

УДК 629.4.014

Применение машинного обучения для анализа технических характеристик и параметров эксплуатируемого подвижного состава на высокоскоростных железнодорожных магистралях

А. Н. Сычугов

Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург, Московский пр., 9

Для цитирования: Сычугов А. Н. Применение машинного обучения для анализа технических характеристик и параметров эксплуатируемого подвижного состава на высокоскоростных железнодорожных магистралях // Бюллетень результатов научных исследований. — 2023. — Вып. 2. — С. 171–180. DOI: 10.20295/2223-9987-2023-2-171-180

Аннотация

Цель: Машинное обучение — это одна из самых популярных технологий в настоящее время, которая находит применение в различных отраслях. Одной из таких отраслей является железнодорожный транспорт, где машинное обучение может существенно улучшить процессы управления локомотивами и электропоездами. Статья посвящена исследованию возможности применения машинного обучения для анализа технических характеристик и параметров подвижного состава на примере электропоездов, эксплуатирующихся на высокоскоростных железнодорожных магистралях. В статье будут рассмотрены механизмы машинного обучения для анализа данных, а также предложены практические рекомендации по использованию машинного обучения для анализа данных на железнодорожном транспорте. **Методы:** Метод k-средних (k-means) — один из наиболее популярных алгоритмов машинного обучения для кластеризации, который позволяет разбить набор данных на k схожих групп или кластеров. Алгоритм основан на поиске центроидов (средних значений) каждого кластера и присваивании объектов к тому кластеру, который имеет наиболее близкий центроид. В зависимости от выбора начальных значений центроидов и параметров алгоритма, разбиение данных на кластеры может быть разным. Алгоритм k-средних является относительно простым в реализации и эффективным алгоритмом кластеризации, который может быть использован в качестве первоначальной оценки для более сложных алгоритмов кластеризации; метод линейной регрессии — это статистический алгоритм, используемый для определения связи между двумя непрерывными переменными. Алгоритм использует линейную функцию, которая описывает связь между зависимой переменной и одной или несколькими независимыми переменными. **Результаты:** Алгоритм обработки и первичного анализа характеристик и параметров высокоскоростного подвижного состава, результаты кластеризации параметров, методика прогнозирования технических характеристик перспективного высокоскоростного подвижного состава. **Практическая значимость:** Показаны методы применения машинного обучения для анализа технических характеристик и параметров эксплуатируемого подвижного состава на высокоскоростных железнодорожных магистралях. Методы позволят обосновать и спрогнозировать необходимые характеристики при разработке технических требований для высокоскоростного транспорта в России. С помощью применения машинного обучения можно повысить точность расчета технико-экономических моделей высокоскоростных магистралей, а в дальнейшем сократить затраты на эксплуатацию подвижного состава.

Ключевые слова: Машинное обучение, подвижной состав, прогнозирование технических характеристик транспорта, анализ данных, высокоскоростной транспорт.

Машинное обучение — это одна из самых популярных технологий в настоящее время, которая находит применение в различных отраслях. Одной из таких отраслей является железнодорожный транспорт, где машинное обучение может существенно улучшить процессы управления локомотивами, электровозами и электропоездами.

Одно из основных применений машинного обучения в железнодорожном транспорте — это оптимизация процессов технического обслуживания и ремонта оборудования. С помощью алгоритмов машинного обучения можно предсказывать возможные отказы и сбои в работе оборудования, а также оптимизировать процессы его технического обслуживания. Инженеры General Electric используют машинное обучение для анализа данных, получаемых от локомотивов, что позволяет предсказывать возможные отказы оборудования [1]. Еще одно применение машинного обучения на железнодорожном транспорте — это управление движением поездов. С помощью алгоритмов машинного обучения можно оптимизировать график движения поездов, учитывая множество факторов, таких как погода, количество поездов, проведение работ на пути и т. д. Это позволяет сократить время в пути и повысить эффективность работы железнодорожного транспорта. Компания Deutsche Bahn использует машинное обучение для оптимизации графика движения поездов [2]. Кроме того, машинное обучение может быть использовано для анализа данных о состоянии инфраструктуры железнодорожного транспорта, такой как состояние рельсов, мостов и тоннелей. Это позволяет оперативно выявлять возможные проблемы и предотвращать аварии. BNSF Railway использует машинное обучение для анализа данных о состоянии своей инфраструктуры [3]. Это далеко не все примеры применения технологий machine learning и artificial intelligence на железнодорожном транспорте. Машинное обучение имеет большой потенциал для применения на железнодорожном транспорте. Рассмотрим пример применения базовых алгоритмов машинного обучения для анализа характеристик и параметров высокоскоростного подвижного состава.

