https://doi.org/10.14669/AM/182808

ESTIMATING THE NUMBER OF ACCIDENTS ON POLAND'S ROADS BASED ON THE KIND OF ROAD

PIOTR GORZELAŃCZYK[1]

Abstract

On Polish highways, a staggering number of individuals pass away each year. The quantity is still quite large even if the value is declining year after year. The value of traffic accidents has greatly decreased due to the epidemic, but it is still quite high. In order to reduce this number, it is required to identify the roads where the majority of accidents occur and to understand the predicted number of accidents in the upcoming years.
The article's goal is to predict how many accidents will occur on Polish roads based on the kind of roads. In order to achieve this, yearly data from the Police's statistics for the years 2001–2021 on the number of accidents in Poland were analyzed, and a prediction for the years 2022–2031 was generated. As is evident, either the number of accidents is rising or it is stabilizing.
Additionally, predictions indicate that given the existing circumstances, a significant rise in the number of accidents on Polish roads may be anticipated. This is especially evident on the nation's growing number of freeways. It should be remembered that the current epidemic distorts the findings.

Keywords: road accident; road type; forecasting; exponential smoothing

1. Introduction

Road accidents are incidents that cause property damage as well as harm or fatality to other drivers. According to the WHO, 1.3 million people die in automobile accidents every year. The average country in the globe has a 3% GDP loss as a result of road accidents. Traffic accidents are the leading cause of death for children and young people between the ages of 5 and 29 [32]. The UN General Assembly desires to see a reduction in traffic accident fatalities and injuries by half by the year 2030.

[1] Transport, Stanisław Staszic State University of Applied Sciences in Piła, Poland,
e-mail: piotr.gorzelanczyk@pwsz.pila.pl, ORCID: 0000-0001-9662-400X

https://doi.org/10.14669/AM/182808

The scope of a traffic collision is a consideration in determining its seriousness. It is essential to estimate accident severity in order for the competent authorities to create traffic safety regulations with the goal of preventing accidents, limiting injuries, fatalities, and property damage [30, 36]. The identification of key factors that affect accident severity is necessary before adopting countermeasures to prevent and lessen accident severity [2]. Yang et al. provide a multi-node Deep Neutral Network (DNN) topology for predicting different levels of harm, death, and property loss. It enables a complete and accurate examination of the severity of traffic accidents [34].

There are several sources for the accident statistics. They are typically gathered and assessed by government representatives utilizing the relevant government agencies. Data is gathered using a variety of sources, including hospital records, insurance company databases, and police reports. The transportation sector is subsequently analyzing data on road accidents on a larger scale [12].

Intelligent transportation systems are currently the most important information source for the study and prediction of traffic incidents. These data may be processed using GPS devices installed on moving vehicles [6]. Roadside microwave vehicle detection systems may record data about moving cars continuously, including speed, volume of traffic, and vehicle type [17]. A license plate recognition system may also be used to collect a lot of traffic data over a controlled period of time [26]. Another potential resource for information on traffic and accidents is social media, albeit due to the inexperience of the reporters, their accuracy may not be adequate [35].

Vilaca et al. [31] conducted a statistical study to assess the gravity of the situation and establish a link between traffic participants and accidents. The study's conclusion recommends increasing the standards for traffic safety laws and putting in place more traffic safety measures.

Bak et al. [3] carried out a statistical analysis of traffic safety in a selected Polish area based on the number of traffic accidents, which is a gauge of the research into accident causes. The study used multivariate statistical analysis to look at the safety factors of those who cause accidents.

The source of accident data to be utilized for analysis depends on the type of traffic issue being addressed. Combining statistical models with extra data from actual driving or other information obtained from intelligent traffic systems increases the accuracy of accident prediction and accident eradication [5].

