Отчет по 5 лабораторной работе

Для выполнение данной лабораторной работы выбран датасет “Average car price in Brazil” на сайте Kaggle. <https://www.kaggle.com/datasets/vagnerbessa/average-car-prices-bazil>

Краткое описание датасета

Каждое наблюдение (строка) соответствует средней цене автомобиля, рассчитанной на месяц отчетного года. Другими словами, в году (year\_of\_reference) есть 12 наблюдений, связанных с одним и тем же автомобилем, однако avg\_price\_brl может отличаться.

Переменные fuel, gear и engine\_size были извлечены из значений модели столбца, так как в оригинале нет столбца, предназначенного для этих значений. Поскольку некоторые значения для модели не содержат информации о объеме двигателя, этот набор данных не содержит всех данных из оригинала FIPE. Кроме того, если «Авт.» в модели нет, предполагается, что машина имеет механическую коробку передач.

Цены рассчитаны FIPE и указаны здесь как оригинальные (в бразильских реалах).

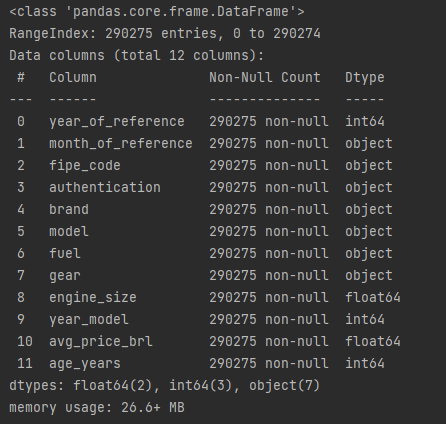
FIPE обновляет информацию ежемесячно. Здесь указанный месяц указан как month\_of\_reference, поскольку соответствующий год имеет переменную с именем year\_of\_reference.

Коды FIPE (fipe\_codes) — это идентификатор модели, используемый на веб-странице FIPE.

Были выбраны следующие классификаторы:

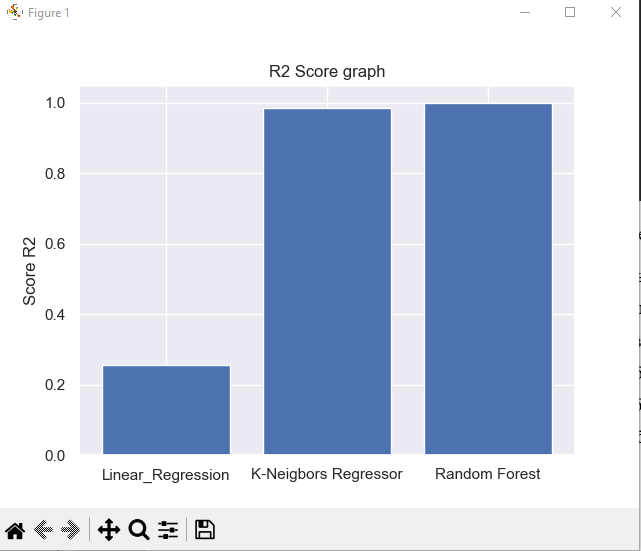
1. Random Forest Regressor
2. K-Neigbors Regressor
3. Linear Regression

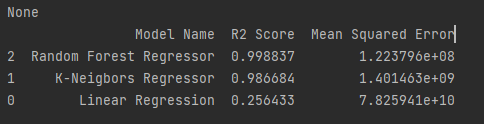
Датасет не требует очистки:



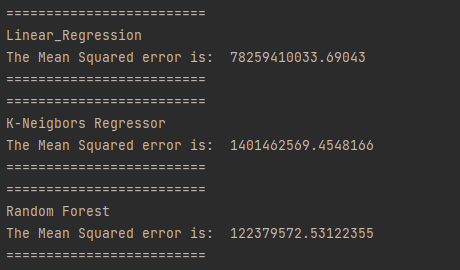
Эффективность обучение и оценки моделей покажет оценка R2:

Оценка R2 – очень важный показатель, который используется для оценки производительности модели машинного обучения на основе регрессии. Он произносится как R в квадрате и также известен как коэффициент детерминации. Суть его работы заключается в измерении количества отклонений в прогнозах, объясненных набором данных. Проще говоря, это разница между выборками в наборе данных и прогнозами, сделанными моделью.





Среднеквадратичная ошибка каждого классификатора:



Таким образом, самым точным и эффективным классификатором является: “Random Forest”, вторым является: “K-Neigbors Regressor”, третьим - “Linear Regression”.