Отчет о выполненной работе

# Тип отчета – информационный.

В рамках задачи предлагалось сделать оценку качества вина на основании признаков из датасета. Задача не конкретно прикладная, а больше фантазийная. Прикладной смысл может быть например для владельцев ресторанов или поставщиков вин – понимать по харрактеристикам вина определять какое закупать, чтобы клиенты оставались довольны качеством, а затраты были оптимальными.

# Аудитория отчета

Аудиторией отчета является преподаватель модуля. Главное что должно быть – это понимание преподавателя, что студент понял как составляется отчет. Поэтому главное структура, а не наполнение и результат.

# Результаты работы:

В ходе работы на практике ознакомились со всеми этапами Crisp-DM за исключением Evaluation и Deployment, поскольку пример учебный и не является прикладной задачей.

Для себя выделил – что необходимо выделить время для освоения базовых конструкций Python для избавления от повторения одинакового кода внутри проекта – но так как это требует отельных трудозатрат по вниманию и времени, в рамках данной работы данным требованием принебрёг специально.

В рамках работы использовались 2 модели – линейная регрессия и случайный лес. Случайный лес показал более хороший результат, но не достаточной для исходного целевого.

Целевое значение было определено 90% до начала выполнения работы по изучению данных, только на бизнес предположении. Но в силу маленького датасета, оказалось недостижим для выполнения. Были попытки оптимизировать модель за счет исключения выбросов и не значимых критериев. Но существенного влияния на качество моделей это не дало, поэтому эту работу можно положить “в стол”. Полезным результатом вижу то, что на этапе сбора данных нужно ставить более реалистичные данные, и от маленького датасета не следует ожидать каких то фундаментальных инсайтов, дающих точность на уровне 85-100%