There are several methods for forecasting the frequency of accidents in the literature. Time series approaches [14, 19] are the most often used methods for estimating the frequency of accidents, but they have the disadvantage of not being able to assess the forecast's accuracy

https://doi.org/10.14669/AM/182808

based on past forecasts and the frequent residual component of autocorrelation. Procházka et al. [24] utilized a multi-seasonality model for forecasting.

The curve-fitting regression models of Al-Madani [1] and Monedero et al. for analyzing the number of fatalities [22], as well as the vector autoregressive model, which has the drawback of requiring many observations of variables to accurately estimate their parameters [33], have all been used to predict the frequency of traffic accidents.

Chudy-Laskowska and Pisula [8] used an autoregressive quadratic trend model, a univariate periodic trend model, and an exponential smoothing model for the presented forecasting issue. A moving average model might also be used to anticipate the issue at hand, but this method suffers from low prediction accuracy, data loss within a sequence, and the inability to account for trends and seasonal changes [16].

Prochozka and Camej [24] used the GARMA technique, which limits the parameter space, to guarantee that the process is stationary. For stationary processes, forecasting often use the ARMA model [25]; for non-stationary processes, it employs the ARIMA or SARIMA model [10, 15]. The advantage of these models is that they provide the examined models a great degree of flexibility, but the disadvantage is that they need a deeper level of research expertise from the researcher than, example, regression analysis [20].

Chudy-Laskowska and Pisula's study [7] used an ANOVA to predict the incidence of traffic accidents. The disadvantage of this approach is that it contains additional presumptions, most notably the presupposition of sphericity, whose failure could lead to inaccurate conclusions [13]. With the use of neural network techniques, the number of auto accidents is also predicted. The disadvantages of neural networks include the need for prior expertise in the field [7], the dependence of the final result on the initial conditions of the network, and the inability to interpret the results conventionally, as neural networks are often referred to as 'black boxes' in which the input is input and the model outputs the results without awareness of the analysis [9].

Using the Hadoop model as a cutting-edge prediction method was done by Kumar et al. [18]. Karlaftis and Vlahogianni [15] used the Garch model to make predictions. An issue with this strategy is its detailed model and complex shape [11, 23]. However, the ADF test [22], which was used by McIlroy [21] and his team, has the issue of not having enough power to identify the autocorrelation of the random component.

The writers of publications have also used data-mining algorithms for prediction, which frequently have the problem of having vast collections of general descriptions [28]. Another illustration of a combination of models is the set of models published by Sebe et al. [27]. Parametric models are also suggested by Bloomfield's work [4].

https://doi.org/10.14669/AM/182808

Based on the above data, the author made forecasts of the number of accidents on Polish roads. Selected exponential models were used to forecast the number of accidents.

2. Materials and methods

There are more and more new vehicles on Polish roads. In Poland, there are currently close to 750 automobiles per 1000 people. This results in either a rise in or a stabilization in the number of traffic accidents (Table 1). Road traffic accident is defined as an incident involving a moving vehicle on the road that results in fatalities or serious injuries, damage to moving objects like equipment or products, or other types of material harm. According to Police data [29], there are the following types of roads in Poland:

- motorway,
- expressway,
- 2 single carriageways,
- single carriageway,
- 1 carriageway 2 directions.

Selected time series models were used in the investigation in Statistica software.

Tab. 1 Number of accidents in Poland from 2007 to 2021, broken down by kind of road [29]

Date/type of road	Motorway	Expressway	2 one-way carriageway	Single carriageway	1 carriageway 2 directions
2007	3699	1468	55379	21427	324332
2008	3873	1371	56257	23507	326965
2009	3828	1508	57539	23063	329335
2010	4253	1845	68557	26827	353425
2011	3902	1692	63292	24645	313109
2012	4454	1725	58536	23175	288645
2013	5443	2444	60769	24671	298432
2014	6369	3114	62871	25364	285259
2015	7011	3751	64476	27481	292457
2016	8646	4739	71175	28791	326927
2017	9836	5882	71568	32326	349604
2018	9775	6542	72656	33386	345698
2019	10559	8082	73960	35680	357460
2020	8632	7856	59077	29680	300413
2021	10593	9999	65029	35254	324560

https://doi.org/10.14669/AM/182808

The Kruskalla–Wallis test was used to analyze how the kind of road affected the number of accidents on the road. The test statistic has a value of 68 and a test probability of p = 0.000. The resultant figure suggests that the assumption of an equal mean level of traffic accidents should be discarded (Figure 1).

