**Отчет: Прогнозирование качества вин**

**1. Загрузка и предварительная обработка данных**

Описание данных:  
Мы используем набор данных о вине, который содержит такие характеристики, как кислотность, сахар, хлориды и алкоголь. Целевой признак - качество вина, выраженное в виде оценок.

Была проведена предобработка данных, а именно:

- проведен исследовательский анализ признаков (проверены типы данных, наличие пропущенных значений, вычислены основные статистики);

- построены гистограммы распределений для каждого признака, что показало различные распределения, большинство из которых – асимметричны (для уточнений распределений нужно провести тесты);

- выполнена корреляция признаков с целью выявить взаимосвязи между ними, используя корреляционную матрицу (например, выявлена положительная корреляция между citric acid и fixed acidity и отрицательная между fixed acidity и ph);

- выполнена нормализация признаков с целью улучшения обучения модели.

**2. Построение и обучение модели**

**Архитектура модели:**  
Для предсказания качества использовалась полносвязная нейронная сеть (Feedforward Neural Network). Была выбрана функция активации между полносвязанными слоями relu. Обучение выполнено методом градиентного спуска (в качестве оптимизатора был выбран Adam – объединяет преимущества других оптимизаторов AdaGrad и RMSProp).

**Обучение и метрики:**  
Для оценки точности модели использовались метрики:

* R2 (коэффициент детерминации).
* MAE (средняя абсолютная ошибка).

**3. Результаты и визуализация**

**График зависимости предсказаний и истинных значений:**  
На графике представлены истинные и предсказанные значения качества вина. Идеальный прогноз лежал бы на красной пунктирной линии, указывающей на идеальное совпадение предсказанных и истинных значений. На текущем графике видны отклонения, что указывает на необходимость улучшения модели. В качестве улучшения можно предложить увеличить количество слоев/нейронов, увеличить количество эпох обучения, изменить learning rate оптимизатора и тд. Также можно попробовать другие модели – типа градиентного бустинга или бэггинга. Применить кросс-валидацию для более надежной оценки модели и предотвращения переобучения

Также была построена гистограмма распределения ошибок, которая показывает, что большинство ошибок имеют небольшое абсолютное значение и подтверждает адекватное качество предсказания. Распределение ошибок является нормальным, что свидетельствует о стабильности модели.