sciendo

LOGI – Scientific Journal on Transport and Logistics
Vol. 14 No. 1 2023 DOI: 10.2478/logi-2023-0004
© 2023 P. Gorzelanczyk and H. Tylicki. This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

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

Piotr Gorzelanczyk[1*] and Henryk Tylicki[1]

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

***Corresponding Author:** Piotr Gorzelanczyk

Received: 18 July 2022; Revised: 16 November; Accepted: 9 January 2023; Published: 14 February 2023

Abstract: The number of road accidents in the world is decreasing year by year. This number has been affected by the pandemic in recent years but is still very high. Therefore, it is necessary to take all possible measures and steps to reduce this number. The objective of the article is to forecast the number of road accidents in Poland depending on the day of the week. For this purpose, the annual data on the number of road accidents in Poland broken down by days of the week were analyzed, and a forecast for the years 2022-2040 was prepared on the basis of the police statistics. The forecast was made using selected models of neural networks. The research results show that a decrease in in the number of road accidents can be expected. However, the obtained results could be affected by the selected number of random samples (training, testing and validation).

Keywords: Road accident, pandemic, forecasting, neural networks

1. Introduction

Every year, a large number of people die in road accidents. According to the WHO, it is about 1.3 million people. Road accidents are the main cause of death of minors and young people aged 5 to 29 [1]. Therefore, the UN General Assembly has decided to halve the number of road fatalities and injuries by 2030.

The data on road accidents can be obtained from various sources. For the purposes of this paper, data collected by government bodies through the relevant government agencies are considered. The collected data were obtained from police reports, insurance databases, or hospital records. The obtained information about road accidents is then processed for the transport sector on a larger scale [2].

Professional literature dealing with this issue provides various methods for predicting the number of accidents, with the most common one being time series methods [3,4]. However, the shortcoming

of this metho is the inability to assess the quality of the forecast based on outdated forecasts and the frequent autocorrelation of the residuals [5]. Procházka et al. [6,7] used a multiple seasonality model for forecasting; in contrast, Sunny et al. [8] chose the Holt-Winters exponential smoothing method. The disadvantage of this method is the inability to introduce exogenous variables into the model [9-11].

Predictions can also be made using vector autoregression models; however, for correct estimates of its parameters, a large number of observations of variables are necessary [12], which is also not always possible to obtain. Other possible methods are autoregression models [13] and curve fitting regression models. These, in turn, require only simple linear relationships and an autoregressive order (assuming that the series are already stationary) [14,15].

Chudy-Laskowska and Pisula used the ANOVA method for similar purposes [16,17]. The shortcoming of this method is that it makes additional assumptions, the violation of which may lead to erroneous conclusions. Another method to predict the number of road accidents is the application of neural network models. The disadvantage of this method is the necessity of having practical knowledge in this area [16,18]. In addition, the output is dependent on the assumption of meeting the initial network conditions and cannot be interpreted in the traditional way because ANN are usually referred to as a black box where you provide the input and the model gives the results without providing any information about the analysis [19]. For the purposes of the article, the authors focus on the last method mentioned.

2. Materials and Methods

As already mentioned, many road accidents occur on Polish roads every year. In recent years, the number of road accidents has decreased thanks to the pandemic, which also affects the predicted value. Despite the pandemic, the number of road accidents is still high; efforts should thus be made to reduce this value and determine the days on which the highest number of road accidents occur. As seen from Figure 1, the lowest number of road accidents happen on Sundays, while the day with the highest number is Friday.