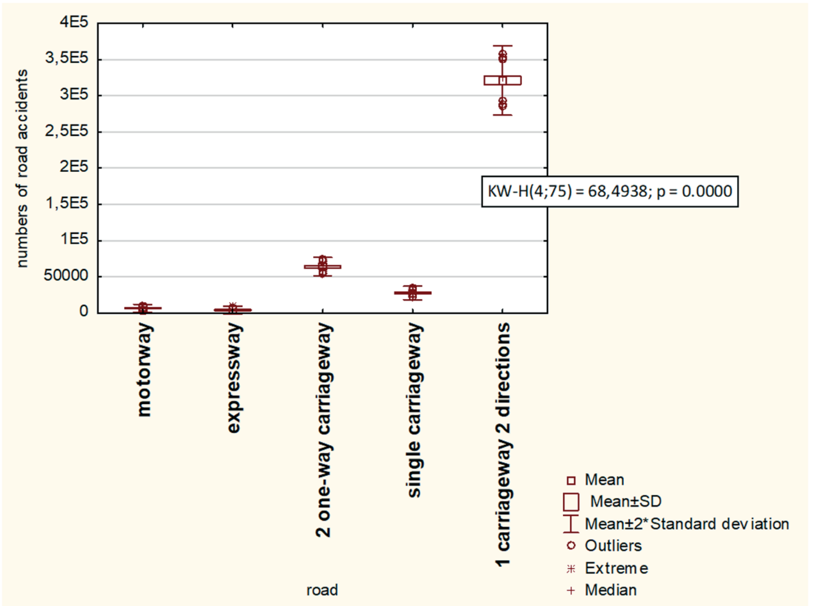

Fig. 1. From 2007 to 2021, the average number of accidents on Poland's roads was compared by kind of road

3. Estimating the number of accidents on the road

For each type of road, the number of accidents was predicted using specific exponential equalization models. The fundamental difference between both approaches is that the weights are chosen using an exponential function, and the time series of the forecast variable is given by a weighted moving average. The Statistica program used to conduct the applied analysis chose these weights in the best possible way.

A weighted average of the recent and historical records was used to predict the number of accidents for each kind of route under study. The model used and the best values for its parameters determine the outcomes of forecasts made using these approaches. Using certain time series models, predictions of the number of accidents on Polish roads by kind of road were created.

https://doi.org/10.14669/AM/182808

Measures of analytical forecasting perfection were calculated using the errors of forecasts that had expired, which were calculated using equations (1–5):

- ME – mean error:

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

- MAE –mean average 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:

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

- MSE – mean square error:

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

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

The mean absolute percentage error was reduced to compare the number of accidents during a pandemic and if it didn't.

Annual Police statistics from 2007 to 2021 were used to anticipate the number of accidents on the route categories under study. In Figures 2 to 6, the forecasted outcomes for the roads are displayed. The various forecasting techniques utilized in the study are denoted by the codes M1, M2,..., Mn. The following are the forecasting methods applied in the study:
M1 – moving average method 2-points,
M2 – moving average method 3-points,
M3 – moving average method 4-points,
M4 – exponential smoothing no trend seasonal component: none,
M5 – exponential smoothing no trend seasonal component: additive,
M6 – exponential smoothing no trend seasonal component: multiplicative,
M7 – exponential smoothing linear trend seasonal component: none HOLTA,
M8 – exponential smoothing linear trend seasonal component: additive,
M9 – exponential smoothing linear trend seasonal component: multiplicative WINTERSA,
M10 – exponential smoothing exponential seasonal component: none,
M11 – exponential smoothing exponential seasonal component: additive,
M12 – exponential smoothing exponential seasonal component: multiplicative,

https://doi.org/10.14669/AM/182808

M13 – exponential smoothing fading trend seasonal component: none,
M14 – exponential smoothing fading trend seasonal component: additive,
M15 – exponential smoothing fading trend seasonal component: multiplicative.

