მე-6 ქართულ-პოლონური საერთაშორისო სამეცნიერო კონფერენცია
სატრანსპორტო ხიდი ევროპა-აზია

6th Georgian-Polish International Scientific Conference
Transport Bridge Europe-Asia

ეძღვნება სტუ-ს 100 წლის იუბილეს

Dedicated to the 100th anniversary of GTU

ISBN 978-9941-8-4775-2

26-28 სექტემბერი, 2022, თბილისი

26-28 September, 2022, Tbilisi

მე-6 ქართულ-პოლონური საერთაშორისო სამეცნიერო კონფერენცია „სატრანსპორტო ხიდი ევროპა-აზია“

6th Georgian-Polish International Scientific Conference "Transport Bridge Europe-Asia"

ეძღვნება სტუ-ს 100 წლის იუბილეს

Dedicated to the 100th anniversary of GTU

26-28 სექტემბერი, 2022, თბილისი

26-28 September, 2022, Tbilisi

ISBN 978-9941-8-4775-2

დაიბეჭდა შპს ,,მწიგნობარი"-ს სტამბაში
კიევის ქუჩა N10

შინაარსი

THE USE OF MACHINE LEARNING TO MANAGE TRANSPORT PROCESSES IN A STEEL MILL ON THE EXAMPLE OF THE COMPANY "HUTA POKÓJ PROFILE"

Artur Budzyński[*1], Aleksander Sładkowski[*1], Paweł Malinowski[*2]
[1]Silesian University of Technology, Katowice, Poland
[2]Huta Pokój Profile sp. z o.o., Ruda Śląska, Poland
[*]E-mail: artur.budzynski@polsl.pl

Summary:
The problem of the valuation of road transport of goods in a steel producing company is presented. The specificity of the management of transport processes in the steelworks is discussed. The problem of steel transportation is presented. The impact of distance on the price of the transport service is described. The article presents two methods of training the model in a way that results directly from the data and in a universal way by adding information about the number of kilometers in each relation. Four machine learning models were compared according to the mean absolute error metric.

Keywords:
freight management, logistic, transport management, steel production, machine learning.

1. INTRODUCTION

Technological advances make it possible to replace man by machine in tasks that require computing power. Machine learning was defined in [1] by Tom Mitchel as the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general principles in large datasets to information filtering systems that automatically learn about users' interests. The process of pricing a transportation service takes time for employees, which is valuable. The company Huta Pokój Profile has been creating the history of metallurgy and steel processing since 1840. It continues this long tradition of a development leader, caring for the local community and focusing on increasing the company's value. The company has the ambition to be a modern and profitable company, integrating and inspiring other entrepreneurs, including steel processors. The company specializes in cold-formed sections of various cross-sectional geometry and very high quality and precision workmanship; hot-rolled sections in a wide range of steel grades, with complex cross sectional geometry and complementary services, in the above-mentioned areas. The company has its own technological facilities, where it can make new or regenerate the existing equipment needed for the production of steel sections. Thanks to this, it has the right equipment on time, as well as the ability to start production according to the customer's needs. The company has a quality and environmental management system compliant with the requirements of PN-EN ISO 9001, PN-EN ISO 14001, as well as a quality management system in research laboratories compliant with the requirements of PN-EN ISO / IEC 17025. They ensure the quality of products and services. laboratories at the highest level, taking into account the surrounding natural environment. The company uses the SOOT auction platform to manage transport processes. The auction system consists directly in an auction and the company with the lowest price performs the service. Machine learning is used to solve transport problems, in [2] the consumption of diesel fuel per 40 tons of vehicles in international transport is presented, in [3] the prediction of demand on the transport exchange based on machine learning solutions is presented. The price depends on many factors. The impact of the gross domestic product on the price is presented in [4]. The publication [5] presents more complex dependencies that affect the price. The Jupyter Notebook [6] is an editor for programming code. Pandas [7] is a data processing programming library. NumPy [8] is the programming library for calculation analysis, Scikit-Learn [9] is the machine learning programming library. Matplotlib [10] and Seaborn [11] are visualization programming libraries.

