

УДК: 004.8

DOI: <https://doi.org/10.47813/2782-2818-2024-4-2-0298-0305>

EDN: QTRIU



Применение метода линейной регрессии для прогнозирования расхода топлива автомобиля

М. А. Зуев, В. М. Шibaев, К. С. Баланев

ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

Аннотация. В данной статье рассматривается использование метода машинного обучения для прогнозирования расхода топлива автомобиля. Для этой задачи применяется линейная регрессия. Описаны этапы подготовки данных, включая обработку данных и устранение пропусков, а также процессы обучения и тестирования модели. Проведен анализ данных и построена визуализация тепловой карты корреляций. Построена модель линейной регрессии на основе признаков, влияющих на расход топлива автомобиля. Проведена оценка эффективности модели с использованием метрик RMSE и R^2 . Основное внимание уделено практическому применению модели для предсказания расхода топлива на основе реальных данных.

Ключевые слова: машинное обучение, линейная регрессия, прогнозирование расхода топлива, обработка данных, корреляционный анализ.

Для цитирования: Зуев, М. А., Шibaев, В. М., & Баланев, К. С. (2024). Применение метода линейной регрессии для прогнозирования расхода топлива автомобиля. *Современные инновации, системы и технологии - Modern Innovations, Systems and Technologies*, 4(2), 0298–0305. <https://doi.org/10.47813/2782-2818-2024-4-2-0298-0305>

Application of the linear regression method to predict vehicle fuel consumption

M. A. Zuev, V. M. Shibaev, K. S. Balanov

National Research University «Moscow Power Engineering Institute», Moscow, Russia

Abstract. This article discusses the use of machine learning to predict vehicle fuel consumption. Linear regression is used for this task. The stages of data preparation are described, including data processing and elimination of omissions, as well as the processes of training and testing the model. The data was analyzed and a visualization of the correlation heat map was built. A linear regression model is constructed based on the features that affect the fuel consumption of the car. The effectiveness of the model was evaluated using the RMSE and R^2 metrics. The main attention is paid to the practical application of the model for predicting fuel consumption based on real data.

Keywords: machine learning, linear regression, fuel consumption forecasting, data processing, correlation analysis.

For citation: Zuev, M. A., Shibaev, V. M., & Balanov, K. S. (2024). Application of the linear regression method to predict vehicle fuel consumption. *Modern Innovations, Systems and Technologies*, 4(2), 0298–0305. <https://doi.org/10.47813/2782-2818-2024-4-2-0298-0305>

ВВЕДЕНИЕ

Машинное обучение представляет собой область искусственного интеллекта, которая занимается разработкой алгоритмов и статистических моделей, позволяющих компьютерам улучшать свои функции на основе данных. В отличие от традиционного программирования, где инструкции прописываются явно, машинное обучение позволяет моделям обучаться на основе примеров и выявлять закономерности в данных. Тем самым машинное обучение является особенно полезным для задач, где существует большая сложность и объем данных, таких как прогнозирование, классификация, и распознавание образов [1].

В данной работе используется обучение с учителем, при котором модель обучается на размеченных данных, где каждому входу соответствует известный выход. Это позволяет модели делать предсказания на новых, невиданных ранее данных.

Линейная регрессия - один из самых простых и широко используемых методов машинного обучения. Этот метод позволяет моделировать зависимость между одной зависимой переменной (в данном случае, расходом топлива) и одной или несколькими независимыми переменными (характеристиками автомобиля). Линейная регрессия предполагает, что между этими переменными существует линейная связь, и цель метода - найти такую прямую (линии регрессии), которая наилучшим образом описывает зависимость данных [1].

Математически модель множественной линейной регрессии представлена уравнением:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (1)$$

где y – зависимая переменная, β_0 – свободный член, $\beta_1, \beta_2, \dots, \beta_n$ – коэффициенты регрессии, x_1, x_2, \dots, x_n – независимые переменные, а ϵ – ошибка модели.

МАТЕРИАЛЫ И МЕТОДЫ

Выбор набора данных

Для исследования используется набор данных "Auto MPG Dataset", который содержит информацию о расходе топлива автомобилями, а также их характеристиках, таких как количество цилиндров, мощность, вес и другие параметры. Набор данных состоит из следующих столбцов:

1. mpg - миль на галлон.
2. cylinders – количество цилиндров.
3. displacement – объем двигателя.
4. horsepower – лошадиные силы.
5. weight – вес.
6. acceleration – ускорение.
7. modelyear – модельный год.
8. carname – модель автомобиля.

Выбор инструментов решения задачи

Для решения задачи линейной регрессии был использован язык программирования Python и среда Google Colab, а также библиотеки [2]:

1. pandas – для чтения данных.
2. numpy – для оценки качества модели.
3. matplotlib и seaborn – для визуализации данных.
4. sklearn – для создания модели линейной регрессии.