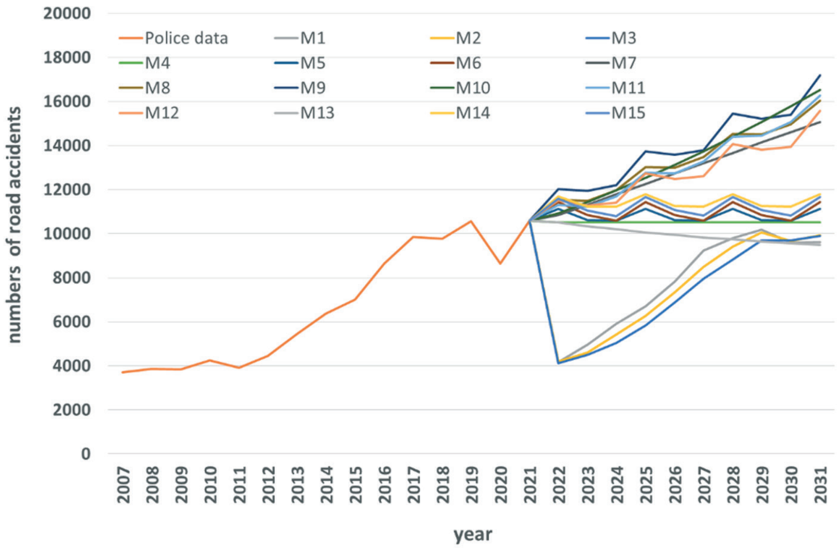

Fig. 2. Forecasting the number of road accidents for the motorway between 2022 and 2031

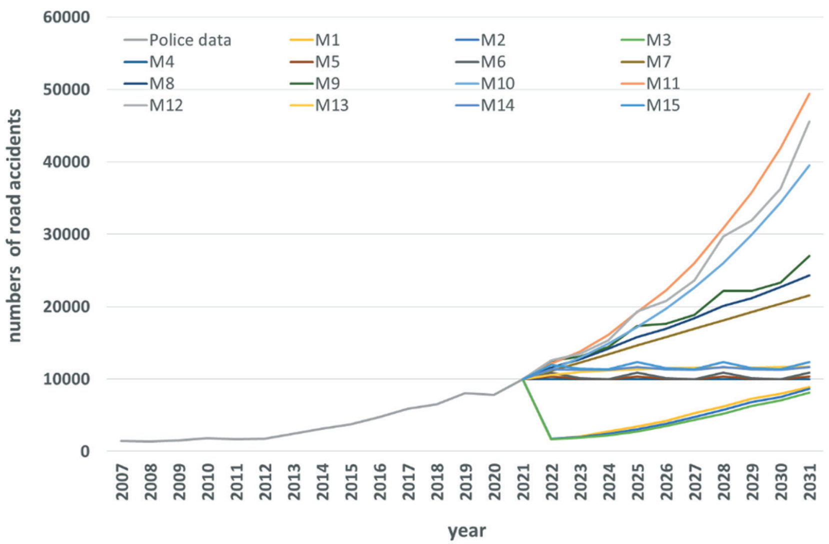

Fig. 3. Forecasting of the number of road accidents for the express road between 2022 and 2031