С целью определения технических характеристик перспективного высокоскоростного подвижного состава выполним анализ параметров эксплуатируемого подвижного состава на высокоскоростных железнодорожных магистралях Европы и Азии. В качестве исходной базы данных используем данные Международного союза железных дорог [4].

Максимальная скорость на маршрутах ВСМ сегодня колеблется между 240–350 км/ч на большинстве основных линий и 200–250 км/ч на модернизированных линиях. Лидером по этому показателю являются поезда серий CRH в Китае, скорость которых может достигать 350 км/ч. Максимальная скорость для новых построенных магистралей должна составлять не менее 300–360 км/ч, с возможностью движения до 400 км/ч. Максимальная скорость должна определяться по коммерческим факторам (время в пути между городами), оценке стоимости

проекта (чрезвычайно высокая скорость может оказаться экономически невыгодной) и стать причиной больших технических сложностей.

В настоящее время можно классифицировать ВСПС на рынке по трем типам:

– оптимальная высокая скорость (более 300 км/ч): работает в основном на выделенных высокоскоростных линиях.

– высокая скорость (240–300 км/ч): самый распространенный тип высокоскоростного поезда, работающего в мире сегодня, и работает в основном на выделенной линии;

– высокая скорость для обычных линий (200–250 км/ч): работает как на выделенной высокоскоростной линии, так и на модернизированной традиционной линии [5].

Необходимость в перспективном высокоскоростном подвижном составе была обусловлена новыми технологиями, развитием систем электрического привода с преобразователями энергии, выполненными на IGBT, которые улучшили энергоэффективность и снизили эксплуатационные расходы. Самый распространенный в настоящее время тип — двигатели переменного тока.

В последние годы активно развиваются технологии по применению синхронных двигателей с постоянными магнитами, которые имеют свои преимущества и недостатки с точки зрения веса, эффективности и управляемости. Линейные двигатели могут стать перспективой в качестве силовой установки, но не будут широко использоваться в основном из соображений стоимости и совместимости с существующей инфраструктурой.

Рекуперативное торможение необходимо для снижения потребления энергии и может даже использоваться вместо механических тормозов для остановки поезда, чтобы уменьшить расходы на техническое обслуживание, однако механические тормоза все равно будут необходимы в качестве резервной системы в случае чрезвычайной ситуации.

Использование технологий для получения оптимального распределения силы тяги и тормозных усилий в поезде может быть необходимо для улучшения сцепления колесо-рельс. Сила сцепления, конечно, почти пропорциональна нагрузке на ось — показателю, который необходимо уменьшать с ростом скорости, для уменьшения динамического воздействия на путь.

При разработке технических требований для высокоскоростного транспорта компания (оператор) опирается на действующие собственные или зарубежные требования. Однако на стадии обоснования инвестиций в проект достаточными и необходимыми данными являются:

- тяговые характеристики;
- максимальная эксплуатационная скорость;
- пассажироместимость;
- исполнение (сосредоточенная или распределенная тяга);
- род тока;

- нагрузка на ось;
- длина поезда.

На основании этого ряда параметров выполняется оценка достижимости целевых параметров проекта, таких как: время в пути, выполнение графика и заданного объема пассажирских перевозок. На этапе обоснования инвестиций или подготовки технических требований машинное обучение можно применить для прогнозирования указанных выше технических характеристик на основании имеющихся данных о подвижном составе. В нашем случае данные представлены в виде pdf-файла, который содержит таблицу с перечислением всех серий ВСПС, которые находились или находятся в эксплуатации, а также операторов подвижного состава и технических характеристик электропоездов.