Подготовка данных

Перед началом создания модели линейной регрессии следует подготовить данные. Методами .head, .info и .describe были выявлены типы данных, выбросы и количество пропущенных значений [3-5]. В столбце horsepower тип данных – object, так как у некоторых моделей машин неизвестен параметр лошадиных сил, и вместо числа хранится вопросительный знак – «?». Были удалены все строки с параметром horsepower равным «?» и данный столбец был преобразован в тип данных int.

Определение коррелирующих параметров

После предварительной обработки данных следует построить тепловую карту корреляции, чтобы определить какие параметры автомобилей влияют на расход топлива (рисунок 1).

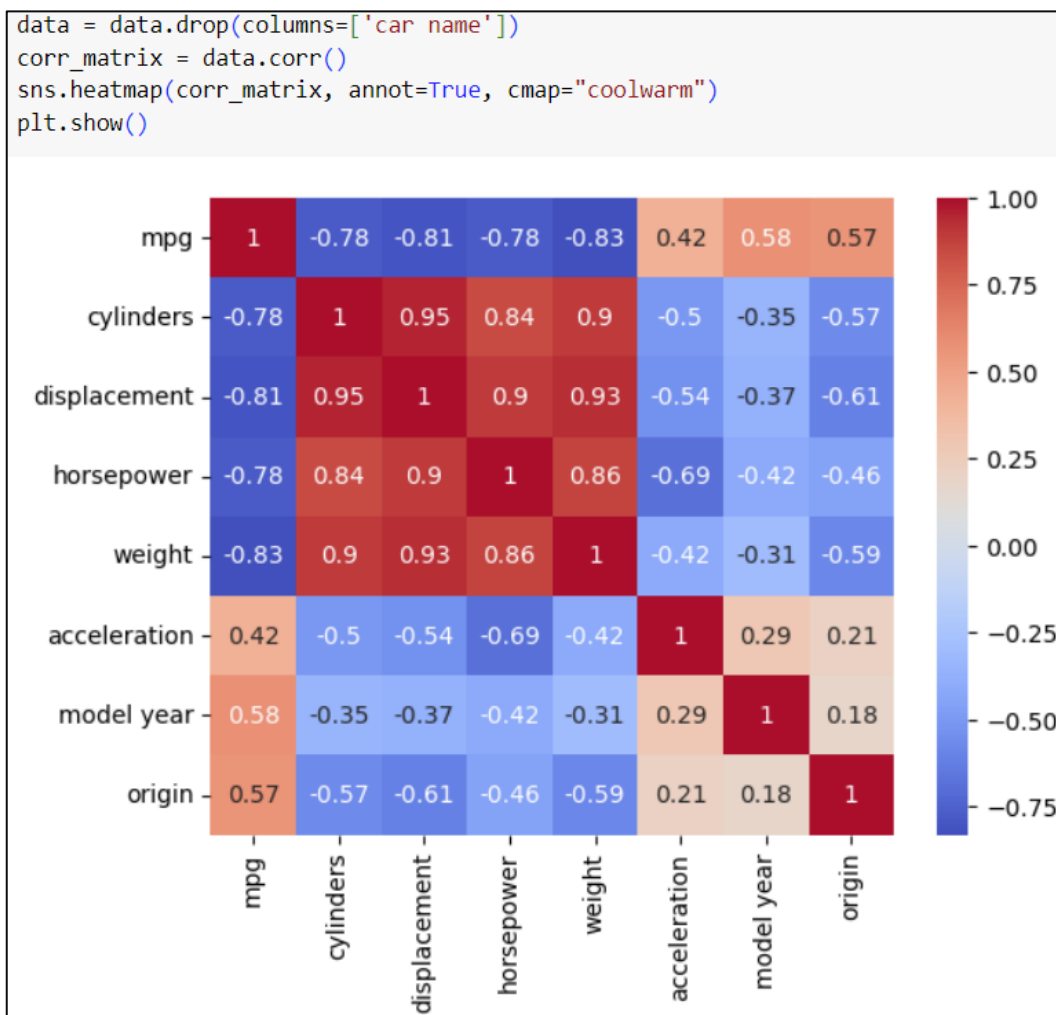


Рисунок 1. Тепловая карта корреляции.

Figure 1. Correlation heat map.

Из тепловой карты корреляции видно, что все параметры автомобиля имеют либо положительную, либо отрицательную корреляцию относительно расхода топлива, это значит, что все имеющиеся параметры так или иначе влияют на расход топлива [6]. Все показатели были использованы для построения модели линейной регрессии.

