Technical Sciences, 2025, 28, 187–197

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Yearbook peer-reviewed scientific journal
ISSN 1505-4675
e-ISSN 2083-4527
TECHNICAL SCIENCES
Homepage: https://czasopisma.uwm.edu.pl/index.php/ts/

DOI: https://doi.org/10.31648/ts.11331

FORECASTING THE NUMBER OF ROAD ACCIDENTS IN POLAND USING TREND MODELS

***Piotr Gorzelanczyk*[1], *Martin Jurkovič*[2]**

[1]ORCID: 0000-0001-9662-400X
Department of Transport
Stanislaw Staszic State University of Applied Sciences in Pila
[2]ORCID: 0000-0001-7673-1350
Faculty of Operation and Economics of Transport and Communication
University of Zilina

Received 31 March 2025, accepted 13 September 2025, available online 17 September 2025.

K e y w o r d s: traffic accident; forecasting; trend models.

A b s t r a c t

Every year a very large number of people die on the roads. From year to year the value decreases, but it is still a very large number. The purpose of this article is to forecast the number of road accidents in Poland. The study was divided into two parts. The first was the analysis of annual data from police statistics on the number of road accidents in Poland in 2000-2021, and on this basis the forecast of the number of road accidents for 2022-2031 was determined. The second part of the study, dealt with monthly data from 2000-2021. Again, the analyzed forecast for the period January 2022 – December 2023 was determined.

Introduction

Road accidents are incidents that cause not only injury or death to road users, but also property damage. According to the WHO, about 1.3 million people die each year as a result of road accidents. In most countries of the world, road

Correspondence: Piotr Gorzelańczyk, Akademia Nauk Stosowanych im. Stanisława Staszica w Pile, ul. Podchorążych 10, 64-920 Piła, e-mail: piotr.gorzelanczyk@ans.pila.pl

accidents account for about 3% of their GDP. Road accidents are the leading cause of death for minors and young people aged 5-29 (*Global status report on road safety 2018* 2018). The UN General Assembly has set an ambitious goal of halving the number of deaths and injuries caused by road accidents by 2030.

The size of a traffic accident is an attribute that determines its severity. Predicting the severity of an accident is important for relevant authorities in developing road safety policies aimed at eliminating accidents, reducing injuries, fatalities and property damage (TAMBOURATZIS et al. 2014, ZHU et al. 2019). Identifying critical factors affecting accident severity is a prerequisite for adopting countermeasures to eliminate and mitigate accident severity (ARTEAGA et al. 2020). YANG et al. (2022) propose a multi-node DNN (Deep Neutral Network) framework for predicting different levels of injury severity, death and property loss. This enables a comprehensive and accurate analysis of the severity of traffic accidents (YANG et al. 2022).

There are several sources of accident data. Most often, they are collected and analyzed by government bodies through the relevant government agencies. Data collection is done through police reports, insurance company databases or hospital records. Partial information on traffic accidents is then processed on a larger scale for the transportation sector (GORZELANCZYK et al. 2020).

For accident data to be relevant, it is necessary to work with several data sources, which must be properly combined. Combining different data sources by consolidating heterogeneous traffic accident data helps increase the accuracy of analysis results (ABDULLAH, EMAM 2016).

VILACA et al. (2017) conducted a statistical study to assess severity, establishing a link between traffic accidents and traffic participants. The result of the study is a proposal to improve traffic safety standards and adopt other policies related to road safety.

BAK et al. (2019) conducted a statistical study of traffic safety in a selected region of Poland based on the number of traffic accidents, an index of accident causation research. The study used multivariate statistical analysis to investigate the safety aspects of those responsible for accidents.

The choice of the source of accident data for analysis depends on the type of traffic problem to be solved. Combining statistical models with other natural driving data or other data obtained through intelligent traffic systems contributes to the accuracy of accident forecasting and contributes to the elimination of accidents (CHAND et al. 2021).

Various methods for forecasting the number of accidents can be found in the literature. For forecasting the number of traffic accidents, time series methods are most often used (HELGASON 2016, LAVRENZ et. al. 2018), which have the disadvantage of not being able to assess the quality of the forecast based on previous forecasts and the frequent residual component of autocorrelation (FORECASTING 2022). PROCHÁZKA et al. (2017) used a multi-seasonality model for forecasting, and SUNNY et al. (2018) used the Holt-Winters exponential smoothing method. Its limitations include the inability to introduce exogenous variables into the model (Dudek 2013, Szmuksta-Zawadzka, Zawadzki 2009).

