

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ**

**ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

**(ДГТУ)**

Факультет «\_\_\_\_\_Медиакоммуникации и мультимедийные технологии\_\_»»

наименование факультета

Кафедра «\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Медиатехнологии \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_»

наименование кафедры

**КОНТРОЛЬНАЯ РАБОТА**

Дисциплина (модуль) «Проект - применение машинного обучения в креативных индустриях»

наименование учебной дисциплины (модуля)

Направление подготовки/специальность \_090402\_\_ \_Информационные системы и технологии\_\_

коднаименование направления подготовки/специальности

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Направленность (профиль) \_\_Интеллектуальные медиатехнологии\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Номер зачетной книжки \_2239923\_ Номер варианта \_23\_ Группа \_\_\_МЗИК21\_\_\_\_

Обучающийся \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_В.А. Юрасов\_\_\_\_\_\_\_\_\_

подпись, дата И.О. Фамилия

Контрольную работу проверил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ доц. каф. МТ И.С. Трубчик

подпись, дата должность, И.О. Фамилия

Ростов-на-Дону

2024

**Введение**

В данной контрольной работе рассматривается тема оценки эффективности рекламных кампаний и разработки модели для оценки этой эффективности на основе данных о взаимодействии пользователей с рекламным контентом. Целью данного проекта является предсказание того, нажмет ли конкретный пользователь Интернета на рекламу, исходя из особенностей пользователя.

В рамках данного проекта будет использоваться синтетический набор рекламных данных, который содержит информацию о взаимодействии пользователей с рекламным контентом, включая информацию о том, нажал ли пользователь на рекламу или нет. Эти данные будут использоваться для разработки модели, способной предсказывать, будет ли пользователь нажимать на рекламу, исходя из его особенностей.

После обучения модели необходимо провести ее оценку и тестирование на отложенных данных. Это позволит оценить ее точность и способность предсказывать, нажмет ли пользователь на рекламу. Результаты оценки модели могут быть использованы для принятия решений в области рекламы и оптимизации рекламных кампаний.

1. Описание исследования предметной области

Первым этапом исследования является сбор данных о взаимодействии пользователей с рекламным контентом. Для этого используются различные источники данных, такие как веб-аналитика, рекламные платформы и социальные сети. Собранные данные представляют собой множество факторов, которые могут влиять на эффективность рекламных кампаний.

Далее следует этап предварительного анализа данных, включающий их очистку, преобразование и подготовку для дальнейшего использования. В этом этапе происходит удаление выбросов, обработка пропущенных значений и преобразование категориальных переменных в числовой формат. Также проводится исследование корреляций между различными факторами и эффективностью рекламных кампаний.

После предварительного анализа данных происходит разработка модели для оценки эффективности рекламных кампаний. В данном исследовании используются методы машинного обучения, такие как линейная регрессия, деревья решений, случайные леса или нейронные сети. Модель обучается на основе собранных данных, учитывая различные факторы, которые могут влиять на эффективность рекламных кампаний.

После обучения модели происходит ее оценка и тестирование на отложенных данных. Для этого используются данные, которые не использовались в процессе обучения модели. Оценка модели включает анализ ее точности, полноты, специфичности и других метрик, которые позволяют оценить ее способность предсказывать эффективность рекламных кампаний.

1. Постановка задачи

Необходимо создать проект, который может оценить, на основе собранного датасета, эффективность рекламной кампании. Для этого мы будем использовать язык программирования Python и библиотеки pandas, numpy, sklearn, plotlib и seaborn. Весь проект будет построен на базе Google Colab, а также размещён на Github.

1. Подготовка датасета

С помощью Google Analytics собираем данные пользователей по следующим параметрам Daily Time Spent on Site, Age, Area, Income, Daily Internet Usage, Ad Topic Line, City, Male, Country и переводим их в формат CSV.