Работу с данными продолжим с помощью блокнотов Colab [6]. Установим и подключим необходимые библиотеки, предварительно разместив наш файл с данными в облачном пространстве Google:

```
#https://github.com/astanin/python-tabulate
pip install tabula-py
import tabula
#https://pandas.pydata.org/
import pandas as pd
from google.colab import drive
drive.mount('/content/gdrive')
```

Проанализируем результат преобразования данных. Для этого с помощью метода `head()` напечатаем первые 5 строк таблицы. Результат выполнения кода представлен на рис. 1.

```
pages = 'all'
df = tabula.read_pdf('/content/gdrive/MyDrive/high_speed.pdf', pages=pages)
result_df = pd.concat(df)
result_df.head()
```

| order | Country\r/Region | Photograph | Owner or Operator | Suppliers | Class | Trainset Formula\rM: Motor Car,\rT: Trailer Car,\rL: Locomotive,\rMB: Motor Bogie,\rTB: Trailer Bogie | Features\rC: Concentrated,\rA: Articulated,\rT: Tilting,\rD: Double Decker | |
|-------|------------------|------------|-------------------|-----------|---------|--|--|------|
| 0 | 1.0 | Austria | NaN | ÖBB | Siemens | "Railjet"\rSiemens Taurus\r(OBB 1216) +\rSieme... | 1L6T | C |
| 1 | 2.0 | Austria | NaN | WestBahn | Stadler | 4010 | 2M4T | NaN |
| 2 | 3.0 | France | NaN | SNCF | Alstom | TGV PSE | 2L8T | C, A |
| 3 | 4.0 | France | NaN | La Poste | Alstom | TGV Postal | NaN | C, A |
| 4 | 5.0 | France | NaN | SNCF | Alstom | TGV Atlantique | 2L10T | C, A |

Рис. 1. Результат первичной обработки информации

В результате предварительной обработки был получен dataframe. Информацию о названии столбцов и типах данных в этих столбцах получим с помощью следующих команд:

```
result_df.info () #Этот метод выводит информацию о DataFrame, включая тип и столбцы индекса, ненулевые значения и использование памяти.
```

```
result_df.columns.values.tolist () #Этот метод выводит информацию список имен столбцов dataframe.
```

Данные представлены следующими типами: float64, object. Часть данных в строках имеет тип данных NaN (Not a Value). Для последующей обработки данных выполним предварительную обработку данных, для этого сначала удалим столбцы, которые не содержат данных (например, столбец ['Photograph']), и для дальнейшего удобства работы с данными переименуем столбцы.

Выполним кластеризацию полученного массива данных. Для этого воспользуемся одним из наиболее широко применяемых алгоритмов машинного обучения — кластеризация K-средних. Программа написана с помощью библиотеки scikit-learn, где реализована улучшенная версия алгоритма k-means++ [7]. Алгоритм k-means++ выбирает начальные центроиды кластера, используя выборку на основе эмпирического распределения вероятностей вклада точек в общую инерцию. Этот метод ускоряет сходимость. Листинг представлен ниже:

```
#https://matplotlib.org/  
import matplotlib.pyplot as plt  
#https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html  
from sklearn.cluster import KMeans  
# https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html  
from sklearn.decomposition import PCA  
kmeans = KMeans(n_clusters=3, random_state=2023).fit(data)  
labels = kmeans.labels_  
data['cluster'] = labels
```

Визуализируем полученный результат (рис. 2) кластеризации и проанализируем полученные результаты.

На рис. 2 получены три кластера. Отметим, что кластеризация с другим числом кластеров не дала хорошего результата. Как отмечалось ранее, весь высокоскоростной подвижной состав условно классифицирован на три класса в зависимости от скорости. В результате машинного обучения получен очень близкий результат:

– подвижной состав со скоростью эксплуатации до 250 км/ч и удельной мощностью от 7 кВт/т до 15 кВт/т;

