данному начальному решению.





Решение кейса



Извлечение данных

- Преобразование типов данных даты и времени

Формирование гипотез

- Сформировано 4 гипотезы фрод-операций*

Проверка гипотез

- Проверка визуализацией данных в виде графиков и группировкой данных по параметрам (вычисление различных характеристик)

Формирование четырехзначного кода признаков каждой операции

- Код-строка из нулей и единиц, где единица-присутствие признака.*

*Сформированные 4 гипотезы фрод операций

- Необычное время
- Необычная локация
- Частота
- Необычные суммы

Определение фрод операции

- Операция, которая в своем коде признаков имеет три единицы является фродом

*Метод составления кода

- Формирование новых колонок в датафрейме с помощью маскирования, а затем суммирование в код.



Решение кейса

Перенос датафрейм в базу данных PostgreSQL

- Был послан запрос на создание таблицы и перенесены данные

Извлечение данных из базы данных

- Извлекаем данные в виде базы с фрод операциями, чтоб проверить целостность данных и закрываем базу данных

Преобразование данных в датафрейм с подробными характеристиками операций

- Преобразование данных из базы данных в датафрейм и слитие датафрейма с исходным датафреймом по индексам

Сохранение фрод операций в csv формат

- Фрод операции в csv

- Процент фродов от всех данных нам операций составил 9 %

Выводы и анализ решения



Прodelав всю вышеперечисленную работу, мы пришли к выводам:

- Фрод операции происходят в необычное время, например ночью
- Фрод операции можно определить, если частота совершаемых клиентов операций нехарактерно быстро возросла
- Фрод операции часто происходят из отдаленных от клиента локаций, в которые раньше переводов не совершалось
- Фрод операции могут совершаться на странные, нехарактерные суммы, но это не является основным показателем

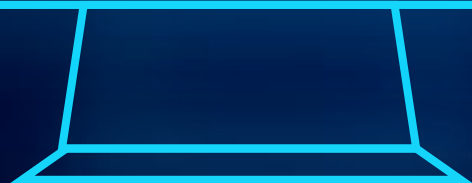
Если в данных совпадают 3-4 признака, значит операция совершена мошенником.





Чему мы научились и что нам дал этот проект?

- Мы познакомились и попробовали работать с json файлом.
- Узнали кто такие фроды и как определять паттерны.
- Попрактиковались в анализе данных при выборе оптимального решения.
- Узнали, что такое ETL-процесс и реализовали его.





Наша команда

Миронов Сергей Игоревич

Программист-аналитик

- Место обучение: Финуниверситет (2 курс ПИ факультета ИТиАБД)
- Опыт работы: Любительские и университетские проекты

Волкова Анна Дмитриевна

Data Miner (Поиск информации)

- Образование: МТИ (2 курс, ОЗПИПо)
- Опыт работы в IT: программист-аналитик в 1С, учебные проекты

Седых София Викторовна

Дизайнер презентации

- Образование: ИТМО (1 курс мега факультета ТИНТ, факультет ФИТИП)
- Опыт работы: Участие в любительских и университетских проектах

Морозов Арсений Владимирович

Координатор проекта

- Образование: МТУСИ (СиСС) первый курс
- Опыт работы: нет





Приложение

Код решения

- [Решение CaseCup_2022](#)

Дополнительные материалы

- [Работа с PostgreSQL в Python](#)
- [Анализ методов обнаруживания мошеннических транзакций в системе ДБО](#)
- [Как банки будут выявлять подозрительные операции](#)
- [Pandas в Python](#)

