https://doi.org/10.3311/PPtr.22203

Periodica Polytechnica Transportation Engineering, 51(4), pp. 431–435, 2023

Forecasting the Number of Road Accidents in Poland Depending on the Day of the Week Using Neural Networks

Piotr Gorzelanczyk[1*]

[1] Department of Transport, Stanislaw Staszic University of Applied Sciences in Pila, Podchorazych 10 Street, 64-920 Pila, Poland
[*] Corresponding author, e-mail: piotr.gorzelanczyk@ans.pila.pl

Received: 13 March 2023, Accepted: 19 May 2023, Published online: 21 June 2023

Abstract

A very large number of people are killed on the roads. Although the value is decreasing from year to year and the pandemic has reduced the number of traffic accidents even further, the number is still very high. For this reason, it is necessary to find out on which days of the week the highest number of traffic accidents occur, and to know the accident forecast for the coming years, so that we can do everything possible to minimize the analyzed value.

The aim of the article is to develop a forecast of the number of road accidents in Poland depending on the days of the week. For this purpose, monthly data on the number of road accidents in Poland by days of the week were analyzed on the basis of Police data. Based on this data, a forecast of the number of road accidents for 2022-2024 was made in Statistica software. Selected neural network models were used to forecast the number of road accidents. The results show that we can still expect a reduction in the number of road accidents on Polish roads, but the prevailing pandemic disturbs the results obtained. The choice of the number of random samples (learning, testing and validation) affects the obtained results.

Keywords

traffic accident, pandemic, forecasting, neural networks, day of the week

1 Introduction

Every year a very large number of people die in road accidents. Based on WHO data, about 1.3 million people die on the roads each year as a result of road accidents. Road accidents are the leading cause of death for young people between the ages of 5 and 29. The UN plans to halve the number of deaths and injuries on the roads by 2030 (WHO, 2018).

We obtain data on the number of traffic accidents from various sources. We can include data collected by government bodies through relevant government agencies. In Poland, the most commonly used ones are data from police statistics and insurance company databases. Partial information on traffic accidents is then processed for the transportation sector on a larger scale (Gorzelańczyk et al., 2020).

Methods for forecasting the number of accidents are developed using various numerical methods. The most commonly used are time series methods (Helgason, 2016; Lavrenz, 2018), which have the disadvantages of not being able to assess the quality of the forecast on the basis of outdated forecasts and frequent autocorrelation of the residual component (Kowalsi, 2005). Procházka et al. (Prochazka and Camaj, 2017; Procházka et al., 2017) used a multiple seasonality model for forecasting, while Sunny et al. (2018) used the Holt-Winters exponential smoothing method, the drawback of which is the inability to introduce exogenous variables into the model (Dudek, 2013; Szmuksta-Zawadzka and Zawadzka, 2009).

We can also use vector autoregression models to forecast the number of traffic accidents. Their disadvantage is the need for a large number of observations of variables in order to correctly estimate their parameters (Wójcik, 2014), which is not always achievable. Duarte Monedero et al. (2021) used autoregressive models, Al-Madani (2018) used regression models with curve fitting. These, in turn, require only simple linear relationships (Mamczur, 2020) and autoregressive order (assuming the series is already stationary) (Piłatowska, 2012).

The ANOVA method (Chudy-Laskowska and Pisula, 2015) can also be used in forecasting the issue at hand. A disadvantage of this method is that it makes

Cite this article as: Gorzelanczyk, P. (2023) "Forecasting the Number of Road Accidents in Poland Depending on the Day of the Week Using Neural Networks", Periodica Polytechnica Transportation Engineering, 51(4), pp. 431–435. https://doi.org/10.3311/PPtr.22203

additional assumptions, the violation of which can lead to erroneous conclusions (Gregorczyk and Swarcewicz, 2012). Another disadvantage of this method is the knowledge of the domain (Chudy-Laskowska and Pisula, 2014), and the forecasting results depend on the assumption of the initial conditions of the network, as well as the inability to interpret in the traditional way, since neural networks are referred to as a black box, in which we provide input data and the model provides results without knowledge of the analysis (StatSoft, Inc., 2011). The latter is the subject of this article.

2 Materials and methods

A very high number of people die on Polish roads. Although the value is decreasing from year to year and the pandemic has reduced the number of road accidents even further, the number is still very high. Compared to the European Union, the number of accidents in Poland is still very high. The lowest number of road accidents occurs on Sundays and the highest on Fridays. For this reason, every effort should be made to reduce this value and identify the days of the week on which the highest number of road accidents will occur (Fig. 1).