Построение модели

Набор данных был разделен на обучающую и тестовую выборки в размерах 0.8 и 0.2 от общего количества данных соответственно. С помощью класса LinearRegression была создана и обучена модель линейной регрессии (рисунок 2).

```
# Подготовка данных
X = data.drop(columns=['mpg'])
y = data['mpg']

# Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Создание и обучение модели
model = LinearRegression()
model.fit(X_train, y_train)
```

Рисунок 2. Создание и обучение модели.

Figure 2. Creating and training a model.

Оценка качества модели

Обучив модель и посчитав относительную ошибку RMSE, которая равна 13.96%, можно сделать вывод, что модель адекватна и может быть использована для предсказания расхода топлива автомобиля (рисунок 3) [6].

```
# Оценка качества модели
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"RMSE: {rmse}")
print(f"MAE: {mae}")
print(f"R^2: {r2}")

# Анализ диапазона значений mpg
min_mpg = data['mpg'].min()
max_mpg = data['mpg'].max()
mean_mpg = data['mpg'].mean()

print(f"Минимальное значение mpg: {min_mpg}")
print(f"Максимальное значение mpg: {max_mpg}")
print(f"Среднее значение mpg: {mean_mpg}")

# Относительная ошибка RMSE
relative_rmse = rmse / mean_mpg
print(f"Относительная RMSE: {relative_rmse:.2%}")

RMSE: 3.2727457003009515
MAE: 2.4197802491974496
R^2: 0.7901500386760351
Минимальное значение mpg: 9.0
Максимальное значение mpg: 46.6
Среднее значение mpg: 23.445918367346938
Относительная RMSE: 13.96%
```

Рисунок 3. Расчет оценки качества модели.

Figure 3. Calculation of the model quality assessment.

Тестовые предсказания

После обучения и оценки модели необходимо провести тестовые предсказания и сравнить их с реальными с помощью визуализации метрик на графике (рисунок 4).