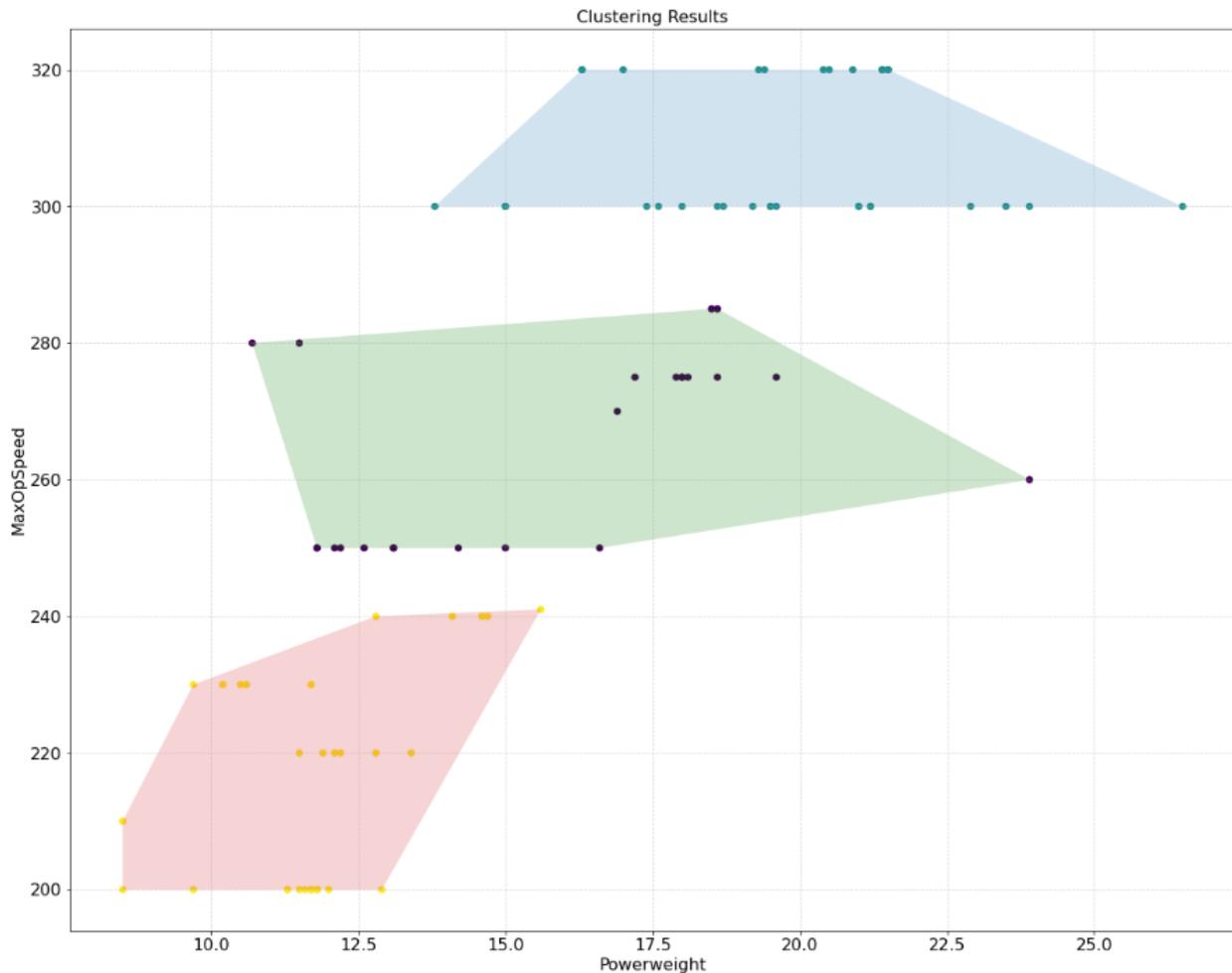


Рис. 2. Результат кластеризации данных

– подвижной состав со скоростью эксплуатации от 250 до 300 км/ч и удельной мощностью от 11 до 20 кВт/т;

– подвижной состав со скоростью свыше 300 км/ч и удельной мощностью от 15 до 25 кВт/т.

Например, мощность электропоезда «Сапсан» составляет 8 МВт, масса поезда версии В1 — 656 тонн, значит, показатель удельной мощности составляет 12,2 кВт/т. Поезд отнесен ко второму кластеру (область зеленого цвета на рис. 2), поскольку эксплуатируется со скоростью 250 км/ч. Согласно полученным результатам, увеличить скорость эксплуатации до 300 км/ч невозможно без увеличения тяговой мощности электрооборудования.