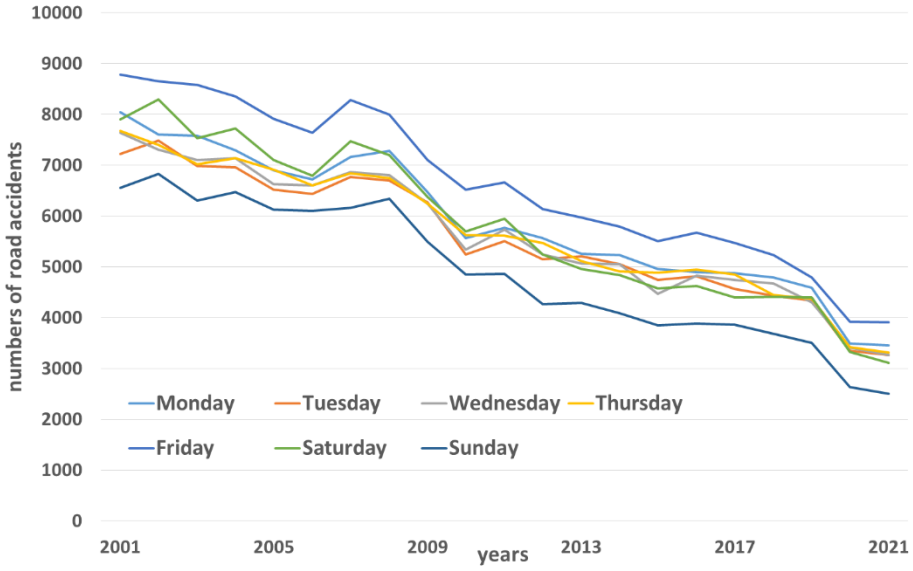


Fig. 1 Number of road accidents in Poland by day of the week in 2009-2021 Source: [20]

Selected models of neural networks were used to forecast the number of road accidents in Poland depending on the day of the week. The advantage of this method is that it copies the behavior of the human brain. Neural networks consist of nodes that have inputs, weights, layers, and outputs. During the tests, the optimal weights were selected using the Statistica software. The prediction made using the aforementioned method depends on the choice of the model and its parameters.

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

- ME – mean error

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

- MAE – mean absolute error

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

- MPE – mean percentage error

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

- MAPE - mean absolute percentage error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - Y_p|}{Y_i}, [\%] \tag{4}$$

- SSE – sum square error

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

where: n is the length of the forecast horizon [-]; Y is the observed value of road traffic accidents [-]; Y_p is the predicted number of road traffic accidents [-].

To forecast the number of road accidents depending on the day of the week, neural network models with the smallest mean percentage error and the mean absolute percentage error were selected.

3. Results

To compare the fluctuations in the number of road accidents in the analyzed period, a statistical test was carried out. The statistical value for the analyzed non-parametric Kruskall-Wallis test is 14.40, with the test probability value being $p < 0.05$. Therefore, the hypothesis stating that the average level of road accidents does not depend on the day of the week can be rejected (see Fig. 2).

Based on the data analysis, it can be concluded that most road accidents occur on Fridays. This is due to the increase in traffic and mobility of people. In turn, the lowest number of road accidents is recorded on Sundays [20].

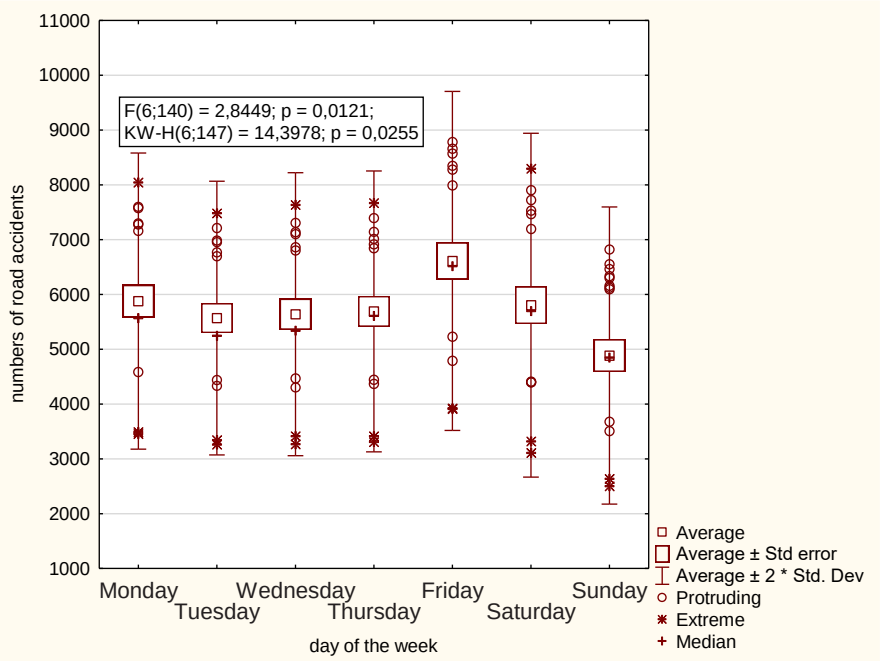


Fig. 2 Average number of road accidents by day of the week in 2000-2021. Source: authors

3.1 Forecasting Road Accidents

Data from the Polish Police for 2001-2021 [20] were used to forecast the daily number of road accidents in Poland, depending on the day of the week. The research was carried out using the Statistica program, assuming two numbers of random samples:

- 70% training, 15% test and 15% validation.
- 80% training, 10% test and 10% validation.