A vector autoregressive model has also been used to predict the number of traffic accidents, the drawback of which is that it requires a large number of observations of variables to correctly estimate their parameters (Wojcik 2014), as well as the autoregressive models of Monedero et al. (2021) for analyzing the number of deaths and Al-Madani (2018), regression models with curve fitting. These, on the other hand, require only simple linear relationships (Mamczur 2022) and an order of autoregression (provided the series are already stationary) (Pilatowska 2012).

Biswas et al (2019) used Random Forest regression to predict the number of traffic accidents. In this case, the data contains groups of correlated features of similar significance to the original data, smaller groups are favored over larger ones (*Las losowy* 2022) and there is instability in the method and peak prediction (Fijorek et al. 2010).

Various types of methods for forecasting the number of road accidents are encountered in the literature. Most commonly, neural networks, trend models, ARIMA, SARIMA, GARMA models are used to forecast their number. However, a comprehensive approach to road safety forecasting is lacking. Most researchers use ‘favourite’ methods to forecast the number of road accidents without considering other methods. Furthermore, an analysis of the factors influencing the number of road accidents is rarely found in the scientific literature. Often individual solutions are treated individually rather than holistically. A holistic approach should take into account not only all these elements, but also the constraints affecting road safety. Taking this into account, the authors used the trend method for forecasting.

Number of road accidents

Every year a very large number of people die on the roads. From year to year the value decreases, but there are still a very high number. The pandemic has reduced the number of road accidents, but the value is still very high. Analyzing the data on an annual and monthly basis, it can be said that there are clear fluctuations with a continuing downward trend. Compared to the European Union, the number of accidents in Poland is still very high. For this reason, every effort should be made to know the forecast of the number of accidents for the coming years (Figs. 1, 2).

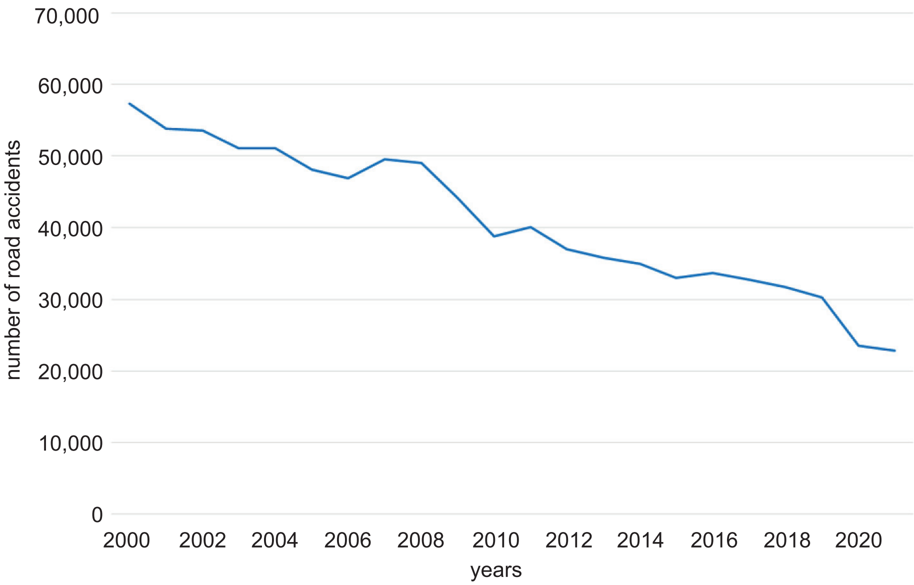

Fig. 1. Number of accidents in Poland from 2000 to 2021 by year
Source: based on Road Accident Statistics (2024).