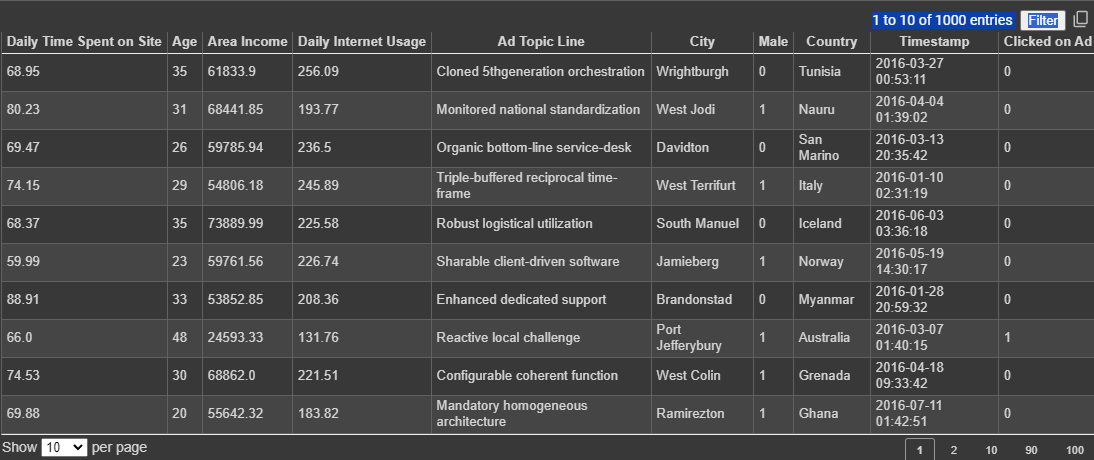


Рисунок 1 – подготовленный датасет

1. Написание кода

Для начала импортируем необходимые библиотеки и сократим их название для удобства написания кода.

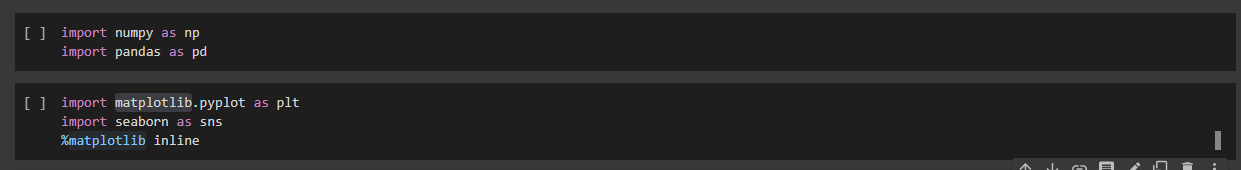


Рисунок 2 – импортированные библиотеки

Далее, используя функцию Pandas.read\_csv укажем путь и импортируем наш датасет в формате CSV и выставим количество отображения данных.

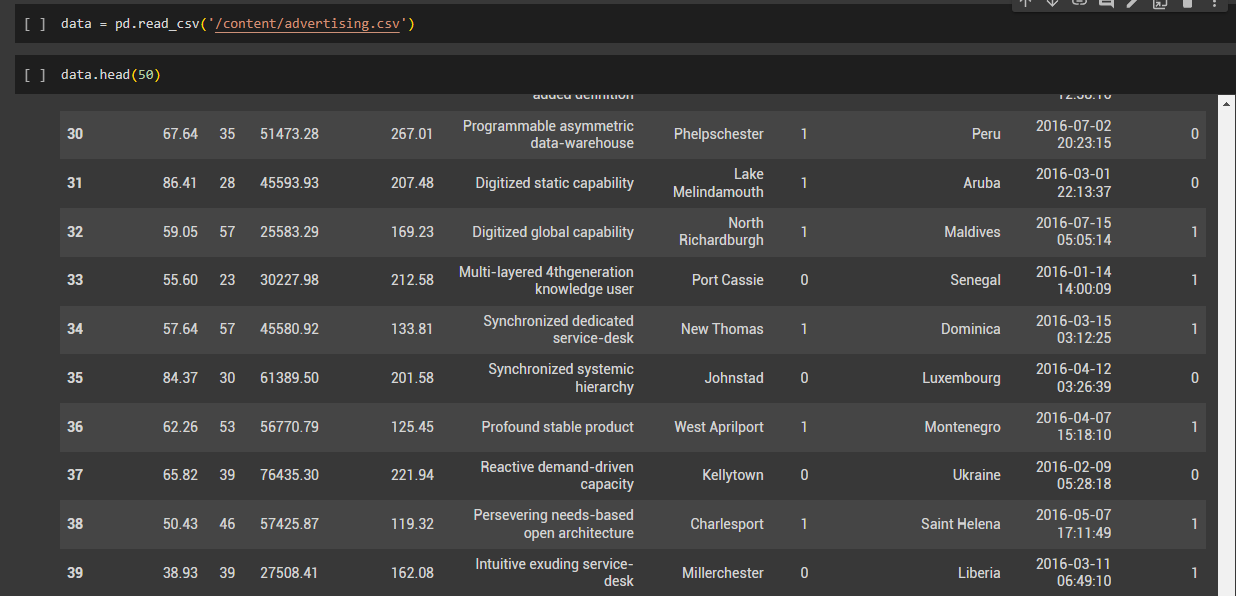


Рисунок 3 – импортированный датасет

Далее мы начинаем анализ данных, чтобы понять определённые закономерности такие как количество пользователей, посетивших сайт и их возраст. Создаем график с помощью функций plotlib.figure и seaborn.distplot.

