

# Описание общего принципа для решения задач

Рысьева Елизавета Антоновна

[GitHub с решением конкретной задачи](#)

## Для проведения разведочного анализа данных можно использовать следующие методы

1. Анализ описательных статистик для каждого столбца таблицы (среднее, медиана, минимум, максимум, стандартное отклонение и т.д.) для оценки распределения данных. `Python: df.describe()`
2. Построение графиков зависимости выходной переменной от каждой из входных переменных для оценки возможной линейной или нелинейной связи между ними.
3. Корреляционный анализ для оценки силы и направления связи между парами переменных. `Python: df.corr()`
4. Анализ выбросов и пропущенных значений для каждого столбца таблицы.

# Для выбора критерия точности модели можно использовать следующие методы

1. Коэффициент детерминации ( $R^2$ ) для оценки объяснительной способности модели.
2. Средняя квадратическая ошибка (MSE) для оценки точности прогнозов.
3. Средняя абсолютная ошибка (MAE) для измерения среднего отклонения предсказанных значений от фактических значений
4. Медианная абсолютная ошибка (MAE) для измерения медианного отклонения предсказанных значений от фактических значений. Эта метрика более устойчива к выбросам, чем средняя абсолютная ошибка .

## Для оценки важности можно использовать следующие подходы

1. В виде таблицы оптимальных признаков с помощью feature selection:
2. Сортировки значений важности признаков по убыванию и вывода графика с помощью библиотеки matplotlib
3. Метод главных компонент (PCA) для выделения наиболее информативных признаков.
4. Методы отбора признаков на основе статистических тестов, например, ANOVA, t-test и др.
5. Методы регуляризации, такие как Lasso и Ridge, для оценки важности признаков на основе их весов в модели.

Для улучшения получившегося результата можно предложить следующие методы

1. Использование алгоритмов машинного обучения, таких как случайный лес, градиентный бустинг для улучшения точности прогнозов.
2. Использование методов ансамблирования, таких как бэггинг и бустинг, для уменьшения разброса прогнозов.