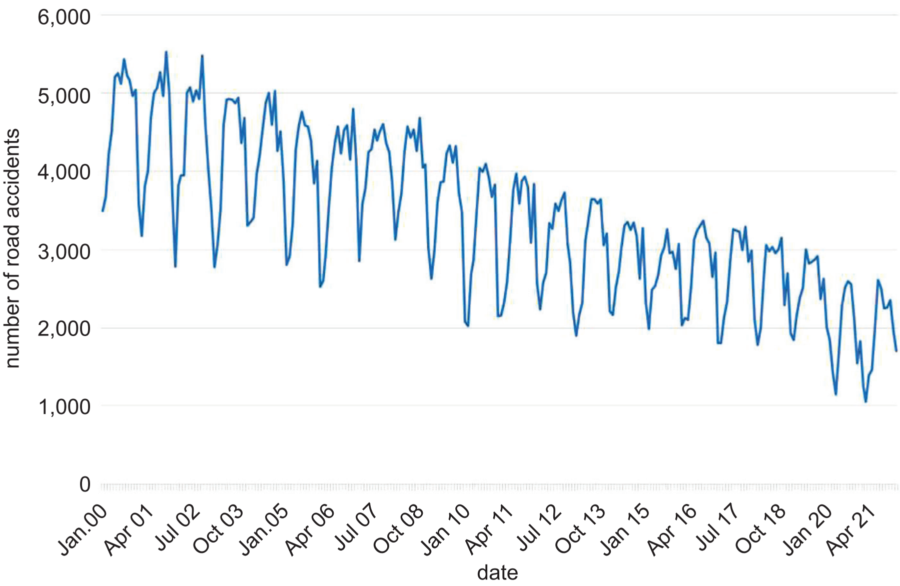

Fig. 2. Number of accidents in Poland from 2000 to 2021 by month
Source: based on Road Accident Statistics (2024).

Forecasting the number of road accidents

In the trend forecasting method, the aim is to determine an appropriate approximation, which is then extrapolated to the forecast periods. The most commonly used in this group are solutions that approximate the trend in road accident values with the following functions: linear, power, logarithmic, polynomial and exponential. The selection of a particular approximation according to the criterion adopted, e.g. the mean value of the fitting error, results from the evaluation of the fit of the results obtained on its basis to the empirical data (BATKO 1984, MARCINIAK 1982, TYLICKI 1993).

The above method gives good results when the changes in the values of the parameters are regular, which can be observed in the case of changes in the value of the number of road accidents. Moreover, in this case, the assumed development trend is assumed to be maintained over the forecast periods. For this reason, various modifications of these solutions can be encountered, e.g. by limiting the trend function estimation process to a narrowed data set (by omitting the oldest observations) or by including at each time point after a fixed time tb, the actual value of the trend to build its approximation. Limitations of the method discussed above include:

– requires access to historical data to effectively forecast future trends. Inappropriate or incomplete data can lead to erroneous forecasts;

– may be less effective in the case of dynamic changes in the market environment, such as economic crises or technological changes;

– sample size limitations may affect the reliability of the results;

– the method is based on the assumption that trends are linear. In reality, they may be non-linear, leading to distortions;

– it requires regular data updates to ensure that forecasts remain valid and accurate.

The trend forecasting method can be used for:

– long-term planning in organisations, helping to predict future needs;

– analysing market trends, which can help companies adjust marketing and sales strategies;

– forecasting revenue and expenditure, which supports budgeting processes;

– improve inventory management by anticipating future demand;

– analyse data and identify emerging trends.

The following trend models were used in forecasting the number of traffic accidents:

– exponential,

– linear,

– logarithmic,

– polynomial of 2nd degree,

– polynomial of 3rd degree,

– polynomial of 4th degree,
– polynomial of 5th degree,
– polynomial of 6th degree,
– potentiometric.

In the first step, the mathematical formula of the analyzed trend models was determined for the analyzed data on an annual and monthly basis. As can be seen, the R-square coefficient, which is a measure of the quality of model fit for annual data, occurs in most cases of good fit except for the model for power, where there is a satisfactory fit. Worse behavior is observed in models for monthly data. In this case, the fit is poor or satisfactory. This is mainly due to the seasonality of the number of traffic accidents (Tab. 1).