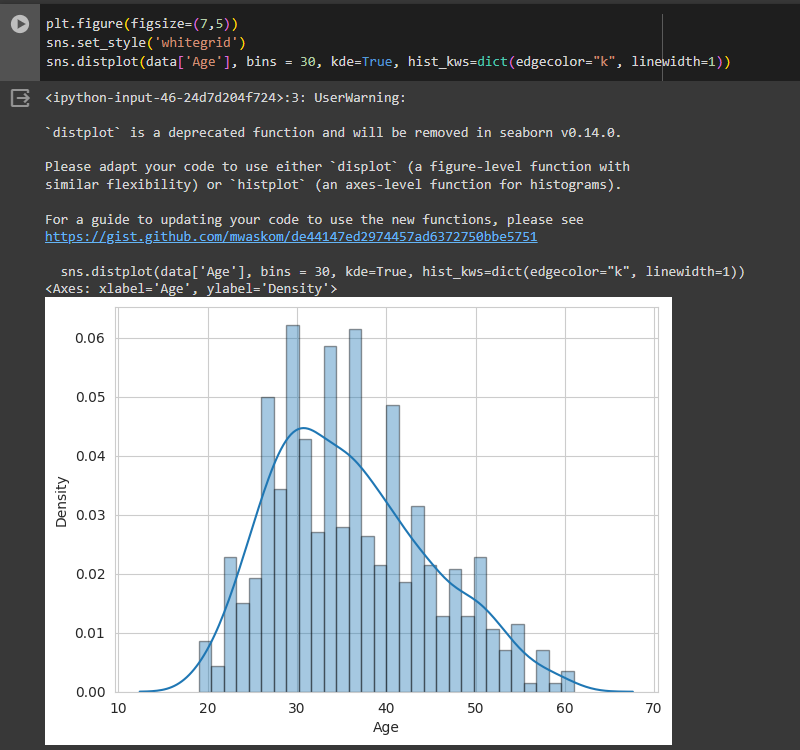


Рисунок 4 – график зависимости количества от возраста пользователей

С помощью функции seaborn.pairplot мы можем построить сразу несколько графиков зависимостей в одном изображении, для наглядного изучения зависимостей