Selected neural network models were used to predict the number of traffic accidents in Poland depending on the day of the week. The advantage of the method used is that it mimics the behavior of the human brain. A neural network consists of nodes that have inputs, weights, variances and outputs. During the study, the optimal weights were adjusted using Statistica and the artificial neural networks module available in it. Multilayer MLP (Multilayer Peceptron) neural networks with one layer of hidden neurons were used for forecasting. In the networks analyzed, the input layer consisted of 12 neurons

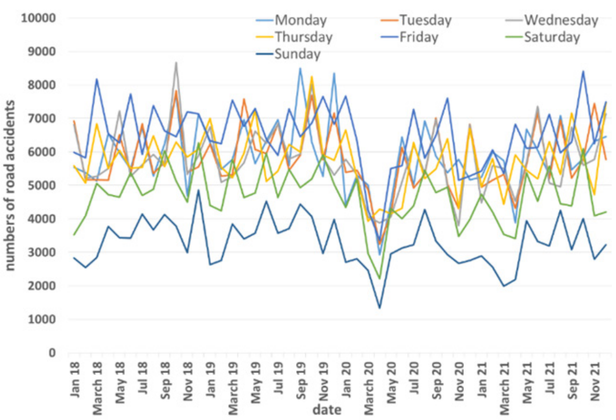

Fig. 1 Number of road accidents in Poland by the day of week in 2018-2021 (Statystyka Policja, 2022)

processing information representing 12 input time series, i.e. the number of traffic accidents (12-month period). The number of neurons in the hidden layer in the analyzed case varied from 2 to 8 neutrons, and the output layer was a single neutron, representing the output values of the time series of the number of traffic accidents. The result of the forecast by the method discussed depends on the choice of the model and its parameters.

The following forecast errors determined from Eqs. (1)–(5) were used to calculate the excellence measures of the analytical forecast:

- ME – mean error

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right), \tag{1}$$

- MAE – mean absolute error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - Y_p\right|, \tag{2}$$

- MPE – mean percentage error

$$\mathrm{MPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - Y_p}{Y_i}, \tag{3}$$

- MAPE – mean absolute percentage error

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - Y_p\right|}{Y_i}, \tag{4}$$

- SSE – sum of squared errors

$$\mathrm{SSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right)^2} \tag{5}$$

where:
n – the length of the forecast horizon,
Y – observed value of road traffic accidents,
Y_p – forecasted value of road traffic accidents.

Neural network models for which the mean percentage error and mean absolute percentage error were the smallest were used to predict the number of traffic accidents.

3 Results

The analysis of the change in the number of traffic accidents by day of the week was examined using the Kruskal-Wallis test. The value of the test statistic is 165.7 with a test probability of $p = 0.000$. The value obtained indicates the rejection of equality of the average level of traffic accidents. On this basis, it can be concluded that the number of traffic accidents analyzed, over the years, shows a systematic decrease in the average level of accidents. In addition, depending on the day of the week, there is a clear variation among accidents, as shown in Fig. 1. On the other hand, it is clear from Fig. 2 that the most traffic accidents occur on Fridays and the fewest on Sundays.

3.1 Forecasting the number of traffic accidents

Forecasting the number of traffic accidents in Poland by the day of the week was conducted using data from the Polish Police from 2007-2021 (Statystyka Policja, 2022). The study was conducted using Statistica software, assuming two random sample sizes:

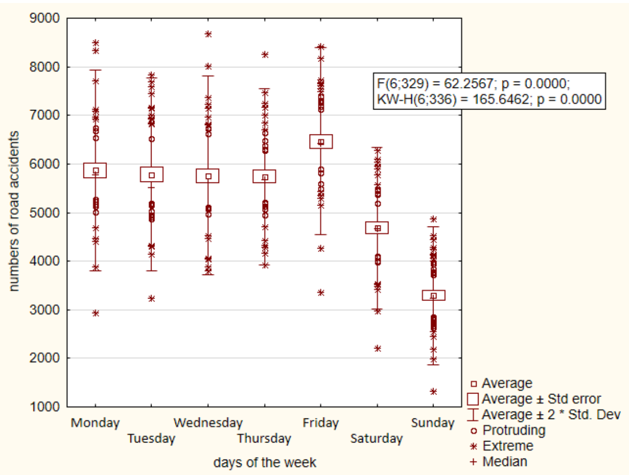

Fig. 2 Average number of traffic accidents by the day of week from 2007 to 2021

- teaching 70%, testing 15% and validation 15%,
- teaching 80%, testing 10% and validation 10%,