https://doi.org/10.14669/AM/182808

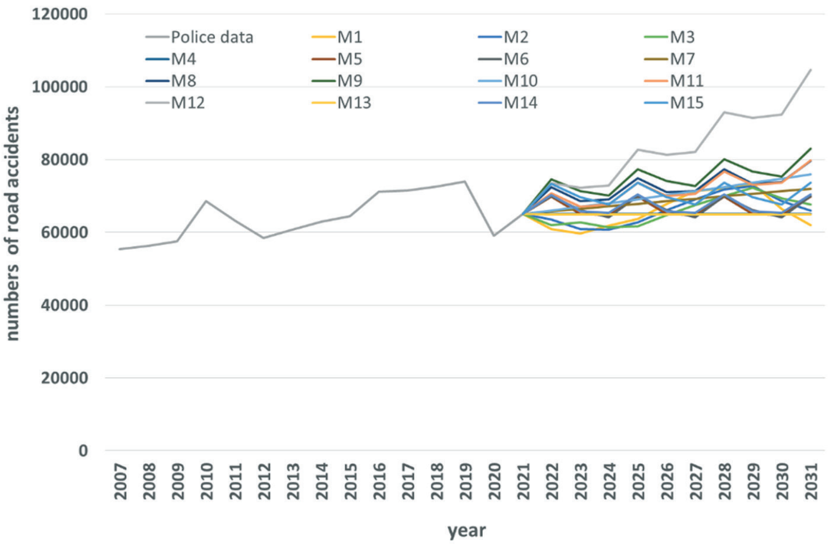

Fig. 4. Forecasting the number of road accidents for a 2 carriageway road from 2022 to 2031

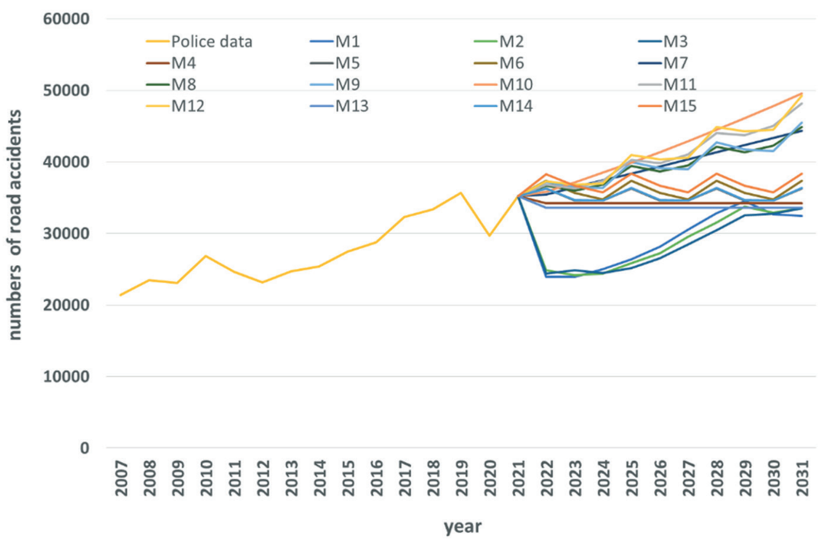

Fig. 5. Forecasting the number of road accidents for a one-way road from 2022 to 2031

