Задание 1: Исследование открытого набора данных.

Критерии оценки:

Выбор набора данных (2 балла)

• Подбор актуального и интересного набора данных. (1 балл)

• Наличие детального объяснения причин выбора данного набора данных. (1 балл)

Описание структуры данных (3 балла)

• Ясное и полное описание структуры данных (количество строк и столбцов). (1 балл)

• Указание типов переменных (например, числовые, категориальные). (1 балл)

• Наличие информации о пропущенных значениях (количество, процент). (1 балл)

Предварительный анализ данных (4 балла)

• Определение и построение графиков распределения для количественных переменных (гистограммы, диаграммы размаха и т.д.). (1 балл)

• Наличие комментариев к графикам, объясняющих ключевые наблюдения. (1 балл)

• Исследование корреляции между переменными с помощью тепловой карты. (1 балл)

• Интерпретация корреляции (например, положительная, отрицательная, сильная или слабая). (1 балл)

Очистка данных (2 балла)

• Правильное устранение пропущенных значений (например, удаление, замена, интерполяция). (1 балл)

• Выявление и обработка выбросов (например, с использованием межквартильного диапазона или стандартного отклонения). (1 балл)

Задание 2: Построение и оценка модели машинного обучения.

Критерии оценки:

Выбор целевой переменной и подготовка данных (2 балла)

• Целевая переменная очевидно определена, и обоснован его выбор (это должно быть основано на целях анализа). (1 балл)

• Данные подготовлены для обучения: нормализация, кодирование категориальных переменных и т.д. (если это необходимо). (1 балл)

Разделение данных (2 балла)

• Данные разделены на обучающую и тестовую выборки правильно (например, 80/20). (1 балл)

• Обоснование выбора пропорции разделения. (1 балл)

Выбор и обучение модели (2 балла)

• Выбор подходящей модели обосновывается с учётом типа задачи (регрессия или классификация). (1 балл)

• Модель обучена корректно, с минимально необходимыми параметрами и испытаниями (например, кросс-валидация). (1 балл)

Оценка производительности модели (2 балла)

• Использованы адекватные метрики для оценки модели (например, MSE, точность, F1-меры). (1 балл)

• Результаты представлены с пояснениями и интерпретациями, включая возможные ограничения. (1 балл)

Задание 3: Визуализация и интерпретация данных.

Критерии оценки:

Измерение влияния факторов (2 балла)

• Корректный выбор факторов, влияющих на целевую переменную. (1 балл)

• Анализ влияния с использованием корректных методов (например, регрессионный анализ, ANOVA). (1 балл)

Создание интерактивных визуализаций (2 балла)

• Использование библиотеки Plotly для создания интерактивных графиков. (1 балл)

• Разнообразие визуализаций (scatter plots, bar charts, line plots и т.д. для разных переменных). (1 балл)

Качество визуализаций (2 балла)

• Четкость и наглядность визуализаций (подписи, заголовки, легенды). (1 балл)

• Эстетика и профессиональный вид графиков. (1 балл)

Презентация результатов (2 балла)

• Структурированная презентация с логическим изложением выводов и рекомендаций. (1 балл)

• Объяснение интерпретации визуализаций и их значения. (1 балл)

Шкала и критерий оценивания

• 22-27 баллов — Задание выполнено полностью, оптимально и без ошибок. Код структурирован, читабелен и соответствует требованиям.

• 16-21 баллов — Задание выполнено почти полностью, имеются незначительные ошибки или недочеты, не влияющие на общую логику.

• 9-15 баллов — Основная часть задания выполнена, но с недочетами, требующими корректировки.

• 5--8 балла — Задание выполнено частично или с заметными ошибками.

• 0-4 балла — Задание выполнено некорректно, нарушена логика или задание выполнено не в полном объеме.