with the following number of teaching networks: 20, 40, 60, 80, 100, 200. Meanwhile, the following functions were considered as neuron activation functions in the hidden and output layers: logistic, linear, exponential and hyperbolic tangent. The minimization of the sum of squares of the errors in the output of the neural network was taken as the learning error functions. The BFGS (Broyden-Fletcher-Goldfarb-Shanno) iterative algorithm was used as the learning algorithm. The calculations in Statistica software used the automatic search module for the best networks. Of the 60 different combinations tested, those for which the MPE error value was the smallest were selected. Tables 1 and 2 show the best-trained neural networks that were used to predict the number of traffic accidents in Poland by the day of the week.

Taking into account the data in Tables 1 and 2, it can be concluded that the number of traffic accidents in Poland, regardless of the day of the week, will decrease from year to year and then stabilize. The results obtained depend on the selection of the random sample size. In the case of the teaching group of 70%, the test group of 15% and the validation group of 15% in proportions (70-15-15), the average percentage error was 0.25%, while for the second sample (80-10-10) the error was 0.44%. It should be added that the results obtained were affected by the pandemic (Figs. 3, 4).

Table 1 Summary of neural network learning for the case of random sample sizes teaching 70%, testing 15% and validation 15%

Day of the week	Network number	Network name	Quality (learning)	Quality (learning)	Quality (validation)	Learning algorithm	Activation (hidden)	Activation (output)	Errors				
									ME	MAE	MPE	MAPE	SSE
Monday	200	MLP 12-2-1	0.693338	0.758002	0.633341	BFGS 27	Exponential	Exponential	89.26906	561.9727	0.06%	10.36%	712.8712
Tuesday	40	MLP 12-2-1	0.590683	0.723670	0.597656	BFGS 5	Linear	Linear	139.8567	455.489	0.21%	8.99%	564.3721
Wednesday	20	MLP 12-3-1	0.534374	0.690442	0.443639	BFGS 6	Logistics	Logistics	106.998	611.9131	0.00%	11.41%	787.9739
Thursday	100	MLP 12-6-1	0.579401	0.584755	0.476653	BFGS 4	Tanh	Tanh	103.2719	572.3108	0.12%	10.46%	773.7889
Friday	60	MLP 12-8-1	0.815785	0.544633	0.512112	BFGS 35	Exponential	Exponential	69.33802	498.2956	0.02%	8.45%	657.7115
Saturday	60	MLP 12-8-1	0.541082	0.498878	0.822789	BFGS 6	Tanh	Tanh	87.01389	478.2827	0.05%	11.00%	614.0998
Sunday	100	MLP 12-7-1	0.709819	0.630756	0.772056	BFGS 18	Exponential	Tanh	21.59163	346.5561	1.29%	10.98%	441.2766
Average:									88.19131	503.5457	0.25%	10.24%	650.2991

Table 2 Summary of neural network learning for the case of random sample sizes teaching 80%, testing 10% and validation 10%

Day of the week	Network number	Network name	Quality (learning)	Quality (learning)	Quality (validation)	Learning algorithm	Activation (hidden)	Activation (output)	Errors				
									ME	MAE	MPE	MAPE	SSE
Monday	40	MLP 12-5-1	0.657814	0.734452	0.602295	BFGS 6	Logistics	Exponential	97.24617	600.8996	0.17%	11.07%	753.8129
Tuesday	100	MLP 12-2-1	0.670918	0.742790	0.631198	BFGS 45	Logistics	Logistics	31.95242	499.4731	0.95%	9.56%	669.8008
Wednesday	60	MLP 12-3-1	0.546019	0.379745	0.667254	BFGS 5	Logistics	Logistics	114.7553	613.4296	0.10%	11.34%	799.914
Thursday	80	MLP 12-2-1	0.566256	0.376886	0.645342	BFGS 5	Tanh	Tanh	131.9346	749.9932	0.05%	13.69%	1042.563
Friday	40	MLP 12-8-1	0.821667	0.677115	0.504405	BFGS 53	Exponential	Linear	55.59144	469.4439	0.00%	7.91%	631.5572
Saturday	20	MLP 12-7-1	0.559942	0.594536	0.899206	BFGS 13	Tanh	Logistics	47.64403	461.3273	0.91%	10.61%	598.7771
Sunday	60	MLP 12-6-1	0.672145	0.744519	0.755360	BFGS 12	Linear	Linear	27.21625	368.8581	0.92%	11.64%	462.1352
Average:									72.33431	537.6321	0.44%	10.83%	708.3657