https://doi.org/10.14669/AM/182808

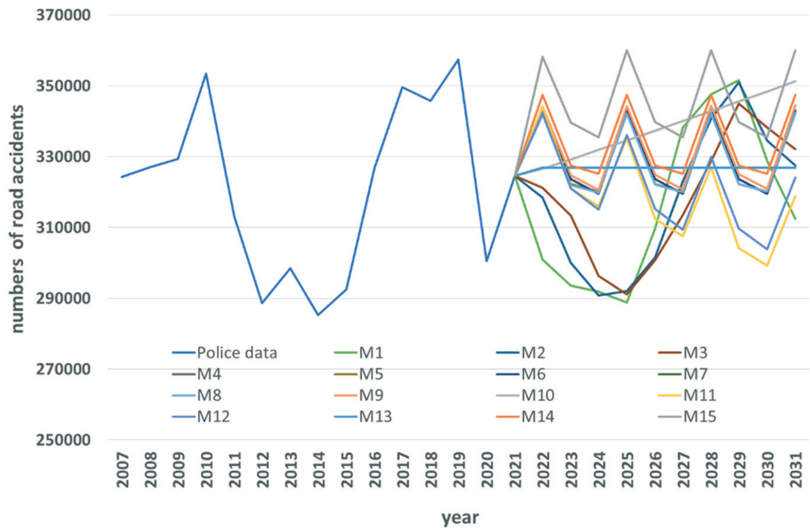

Fig. 6. Projected number of road accidents for road 1 carriageway 2 directions in 2022–2031

The aforementioned information leads to the conclusion that not all approaches utilized in the case study are successful. The best forecasting techniques for each road were determined to be the following:

- motorway – M13,
- expressway – M8,
- 2 single carriageway road – M9,
- single carriageway – M12,
- 1 carriageway 2 directions – M14.

The data acquired allow it to be determined that the strategy chosen depends on the type of route being studied. The linear trend and fading trend techniques consistently produced the smallest MAPE error. Based on this, a forecast of the number of traffic accidents according to the types of roads that were analyzed was made, as shown in Figure 7. The results of the forecast errors are reported in Table 2. The findings indicate that there will likely be more accidents in the future. On highways, this is particularly important. The results have been altered as a result of the epidemic, it should be highlighted. The selection of an efficient forecasting method is evidenced by an error value of 1% at most, with the exception of highways at 3.5%.

https://doi.org/10.14669/AM/182808

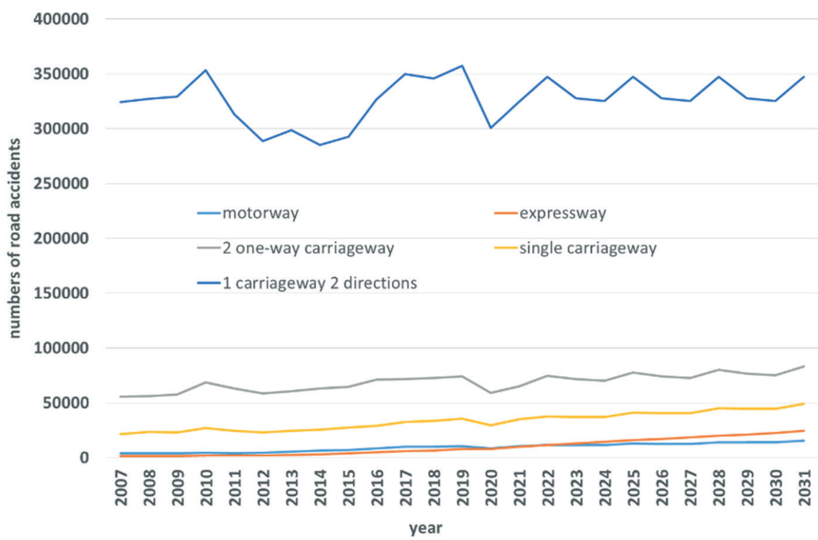

Fig. 7. Optimal projected number of road accidents according to the analysed road types in the period 2022–2031

Tab. 2. Forecasting errors for the best forecasting methods

Road type/forecast error	ME	MPE	MSE	MAPE [%]	MAE [%]
motorway	296.02	678.91	781259.30	3.56	9.39
expressway	99.86	410.92	406005.50	0.69	10.22
2 single carriageway road	471.47	3832.82	23170378.00	0.31	5.98
single carriageway	8.33	1922.17	5491637.00	0.76	6.86
1 carriageway 2 directions	0.41	16.51	410.13	0.03	5.19

4. Conclusions

With the help of the Statistica application, an exponential equalization approach was used to anticipate the number of accidents in Poland. In order to reduce the mean absolute error and the mean absolute percentage error, the algorithm determined the weights that would be most effective.

