TRANSPORT TECHNOLOGIES

Vol. 3, No. 2, 2022

Artur Budzyński, Aleksander Sładkowski
Silesian University of Technology
8, Krasińskiego Str., Katowice, 40-019, Poland

https://doi.org/10.23939/tt2022.02.001

FORECASTING FUEL CONSUMPTION IN MEANS OF TRANSPORT WITH THE USE OF MACHINE LEARNING

Summary. *Transport is a key factor influencing greenhouse gas emissions. In relation to this, the issues and challenges facing the transport industry were presented. The issues of challenges for the transport industry related to the European Green Deal were discussed. It discussed how the transport system is critical for European companies and global supply chains. The issues related to the exposure of society to costs are presented: greenhouse gas emissions and pollution. The article deals with the issues of managing transport processes in an enterprise. It was decided to raise the topic of fuel consumption in means of transport. Based on a review of the scientific literature, 3 categories of features are indicated: the vehicle characteristics, the driver's characteristics, and the route's impact on fuel consumption. The study is based on actual data from the archives of the GPS vehicle monitoring system. Data was collected on 1890 routes operated between May 30, 2020, and May 31, 2021. The routes were performed by twenty-nine drivers and 8 vehicles. The vehicles are 40-ton road sets consisting of a tractor unit and a semi-trailer. The analysis of factors influencing fuel consumption is presented. The methodology for conducting feature engineering is described. The benefits of using the method of reducing fuel consumption are presented. The possibilities of using the methods of forecasting electricity and hydrogen consumption in various means of transport, including public transport, where indicated. The data is processed using the Pandas library. The models are compared according to the MAE success measure. The application of methods of working with large data sets is presented. The calculations are made with the help of the NumPy library. Data visualization is done with Matplotlib and Seaborn. Scikit-Learn models are used.*

Key words: *transport, transport management, machine learning, modeling, fuel consumption.*

1. INTRODUCTION

Reducing the consumption of carbon monoxide emissions is a major challenge for the transport industry. Climate change and environmental degradation pose a threat to Europe and the rest of the world. The European Green Deal Action Plan was designed to meet these challenges. Applying the plan will transform the EU into a modern, resource-efficient and competitive economy with a sustainable transport system. The economy is set to reach zero net greenhouse gas emissions in 2050, decoupling economic growth from resource use, which no one or no region will be left behind. Reducing net greenhouse gas emissions by at least 55 % by 2030 compared to the 1990-year level is a goal that transport policy is also part of. Achieving these goals requires the cooperation of organizations in many areas. The scientific area requires involvement in research in various disciplines. In addition to developing modern technologies, responsible management of the resources that current technology uses is essential. The management of transport processes affects the consumption of resources. Good management needs knowledge based on scientific evidence. Fuel consumption is one of the main costs of road transport. Actions to optimize fuel consumption are desirable. It is dictated by both ecological and economic factors. Artificial intelligence and machine learning solutions help solve problems in many industries. One of them is broadly understood transport. Scientists around the world continue their efforts to replace people in tiresome repetitive tasks with machines. Such activities are beneficial in the context of sustainable development and care for human health. Replacing people with machines allows reducing the amount of time spent by them in front of the monitor, often in an unhealthy position. This allows people to focus on tasks that require natural intelligence such as building relationships. For people involved in the management of transport processes, it is important to build a relationship with the customer or carrier.

2. RESEARCH STATEMENT

The problem is the number of features that a person, managing transport processes, must consider. Making forecasts with so many factors in mind is difficult. There are studies on forecasting fuel consumption. The problem is different due to the type of vehicle and the specifics of the work. This study addresses the issue of 40-tonne vehicles that drive several hundred kilometers each day. The problem is energy and fuel consumption. The impact of greenhouse gas emissions has a detrimental effect on the environment. On the other hand, the growing cost of fuels translates into the price of goods transported by road transport. The costs of the transport service translate into increased prices in stores. Considering every point of view, the problem of wear is worth the efforts of scientists from many fields. The issues concern research on the use of modern means of transport and the use of existing ones. The problem is the difficulty in finding bottlenecks in the operation of vehicles and drivers. The amount of data generated by vehicle monitoring systems are large and difficult to effectively analyze using standard statistical methods.