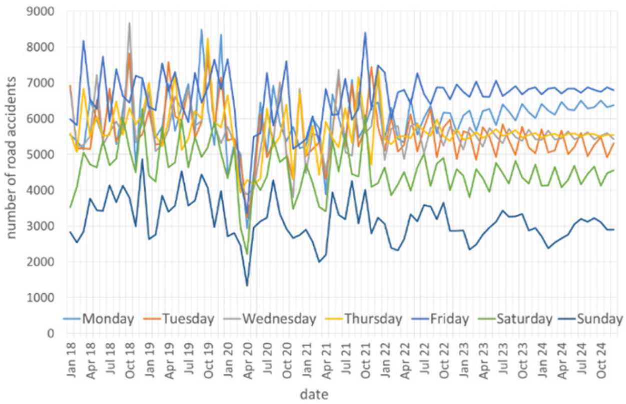

Fig. 3 Projected number of traffic accidents for 2021-2024 for the 70-15-15 test group

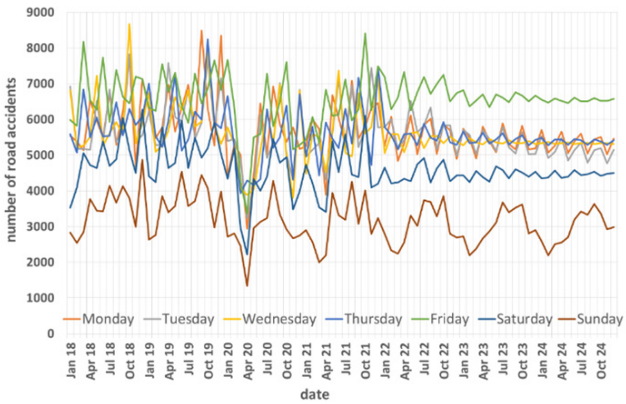

Fig. 4 Projected number of traffic accidents for 2021-2024 for the 80-10-10 test group

4 Summary

A neural network method was used to predict the number of accidents in Poland by the day of the week using Statistica software. The weights used were estimated by the program in such a way as to minimize the average absolute error and the average absolute percentage error.

Based on the data obtained, it can be concluded that the pandemic caused a reduction in the number of traffic accidents, and in the coming years it can be expected that the number of traffic accidents will continue to decrease and then stabilize. In addition, the pandemic further may change the value of the results obtained. The calculated forecast errors prove the accuracy of the models used.

Taking into account the obtained forecasts, measures should be taken to further minimize the number of traffic accidents. Such measures have already been implemented in Poland, as of January 1, 2022, by increasing penalties for traffic offenses on Polish roads. The pandemic, which drastically altered the number of road accidents, certainly had an impact on disturbing the obtained results of the study.

In further studies, the authors plan to take into account more factors affecting accident levels and use various statistical methods to determine the number of traffic accidents. These may include traffic volume, weather conditions or the age of the accident perpetrator.

References

Al-Madani, H. M. N. (2018) "Global road fatality trends' estimations based on country-wise microlevel data", Accident Analysis Prevention, 111, pp. 297–310.
https://doi.org/10.1016/j.aap.2017.11.035

Chudy-Laskowska, K., Pisula, T. (2014) "Prognoza liczby wypadków drogowych w Polsce" (Forecast of road accidents in Poland), Logistyca, 6, pp. 2710–2722. (in Polish)

Chudy-Laskowska, K., Pisula, T. (2015) "Prognozowanie liczby wypadków drogowych na Podkarpaciu" (Forecasting of the Number of Road Accidents in the Subcarpathian Region), Logistyka, 4, pp. 2782–2796. (in Polish)

Duarte Monedero, B., Gil-Alana, L. A., Valbuena Martínez, M. C. (2021) "Road accidents in Spain: Are they persistent?", IATSS Research, 45(3), pp. 317–325.
https://doi.org/10.1016/j.iatssr.2021.01.002

Dudek, G. (2013) "Forecasting Time Series with Multiple Seasonal Cycles Using Neural Networks with Local Learning", In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., Zurada, J. M. (eds.) Artificial Intelligence and Soft Computing, Springer, pp. 52–63. ISBN 978-3-642-38657-2
https://doi.org/10.1007/978-3-642-38658-9_5