Table 1

Trend models

Data / model	Annual data	Monthly data
Exponential	$y = 62{,}101e^{-0.039x}$	$y = 4{,}973.7e^{-0.003x}$
	$R^2 = 0.9357$	$R^2 = 0.5795$
Linear	$y = -1{,}522.6x + 58{,}377$	$y = -10.331x + 4{,}774.4$
	$R^2 = 0.9633$	$R^2 = 0.6129$
Logarithmic	$y = -110{,}68 \ln(x) + 65.252$	$y = -693.4 \ln(x) + 6{,}588.4$
	$R^2 = 0.8159$	$R^2 = 0.4386$
Polynomial of 2nd degree	$y = -9.1172x^2 - 1{,}313x + 57{,}538$	$y = -0.0069x^2 - 8.5113x + 4{,}693.7$
	$R^2 = 0.9644$	$R^2 = 0.6141$
Polynomial of 3rd degree	$y = -0.5932x^3 + 11.347x^2 - 1{,}505.4x + 57{,}948$	$y = 5E\text{-}05x^3 - 0.027x^2 - 6.3755x + 4{,}646.1$
	$R^2 = 0.9645$	$R^2 = 0.6144$
Polynomial of 4th degree	$y = -0.4175x^4 + 18.611x^3 - 277.01x^2 + 47.322x + 55{,}808$	$y = -2E\text{-}06x^4 + 0.0012x^3 - 0.2201x^2 + 5.0451x + 4{,}492.2$
	$R^2 = 0.9666$	$R^2 = 0.6168$
Polynomial of 5th degree	$y = -0.1819x^5 + 10.041x^4 - 197.83x^3 + 1{,}657.7x^2 - 6{,}910.8x + 62{,}800$	$y = -3E\text{-}08x^5 + 2E\text{-}05x^4 - 0.004x^3 + 0.296x^2 - 14.641x + 4{,}670.7$
	$R^2 = 0.9782$	$R^2 = 0.6194$
Polynomial of 6th degree	$y = 0.0072x^6 - 0.6774x^5 + 23.127x^4 - 362.92x^3 + 2667.2x^2 - 9{,}572.3x + 64{,}908$	$y = 5E\text{-}11x^6 - 7E\text{-}08x^5 + 3E\text{-}05x^4 - 0.0057x^3 + 0.4125x^2 - 17.763x + 4{,}691.1$
	$R^2 = 0.9787$	$R^2 = 0.6194$
Potentiometric	$y = 72{,}375x^{-0.274}$	$y = 8{,}460.4x^{-0.209}$
	$R^2 = 0.7341$	$R^2 = 0.386$

Source: own data.

Then, using the data in Table 1, the projected number of traffic accidents was determined. For annual data it was 2022-2031, while for monthly data it was January 2022 – December 2023. The forecast in this case was based on a weighted average of current data and historical series values. The result of the forecast using this method depends on the choice of model and its fit (Figs. 3, 4).

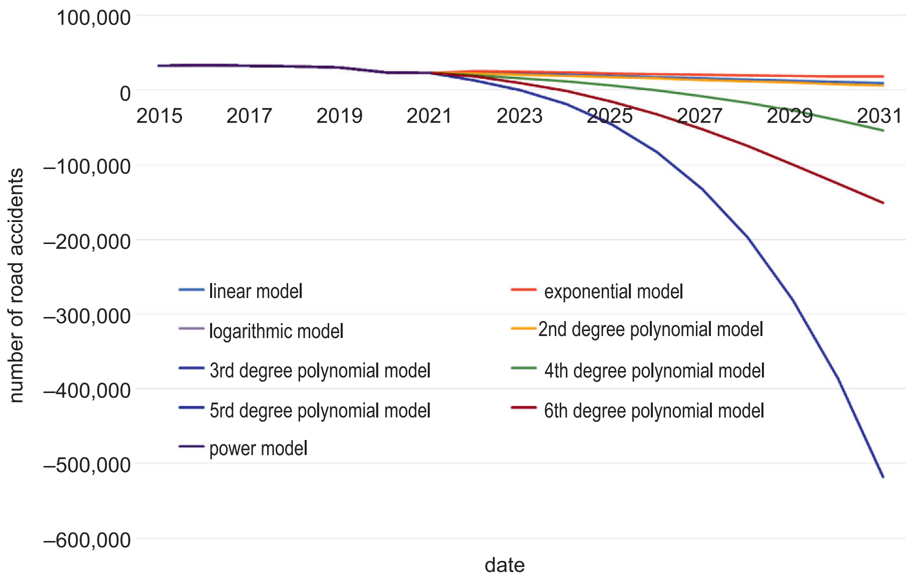

Fig. 3. Projected number of road accidents in 2022-2031
Source: own data.