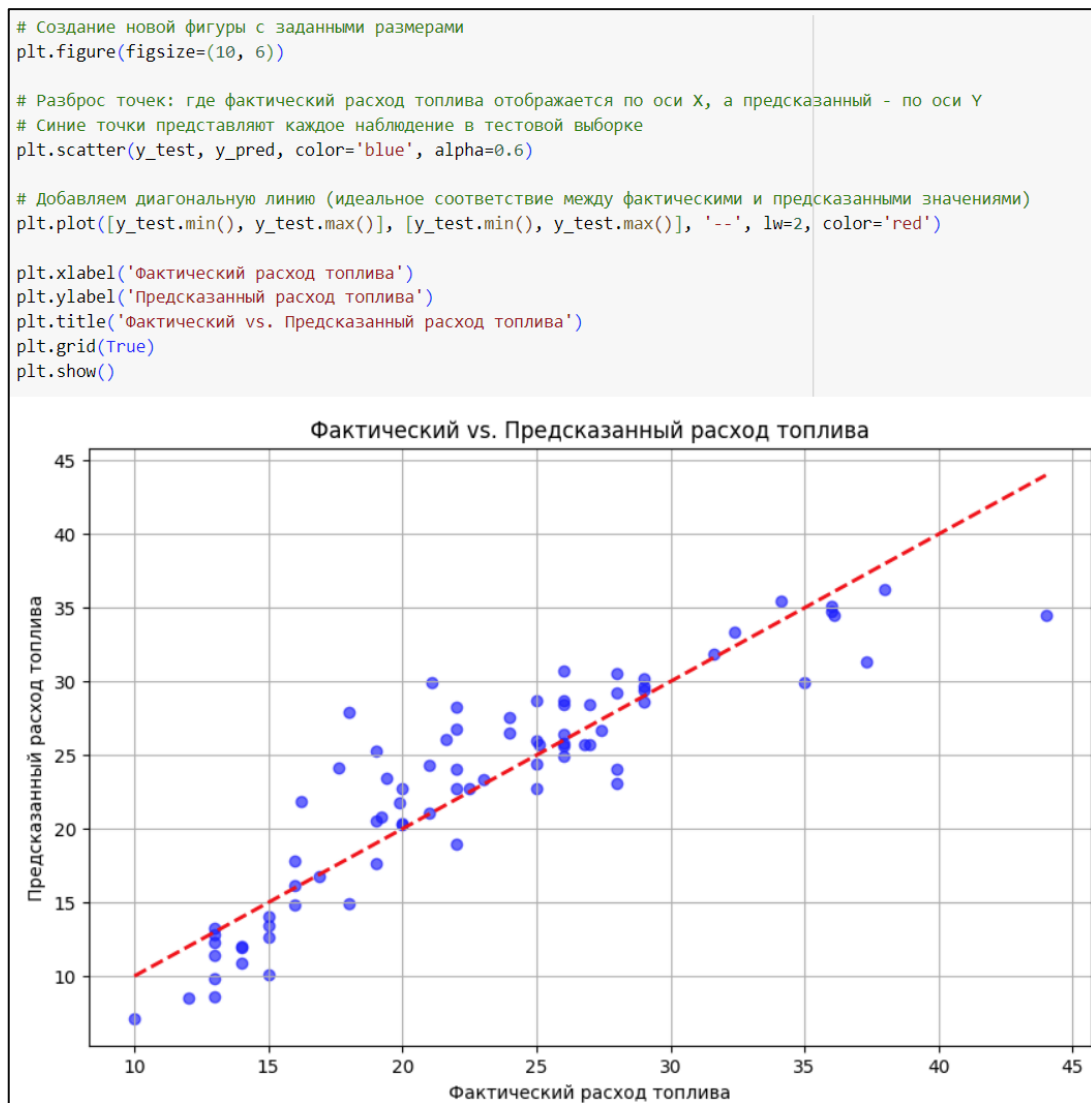


Рисунок 4. Тестирование модели.

Figure 4. Testing the model.

ЗАКЛЮЧЕНИЕ

Результаты показали, что метод линейной регрессии может быть эффективно использован для прогнозирования расхода топлива автомобиля. Тем не менее, для улучшения точности модели возможно использование более сложных методов машинного обучения, таких как деревья решений, случайные леса или нейронные сети [7].

В данной статье был рассмотрен метод линейной регрессии для прогнозирования расхода топлива автомобиля на основе различных характеристик. Проведенный анализ данных и визуализация результатов показали, что данный метод может быть полезен для решения задач в автомобильной индустрии.

СПИСОК ЛИТЕРАТУРЫ

- [1] Кун М., Джонсон К. Предиктивное моделирование на практике. Springer; 2019. 640.
- [2] Лутц М. Изучаем Python. М.: Эксмо; 2019. 1200.
- [3] Педрегоса Ф., Варуко Г., Scikit-learn: машинное обучение на Python. Журнал исследований машинного обучения. 2011; 12: 2825-2830.
- [4] Негодин В.А. Машинное обучение в языке программирования Python. Форум молодых ученых. 2019; 8(36): 201-203.
- [5] Зулунов Р.М., Солиев Б.Н. Использование Python для искусственного интеллекта и машинного обучения. Al-Farg' oniy avlodlari. 2023; 1(3): 18-24.
- [6] Murphy K.P. Machine Learning: A Probabilistic Perspective. MIT Press; 2012. 1067.
- [7] Гудфеллоу И., Бенгио Й., Курвилл А. Глубокое обучение. М.: MIT Press; 2016. 775.

REFERENCES

- [1] Kun M., Dzhonson K. Prediktivnoe modelirovanie na praktike. Springer; 2019. 640. (in Russian)
- [2] Lutc M. Izuchaem Python. M.: Eksmo; 2019. 1200. (in Russian)
- [3] Pedregosa F., Varuko G., Scikit-learn: mashinnoe obuchenie na Python. Zhurnal issledovaniy mashinnogo obucheniya. 2011; 12: 2825-2830. (in Russian)
- [4] Negodin V.A. Mashinnoe obuchenie v yazyke programmirovaniya Python. Forum molodyh uchenyh. 2019; 8(36): 201-203. (in Russian)
- [5] Zulunov R.M., Soliev B.N. Ispol'zovanie Python dlya iskusstvennogo intellekta i mashinnogo obucheniya. Al-Farg' oniy avlodlari. 2023; 1(3): 18-24. (in Russian)
- [6] Murphy K.P. Machine Learning: A Probabilistic Perspective. MIT Press; 2012. 1067.
- [7] Gudfellou I., Bengio J., Kurvill A. Glubokoe obuchenie. M.: MIT Press; 2016. 775. (in Russian)

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Зуев Максим Алексеевич, студент, кафедра «БИТ», инженерно-экономический институт, направление «Прикладная информатика», ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

Maxim Zuev, student, Department of "BIT", Institute of Engineering and Economics, direction "Applied Informatics", National Research University «Moscow Power Engineering Institute», Moscow, Russia

Шibaев Владимир Михайлович, студент, кафедра «БИТ», инженерно-экономический институт, направление «Прикладная информатика», ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

Vladimir Shibaev, student, Department of "BIT", Institute of Engineering and Economics, direction "Applied Informatics", National Research University «Moscow Power Engineering Institute», Moscow, Russia

Баланев Кирилл Сергеевич, ассистент, ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

Kirill Balanев, assistant Professor, National Research University «Moscow Power Engineering Institute», Moscow, Russia

Статья поступила в редакцию 14.06.2024; одобрена после рецензирования 26.06.2024; принята к публикации 27.06.2024.

The article was submitted 14.06.2024; approved after reviewing 26.06.2024; accepted for publication 27.06.2024.