Аналогичным образом можно выполнить кластеризацию по другим признакам. Полученный результат позволяет судить о зависимости необходимой удельной и абсолютной мощности электропоезда от планируемой скорости для перспективных ВСМ. Уточнить показатель удельной мощности можно с помощью применения машинного обучения с учителем. Наиболее простым примером является модель линейной регрессии. Для изучаемого датафрейма входными данными

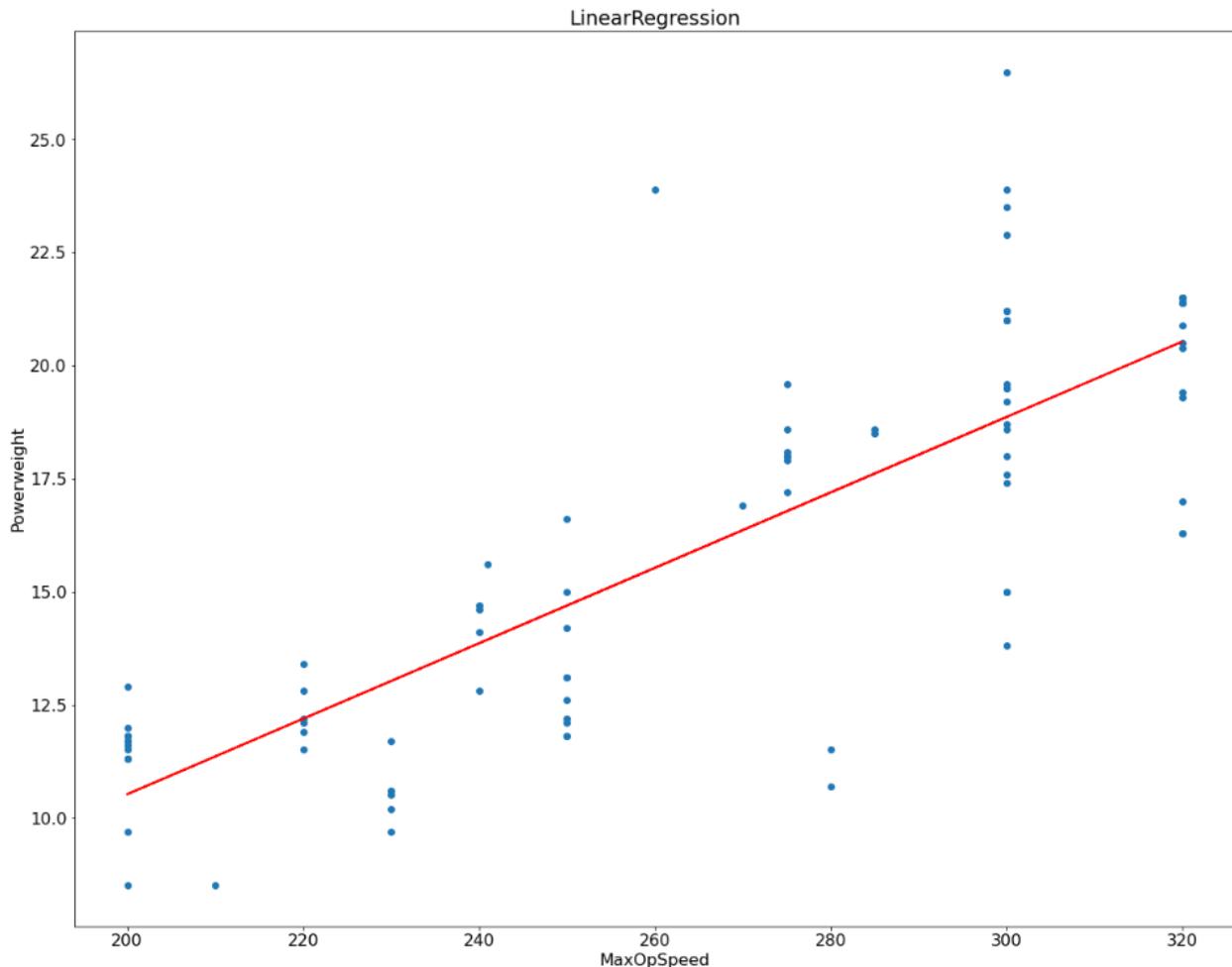


Рис. 3. Результат построения линейной регрессии

объявим скорость эксплуатации, а откликом является удельная мощность. Решение такой задачи позволяет «предсказать» значение удельной мощности, зная скорость эксплуатации. Точность модели оценим стандартной R^2 статистики [8]. Программная реализация представлена ниже:

```
#https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LinearRegression.html
```