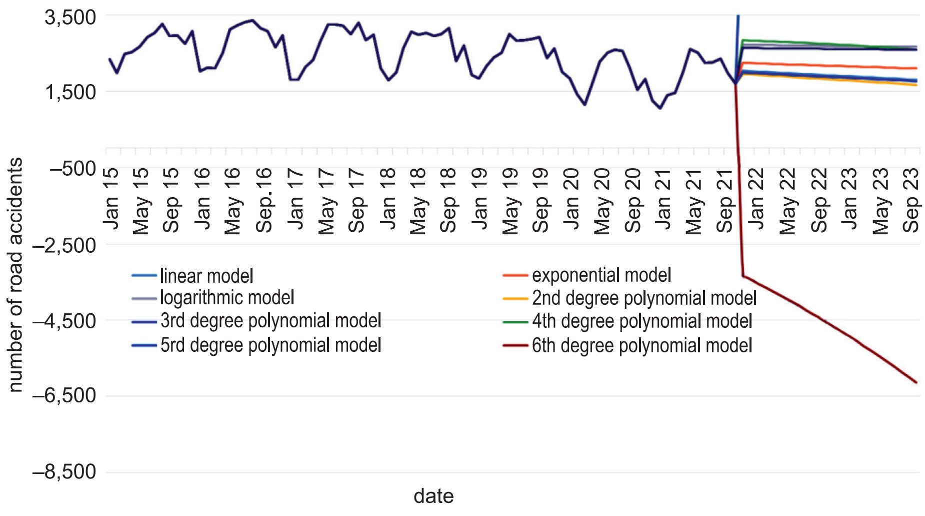

Fig. 4. Projected number of road accidents in 2022-2023
Source: own data.

Figures 4 and 5 show that it is possible to use all the adopted trend models to forecast the number of traffic accidents. For this reason, the following lagged forecast errors determined from equations (1-5) were used to calculate measures of analytical excellence of the forecasts:

– ME – mean error

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

– MAE – mean average 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

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

– MSE – mean square error

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

where:

n – the length of the forecast horizon,

Y – observed value of road accidents,

Y_p – forecasted value of road accidents.

For forecasting the number of traffic accidents, trend models were selected for which the mean percentage error and mean absolute percentage error were the smallest. On this basis, it was found that for the annual data, the best fit was the 2nd degree polynomial model, for which the MAPE error was 0.27%. On the other hand, for the data, the 3rd degree polynomial model, for which the MAPE error was 4.19%. On this basis, the projected number of accidents for the following years was determined on a monthly and annual basis (Figs. 5, 6). Based on Figures 4 and 5, it can be expected that the number of traffic accidents will continue to decline in the coming years. It should be noted that the pandemic has caused significant changes in the forecasts.

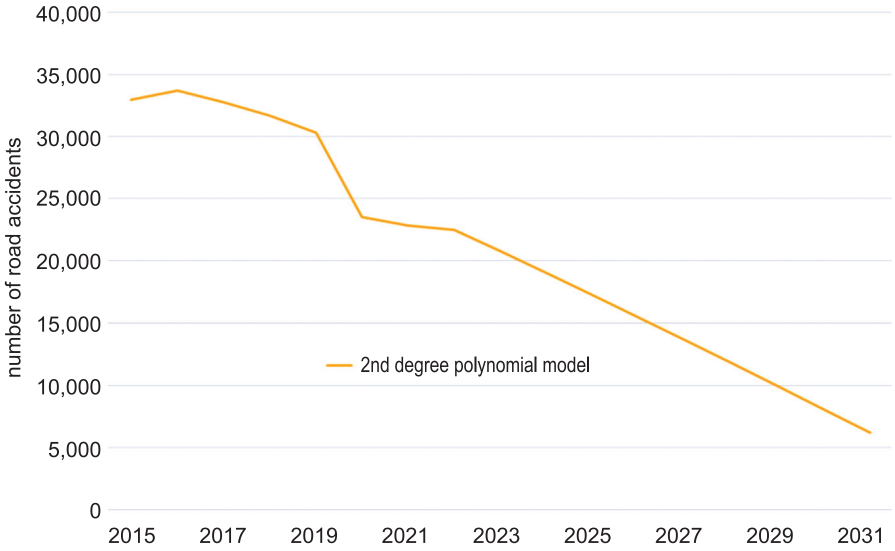

Fig. 5. Projected number of road accidents for 2022-2031
Source: own data.

