

Согласовано 22 Мирнов С.М.

Модуль А. Парсинг и предобработка данных

Описание модуля А:

В рамках данного модуля необходимо загрузить данные для обработки и обучения из различных источников. Изначально данные могут храниться в файлах или других источниках данных, таких как базы данных, JSON, XML или коллекции в памяти. В рамках данного модуля требуется выполнить загрузку данных, предварительную обработку и формирование структуры данных.

Для проведения дальнейшего исследования динамики и структуры показателей, характеризующих эффективность функционирования компании за определенный период времени, необходимо задать функции для расчета и анализа основных метрик.

Исходные файлы:

- 1) visitation.csv, orders.csv, expenses.csv (Исходные данные)
- 2) Description.docx (Описание данных)

Результаты работы:

- 1) Data.zip (Предобработанные данные)
- 2) Report_C1.html, Report_C1.ipynb (Отчет о проделанной работе)
- Readme.txt (Дополнительные комментарии)

Задания:

1.1 Загрузка данных

Создать базу данных Users, содержащую структуру, согласно исходным файлам.

Выполнить подключение к базе данных Users. Извлечь необходимые для исследования данные из базы данных (visitation, orders, expenses).

Предусмотреть обработку исключений при загрузке данных.

Все данные необходимо представить в виде DataFrame — объекта для представления табличных данных в Pandas.

1.2 Предварительная обработка данных

Выполнить предварительную обработку загруженных данных. Провести анализ и обработку пропусков, дубликатов, аномальных значений. Привести данные к приемлемому формату. Сохранить данные в формате csv и обновить данные в используемой базе данных для дальнейшего исследования.

1.3 Задание функций расчета ключевых показателей

Задать функцию для формирования набора данных с профилями пользователей. Набор должен включать поля: дата и время начала сессии при первом посещении каждого пользователя, устройство, регион, рекламный источник. Набор необходимо дополнить информацией по дате

первого посещения и о средней стоимости привлечения пользователей в день.

Определить функции для расчёта прибыли от клиентов и коэффициента окупаемости вложений (LTV, ROI), расчёта удержания и конверсии.

1.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в форматах .ipynb и .html, а также исходников с возможностью перекомпиляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.

Необходимо вычислить такие характеристики, как заинтересованность клиента в продукте (DAU, WAU и MAU), количество времени, проведенное в приложении, распределение привлеченных пользователей с разбивкой по различным показателям (страна, устройство, канал привлечения). Аналитическая панель должна содержать анализ продаж для онлайн-торговли.

Визуализируйте динамику полученных метрик во времени. В дашборде должны быть отражены ключевые сводные показатели.

После выполнения пунктов 2.1 и 2.2 необходимо продемонстрировать выполненную работу экспертам

2.3 Создание профилей пользователей

Используя ранее определенную функцию построить набор данных с профилями пользователей. Определить:

- Из каких стран приходят посетители? Какие страны дают больше всего платящих пользователей?
- Какими устройствами они пользуются? С каких устройств чаще всего заходят платящие пользователи?
- По каким рекламным каналам шло привлечение пользователей? Какие каналы приносят больше всего платящих пользователей?

2.4 Маркетинговый анализ

Для проведения маркетингового анализа необходимо рассчитать:

- Сколько денег потратили (всего / на каждый источник / по времени)?
- Сколько в среднем стоило привлечение одного покупателя из каждого источника?

Оцените окупаемость рекламы для привлечения пользователей. С помощью ранее разработанных функций проанализируйте:

- Общую окупаемость рекламы;
- Окупаемость рекламы с разбивкой по устройствам;
- Окупаемость рекламы с разбивкой по странам;
- Окупаемость рекламы с разбивкой по рекламным каналам.

Изучите конверсию и удержание с разбивкой по устройствам, странам, рекламным каналам.

Дайте интерпретацию полученным результатам.

2.5 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в форматах .ipynb и .html, а также исходников с возможностью перекомпиляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.

Согласовано г.г. Вернов Е.И.

Модуль В. Построение, обучение и оптимизация модели

Описание модуля В:

Необходимо выполнить распределение текущих и возможных клиентов компании на группы (сегменты) по определённым схожим признакам. Также в данном модуле требуется осуществить выбор алгоритма для прогнозирования стоимости привлечения пользователей, построить модель и провести оптимизацию полученной модели машинного.

Исходные файлы:

- 1) Data.zip (Результаты предыдущей сессии)

Результаты работы:

- 1) Data.zip (Предобработанные данные)
2) Report_C3.html, Report_C3.ipynb (Отчет о проделанной работе)
Readme.txt (Дополнительные комментарии)

Задания:

3.1 Сегментация пользователей

Сегментируйте пользователей на основе RFM анализа. Исследуйте поведение пользователей из различных групп (уровни удержания клиентов, количество совершенных покупок, время между покупками и т.д.). Проверьте гипотезу различия конверсии между пользователями всех полученных сегментов для оценки функционала приложения. Дайте интерпретацию полученных результатов. Сделайте выводы для оптимизации работы приложения.

3.2 Построение регрессионной модели

На основе проведенного анализа выполнить и обосновать выбор наиболее оптимальных алгоритмов машинного обучения в соответствии со спецификой решаемой задачи. Добавьте сегмент целевой аудитории в качестве атрибута в набор данных для более точной работы выбранной модели.