2. DATA

Fig. 1 shows the basic information about the data set. In the raw data, the names of the columns are in Polish. For a better understanding of the data, they were translated into English. The data is described by 11 features. 10 features are described as object type, 1 feature as int64. The collection consists of 3354 cases, all data is complete. There are no empty values. The object variables will need to be processed to be useful for modeling.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3354 entries, 0 to 3353
Data columns (total 11 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   STATUS                          3354 non-null   object
 1   AUCTION NUMBER                  3354 non-null   object
 2   ROUTE                           3354 non-null   object
 3   TRANSPORT                       3354 non-null   object
 4   START DATE (PLAN)               3354 non-null   object
 5   FINISH DATE (PLAN)              3354 non-null   object
 6   REQUIRED MEANS OF TRANSPORT     3354 non-null   object
 7   OFFER AMOUNT                    3354 non-null   object
 8   REMAINING TIME                  3354 non-null   object
 9   END DATE                        3354 non-null   object
 10  OFFERS                          3354 non-null   int64
dtypes: int64(1), object(10)
memory usage: 288.4+ KB
```

Fig. 1. Information about the dataset

Tab. 1 shows a random 5 cases. Status has two unique values. It is "out of time" and "interrupted". The analysis should take those that took place and have the status "out of time". The auction number is a unique number assigned to each auction. This is exactly 3354 unique values. There are 242 unique routes among the data. In the collection, there are signs "Ruda Śląska" and "Ruda Śląska 9", which differ in the type of product in one company. It was decided to combine these places, which resulted in reducing the number of unique values to 229. START DATE and END DATE (PLAN) determine when the auction is to start and end. When an offer arrives near the end of the bidding time, it is automatically extended for a specified period of time. END DATE determines when the auction ended. REQUIRED MEANS OF TRANSPORT contains 2 unique values. They mean the same thing - a standard tarpaulin semi-trailer. OFFERS determine how many offers have been proposed. The auction time feature has been created, which is the difference between the scheduled end and the start of the auction. The shortest auction lasted 2 minutes, and the longest 6 hours and 43 minutes. The average is 9 minutes and 31 seconds. The median is 5 minutes. Making predictions on the basis of unique relations is used when prediction is needed on a route on which transport has already taken place. The problem occurs when you need a quote on a new route. This is a common occurrence for a steel producer. The producer must often include the transport price in the purchase price at the customer's request.