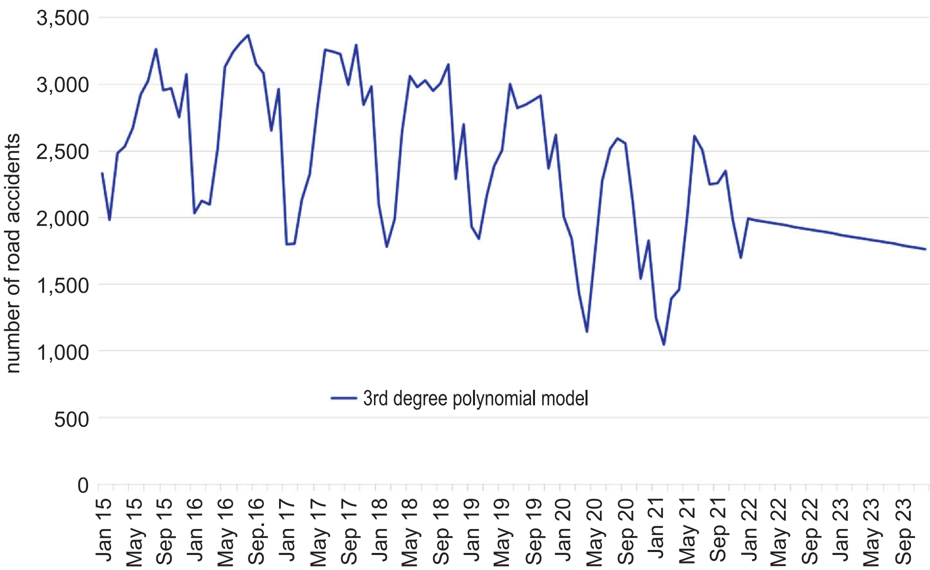

Fig. 6. Projected number of road accidents for 2022-2023
Source: own data.

Conclusions

The forecast of the number of accidents in Poland was determined based on selected trend models using Excel. The results show that we can still expect a decrease in the number of traffic accidents in the coming years. It should be noted that the pandemic skewed the results obtained, and in the event of its continuation and the introduction of traffic restrictions, the proposed model may prove inadequate. The error value of up to 4% may indicate the choice of an effective forecasting method. As can be seen, trend models do not perform well in forecasting the monthly number of traffic accidents with seasonality. In contrast, for annual data, the results are satisfactory.

The forecasted number of traffic accidents obtained in the article can be used in the future to formulate further measures to minimize the number of accidents in the analyzed country. These measures may include, for example, the introduction of higher penalties for traffic offenses on Polish roads from January 1, 2022.

In his further research, the author plans to take into account more factors influencing the number of accidents in Poland and use other methods to forecast the number of road accidents. We can take into account traffic volume, the day of the week or the age of the accident perpetrator, among others.

References

Abdullah E., Emam A. 2016. *Traffic accidents analyzer using big data*. 2015 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, p. 392-397. https://doi.org/10.1109/CSCI.2015.187

Al-Madani H. 2018. *Global road fatality trends' estimations based on country-wise microlevel data*. Accident Analysis & Prevention, 111: 297-310. https://doi.org/10.1016/j.aap.2017.11.035

Arteaga C., Paz A., Park J. 2020. *Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach*. Safety Science, 132, art. no 104988, https://doi.org/10.1016/j.ssci.2020.104988

Bąk I., Cheba K., Szczecińska B. 2019. *The statistical analysis of road traffic in cities of Poland*. Transportation Research Procedia, 39: 14-23. https://doi.org/10.1016/j.trpro.2019.06.003

Batko W. 1984. *Methods of synthesis of predictive diagnoses in technical diagnostics*. Habilitation dissertation. Akademia Górniczo-Hutnicza, Kraków.