3. RELEVANCE OF THE STUDY

The study is important to improve the management of transport processes. The development of a method of selecting features for a machine-learning model will allow the model to be trained efficiently. This is important for management. Operators can save time and spend it on tasks that require natural intelligence. The importance of research is significant in the context of the efforts of scientists around the world struggling with climate change. Conducting this experiment based on facts about factors influencing fuel consumption has made a significant contribution to the development of research.

4. AIM AND THE TASKS OF THE STUDY

The goal is to train a model that will forecast fuel consumption by means of transport. To achieve this goal, it is essential to develop a scientific method for selecting features and conducting a feature engineering process. The article deals with research consisting of the following stages: gathering data, data preparation, data wrangling, data analysis, and training model. Gathering data is the first stage in the machine learning life cycle. The purpose of this step is to find and control all data problems. At this point, it is necessary to find the different data sources. It is one of the most important stages in the life cycle. The amount and quality of the collected data will decide the performance of the output. The more data, the more correct the forecast will be. After doing the above, a consistent set of data that will be used in the next stages of the machine learning life cycle are got. In practice, the quality of the final models depends much more on the quality of the prepared data than on the choice of the model itself and its optimization. After collecting the data, it is needed to be prepared for the next steps. At this stage, the data is put in the right place and prepared for use in machine learning. At this point, it first combines all the data and then randomizes their order. This step can be divided into two processes: data exploration and data pre-processing for analysis It serves to understand the nature of the data we need to work with. Data research allows for understanding the characteristics, format, and quality of the data. A better understanding of the data leads to better results. Thus, you find correlations, general trends, and outliers. Data processing is the process of cleaning and converting raw data directly from the GPS vehicle tracking system into a usable format. Only based on excellent quality data can an excellent quality analysis and forecast be made. Various problems can arise with the collected data, such as missing values that are not specified, values that are meaningless or undefined; duplicate data; invalid values that are outside the desired range; spelling mistakes, word permutation; noise, or contradiction of information. Data cleansing is a necessary step as the value of data depends not only on the amount of data but also on the quality of the information collected. There are many processing methods that can improve the quality of the data. Data Analysis as a stage in machine learning consists in analyzing data that has already been cleaned and prepared. The goals of this step are selecting the analytical methods of the algorithms; creating machine learning models; checking the results of models. Usually, for each problem, there are several algorithms that can be used. Often it is necessary to use the trial-and-error method. More often than choosing an algorithm is working with features for specific data.

5. ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

Scientists in [1] built models for fuel consumption prediction based on data from a smartphone and the OBD (On-Board Diagnostic) system in taxis. The following models were used to model the forecasts: carrier vector regression, backpropagation errors, carrier vector regression, random forests, and neural networks. The inputs were: average speed, average acceleration, average deceleration, percentage of acceleration time, and percentage of deceleration time. The best model was a random forest.

Artificial neural networks have been used to predict fuel consumption in vehicles operating in energy companies. The fleet of vehicles includes maintenance eliminating downtime, service, and supervision. The input data to the network consisted of the number of cylinders, displacement volume, number of valves, model, and weight of the vehicle [2].

Scientists from Turkey have built 3 models to predict instantaneous and overall fuel consumption. Several methods were selected for the study: multiple linear regression, support vector machine and artificial neural network. Fuel type, swept volume, frontal area, vehicle weight, distance, and average speed were used as input data. The support vector machine worked best [3].

In [4], scientists built predictive models based on large data sets. The input data included: car segment, make, model, year, gearbox type, engine capacity, driving time, and distance. The models selected for the experiment carrier vector regression and artificial neural networks turned out to be better. In [5] the prediction of diesel oil consumption in heavy goods vehicles was made.