The following number of training networks were used: 20,40,60,80,100,200, for which the MPE error value was minimal (Tab. 1,2).

Table 1 Overview of neural network training for the case of random sample sizes: training 70%, testing 15% and validation 15%. Source: authors

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	20	MLP 10-5-1	0.303011	0.612736	0.748684	BFGS 8	Logistic	Tanh	107.085846	195.5969	3.28%	5.02%	315.9813
Tuesday	60	MLP 10-2-1	0.342640	0.563709	0.739524	BFGS 8	Tanh	Linear	72.2116552	146.4988	2.21%	3.76%	219.4247
Wednesday	40	MLP 10-3-1	0.337755	0.626536	0.741267	BFGS 5	Logistic	Tanh	50.7250401	274.3396	2.16%	6.71%	332.6728
Thursday	100	MLP 10-5-1	0.317767	0.529929	0.799875	BFGS 18	Tanh	Tanh	102.816062	116.5875	2.98%	3.23%	244.0794
Friday	40	MLP 10-3-1	0.349338	0.542171	0.747112	BFGS 7	Logistic	Tanh	12.2952013	174.7906	0.61%	3.37%	236.4118
Saturday	40	MLP 10-3-1	0.343978	0.541514	0.701967	BFGS 5	Exponential	Tanh	210.200154	904.7954	0.55%	18.41%	1856.928
Sunday	40	MLP 10-7-1	0.311755	0.507014	0.644922	BFGS 5	Tanh	Tanh	107.128912	236.475	4.16%	7.33%	298.0765

Table 2 Overview of neural network training for the case of random sample sizes: training 80%, testing 10% and validation 10%. Source: authors

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 10-4-1	0.495863	0.00	0.00	BFGS 58	Tanh	Exponential	29.0304892	49.4298	0.83%	1.24%	92.61942
Tuesday	60	MLP 10-8-1	0.501535	0.00	0.00	BFGS 2	Exponential	Tanh	62.1471235	385.8246	0.49%	9.32%	480.5871
Wednesday	200	MLP 10-6-1	0.538686	0.00	0.00	BFGS 12	Exponential	Tanh	0.02486799	14.61703	0.02%	0.34%	18.11308
Thursday	100	MLP 10-5-1	0.494829	0.00	0.00	BFGS 7	Exponential	Linear	18.4522623	146.2379	0.17%	3.44%	192.717
Friday	20	MLP 10-4-1	0.491451	0.00	0.00	BFGS 10	Exponential	Logistic	28.7569857	198.3383	1.10%	3.97%	254.1246
Saturday	200	MLP 10-8-1	0.430994	0.00	0.00	BFGS 4	Exponential	Linear	6.51214621	277.7761	0.60%	6.71%	316.5582
Sunday	100	MLP 10-3-1	0.462087	0.00	0.00	BFGS 6	Logistic	Tanh	31.3387081	232.5678	0.64%	6.97%	301.0297

Based on the results presented, it can be concluded that the number of traffic accidents in Poland, regardless of the day of the week, will show a year-on-year decrease. However, the results obtained depend on the selection of the random sample size. Increasing the percentage of the networks in training dataset relative to the test and validation group minimizes the mean percentage error. In the case of the training dataset accounting for 70%, the testing dataset 15% and the validation dataset 15% (70-15-15), the error was 2.28%, while for the second sample (80-10-10) the error was 0.55% (Fig. 3,4).

Based on the obtained data on the predicted number of traffic accidents depending on the day of the week, it can be concluded that the number of traffic accidents can still be expected to decrease and then stabilize in the coming years. It should be added that the results obtained are influenced by the pandemic [21].