The study's findings suggest that the number of traffic accidents would likely be around the same as they were prior to the epidemic. The exception is the number of highways, which is growing along with the amount of accidents that occur on them. The results have been altered as a result of the epidemic, it should be highlighted. The selection of an efficient forecasting technique may be demonstrated by an error value of no more than 3.55%.

https://doi.org/10.14669/AM/182808

Future methods to reduce the number of accidents on roads, particularly on motorways, can be developed using the anticipated number of road accidents obtained in this article. These actions may, for instance, begin on January 1, 2022, with the imposition of harsher fines for traffic infractions on Polish roads.

In his next research, the author intends to take into consideration other variables affecting accident rates in Poland. Examples of these can be the amount of traffic, the day of the week, or the age of the accident's perpetrator. In addition, the author plans to analyse the impact of the number of vehicles, traffic volume and vehicle type on the number of road accidents in the following studies

5. References

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

[2] Arteaga C, Paz A, Park J. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. Safety Science. 2020;32:104988. https://doi.org/10.1016/j.ssci.2020.104988.

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

[4] Bloomfield P. An exponential model in the spectrum of a scalar time series. Biometrika. 1973;60(2):217–226. https://doi.org/10.1093/biomet/60.2.217.

[5] Chand A, Jayesh S, Bhasi AB. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. Materials Today: Proceedings. 2021;7(15):5135–5141. https://doi.org/10.1016/j.matpr.2021.05.415.

[6] Chen C. Analysis and forecast of traffic accident big data. ITM Web Conference. 2017;12:04029. https://doi.org/10.1051/itmconf/20171204029.

[7] Chudy-Laskowska K, Pisula T. Forecasting the number of road accidents in Subcarpathia. (Prognozowanie liczby wypadków drogowych na Podkarpaciu. In Polish). Logistics. 2015;4:2782–2796.

[8] Chudy-Laskowska K, Pisula T. Forecastof Road Accidents in Poland. (Prognoza liczby wypadków drogowych w Polsce. In Polish). Logistics. 2014;6:2710–2721.

[9] Data mining techniques StatSoft: https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html (accessed on 23.05.2023).

[10] Dutta B, Barman MP, Patowary AN. Application of Arima model for forecasting road accident deaths in India. International Journal of Agricultural and Statistical Sciences. 2020;16(2):607–615.

[11] Fiszeder P. GARCH class models in empirical financial research. (Modele klasy GARCH w empirycznych badaniach finansowych. In Polish). Scientific Publishers of the Nicolaus Copernicus University. Torun, 2009.

[12] Gorzelańczyk P. Change in the Mobility of Polish Residents during the Covid-19 Pandemic. Communications - Scientific letters of the University of Zilina. 2022;24(3):A100–111. https://doi.org/10.26552/com.C.2022.3.A100-A111.

[13] Gregorczyk A, Swarcewicz M. Analysis of variance in a repeated measures design to determine the effects of factors affecting linuron residues in soil. Polish Journal of Agronomy. 2012;11:15–20.

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

https://doi.org/10.14669/AM/182808

[15] Karlaftis M, Vlahogianni E. Memory properties and fractional integration in transportation time -series. Transportation Research Part C: Emerging Technologies. 2009;7(4):444–453. https://doi.org/10.1016/j.trc.2009.03.001.

[16] Kashpruk N. Comparative research of statistical models and soft computing for identification of time series and forecasting. (Badania porównawcze modeli statystycznych i obliczeń miękkich dla identyfikacji ciągów czasowych i prognoz. In Polish). Opole University of Technology. Opole, 2010.