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
X = np.array(result_df['Powerweight']).reshape((-1, 1))
y = result_df['MaxOpSpeed']
model = LinearRegression().fit(X, y)
```

Результат обучения представлен на рис. 3.

Значение метрики линейной регрессии $R^2 = 0,66$ означает, что 66 % дисперсии зависимой переменной являются объяснимым влиянием независимых

переменных, включенных в модель. Таким образом, это может быть интерпретировано как довольно сильная связь между переменными, но также указывает на наличие других факторов, которые не были включены в модель и влияют на зависимую переменную. К таким факторам следует отнести систему энергоснабжения линии ВСМ (постоянный/переменный ток), регион эксплуатации (равнинный/горный) и другие факторы.

В рамках выполненных исследований использовался очень небольшой массив открытых данных с ограниченным набором параметров. Тем не менее даже на таком небольшом объеме удалось выполнить кластеризацию технических характеристик, построить модель линейной регрессии. Располагая большими объемами данных об эксплуатационных и технических характеристиках подвижного состава, можно исключить многие человеческие ошибки, принимать взвешенные решения, основанные на данных и опыте компьютерных моделей. Такое применение машинного обучения позволяет снизить риски и оптимизировать работу железнодорожного транспорта.

Уже в ближайшей перспективе машинное обучение может быть применено для решения важнейших задач российских железных дорог. Выделим задачи, наиболее актуальные на сегодняшний день:

1. Повышение безопасности движения

Системы машинного обучения могут анализировать данные инфраструктуры, включая препятствия, нарушения целостности железнодорожного полотна и другие факторы, которые могут привести к сходу подвижного состава, и заблаговременно предупредить машиниста или автоматически остановить поезд.

2. Энергетический менеджмент/управление графиком движения поездов

Системы машинного обучения могут анализировать данные о поездах, включая их скорость, время и местоположение, а также другие факторы. На основе этих данных компьютер может создать оптимальный график движения поездов, что не только повышает эффективность перевозок, но и снижает затраты на топливо.

3. Организация ремонта по фактическому состоянию

Важной задачей является прогнозирование вероятности отказов локомотивов. Применение машинного обучения позволяет собрать и проанализировать огромное количество данных, включая данные об отказах, технические характеристики, история (журнал) о выполненных ремонтах и обслуживаниях, эксплуатационные данные и т. д. Анализ этих данных позволяет создать модель, способную определять по событию, какая деталь или узел может выйти из строя.

Библиографический список

1. Machine Learning // General Electric. — URL: <https://www.ge.com/research/technology-domains/artificial-intelligence/machine-learning> (дата обращения: 21.03/2023).

2. Artificial intelligence at DB // Deutsche Bahn. — URL: https://www.deutschebahn.com/en/artificial_intelligence-6935068 (дата обращения: 21.03.2023).
3. Artificial Intelligence // BNSF Railway URL: <https://www.bnsf.com/news-media/railtalk/innovation/artificial-intelligence.html> (дата обращения: 21.03/2023).
4. High-speed Data // INTERNATIONAL UNION OF RAILWAYS (UIC). — URL: <https://uic.org/passenger/highspeed/article/high-speed-data-and-atlas> (дата обращения: 21.03.2023).
5. Highspeed // INTERNATIONAL UNION OF RAILWAYS (UIC). — URL: <https://uic.org/passenger/highspeed/> (дата обращения: 21.03/2023).
6. Colab // Colab. — URL: <https://colab.research.google.com/> (дата обращения: 21.03/2023).
7. Sklearn.cluster.KMeans // Scikit-learn. — URL: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html> (дата обращения: 21.03.2023).
8. Sklearn.linear_model.LinearRegression // Scikit-learn URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html (дата обращения: 21.03.2023).

Дата поступления: 06.05.2023

Решение о публикации: 25.05.2023

Контактная информация:

СЫЧУГОВ Антон Николаевич — аспирант; sychugov@pgups.ru

Application of Machine Learning to Analyze the Technical Characteristics and Parameters of the Rolling Stock in Operation on High-Speed Railway Lines

A. N. Sychugov

