Ярошевич В.А.

ФИТУ 3 курс

090302-ИСТа-о22

Практическая №17

Multiple Linear Regression

Цель работы: использовать scikit-learn для множественной линейной регрессии, создание, оценивание и использование модели для прогнозирования неизвестного значения.

Multiple Linear Regression (множественная линейная регрессия) — это статистический метод, который используется для моделирования зависимости между одной зависимой переменной (обычно обозначается как y y y) и несколькими независимыми переменными (обычно обозначаются как x1,x2,...,xn x_1, x_2, ..., x_n x1,x2,...,xn). Цель этого метода — найти линейное уравнение, которое наилучшим образом описывает эту зависимость.

Загружаем файл, содержащий показатели расхода топлива для конкретной модели и предполагаемые выбросы углекислого газа для новых малотоннажных автомобилей для розничной продажи в Канаде. Построим график зависимости значений выбросов от объема двигателя:

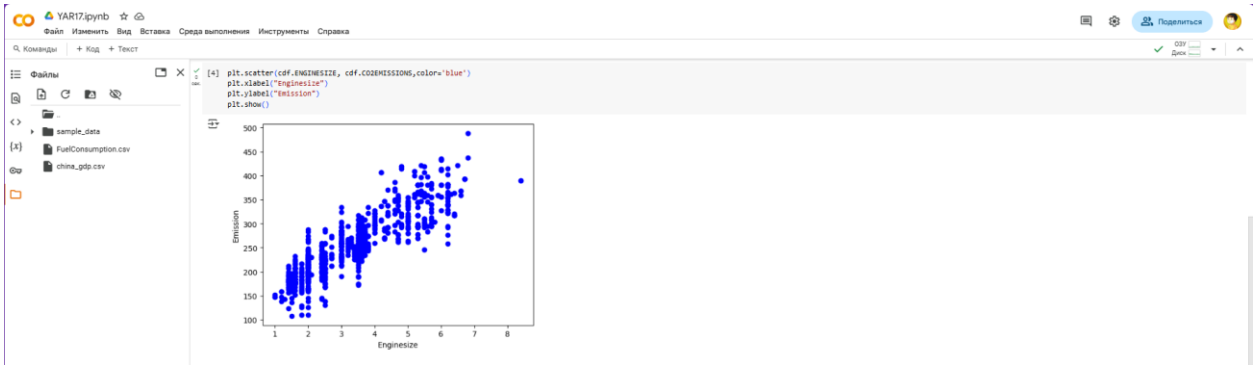

Рисунок 1 - Модель и коэффициенты

Рисунок 1.1 – Данные точности

Далее практическое задание. Нам необходимо изменить один из параметров, на которых строится модель и после оценивается. А конкретно

посмотреть, какие будут значения при «**FUELCONSUMPTION_CITY**» и «**FUELCONSUMPTION_HWY**»:

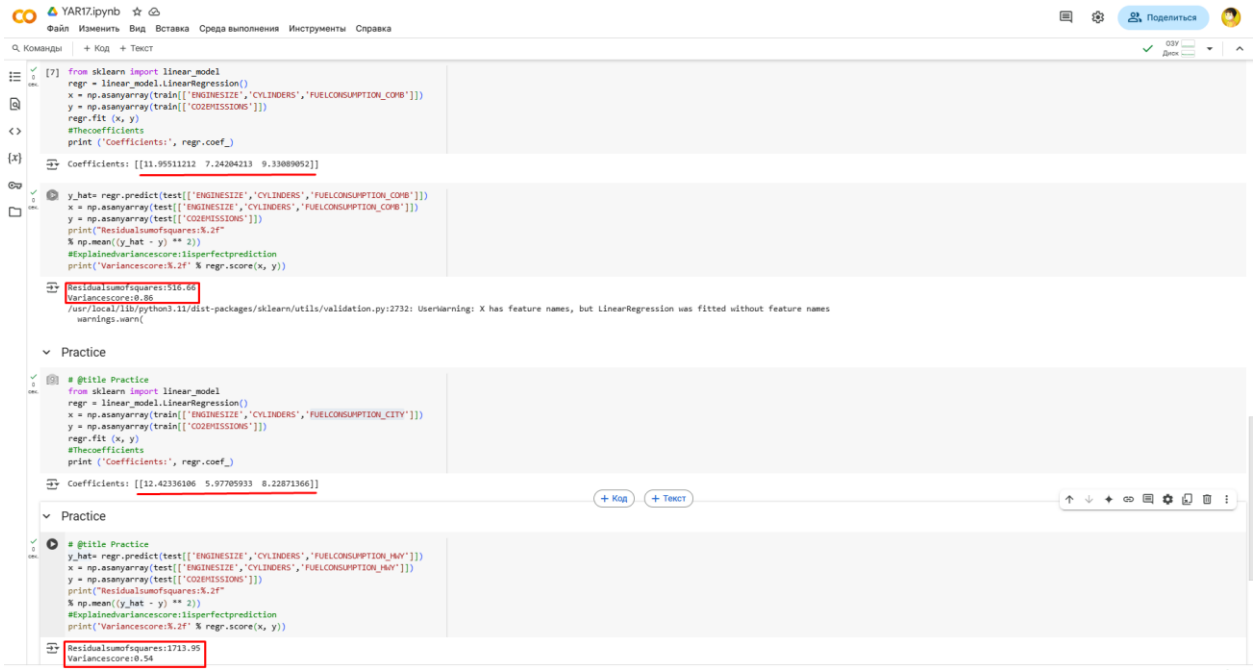

Рисунок 1.2 - Коэффициенты и новые точности

Исходя из данных с рисунков мы видим, что точность при «городской» переменной немного лучше, в то время как при «шоссе» она заметно хуже.

Вывод: в результате практической работы мы реализовали scikit-learn для множественной линейной регрессией. Создали и оценили собственную модель для прогнозирования неизвестного значения. И в процессе оценки поняли, что лучше проверять все переменные, чтобы найти самую точную.