Gorzelańczyk, P., Pyszewska, D., Kalina, T., Jurkovič, M. (2020) "Analysis of road traffic safety in the Pila poviat", Scientific Journal of Silesian University of Technology. Series Transport, 107, pp. 33–52.
https://doi.org/10.20858/sjsutst.2020.107.3

Helgason, A. F. (2016) "Fractional integration methods and short time series: Evidence from a simulation study", Political Analysis, 24(1), pp. 59–68.
https://doi.org/10.1093/pan/mpv026

Kowalski, L. (2005) "Prognozowanie na podstawie szeregów czasowych" (Forecasting based on time series), [pdf] Forecasting and Simulation, Available at: http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf [Accessed: 17 April 2022] (in Polish)

Lavrenz, S. M., Vlahogianni, E. I., Gkritza, K., Ke, Y. (2018) "Time series modeling in traffic safety research", Accident Analysis & Prevention, 117, pp. 368–380.
https://doi.org/10.1016/j.aap.2017.11.030

Mamczur, M. (2020) "Jak działa regresja liniowa? I czy warto ją stosować?" (How does linear regression work? And is it worth using?), Miroslaw Mamczur: A blog about data science, AI, machine learning and data visualization, Machine learning, Aug., 31. [online] Available at: https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/ [Accessed: 17 April 2022] (in Polish)

Piłatowska, M. (2012) "The choice of the order of autoregression depending on the parameters of the generating model", Econometrics, 4(38), pp. 16–35.

Prochazka, J. Camaj, M. (2017) "Modelling the number of road accidents of uninsured drivers and their severity", In: Proceedings of the 32nd International Academic Conferences, Geneva, Switzerland, p. 217. ISBN 978-80-87927-39-7
https://doi.org/10.20472/IAC.2017.032.035

Procházka, J., Flimmel, S., Čamaj, M., Bašta, M. (2017) "Modelling the Number of Road Accidents", In: 20th International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics 2017, Szklarska Poręba, Poland, pp. 355–364. ISBN 978-83-7695-693-0
https://doi.org/10.15611/amse.2017.20.29

Gregorczyk, A., Swarcewicz, M. (2012) "Analiza wariancji w układzie powtarzanych pomiarów do określenia efektów czynników wpływających na pozostałości linuronu w glebie" (Repeated Measures ANOVA to Determine the Effects of Factors Affecting Herbicide Residues in Soil), [pdf] Polish Journal of Agronomy, 11, pp. 15–20. Available at: https://www.iung.pl/PJA/wydane/11/PJA11_3.pdf [Accessed: 17 April 2022] (in Polish)

StatSoft, Inc. (2011) "Techniki zgłębiania danych (data mining)" (Data mining techniques) StatSoft Electronic Statistics Textbook, [online] Available at: https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html [Accessed: 17 April 2022] (in Polish)

Statystyka Policja "Wypadki Drogowe - Raporty Roczne" (Road Accidents - Annual Reports), [online] Available at: https://statystyka.policja.pl/st/ruch-drogowy/76562,Wypadki-drogowe-raporty-roczne.html?search=955207048705 [Accessed: 17 April 2022] (in Polish)

Sunny, C. M., Nithya, S., Sinshi, K. S., Vidya Vinodini, M. D., Aiswaria Lakshmi, K. G., Anjana, S., Manojkumar, T. K. (2018) "Forecasting of Road Accident in Kerala: A Case Study", In: 2018 International Conference on Data Science and Engineering (ICDSE), Kochi, India, pp. 1–5.
https://doi.org/10.1109/ICDSE.2018.8527825

Szmuksta-Zawadzka, M., Zawadzka, J. (2009) "O prognozowaniu na podstawie modeli Holta-Wintersa dla pełnych i niepełnych danych" (Forecasting on the basis of Holt-Winters models for complete and incomplete data), Research Papers of the Wrocław University of Economics, 38, pp. 85–99. (in Polish)

WHO (2018) "Global status on road safety 2018", World Health Organization, Geneva, Switzerland. ISBN 9789241565684 [online] Available at: https://www.who.int/publications/i/item/9789241565684 [Accessed: 17 April 2022]

Wójcik, A. (2014) "Modele wektorowo-autoregresyjne jako odpowiedź na krytykę strukturalnych wielorównaniowych modeli ekonometrycznych" (Vector Autoregression Models (VAR) – Response to Criticism Structural Econometric Models), Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach, 193, pp. 112–128. (in Polish)