The Jupyter Notebook [6] is an editor for programming code. The authors' goal is to make the tool available to multiple audiences. Jupyter is a notebook that allows users to experiment with code. It is available in open access. Pandas [7] is a data-processing programming library. NumPy [8] is the programming library for calculation analysis, and Scikit-Learn [9] is the machine learning programming library. Matplolib [10] and Seaborn [11] are visualization programming libraries. GitHub [12] is a repository for programming projects. The above methods are shown in Fig. 1. Each method was assigned the number of citations in the Scopus and ResearchGate database as of May 15, 2022. It was decided to also quote from Researchgate as not all articles are indexed in Scopus.

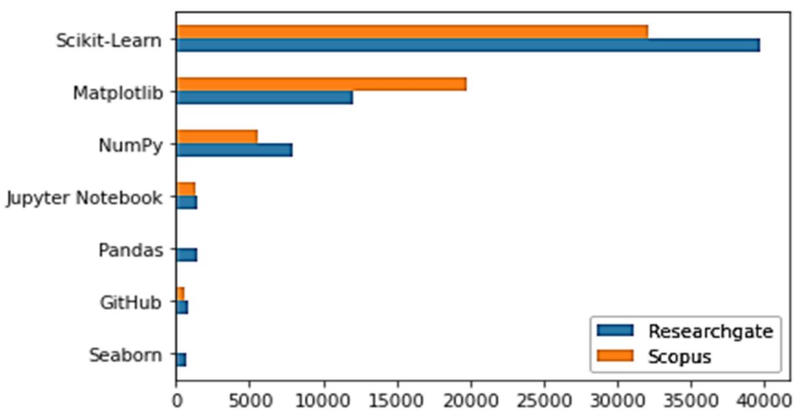

Fig. 1. The number of method citations in Scopus and Researchgate

6. PRESENTATION OF BASIC MATERIAL

The statistical method, including regression and correlation analysis, is used in the project. An experimental method was used in the project. The experiments consist of checking which model will make the lesser error. Data processing and model training was performed in the Jupyter Notebook editor and the Python programming language. The methods of using Python are described in [13] The import system, the Python code in one module accesses the code in another module through the import process. Programming libraries: Pandas [14] for data processing, NumPy [15] for calculation analysis, Scikit-Learn [16] for machine learning, Matplotlib, and Seaborn for visualization are used to complete the project. ELI5 is a library that allows checking what features were important for the model. It is compatible with many frameworks. ELI5 is a Python library that allows visualizing and checking which features have influenced the model.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1890 entries, 0 to 1889
Data columns (total 24 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   PLATE                  1890 non-null   object
 1   DRIVER                 1890 non-null   object
 2   START_DATE             1890 non-null   object
 3   START_PLACE            1890 non-null   object
 4   END_DATE               1890 non-null   object
 5   END_PLACE              1890 non-null   object
 6   DRIVING_TIME           1890 non-null   object
 7   STOP_TIME              1890 non-null   object
 8   CAN_DISTANCE           1890 non-null   float64
 9   GPS_DISTANCE           1890 non-null   int64
 10  USED_FUEL              1890 non-null   float64
 11  MEAN_USED_FUEL_FHKM    1890 non-null   float64
 12  MEAN_SPEED             1890 non-null   int64
 13  MAX_SPEED              1890 non-null   int64
 14  START_MILEAGE          1890 non-null   int64
 15  END_MILEAGE            1890 non-null   int64
 16  CARGO_WEIGHT           1890 non-null   float64
 17  DRIVER_YEAR_BORN       1890 non-null   float64
 18  QTY_DRIVERS            1890 non-null   int64
 19  TRUCK_YEAR_PRODUCTION  1890 non-null   float64
 20  KW                     1890 non-null   float64
 21  DM3                    1890 non-null   float64
 22  TRUCK_WEIGHT           1890 non-null   float64
