
Вы – управляющий сетью элитных фитнес-клубов, которые всегда были на острие передовых технологий в области спортивной медицины и аналитики. Но всё меняется, когда в один из ключевых филиалов неожиданно перестаёт поступать поток новых клиентов. Выясняется, что конкуренты внедрили «умную» систему персональных тренировок на основе машинного обучения, которая прогнозирует сожжённые калории с невероятной точностью и тем самым привлекает посетителей всё сильнее.

Ваша цель – взять реванш и доказать, что ваш фитнес-центр не зря славится своими инновационными решениями. Вы решаете провести масштабное исследование исторических данных о тренировках, разработать собственную модель предсказания сожжённых калорий и встроить результаты в удобный дашборд для руководства. Однако времени катастрофически мало – конкуренты уже дышат в спину, а инвесторы требуют конкретных цифр.

В невероятной атмосфере срочности и высоких ставок вам предстоит:

1. Провести глубокий анализ данных (EDA), чтобы разобраться, какие факторы влияют на калорийный расход.
2. Разработать машинную модель, которую трижды улучшить, доказывая, что вы движетесь в правильном направлении.
3. Создать интерактивный дашборд для наглядной демонстрации результатов руководству.

Но есть одно важное условие: финальные предсказания модели для дашборда использовать не нужно – достаточно показать там итоги исследований и потенциал ваших новых аналитических подходов.

Техническое задание

Часть 1. Exploratory Data Analysis (EDA)

Цель исследования

Провести комплексный анализ фитнес-данных, выявить ключевые закономерности и факторы, влияющие на сожжённые калории, чтобы получить конкурентное преимущество за счёт персонализированных тренировок.

Задачи

1. Анализ распределений:

- Постройте графики распределения для числовых переменных
- Изучите распределение по категориям (пол, тип тренировки)

2. Проверка данных:

- Определите наличие пропущенных значений, выбросов и аномалий.
- Примите решение, как корректировать или учитывать эти пропуски и аномалии при дальнейшем анализе.

3. Корреляционный анализ:

Исследуйте взаимосвязи между признаками.

- Используйте карту корреляций или аналогичный инструмент для визуализации.

4. Гипотезы и инсайты:

- Сформулируйте предположения, какие факторы наиболее влияют на сжигание калорий.
- Опишите возможные сценарии оптимизации тренировок (например, «Если увеличить средний пульс до X, то при той же продолжительности мы сожжём на Y% калорий больше»).

Ожидаемый результат

- Читаемые и понятные графики, таблицы и статистика.
- Аналитический отчёт с основными выводами, гипотезами и рекомендациями для руководства фитнес-клуба, оформленный так, чтобы было ясно, что вы провели глубокий анализ и готовы к дальнейшим шагам.

Часть 2. Моделирование и Машинное Обучение

Задача

Разработать и задокументировать ML-пайплайн для предсказания колонки

Calories_Burned с использованием регрессионной модели. **Важно:** продемонстрировать три итерации улучшения метрики. Способы улучшения – на ваше усмотрение (обработка выбросов, добавление новых признаков, разные алгоритмы, подбор гиперпараметров и т. д.).

При этом окончательные прогнозы модели **не** будут использоваться в дашборде – его цель лишь отразить результаты EDA и сам факт наличия ML-подхода. Однако показать, что метрики по модели действительно улучшались на каждом шаге, необходимо.

Обязательные требования

- Ясное описание выбранной метрики (RMSE, MAE и пр.) и почему с точки зрения бизнеса это важно.
 - Документирование всех этапов с пояснениями – от идей до финальных результатов.
 - Код в Jupyter Notebook или аналогичном формате с комментариями и описаниями, куда может заглянуть любой коллега и понять логику.
 - Опишите, какие шаги нужно пройти, чтобы выпустить модель в прод
 - Опишите в отчете, на какие метрики после запуска модели вы будете смотреть, контролируя, что модель функционирует нормально
-

Часть 3. Создание Интерактивного Дашборда

Задача

Создать интерактивный дашборд для наглядного представления результатов исследования перед руководством. Руководство хочет понимать общие закономерности тренировок, особенности клиентов, динамику ключевых метрик.

Требования к дашборду

1. Основные результаты EDA:

- Графики распределений, динамика тренировок по типам, факторам пульса и т. д.

2. Фильтры для сегментации данных:

- По возрасту, типу тренировки, продолжительности сессии и другим ключевым признакам, которые вы считаете важными.
- Эти фильтры должны помогать менеджерам быстро увидеть срезы данных и принимать решения.

3. Визуализации:

- Графики, диаграммы, таблицы – всё должно быть представлено наглядно и доступно даже для пользователей без технического бэкграунда.
- Если данные содержат географическую привязку (локации залов) – можно добавить интерактивную карту.

Ожидаемый результат

- Полноценный дашборд в Yandex DataLens.
-

Формат сдачи и дополнительные указания

1. Единый отчёт

- Все этапы проекта (EDA, модель, дашборд) описываются последовательно в одном документе или ноутбуке.
- Должна быть логика повествования: от постановки задачи до итогового представления решения.

2. Презентация и короткое видео

- На базе итогового отчёта подготовьте презентацию.
- Запишите видео длительностью не более 10 минут, чтобы кратко показать основные инсайты анализа, дашборд, объяснить основные шаги проведенного анализа
- Указать обязанности каждого члена команды

3. Документирование

- Каждый график должен иметь подпись, поясняющую, что на нём изображено и какие выводы можно сделать.
- Код должен сопровождаться комментариями и объяснениями ключевых шагов.
- Если используете что-то сверх пройденного курса (дополнительные библиотеки, математические приёмы и т. д.), дайте ссылки или краткое описание.

4. **Из чего складывается оценка:**

- **40%:** отчет (презентация с EDA, описанием шагов, расчетные файлы)
- **60%:** защита проекта (отчетное видео с презентацией и дашбордом)

5. **Критерии оценивания**

- **До 3 баллов:** за качество EDA (глубина анализа, визуализации, проверка гипотез).
 - **До 3 баллов:** за ML-пайплайн (масштабирование, feature engineering, подбор гиперпараметров) и демонстрацию улучшения метрик (3 итерации).
 - **До 2 баллов:** за интерактивный дашборд (доступность, структурированность).
 - **До 2 баллов:** за потенциальную пользу для бизнеса (инсайты по привлечению и удержанию спортсменов)
 - **Итого максимум 10 баллов.**
-

Удачи!