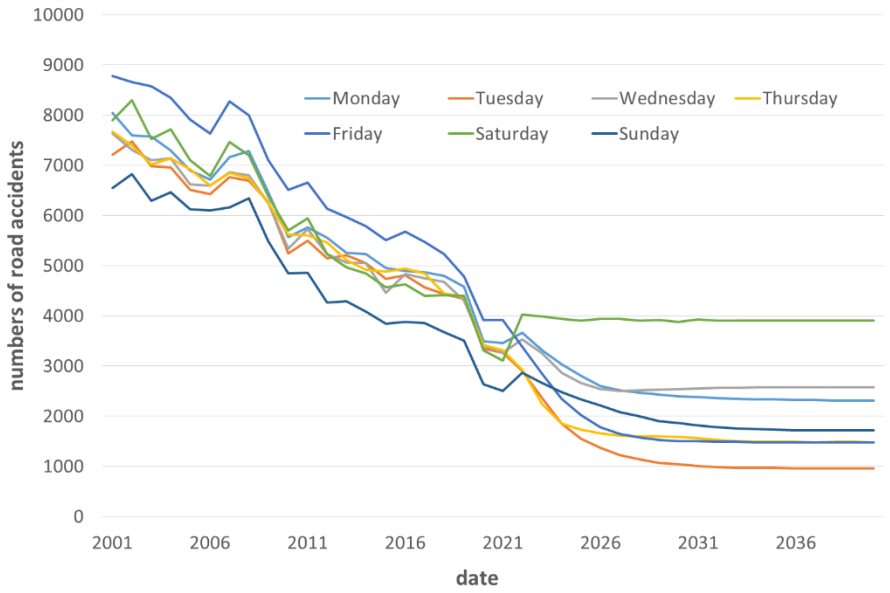


Fig. 3 Predicted number of traffic accidents for 2022-2040 for the 70-15-15 test group. Source: authors

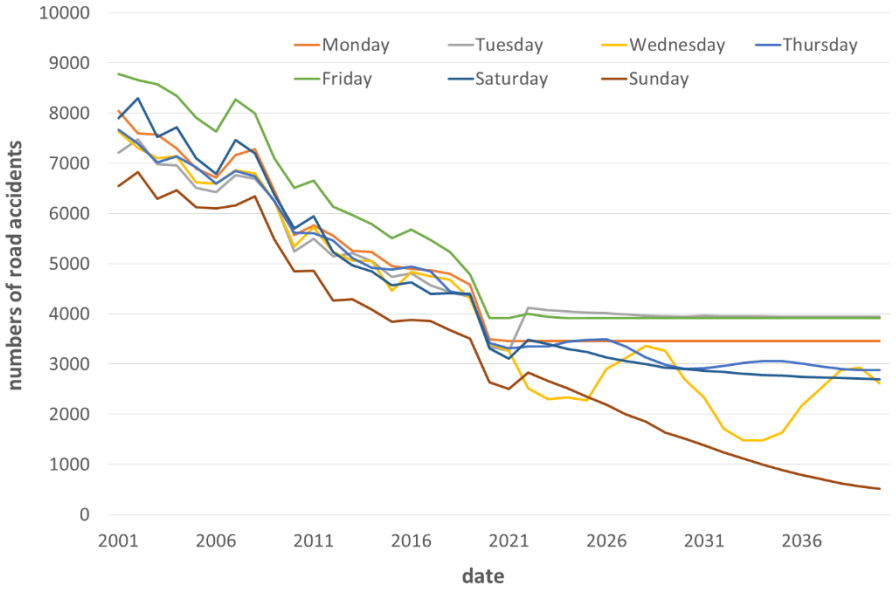


Fig. 4 Predicted number of traffic accidents for 2022-2040 for the 80-10-10 test group Source: authors

4. Conclusion

The number of accidents in Poland depending on the day of the week was predicted using the method of neural networks. The study was conducted in the Statistica environment. The weights used in the study were estimated by the program in such a way as to minimize the mean absolute error and mean absolute percentage error.

Based on the data obtained, it can be concluded that the pandemic has caused a reduction in traffic accidents, and further decrease in traffic accidents and stabilization can be expected in the coming years. However, another pandemic could change the value of the obtained results. The calculated forecast errors prove the accuracy of the models used.

Based on the obtained forecasts, measures should be taken to further reduce the number of traffic accidents. Such measures may include the imposing higher fines for traffic offenses on Polish roads from January 1, 2022. The pandemic, which has dramatically changed the number of road accidents, certainly distorted the obtained results of the study.

In further studies, the authors plan to take into account more factors affecting accident levels and use different statistical methods to determine the number of road accidents. These could include traffic volume, weather conditions or the age of the person at fault, as well as exponential methods for determining the number of traffic accidents.

References

[1] The Global status on road safety. (2018, June). The Global status on road safety 2018. World Health Organization. Retrieved June 5, 2022, from https://www.who.int/publications/i/item/9789241565684

[2] Gorzelanczyk, P., Pyszewska, D., Kalina, T. & Jurkovic, M. (2020) Analysis of road traffic safety in the Pila poviat. Scientific Journal of Silesian University of Technology. Series Transport 107, 33-52. DOI: 10.20858/sjsutst.2020.107.3.