 23  MAX_TRUCK_TOTALWEIGHT  1890 non-null   float64
dtypes: float64(10), int64(6), object(8)
memory usage: 354.5+ KB
```

Fig. 2. Information about the dataset

The data for analysis were collected from the archive system of the GPS monitoring system of the road transport company. The date range is from May 30, 2020, to May 31, 2021. Twenty-nine drivers took part in the survey. Data saved in xls format was converted to csv format and loaded into Jupyter Notebook using the read function of the Pandas library. Fig. 2 shows information about a dataset. The data is presented in a DataFrame using the info function. The data marked with the black rectangle shows the basic information. The dataset consists of 1890 rows labeled with sequential numbers from 0 to 1889 in 24 columns. The data marked in the red frame means the next number. It is important in the next stage of data processing. Numbering starts from 0. The data in the green frame are the names of the columns. PLATE stands for a unique registration number for each vehicle. DRIVER stands for a unique name and surname for each driver. START_DATE is the start date of the drive consisting of a year, month, day, hour, and minute. START_PLACE is the address of the place of loading. END_DATE is the end date of the drive consisting of a year, month, day, hour, and minute. END_PLACE is the address of the place of unloading. DRIVING_TIME means how much driving time was between the start and end of the route. STOP_TIME means how much time the stop was between the start and the end of the route. CAN_DISTANCE is the distance traveled between the start and end of the route according to the tachograph. USED_FUEL is the amount of fuel consumed on the route expressed in cubic meters. This is the target variable. MEAN_USED_FUEL_FHKM is the average fuel consumption per kilometer. MEAN_SPEED is the average speed of the route. MAX_SPEED is the maximum speed on the route. START_MILEAGE marks the mileage of the vehicle at the start of the route. END_MILEAGE means the mileage of the vehicle at the end of the route. CARGO_WEIGHT is the weight of the goods. DRIVER_YEAR_BORN is the driver's year of birth. QTY_DRIVERS is the number of drivers in the crew. TRUCK_YEAR_PRODUCTION is the year of manufacture of the vehicle. KW is the vehicle power expressed in kilowatts. DM3 is the engine capacity. TRUCK_WEIGHT is the curb weight of the vehicle. MAX_TRUCK_TOTALWEIGHT is the gross vehicle weight. The data marked in the blue frame means the amount of non-null data. In this case, all data is complete. The data marked in the yellow frame indicates the type of the variable. The data in gray are summarized. The counted number of variable types in columns and the memory used appear here.