Based on this, it was decided to build a based model on the number of kilometers and relations between countries. Table 2 shows the average rates per kilometer depending on the relation. The transports apply only to transports commissioned via the auction platform by the Huta Pokój Profile company. The company is both a shipper and unloader. There are 21 unique relationships. The rate for transport depends on the place of loading and unloading. The price in a given relationship is influenced by the dependencies of supply and demand. The data in the table is sorted from cheapest to most expensive. The most expensive rates are in domestic relations after Poland and in exports to countries that are close. The most expensive relation between domestic transports in Poland was € 4.91 per km. The steelworks has many contractors in the Silesiad Voivodeship. Transports to cities such as Dąbrowa Górnicza or Katowice constitute a significant part. These are short routes up to 50 km. Short deliveries also apply to Czechs and the city of Ostrava. The cheapest rates are for import to Poland. Based on this, it can be concluded that short routes are more

expensive. The time taken by loading and unloading processes increases the rate on short journeys. The company has definitely more export than import relations. The company exports goods to 17 countries, and imports to 3. The company receives a significant part of the goods using rail transport. Transports in exports from Poland are more expensive than in imports. It depends on many factors, such as the supply-demand relationship [12]. The company that provides services for Huta Pokój Profile, after the completion of the order, performs subsequent transports on its own. What happens after unloading the goods with the vehicle is not monitored and is not subject to research. Huta does not provide transports with its own fleet.

Tab. 1

Random 5 sample auctions

	STATUS	AUCTION NUMBER	ROUTE	TRANSPORT	START DATE (PLAN)	FINISH DATE (PLAN)	REQUIRED MEANS OF TRANSPORT	OFFER AMOUNT	REMAINING TIME	END DATE	OFFERS
1294	Licytacja po czasie	Licytuj 41/A/03/2022/HPP	Ruda Śląska [PL] Hannover [DE]	TR2020/040	11.03.2022 11:58	11.03.2022 12:03	Naczepa std. (13,6m - plandeka)	1 200,00 EUR	Czas minął	11.03.2022 12:03	1
3043	Licytacja po czasie	Licytuj 119/A/06/2021/HPP	Ruda Śląska [PL] Kalben [DE]	TR2020/027	11.06.2021 13:39	11.06.2021 13:44	Naczepa std. (13,6m - plandeka)	700,00 EUR	Czas minął	11.06.2021 13:47	1
2720	Licytacja po czasie	Licytuj 183/A/07/2021/HPP	Ruda Śląska 9 [PL] Poznań [PL]	TR2021/002	20.07.2021 14:22	20.07.2021 15:15	Naczepa std. (13,6m - plandeka)	---	Czas minął	20.07.2021 15:15	0
490	Licytacja po czasie	Licytuj 95/A/08/2020/HPP	Ruda Śląska [PL] Frenstat p/Radhostem [CZ]	TR2020/005	27.08.2020 12:53	27.08.2020 13:00	Naczepa std. (13,6m - plandeka)	220,00 EUR	Czas minął	27.08.2020 13:00	1
126	Licytacja po czasie	Licytuj 205/A/10/2020/HPP	Ruda Śląska 9 [PL] Tubize [B]	TR2020/019	30.10.2020 09:47	30.10.2020 10:31	Naczepa std. (13,6m - plandeka)	---	Czas minął	30.10.2020 10:31	0

Tab. 2

Average rate of transports ordered by Huta Pokoj Profile

Number	Country Load Place	Country Delivery Place Place	Mean Price [€/km]
1	Denmark	Poland	0.79
2	Austria	Poland	0.86
3	Germany	Poland	0.94
-	-	-	-
19	Poland	Slovenia	1.99
20	Poland	Czech	2.04
21	Poland	Poland	4.91

