Слайд№1. Коллеги, добрый день! Меня зовут Роман! Вашему вниманию представляется моя дипломная работа по созданию модели по определению качества вина по заданным характеристикам.

Слайд №2. Постановка задачи состоит в создании модели предсказания качества вина по заданным характеристикам. Этот датасет связан с вариантами португальского красными вина "Vinho Verde". Датаесет описывает количество различных химических веществ, присутствующих в вине, и их влияние на его качество.

Слайд №3 Описание решения. Импортируем библиотеки, необходимые для создания модели. Подгружаем датасет. Всего представлено 11 химических признаков и целевая переменная оценки качества вина. Пропусков нет, все признаки числовые. Далее добавление новых признаков (Feature engineering). Далее разведывательный анализ данных (EDA). Из распределения целевой переменной видно, что представленные образцы имеют оценки от 3 до 8, в основном оценки 5 и 6 (83%). Из матрицы корреляции можно сделать вывод, что скоррелированных между собой признаков нет. С помощью теста Стьюдента определили, что только три признака серьезно оказывают влияние на оценку. Далее подготовка данных (использовал кросс-валидацию) и непосредственно сам Machine Learning.

Слайд №4. Для оценки модели я использовал метрику МАЕ. Напомню это среднее арифметическое модуля отклонения предсказанного значения от реального. Также на слайде видно какие именно модели я использовал.

Слайд №5. Наилучший показатель метрики является модель Random Forest Regressor

с подбором гиперпараметров.

Мою модель можно применить на производстве вин.