Table 1 shows the statistical analysis of the numerical features. The following features mean: $\bar{x}$ – mean; σ – standard deviation; V – coefficient of variation; q2 – median; min. – minimum; max. – maximum; q1 – 1 quartile; q3 – 3 quartile; q – quartile difference; V_q – positional coefficient of variation. Based on scientific knowledge, the key data for modeling on distance and fuel consumption are assessed as possible and correct. Vehicle speed data is unrealistic and there are errors in reading the device. No negative speed is possible, neither negative nor exceeding 600 [km/h]. This information is important for modeling. A method must be found to solve this problem by, for example, discarding extreme values. Data on drivers and vehicles vary. This will allow the modeling to examine the influence of individual features and their configuration.

Table 1

Data set statistics

Feature	$\bar{x}$	σ	V	q2	Min.	Max.	q1	q3	q	V_q
Distance tachograph [km]	466.82	200.03	42.85	493.17	0	981.7	331.06	633.34	151.14	30.64
Distance GPS [km]	448.74	196.64	43.82	472	1	952	313	612	149.5	31.67
Fuel [dm^3]	139.26	60.95	43.76	145.37	1.5	355.22	98.21	182.14	41.96	28.87
Fuel [dm^3/100km]	30.65	6.60	21.54	30.24	0	121.59	26.5	33.79	3.64	12.05
Mean speed [km/h]	61.81	28.66	46.37	70	-297	649	59	76	8.5	12.14
Max speed [km/h]	94.09	6.26	6.65	93	0	114	92	96	2	2.15
Start mileage [1000 km]	446.12	301.5	67.58	351.76	61.16	1111.64	199.34	639.77	220.21	62.60
End mileage [1000. km]	446.6	301.48	67.51	351.98	61.71	1111.65	200.03	640.31	220.14	62.54
Cargo weight [t]	14.00	0.42	3.03	14	0	22	14	14	0	0
Driver year born	1974.93	2.71	0.14	1975	1964	1982	1974	1975	0.5	0.03
Quantity drivers	1.01	0.08	8.09	1	1	2	1	1	0	0
Truck year production	2016.96	1.46	0.07	2017	2014	2019	2016	2018	1	0.05
Power [kW]	370.04	14.72	3.98	360	360	405	360	375	7.2	2.08
Truck weight [t]	2168.32	3354.33	154.7	8.46	7.83	8310	8.45	2339.39	1165.47	13751.82
Max truck total weight [t]	20.47	0.75	3.68	21	19	21	20.36	21	0.32	1.53

Fig. 3 shows the correlation matrix of numerical features. Matrix made with the use of heatmaps from the Seaborn library. The number of kilometers from GPS does not correspond to the number of kilometers according to the tachograph. The data from the tachograph is more correlated with the average fuel consumption, which may indicate the inaccuracy of the GPS system. The longer the route is the lower the fuel consumption per km. This is due to more fuel consumption over shorter distances when the engine is underheated. Also on long journeys, expressways and highways are used more often, on which traffic is smoother, which reduces fuel consumption. The routes with higher average and maximum speeds have lower fuel consumption. Vehicles with higher mileage have higher fuel consumption. Vehicles with higher mileage are used for shorter journeys. Younger drivers use more fuel. Double casts are more often used on longer routs. Older cars have lower fuel consumption on average. Vehicles with more capacity have lower fuel consumption. Younger vehicles have a larger engine size. Vehicles with more power have a higher mileage. Vehicles with more power have lower fuel consumption. Younger vehicles have a higher total payload.

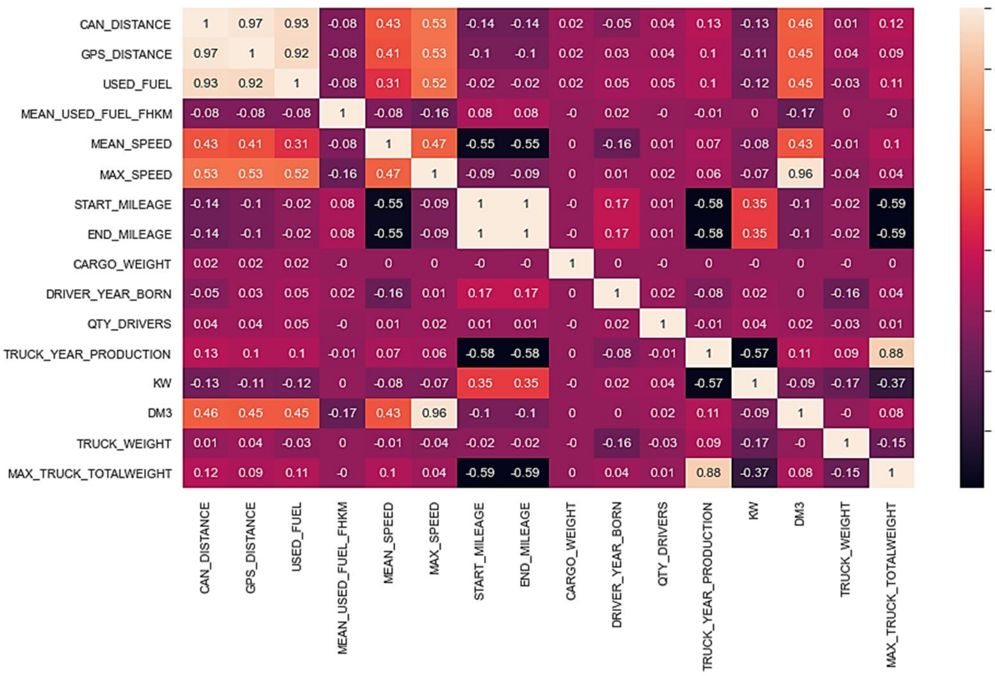

Fig. 3. Matrix of correlation of numerical features