The graph in Fig. 2 shows the dependence of the rate per kilometer on the distance. The outliers at distances between 0 and 100 km are clearly visible. Above this value, the curve flattens out. This confirms the thesis that short routes are more expensive per kilometer of route. This also explains why the national coverage was the most expensive. The company often delivers goods to customers located in the Silesia region. It is also worth noting that steel as a product for road transport is problematic due to the time-consuming loading and unloading.

3. MODEL

Based on the above analysis, it was decided to train the models in 2 versions. The baseline is trained based on the relationships that are in the training data. This model can provide a price prediction on the selected relationship. To train the universal version, the number of kilometers in each country was assigned to each relation. In this version, the model is universal. You can get a price forecast in a relation that has never been performed before. The comparison of the above-described methods and the 4 models: DecisionTreeRegressor, RandomForestRegressor, ExtraTreesRegressor, GradientBoostingRegressor is shown in fig 3. All models are regression models. Decision trees are high execution speed [13]. In RandomForestRegressor classes each tree in the ensemble is built from a sample drawn with replacement from the training set. Furthermore, when splitting each node during the construction of a tree, the best split is found either from all input features or a random subset of size max of features. The purpose of these two sources of randomness is to decrease the variance of the forest estimator. Indeed, individual decision trees typically exhibit high variance and tend to overfit. The injected randomness in forests yield decision trees with somewhat decoupled prediction errors. By taking an average of those predictions, some errors can cancel out. Random forests achieve a reduced variance by combining diverse trees, sometimes at the cost of a slight increase in bias. In practice the variance reduction is often significant hence yielding an overall better model. ExtraTreeRegressor is algorithm that combines the attribute randomization of Random Subspace with totally random selection of the cut-point [14]. In ExtraTreesRegressor randomness goes one step further in the way splits are computed. As in random forests, a random subset of candidate features is used, but instead of looking for the most discriminative thresholds, thresholds are drawn at random for each candidate feature and the best of these randomly-generated thresholds is picked as the splitting rule. GradientBoostingRegressor builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage, a regression tree is fit on the negative gradient of the given loss function. The use of these models for building a publicly available web application is presented in [15].

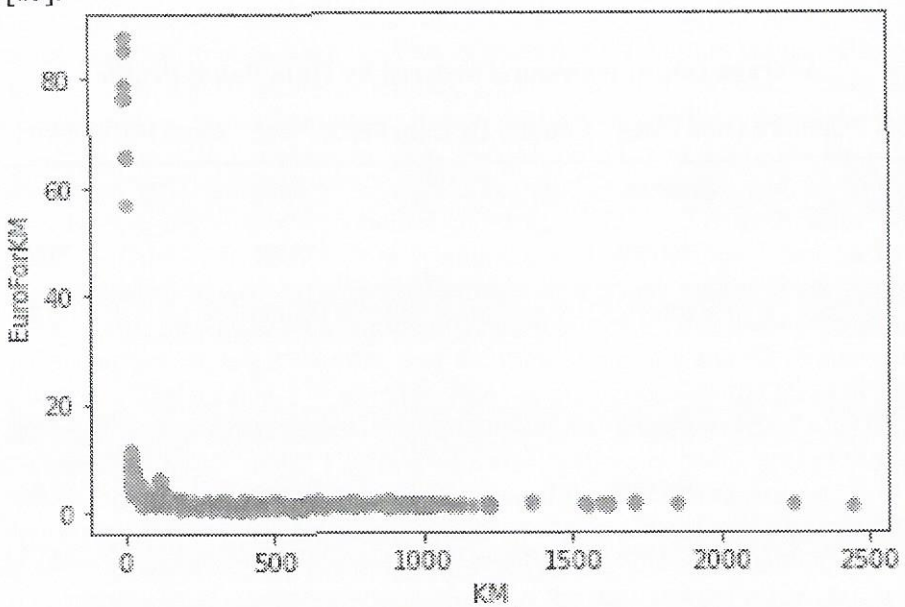

Fig. 2. The dependence of the rate per kilometer on the distance

The error in the base method is smaller than in the universal one. The universal method has a greater error and is more practical. The best model in the base method is ExtraTreesRegressor with an error of 28.76. The best model in the universal method is the RandomForestRegressor with an error of 35.58. Fig. 4 shows the top 20 features for the best model in the RandomForestRegressor universal method. KM means the number of kilometers in total; DE_KM - number of kilometers in Germany; CZ_KM - number of kilometers in the Czech Republic; PL_KM - number of kilometers in Poland; DK_KM - number of