[3] Helgason, A. (2016) Fractional integration methods and short Time series: evidence from a simulation study. Political Analysis 24(1), 59–68. DOI: 10.1093/pan/mpv026.

[4] Lavrenz, S., Vlahogianni, E., Gkritza, K. & Ke, Y. (2018) Time series modeling in traffic safety research, Accident; Analysis and Prevention 117, 368–380. DOI: 10.1016/j.aap.2017.11.030.

[5] Forecasting based on time series (2022) Forecasting based on time series (in Polish). Retrieved June 5, 2022, from http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf

[6] Prochazka, J. & Camaj, M. (2017) Modelling the number of road accidents of uninsured drivers and their severity. Proceedings of International Academic Conferences 5408040, International Institute of Social and Economic Sciences.

[7] Procházka, J., Flimmel, S., Čamaj, M. & Bašta, M. (2017) Modelling the Number of Road Accidents. Publishing house of the University of Economics in Wrocław. Wrocław 2017. DOI:10.15611/amse.2017.20.29.

[8] Sunny, C.M., Nithya, S., Sinshi, K.S., Vinodini, V.M.D., Lakshmi, A.K.G., Anjana, S. & Manojkumar, T.K. (2018) Forecasting of Road Accident in Kerala: A Case Study.

Environmental Science, Computer Science. 2018 International Conference on Data Science and Engineering (ICDSE). DOI: 10.1109/ICDSE.2018.8527825.

[9] Dudek, G. (2013) Exponential smoothing models for short-term power system load forecasting. Energy market No. 3 (106). Retrieved June 5, 2022, from Research Gate database on the World Wide Web: https://www.researchgate.net/publication/286598384_Exponential_smoothing_models_for_short-term_load_forecasting

[10] Dudek, G. (2013). Forecasting Time Series with Multiple Seasonal Cycles Using Neural Networks with Local Learning. In: Artificial Intelligence and Soft Computing. ICAISC 2013. Lecture Notes in Computer Science, 7894. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-38658-9_5.

[11] Szmuksta-Zawadzka, M. & Zawadzki, J. (2009). Forecasting on the basis of Holt-Winters models for complete and incomplete data. Wrocław, Poland: Research papers of the Wrocław University of Economics.

[12] Wójcik, A. (2014). Autoregressive vector models as a response to the critique of multi-equation structural econometric models. Katowice, Poland: Publishing house of the University of Economics in Katowice. Vol. 193.

[13] Monederoa, B.D., Gil-Alanaa, L.A. & Martínezaa, M.C.V. (2021) Road accidents in Spain: Are they persistent?. International Association of Traffic and Safety Sciences 45(3), 317-325. DOI: 10.1016/j.iatssr.2021.01.002.

[14] Mamczur, M. (2022, June) Machine learning How does linear regression work? And is it worth using? Retrieved June 5, 2022, from https://miroslawmamczur.pl/jak-dziala-regresja-Linear-i-czy-warto-ja-stosowac/

[15] Piłatowska, M. (2012) The choice of the order of autoregression depending on the parameters of the generating model. Econometrics 4(38). ISSN 1507-3866.

[16] Chudy-Laskowska, K. & Pisula, T. (2015) Forecasting the number of road accidents in Podkarpacie. Logistics 4.

[17] Chudy-Laskowska, K. & Pisula, T. (2014) Forecast of the number of road accidents in Poland. Logistics 6.

[18] Wrobel, M.S. (2017, June). Application of neural fuzzy systems in chemistry. Retrieved June 5, 2022, from https://rebus.us.edu.pl/bitstream/20.500.12128/5266/1/Wrobel_Zastosowanie_neuronowych_systemow.pdf

[19] Data mining techniques (2022, June) Retrieved June 5, 2022, from StatSoft https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html

[20] Statistic Road Accident (2022, June) Retrieved June 5, 2022, from https://statystyka.policja.pl/

[21] Gorzelańczyk, P., Jurkovič, M., Kalina, T. & Mohanty, M. (2022). Forecasting the road accident rate and the impact of the covid 19 on its frequency in the polish provinces. Communications - Scientific letters of the University of Zilina 24(4), A216-231. DOI: 10.26552/com.C.2022.4.A216-A231.