Transformed data were entered into the models. The models selected for the experiment are developed based on various statistical methods. The selected models are used in many scientific works. Fig. 4 shows a comparison of the 4 models according to mean absolute error (MAE). In the experiment, the model with the smallest error was ExtraTreesRegressor with MAE = 1.43. The RandomFores Regressor had an MAE = 1.48 and the GradientBoostingRegressor had an MAE = 1.68. DecisionTreeRegressor had an MAE = 2.02, LinearRegression had an MAE = 2.09. The worst was MLPRegressor with an MAE = 8.32.

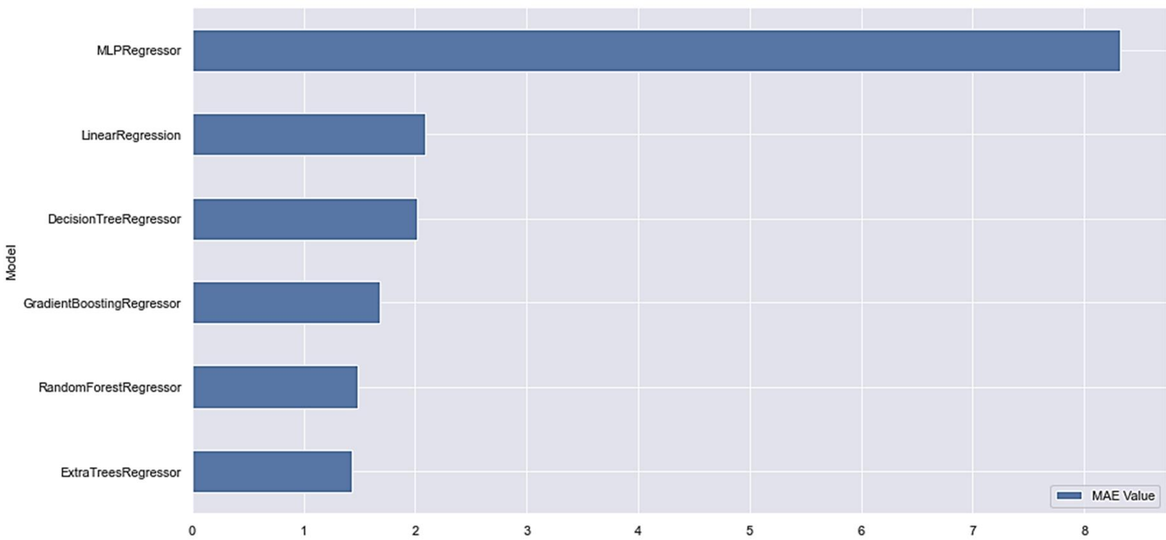

Fig. 4. Models comparison

Fig. 5 shows the main key features for the top ExtraTreesRegressor model. The model predicts fuel consumption in [dm^3/100km]. The distance-related characteristics are nevertheless listed in the most important ones. This is since fuel consumption on short distances is greater than on long distances. Median Fuel for END_PLACE is a feature that denotes the median fuel consumption for vehicles reaching each specific destination. Mean Fuel for END_PLACE is a feature that represents the average fuel consumption for vehicles reaching each specific destination. Median Fuel for START_PLACE is a feature that denotes the median fuel consumption for vehicles leaving each specific location. Mean Fuel for START_PLACE is a feature that represents the average fuel consumption for vehicles leaving each specific location. Median Fuel for DRIVER is the median fuel consumption for each driver. Mean Fuel for DRIVER is the average fuel consumption for each driver. Mean Fuel for Plate is the average fuel consumption for each vehicle. START_DATE_WEEKDAY is the day of the week the route starts. END_DATE_WEEKDAY is the day of the week end of the route. START_DATE_HOUR is the start time of the route. END_DATE_HOUR is the end time of the route. END_DATE_DAY is the end date of the route. START_DATE_DAY is the day the route started. Median Fuel for Plate is the median fuel consumption for each vehicle. START_DATE_WEEK_OF_YEAR is the week of the route start. END_DATE_DAY_OF_YEAR is the day of the year in which the route ends END_DATE_WEEK_OF_YEAR is the week of the route end year END_DATE_MONTH is the month of the end of the route. START_DATE_DAY_OF_YEAR is the day of the year the route starts. START_DATE_MONTH is the month of the start of the route. START_DATE_YEAR is the year in which the route started. END_DATE_YEAR is the end year of the route.

Weight	Feature
0.4800 ± 0.5656	CAN_DISTANCE
0.3029 ± 0.5057	GPS_DISTANCE
0.0586 ± 0.2343	DRIVING_TIME
0.0333 ± 0.0291	Median Fuel for END_PLACE
0.0305 ± 0.1565	STOP_TIME
0.0303 ± 0.0304	Mean Fuel for END_PLACE
0.0225 ± 0.0254	Median Fuel for START_PLACE
0.0191 ± 0.0230	Mean Fuel for START_PLACE
0.0030 ± 0.0103	Median Fuel for DRIVER
0.0028 ± 0.0117	Mean Fuel for DRIVER
0.0023 ± 0.0201	Mean Fuel for Plate
0.0022 ± 0.0088	MEAN_SPEED
0.0015 ± 0.0111	END_MILEAGE
0.0011 ± 0.0075	START_MILEAGE
0.0010 ± 0.0051	DRIVER_YEAR_BORN
0.0010 ± 0.0017	START_DATE_WEEKDAY
0.0009 ± 0.0017	END_DATE_WEEKDAY
0.0007 ± 0.0038	START_DATE_HOUR
0.0006 ± 0.0009	MAX_SPEED
0.0006 ± 0.0008	END_DATE_HOUR
0.0005 ± 0.0007	END_DATE_DAY
0.0005 ± 0.0007	START_DATE_DAY
0.0004 ± 0.0008	TRUCK_WEIGHT
0.0004 ± 0.0009	Median Fuel for Plate
0.0004 ± 0.0005	START_DATE_WEEK_OF_YEAR
0.0004 ± 0.0006	END_DATE_DAY_OF_YEAR
0.0003 ± 0.0005	END_DATE_WEEK_OF_YEAR