kilometers in Denmark; NL_KM - the number of kilometers in the Netherlands; SK_KM - number of kilometers in Slovakia; RELATION_MEAN - means the average price in relation between countries.

4. IMPLEMENTATION

It can be assumed that the knowledge of IT methods among employees managing transport processes is not at a high level. To solve this problem, it is proposed to put a trained model in an application that will have an affordable level of service. Deploying a machine learning model to production is a never-ending process. Fig. 5 shows the machine learning life cycle. Implementation in a way that the employee can use will have a positive effect on further research. A survey of employees using the model will allow us to validate the model. It will also be possible to compare how much less error the model makes. The Flask framework can be used to perform the implementation.

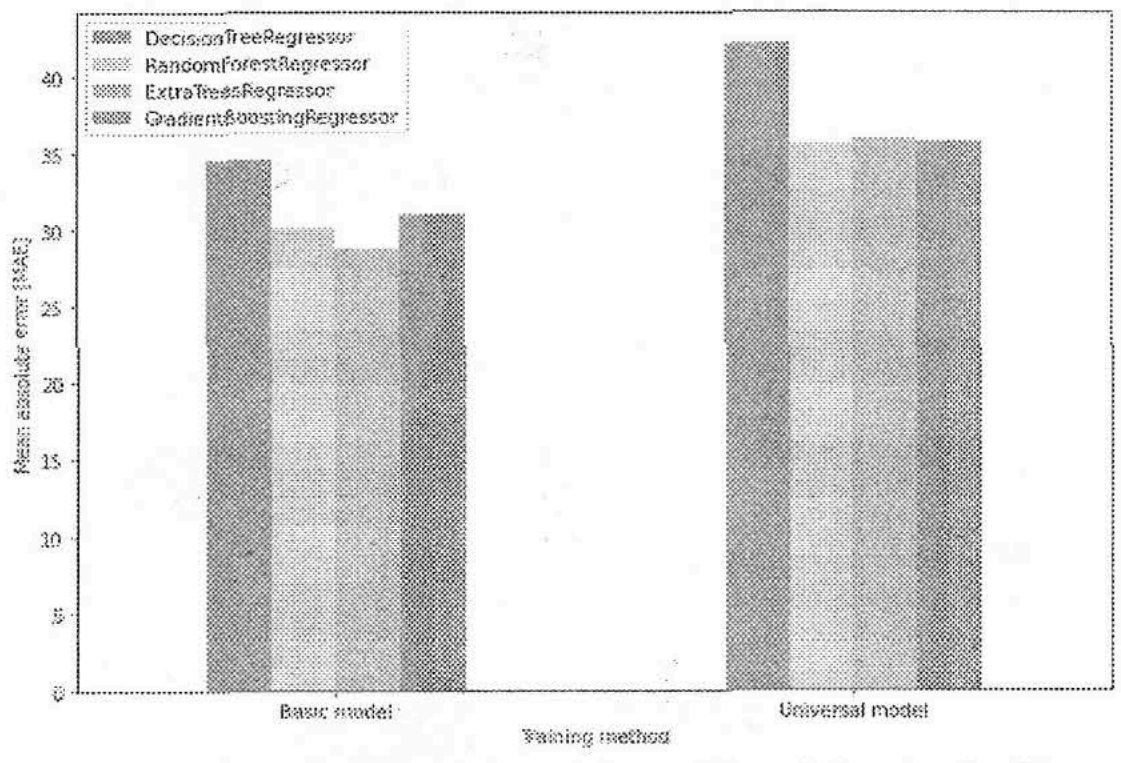

Fig. 3. Comparison of 4 models and 2 training methods

Weight	Feature
0.5366 ± 0.6470	RELATION_MEAN
0.3903 ± 0.6585	KM
0.0129 ± 0.0262	DE_KM
0.0095 ± 0.0108	CZ_KM
0.0094 ± 0.0073	ROUTE_FACTORIZED
0.0089 ± 0.0086	START DATE (PLAN) YEAR
0.0089 ± 0.0087	FINISH DATE (PLAN) YEAR
0.0040 ± 0.0073	PL_KM
0.0024 ± 0.0018	OFFERS
0.0020 ± 0.0043	RELATION_FACTORIZED
0.0020 ± 0.0068	COUNTRY_DELIVERY_PLACE
0.0016 ± 0.0018	START DATE (PLAN) DAY OF YEAR
0.0016 ± 0.0093	DK_KM
0.0014 ± 0.0017	FINISH DATE (PLAN) DAY OF YEAR
0.0012 ± 0.0035	NL_KM
0.0010 ± 0.0016	FINISH DATE (PLAN) WEEK
0.0010 ± 0.0021	SK_KM
0.0010 ± 0.0014	START DATE (PLAN) WEEK
0.0006 ± 0.0012	START DATE (PLAN) DAY
0.0005 ± 0.0011	FINISH DATE (PLAN) MONTH
	... 21 more ...

Fig. 4. Top 20 Features for RandomForestRegressor in the universal method

5. CONCLUSIONS

Machine learning has applications in improving the management of a steel producer's transport processes. Short routes are more expensive per kilometer than longer routes. Providing prices in one currency on a regular basis will improve analytical processes in the future. The error in the base method is smaller than in the universal one. The universal method has a greater error and is more practical. The best model in the base method is ExtraTreesRegressor with an error of 28.76. The best model in the universal method is the RandomForestRegressor with an error of 35.58. Models can be trained using various methodologies. The kilometer-based training method is more universal and is applicable to the valuation of new transport relationships. An example of the implementation of research results to solve a problem in an enterprise was presented.

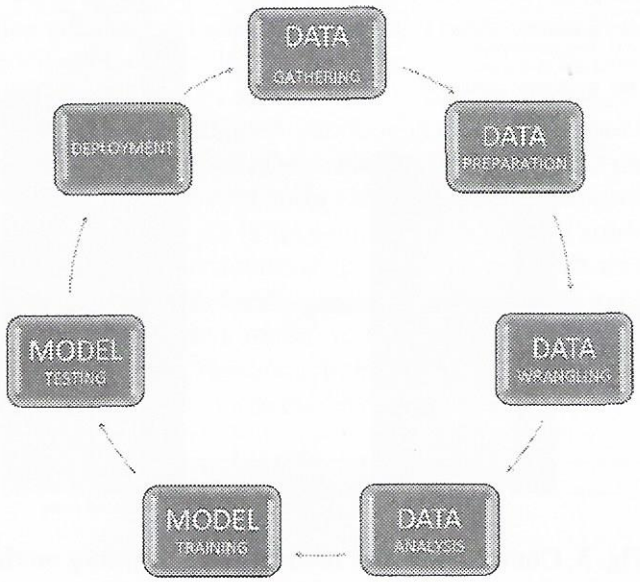

Fig. 5. The life cycle of a machine learning model

Acknowledgments

The project was created thanks to cooperation with the company Huta Pokój Profile. Company made the accident data available for investigation. Publication supported by the Excellence Initiative Research University programme of the Silesian University of Technology, 2021.

References:

1. Mitchell, T. *Machine learning.* McGraw Hill Burr Ridge. 1997;
2. Budzyński A., Sładkowski A. The use of machine learning to predict diesel fuel consumption in road vehicles. In: *Letnik, T. (ed.) 19th European Transport Congress: European Green Deal Challenges and Solutions for Mobility and Logistics in Cities.* Maribor: Zum urbanizem, planiranje, projektiranje d.o.o. 2021. P. 207-222;
3. Budzyński A., Sładkowski A. Forecasting road transport demand with use machine learning. In: *Proceedings of the International Conference on Problems of Logistics, Management and Operation in the East-West Transport Corridor (PLMO).* Baku: Institute of Control Systems of ANAS. 2021. P. 59-63;

4. Budzyński A. Use dependencies between freight prices and economic factor as a solution in improve efficiency work in road transport. In: *Sładkowski, A. (ed.) XII Int. Sci. Conf. & IX Int. Symposium of Young Researches „Transport Problems'2020".* Conference Proceedings, 2020;
5. Budzyński A., Sładkowski A. Predykcja cen w transporcie drogowym z wykorzystaniem sztucznych sieci neuronowych. In: *Sładkowski A. (ed.) XIII Int. Sci. Conf. & X Int. Symposium of Young Researches „Transport Problems'2021". Conference Proceedings.* Katowice: Silesian University of Technology. Faculty of Transport and Aviation Engineering. 2021. P. 137-151;
6. Kluyver T., Ragan-Kelley B., Perez F., Granger B., Bussonnier M., Frederic J., Kelley K., Hamrick J., Grout J., Corlay S., Ivanov P., Avila D., Abdalla S. *Jupyter notebooks – a publishing format for reproducible computational workflows.* Positioning and Power in Academic Publishing: Players, Agents and Agendas, 2016;
7. McKinney W. Data structures for statistical computing in python. *Proc. of the 9th Python in Sci. Conf. (SCIPY).* 2010. P. 56–61;
8. van der Walt S, Colbert C., Varoquaux G. The NumPy Array: A structure for efficient numerical computation. *Computing in Science & Engineering.* 2011. Vol. 13. No. 2. P. 22 – 30;
9. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel Q., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., Louppe G. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research.* 2011. Vol. 12. P. 2825-2830;
10. Hunter J. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering.* 2007. Vol. 9. No. 3. P. 90–95;
11. Waskom M.L. Seaborn: statistical data visualization. *Journal of Open Source Software.* 2021. Vol. 6(60):3021. P. 1-4;
12. Budzyński A., Sładkowski A. The use of machine learning to improve road transport management processes. In: *Transport Problems 2022. XIV International scientific conference, XI International symposium of young researchers. Proceedings / Sładkowski A. (ed.).* 2022. P. 820-829;
13. Quinlan J.R. Induction of decision trees. *Machine Learning.* 1986. Vol. 1(1). P. 81-106;
14. Geurt P., Ernst D., Wehenkel L. Extremely randomized trees. *Machine Learning.* 2006. Vol. 63(1). P. 3-42;
15. Budzyński A., Sładkowski A. Implementation machine learning model into web application for freight price prediction in road transport. In: *Horváth B., Horváth G. (eds.) XX. European Transport Congress & XII. International Conference on Transport Sciences.* University of Győr, Department of Transport. 2022. P. 132-139.