

ФИЦ

**моя
смена.**

Кейс

Прогнозирование бизнес драйверов

ФИЦ: хакатон

2024



USE TECH

SENSE

IT-ONE

**FOGSTREAM
PAT-AIR
THERMOPAT**

EINSPARK

ISS
Информационные
системы и технологии

Описание кейса

Прогнозы нужны не только метеорологам. Правительства прогнозируют экономический рост. Ученые пытаются предсказать численность населения. А предприятия прогнозируют спрос на продукцию – обычная задача специалистов по анализу данных. Прогнозы особенно актуальны для обычных магазинов, будь то продуктовые магазины или торговые точки сотовых операторов.

В частности, прогнозы помогают рассчитать потребность в персонале. Как раз этим и занимается наша компания. Мы формируем графики работы сотрудников магазинов, основываясь на прогнозах бизнес-драйверов.

Под "бизнес-драйверами" мы понимаем ключевые показатели, которые определяют финансовую и операционную деятельность торговой точки. Типичными примерами бизнес-драйверов являются трафик (количество вошедших человек), количество чеков, количество проданных товаров и тд. Если прогнозировать немного больше, то придется вывести больше людей на смену и увеличится фонд оплаты труда. Если прогнозировать немного меньше, то возникнет нехватка персонала, что приведет к очередям и потере дохода. Более точное прогнозирование, благодаря машинному обучению, может помочь ритейлерам угодить клиентам, формируя гибкие графики работы.

Задача

Необходимо разработать решение для задачи прогнозирования временных рядов бизнес-драйверов и произвести прогноз на следующий календарный месяц. Лучшие решения послужат дополнением к уже существующему решению.

Стек

Python

Pandas

Любые библиотеки

Вводные данные

CSV-файлы, содержащие исторические данные по бизнес-драйверам.

Все данные представлены с гранулярностью 1 день. [**ССЫЛКА**](#)

1. data1.csv - файл с данными по количеству чеков из торговой точки продуктовой розницы.
2. data2.csv - файл с данными по трафику (числу вошедших людей) торговой точки непродуктовой розницы (сотовые операторы, магазины одежды, магазины игрушек, DIY).
3. data3.csv - файл с данными по трафику (числу вошедших людей) торговой точки непродуктовой розницы (сотовые операторы, магазины одежды, магазины игрушек, DIY).
4. data4.csv - файл с данными по трафику (числу вошедших людей) торговой точки непродуктовой розницы (сотовые операторы, магазины одежды, магазины игрушек, DIY).
5. data5.csv - файл с данными по трафику (числу вошедших людей) торговой точки непродуктовой розницы (сотовые операторы, магазины одежды, магазины игрушек, DIY).

Помимо представленных данных можно использовать любые дополнительные данные из открытых источников.

Предполагаемый результат

Решение должно представлять собой Jupyter-Notebook содержащий код предлагаемого решения.

Отдельно нужно сформировать 5 csv-файлов с прогнозами. На вход вам доступно 5 csv файлов с историческими данными в виде таблиц из 2-х колонок: date (дата измерения), value (измеренное значение). На выходе у вас должно быть сформировано 5 прогнозов по каждому ряду в том же формате (date,value). Прогноз представляет собой предсказанные значения по каждому дню.

В Jupyter-Notebook должны запускаться все ячейки, а также приложена необходимая документация и зависимости. Например, если вы используете вручную созданные признаки, то необходимо описать логику их создания.

Оцениваться будет комплексный подход к решению, от этапа предобработки данных до формирования итогового прогноза.

Оценка

Качество прогнозирования будет оцениваться по метрике RMSE. Результаты метрик будут отранжированы с 1 по 10 место по каждому ряду. За каждый прогноз присуждаются баллы в соответствии с местом в рейтинге. Чем выше позиция в рейтинге по каждому ряду, тем больше баллов. За 1-е место участник получает 10 баллов за 10-е место - 1 балл. Таким образом максимум можно набрать 50 баллов. Участники, набравшие больше всех баллов проходят в финал, где будет оцениваться подход к решению.

Дополнительные +

1. Отчет о качестве прогнозирования.
2. Визуализация.
3. Отчет feature-importance.

Ограничения

Так как нам приходится работать с более чем 10000 рядов, очень важна скорость формирования прогноза.

Ограничение на скорость работы алгоритма: на формирование прогноза на один временной ряд должно отводиться не более 4-х минут (240 секунд). Этот аспект также будет учитываться при оценке решений.

Решение должно быть воспроизводимым.

Критерии оценивания

Критерий	Описание	Шкала
Работоспособность прототипа	Оценка работоспособности решения. Разработанное решение должно функционировать, запускаться и выдавать требуемый результат. Мелкие ошибки допустимы, а также частично реализованный функционал.	1-10
Соответствие функциональным требованиям	Проверяется соответствие созданного решения поставленной задаче в описании кейса, в полной мере ли реализована задача. Насколько предложенное решение закрывает выявленную проблему; Дополнительный функционал решения.	1-10
Технологичность	Оценивается качество исходного кода и подходы к разработке. Уникальность технического решения.	1-10
Презентация	Оценивается выступление команды по следующим критериям: структура презентации и выступления, питч.	1-10
Потенциал	Оценивается возможность использования разработанного прототипа в реальной практике, а также возможность развития предложенного решения. Ноу-хау.	1-10

Контакты

ФИЦ

Мария

Куратор хакатона

👉 [@mashriya](#)

Дарья

Куратор хакатона

👉 [@danyaignatova](#)

Варвара

Координатор кейса

👉 [@kvarja](#)