0.0003 ± 0.0005	END_DATE_MONTH
0.0003 ± 0.0005	START_DATE_DAY_OF_YEAR
0.0003 ± 0.0005	START_DATE_MONTH
0.0003 ± 0.0006	KW
0.0003 ± 0.0014	QTY_DRIVERS
0.0003 ± 0.0005	TRUCK_YEAR_PRODUCTION
0.0002 ± 0.0004	MAX_TRUCK_TOTALWEIGHT
0.0001 ± 0.0003	START_DATE_YEAR
0.0001 ± 0.0004	END_DATE_YEAR

Fig. 5. Top model features

7. CONCLUSIONS AND FUTURE RESEARCH PERSPECTIVES

Machine learning can be successfully used to forecast fuel consumption in means of transport. In the experiment, the model with the smallest error was ExtraTreesRegressor with MAE = 1.43. Distance has the greatest impact on fuel consumption. The GPS system does not show accurate data. Fuel consumption may depend on the day of the week and the related changes in vehicle traffic. Combustion may be month dependent and the effect of different temperatures on combustion. Younger drivers use more fuel. Fuel consumption depends on the vehicle – the year of production, engine capacity and power. The solution is beneficial to be introduced by entrepreneurs because, in addition to the positive impact on the environment, they can save fuel. The proposed method is one of the solutions that will help the transport industry face the challenges related to the green deal. It can be used to forecast hydrogen or current use.

In future studies, it is proposed to make models for various means of transport and various fuels using this method. It can be particularly applicable to the management of a fleet of public transport vehicles.

References

1. Yao, Y., Zhao, X., Liu, C., Rong, J., Zhang, Y., Dong, Z., & Su, Y. (2020). Vehicle fuel consumption prediction method based on driving behavior data collected from smartphones. *Journal of Advanced Transportation*, *2020*. 1–11. doi: 10.1155/2020/9263605 (in English).

2. Zargarnezhad, S., Dashti, R., & Ahmadi, R. (2019). Predicting vehicle fuel consumption in energy distribution companies using ANNs. *Transportation Research Part D: Transport and Environment*, *74*, 174–188. doi: 10.1016/j.trd.2019.07.020 (in English).

3. Çapraz, A. G., Özel, P., Şevkli, M., & Beyca, Ö. F. (2016). Fuel consumption models applied to automobiles using real-time data: A comparison of statistical models. *Procedia Computer Science*, *83*, 774–781. doi: 10.1016/j.procs.2016.04.166 (in English).

4. Moradi, E., & Miranda-Moreno, L. (2020). Vehicular fuel consumption estimation using real-world measures through cascaded machine learning modeling. *Transportation Research Part D: Transport and Environment*, *88*, 102576. doi: 10.1016/j.trd.2020.102576 (in English).

5. Budzyński, A., & Sładkowski A. (2021). The use of machine learning to predict diesel fuel consumption in road vehicles. *19th European Transport Congress of the EPTS Foundation e.V. European Green Deal Challenges and Solutions for Mobility and Logistics in Cities,* pp. 207–221 (in English).

6. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., & et al. (2016). *Jupyter Notebooks-a publishing format for reproducible computational workflows, 2016*, 87–90 (in English).

7. McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, *445*(1), pp. 51–56 (in English).

8. Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in science & engineering*, *13*(2), 22–30. doi: 10.1109/MCSE.2011.37 (in English).

9. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & et al. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, *12*, 2825–2830 (in English).

10. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, *9*(03), 90–95. doi: 10.1109/MCSE.2007.55 (in English).

11. Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. doi: 10.21105/joss.03021 (in English).

12. Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 1277–1286. doi: 10.1145/2145204.2145396 (in English).

13. Python 3. Retrieved from: https://docs.python.org/3lastaccessed2022/10/03 (in English).

14. Pandas. Retrieved from: https://pandas.pydata.org/docs/last accessed2022/10/03 (in English).

15. NumPy. Retrieved from: https://numpy.org/doc/stable/last accessed2022/10/03 (in English).

16. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., & et al. (2013). API design for machine learning software: experiences from the scikit-learn project. *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases,* pp. 1–15. doi: 10.48550/arXiv.1309.0238 (in English).

Received 12.10.2022; Accepted in revised form 20.10.2022.

ПРОГНОЗУВАННЯ СПОЖИВАННЯ ПАЛИВА РІЗНИМИ ВИДАМИ ТРАНСПОРТУ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Анотація. *Транспорт є ключовим чинником, який впливає на викиди парникових газів. У зв'язку з цим, наведено проблеми та виклики, з якими зустрічається транспортна галузь. Розглянуто питання транспортної галузі, пов'язані з Європейською зеленою угодою. Обговорено, наскільки транспортна система є важливою для європейських компаній та глобальних ланцюгів постачання. Проаналізовано також питання, які мають вплив на суспільство з точки зору витрат коштів, зокрема викиди парникових газів та забруднення довкілля. У статті висвітлено матеріали управління транспортними процесами на підприємстві. Прийнято рішення дослідити витрати палива видами транспорту. На основі огляду літературних джерел, визначено 3 категорії характеристик: характеристики автомобілів, водіїв, а також вплив маршруту на витрати палива. Дослідження виконано на основі даних архівів GPS* *системи моніторингу автомобілів. Вони зібрані на 1890 маршрутах, які здійснювали рух між 30 травня 2020 року та 31 травня 2021 року. На маршрутах працювали 29 водіїв та 8 транспортних засобів. Транспортні засоби – це 40-тонні тягачі з напівпричепами. Наведено аналіз чинників, які впливають на споживання палива. Описано методику отриманих інженерних функцій. Описано переваги методу зменшення споживання палива. Вказано на можливості використання методів прогнозування витрати енергії та водню на різних видах транспорту, включно з громадським транспортом. Дані опрацьовано з використанням бібліотеки "Pandas". Порівняння моделей виконано з використанням середньої абсолютної похибки. Представлено застосування методів роботи з великими наборами даних. Розрахунки проведено з допомогою бібліотеки "NumPy". Візуалізація даних – за допомогою моделей "Matplotlib" та "Seaborn. Scikit-Learn".*

Ключові слова: *транспорт, управління транспортом, машинне навчання, моделювання, споживання палива.*