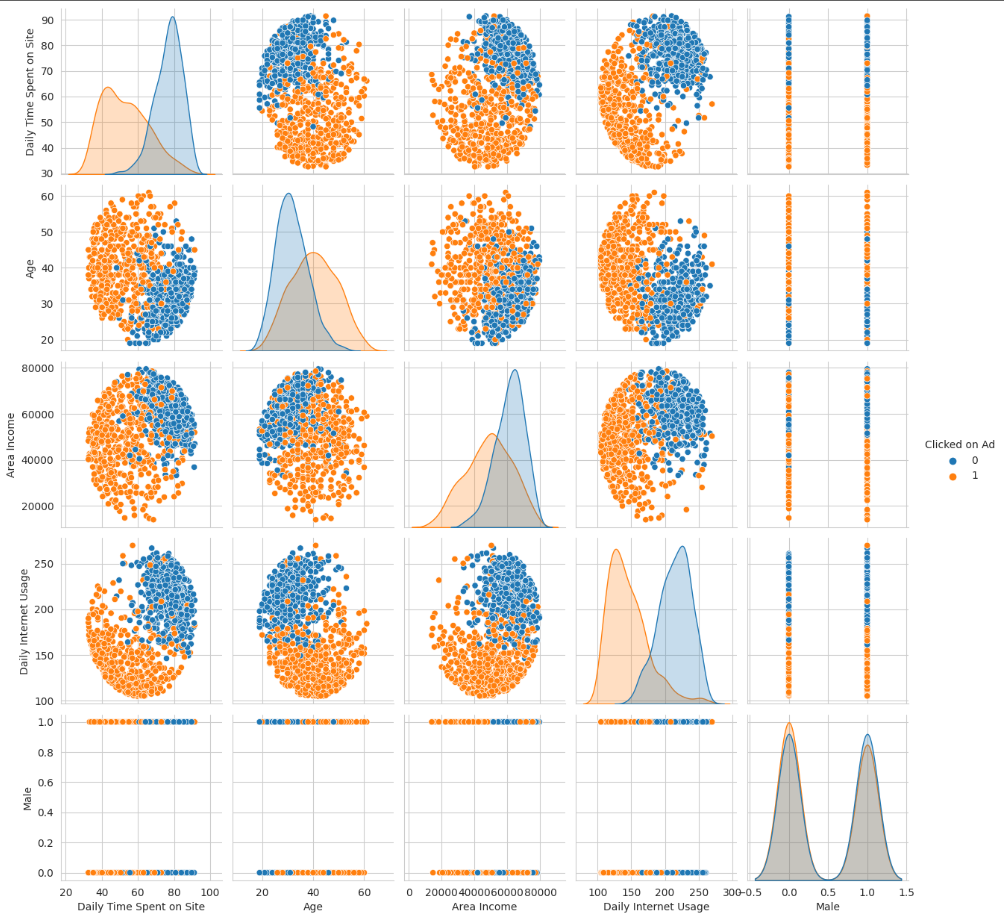


Рисунок 5 – графики зависимостей по всем параметрам

1. Очистка данных

Перед отчисткой проведём проверку на потерянные данные с помощью построения тепловой карты

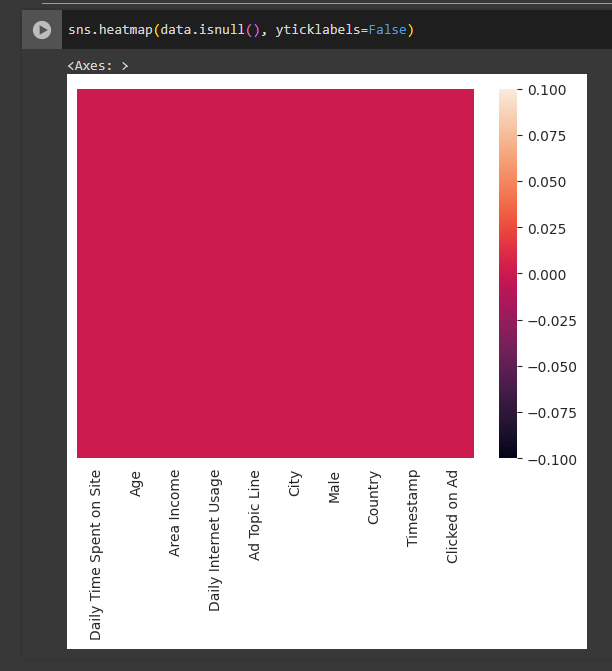


Рисунок 6 – тепловая карта потери данных

Учитывая, что все тепловая карта одноцветная и красная, что соответствует значению 0.000, можем заключить, что потери данных нет.

Теперь, для того чтобы отчистить данные, отбросим нечисловые параметр Ad topic line, параметры City и Country заменим фиктивными переменными с числовыми значениями, а Timestamp разобьём на время и месяц



Рисунок 7 – обработка данных

Далее мы отбрасываем уже более не нужные нам параметры Ad topic line, city, country, timestamp, clicked on ad и у нас выходит следующий датасет

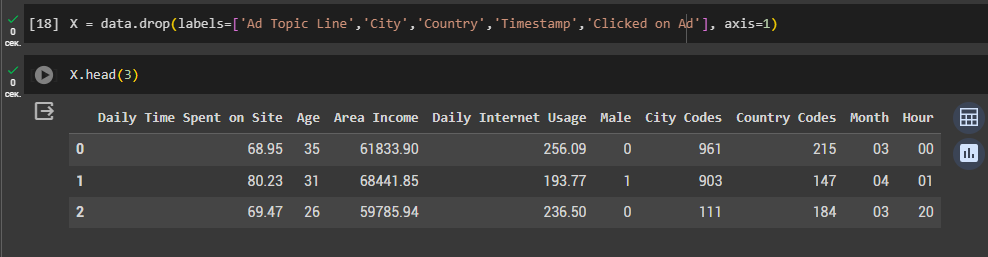


Рисунок 8 – отчищенный датасет

Теперь, когда мы отчистили датасет и параметры стали числовыми, можем переходить к обучению алгоритма

1. Обучение алгоритма

На этом этапе мы разбиваем наши данные с помощью функции sklearn.model\_selection import train\_test\_split на тестовый и обучающий набор, при этом данные clicked on ad мы заносим под y, а наш отчищенный датасет под x.