Необходимо разделить выборку на обучающую и тестовую подвыборки. Первая должна быть использована для настройки весовых коэффициентов, а вторая для проверки корректности настроенной модели.

Проведите обучение моделей, основанных на различных алгоритмах (не менее трех). Сделайте прогноз стоимости привлечения пользователей на валидационной выборке. Выполните оценку моделей разной степени сложности в соответствии с выбранной метрикой. Определите модель, показавшую лучшее качество.

3.3 Оптимизация модели

Путём преобразования набора данных, добейтесь более точной работы выбранной модели. Опишите приемы генерации новых данных и

результаты, к которым они привели, рассматривая все ранее определенные показатели качества. Оцените качество полученной модели, сделайте вывод.

Необходимо оптимизировать полученную модель под решаемую задачу, настраивая гиперпараметры. Выполните оценку данной модели после настройки гиперпараметров. Построить для данной модели кривые валидации и обучения, интерпретируйте полученные результаты.

3.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в форматах .ipynb и .html, а также исходников с возможностью перекомпиляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.

Согласовано 22. Мерников Е.М.

Модуль Г. Разработка программного продукта

Описание модуля Г:

В данном модуле предстоит выполнить прогнозирование прибыли/расходов компании для контрольной выборки с помощью полученной модели. Также необходимо выполнить развертывание модели машинного обучения в рабочей среде в качестве API. Программный интерфейс должен позволить приложению пользователя получать доступ к модели для прогнозирования прибыли компании. Разработать приложение с графическим интерфейсом, которое должно с помощью API генерировать прогнозы по новым данным в режиме реального времени.

Исходные файлы:

- | | |
|----------------|--------------------------------|
| 1) Data.zip | (Результаты предыдущей сессии) |
| 2) Control.csv | (Контрольная выборка) |

Результаты работы:

- | | |
|------------------------------------|------------------------------|
| 1) Data.zip | (Предобработанные данные) |
| 2) Report_C4.html, Report_C4.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt | (Дополнительные комментарии) |

Задания:

4.1 Предсказание целевой переменной

Необходимо выполнить прогнозирование для контрольной выборки с помощью полученной модели. Запишите ответы, содержащие соответствующие атрибуты и предсказания целевой переменной, в файл.

4.2 Разработка прикладного решения

Разработайте программный интерфейс для итоговой модели машинного обучения. API должен позволить приложению пользователя получать доступ к модели регрессии для прогнозирования расходов на привлечение клиентов.

Разработать приложение с графическим интерфейсом, которое должно с помощью разработанного API генерировать прогнозы по новым данным в режиме реального времени.

Приложение должно предоставлять справку по имеющимся командам и их параметрам.

4.3 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в форматах .ipynb и .html, а также исходников с возможностью перекомпиляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.



Модуль Д. Разработка средств интеграции и поддержки готового решения

Описание модуля Д:

Для разработанного приложения и API составьте пользовательскую документацию, представляющую собой руководство пользователя. Руководство должно описывать каждую функцию программы, а также шаги, которые нужно выполнить для использования этой функции. Пользовательская документация должна предоставлять инструкции о том, что делать в случае возникновения проблем. Очень важно, чтобы документация не вводила в заблуждение и была актуальной. Руководство должно иметь чёткую структуру. Необходимо создать презентацию, охватывающую все результаты выполнения задания.

Исходные файлы:

- | | |
|-------------|--------------------------------|
| 1) Data.zip | (Результаты предыдущей сессии) |
|-------------|--------------------------------|

Результаты работы:

- | | |
|--------------------|---------------------------------------|
| 1) Report.doc | (Техническая документация по системе) |
| 2) Rucovodstvo.doc | (Руководство пользователя) |
| 3) Project.pptx | (Презентация) |

Задания:

5.1 Разработка технической документации по системе

При создании программы, одного лишь кода, недостаточно. Должен быть предоставлен некоторый текст, описывающий различные аспекты того, что именно делает код. Такая документация должна быть включена непосредственно в исходный код или предоставляется вместе с ним.

Необходимо разработать техническую документацию для определения и описания API. Документация должна включать описание ресурса, конечные точки и методы, параметры запроса. А также пример полного запроса и ожидаемого ответа.

Для стилизового оформления кода на языке Python использовать правила написания кода по PEP8.

5.2 Пользовательская документация

Для разработанного приложения и API составьте пользовательскую документацию, представляющую собой руководство пользователя. Руководство должно описывать каждую функцию программы, а также шаги, которые нужно выполнить для использования этой функции. Пользовательская документация должна предоставлять инструкции о том, что делать в случае возникновения проблем. Очень важно, чтобы документация не вводила в заблуждение и была актуальной. Руководство должно иметь чёткую структуру.

5.3 Презентация результатов работы

Необходимо создать презентацию, охватывающую все результаты выполнения задания. В ней должно быть указано ёмкое описание результатов работы с обоснованием выбора того или иного решения. Так же в презентации необходимо отразить скриншоты результатов своей работы. Разрабатывать презентацию рекомендуется в Power Point или аналогичной среде. Опишите перспективы улучшения Вашего решения.

5.4 Устный доклад

Подготовить устный доклад по результатам своей работы, включающие основные результаты по каждому модулю и выводы (не более 5 минут). Устные представления докладов – за 60 мин до окончания сессии.