Emperor Alexander I St. Petersburg State Transport University, 9, Moskovsky pr., Saint Petersburg, 190031, Russian Federation

For citation: Sychugov A. N. Application of Machine Learning to Analyze the Technical Characteristics and Parameters of the Rolling Stock in Operation on High-Speed Railway Lines. *Bulletin of scientific research results*, 2023, iss. 2, pp. 171–180. (In Russian) DOI: 10.20295/2223-9987-2023-2-171-180

Summary

Purpose: Machine learning is one of the most popular technologies at the present time, which finds application in various industries. One such industry is rail transport, where machine learning can significantly improve the management of locomotives and electric trains. The article is devoted to the study of the possibility of using machine learning to analyze the technical characteristics and parameters of the rolling stock using electric trains operating on high-speed railway lines as an example. The article discusses machine learning mechanisms for data analysis, as well as offers practical recommendations for using machine learning for data analysis on railway transport. **Methods:** The k-means method is one of the most popular machine learning algorithms for clustering, which allows to split a dataset into k similar groups or clusters. The algorithm is based on finding the centroids (mean values) of each cluster and assigning objects to the cluster that has

the closest centroid. Depending on the choice of the initial values of the centroids and the parameters of the algorithm, the partitioning of data into clusters can be different. In addition, k-means is a relatively easy-to-implement and efficient clustering algorithm that can be used as an initial estimate for more complex clustering algorithms; The linear regression method is a statistical algorithm used to determine the relationship between two continuous variables. The algorithm uses a linear function that describes the relationship between the dependent variable and one or more independent variables. **Results:** Algorithm for processing and primary analysis of the characteristics and parameters of high-speed rolling stock, the results of parameter clustering, a technique for predicting the technical characteristics of prospective high-speed rolling stock. **Practical significance:** The methods of application of machine learning for the analysis of technical characteristics and parameters of the operated rolling stock on high-speed railway lines are shown. The methods will allow to substantiate and predict the necessary characteristics when developing technical requirements for high-speed transport in Russia. Using machine learning, it is possible to improve the accuracy of calculating the technical and economic models of high-speed highways, and further reduce the cost of operating rolling stock.

Keywords: Machine learning, rolling stock, predicting the technical characteristics of transport, data analysis, high-speed transport.

References

1. Machine Learning. General Electric. Available at: <https://www.ge.com/research/technology-domains/artificial-intelligence/machine-learning> (accessed: March 21, 2023).
2. Artificial intelligence at DB. Deutsche Bahn. Available at: https://www.deutschebahn.com/en/artificial_intelligence-6935068 (accessed: March 21, 2023).
3. Artificial Intelligence. BNSF Railway. Available at: <https://www.bnsf.com/news-media/railtalk/innovation/artificial-intelligence.html> (accessed: March 21, 2023).
4. High-speed Data. INTERNATIONAL UNION OF RAILWAYS (UIC). Available at: <https://uic.org/passenger/highspeed/article/high-speed-data-and-atlas> (accessed: March 21, 2023).
5. Highspeed. INTERNATIONAL UNION OF RAILWAYS (UIC). Available at: <https://uic.org/passenger/highspeed/> (accessed: March 21, 2023).
6. Colab. Colab. Available at: <https://colab.research.google.com/> (accessed: March 21, 2023).
7. Sklearn.cluster.KMeans. Scikit-learn. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html> (accessed: March 21, 2023).
8. Sklearn.linear_model.LinearRegression. Scikit-learn URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html (accessed: March 21, 2023).

Received: May 06, 2023

Accepted: May 25, 2023

Author's information:

Anton N. SYCHUGOV — Postgraduate Student; sychugov@pgups.ru