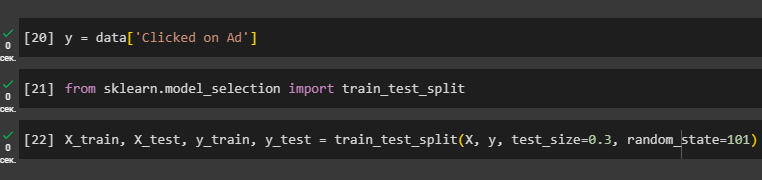


Рисунок 9 – разделение данных на обучающий и тестовые набор

Далее мы импортируем модель линейной регрессии и запускаем её обучение

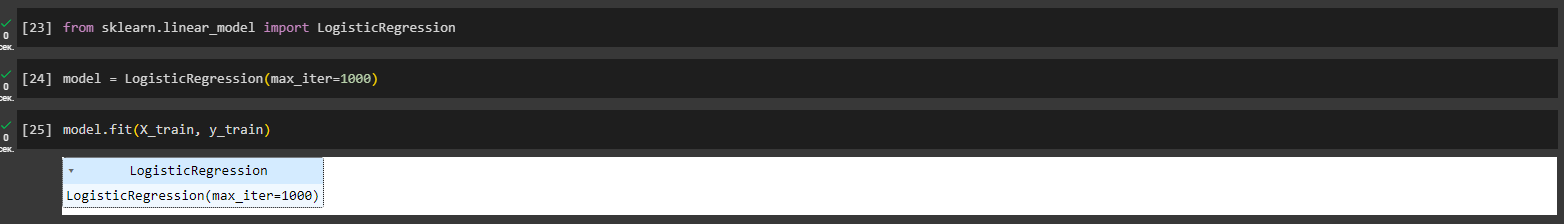


Рисунок 10 – обучение алгоритма логической регрессии

1. Прогнозы и оценки

Теперь с помощью библиотеки sklearn, мы можем спрогнозировать наши тестовые данные и выдать оценку модели по точности, полноте, f1 и количеству считаных объектов класса.

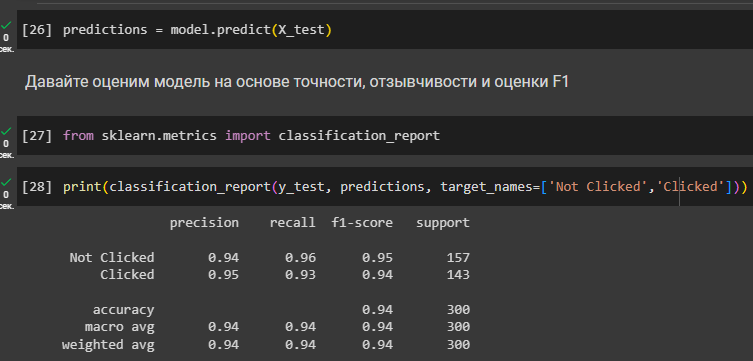


Рисунок 11 – оценка модели

Из данной оценки можно понять, что все значения больше 0.93, что является хорошим результатом.

Вывод

В данной контрольной работе была рассмотрена разработка модели для оценки эффективности рекламных кампаний на основе данных о взаимодействии пользователей с рекламным контентом. Целью работы было предсказание реакции пользователей на рекламу и определение факторов, влияющих на эффективность рекламных кампаний.

Результаты оценки модели показали ее способность предсказывать эффективность рекламных кампаний с высокой точностью. Модель позволяет оценить, насколько успешно рекламная кампания привлекает внимание и вызывает реакцию у целевой аудитории.

Разработанная модель имеет практическую значимость для маркетинговых специалистов и рекламных агентств. Она позволяет оптимизировать рекламные стратегии, улучшить эффективность рекламных кампаний и повысить вовлеченность пользователей с рекламным контентом. Модель может быть использована для принятия обоснованных решений в области рекламы и маркетинга, а также для оптимизации распределения рекламного бюджета.

# Список литературы

1. Метрики в задачах машинного обучения [Электронный ресурс], 14.01.2024 URL: <https://habr.com/ru/companies/ods/articles/328372/>
2. Precision и recall. Как они соотносятся с порогом принятия решений? [Электронный ресурс], 14.01.2024 URL: <https://habr.com/ru/articles/661119/>
3. Документация библиотеки Seaborn [Электронный ресурс], 14.01.2024 URL: <https://seaborn.pydata.org/>
4. Рашка Себастьян, Логунов А.В. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения: Практическое пособие Москва: ДМК Пресс, 2017
5. Ссылка на проект Github https://github.com/VadimU1/EducationAdProj