Biswas A.A., Mia J., Majumder A. 2019. *Forecasting the number of road accidents and casualties using random forest regression in the context of Bangladesh*. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). http://dx.doi.org/10.1109/ICCCNT45670.2019.8944500

Chand A., Jayesh S., Bhasi A.B. 2021. *Road traffic accidents: An overview of data sources, analysis techniques and contributing factors*. Materials Today: Proceedings, 47(15): 5135-5141. https://doi.org/10.1016/j.matpr.2021.05.415

Dudek G. 2013. *Forecasting time series with multiple seasonal cycles using neural networks with local learning*. In: *Artificial Intelligence and Soft Computing*. Eds. L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L.A. Zadeh, J.M. Zurada. ICAISC 2013. Lecture Notes in Computer Science, 7894. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38658-9_5

FIJOREK K., MRÓZ K., NIEDZIELA K., FIJOREK D. 2010. *Forecasting hourly electricity prices in polish day-ahead market: Data mining approach*. Rynek Energii, 91(6): 46-50.

Global status report on road safety 2018. 2018. World Health Organization. Retrieved from https://www.who.int/publications/i/item/9789241565684

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. Transport. 107: 33-52. https://doi.org/10.20858/sjsutst.2020.107.3

HELGASON A. 2016. *Fractional integration methods and short Time series: evidence from asimulation study*. Political Analysis, 24(1): 59-68. Retrieved from http://www.jstor.org/stable/24573204

KOWALSKI L. 2005. *Prognozowanie na podstawie szeregów czasowych*. Prognozowanie i Symulacje. Retrieved from http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf

Las losowy. 2022. Wikipedia. Wolna encyklopedia. Retrieved from https://pl.wikipedia.org/wiki/Las_losowy

LAVRENZ S., VLAHOGIANNI E., GKRITZA K., KE Y. 2018. *Time series modeling in traffic safety research*. Accident Analysis & Prevention, 117: 368-380. https://doi.org/10.1016/j.aap.2017.11.030

MAMCZUR M. 2022. *Jak działa regresja liniowa? I czy warto ją stosować?* Blog o data science, AI, uczeniu maszynowym i wizualizacji danych. https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/

MARCINIAK J. 1982. *Diagnostyka techniczna kolejowych pojazdów mechanicznych*. WKiŁ, Warszawa.

MONEDEROA B.D., GIL-ALANAA L.A., MARTÍNEZAA M.C.V. 2021. *Road accidents in Spain: Are they persistent?* IATSS Research, 45(3): 317-325. https://doi.org/10.1016/j.iatssr.2021.01.002

PIŁATOWSKA M. 2012. *The choice of the order of autoregression depending on the parameters of the generating model*. Econometrics, 4(38).

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.

Road Accident Statistics. 2024. Retrieved from https://statystyka.policja.pl/

SUNNY C.M., NITHYA S., SINSHI K.S., VINODINI V.M.D. LAKSHMI A.K.G., ANJANA S., MANOJKUMAR T.K. 2019. *Forecasting of road accident in Kerala: A case study*. Environmental Science, Computer Science. 2018 International Conference on Data Science and Engineering (ICDSE). https://doi.org/10.1109/ICDSE.2018.8527825

SZMUKSTA-ZAWADZKA M., ZAWADZKI J. 2009. *Forecasting on the basis of Holt-Winters models for complete and incomplete data*. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 38: 85-99.

TAMBOURATZIS T., SOULIOU D., CHALIKIAS M., GREGORIADES A. 2014. *Maximising accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees*. Journal of Artificial Intelligence and Soft Computing Research, 4(1): 31-42. http://dx.doi.org/10.2478/jaiscr-2014-0023

TYLICKI H. 1993. *Optimisation of methods of forecasting reliable operation of equipment*. Zagadnienia Eksploatacji Maszyn, 3.

VILAÇA M., SILVA N., COELHO M.C. 2017. *Statistical analysis of the occurrence and severity of crashes involving vulnerable road users*. Transportation Research Procedia, 27: 1113-1120. https://doi.org/10.1016/j.trpro.2017.12.113

WÓJCIK A. 2014. A*utoregressive vector models as a response to the critique of multi-equation structural econometric models*. Publishing house of the University of Economics, Katowice.

YANG Z. ZHANG W., FENG J. 2022. *Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework*. Safety Science, 146: 105522. https://doi.org/10.1016/j.ssci.2021.105522

ZHU L., LU L., ZHANG W., ZHAO Y., SONG M. 2019. *Analysis of accident severity for curved roadways based on Bayesian networks*. Sustainability, 11(8): 2223. https://doi.org/10.3390/su11082223