[17] Khaliq KA, Chughtai O, Shahwani A, Qayyum A, Pannek J. Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing. Electronics. 2019;8(8):896. https://doi.org/10.3390/electronics8080896.

[18] Kumar S, Viswanadham V, Bharathi B. Analysis of road accident. IOP Conference Series Materials Science and Engineering. 2019;590(1):012029. https://doi.org/10.1088/1757-899X/590/1/012029.

[19] Lavrenz S, Vlahogianni E, Gkritza K, Ke Y. Time series modeling in traffic safety research. Accident Analysis & Prevention. 2018;117:368–380. https://doi.org/10.1016/j.aap.2017.11.030.

[20] Łobejko S. Time series analysis and forecasting with SAS. (Analiza i prognozowanie szeregów czasowych z wykorzystaniem narzędzi SAS. In Polish). Main business school in Warsaw. Warsaw, 2015.

[21] McIlroy RC, Plant KA, Hoque MS, Wu J, Kokwaro GO, Nam VH, et al. Who is responsible for global road safety? A cross-cultural comparison of actor maps. Accident Analysis & Prevention. 2019;122:8–18. https://doi.org/10.1016/j.aap.2018.09.011.

[22] Monedero DB, Gil-Alana L, Martínez MCV. Road accidents in Spain: Are they persistent? IATSS Research. 2021;45(3):317–325. https://doi.org/10.1016/j.iatssr.2021.01.002.

[23] Perczak G, Fiszeder P. The GARCH Model – the Application of Additional Information about Low and High Prices (Model GARCH – wykorzystanie dodatkowych informacji o cenach minimalnych i maksymalnych. In Polish). Bank and Credit. 2014;45(2):105–132.

[24] Prochazka J, Camaj M. Modelling the number of road accidents of uninsured drivers and their severity. Proceedings of International Academic Conference. Geneva, 2017. https://doi.org/10.20472/IAC.2017.032.035.

[25] Procházka J, Flimmel S, Čamaj M, Bašta M. Modelling the Number of Road Accidents. Publishing house of the University of Economics in Wrocław. Wrocław, 2017:355–364. https://doi.org/10.15611/amse.2017.20.29.

[26] Rajput H, Som T, Kar S. An automated vehicle license plate recognition system Computer (Long Beach. Calif). 2015;48(8):56–61. https://doi.org/10.1109/MC.2015.244.

[27] Sebego M, Naumann RB, Rudd RA, Voetsch K, Dellinger AM, Ndlovu C. The impact of alcohol and road traffic policies on crash rates in Botswana, 2004–2011: A time-series analysis. Accident Analysis & Prevention. 2014;70:33–39. https://doi.org/10.1016/j.aap.2014.02.017.

[28] Shetty P, Sachin PC, Kashyap VK, Madi V. Analysis of road accidents using data mining techniques. Inernational Research Journal of Engineering and Technology. 2017;04.

[29] Statistic Road Accident: https://statystyka.policja.pl/ (accessed on 23.05.2023).

[30] Tambouratzis T, Souliou D, Chalikias M, Gregoriades A. 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. 2014;4(1):31–42. https://doi.org/10.2478/jaiscr-2014-0023.

[31] Vilaça M, Silva N, Coelho MC. Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. Transportation Research Procedia. 2017;27:1113–1120. https://doi.org/10.1016/j.trpro.2017.12.113.

[32] WHO. The Global status on road safety 2023. Geneva: https://iris.who.int/bitstream/handle/10665/375016/9789240086517-eng.pdf?sequence=1 (accessed on 06.12.2024).

https://doi.org/10.14669/AM/182808

[33] Wójcik A. Vector Autoregression Models (VAR) – Response to Criticism Structural Econometric Models (Modele wektorowo-autoregresyjne jako odpowiedź na krytykę strukturalnych wielorównaniowych modeli ekonometrycznych. In Polish). Economic Studies. 2014;193:112–128.

[34] Yang Z, Zhang W, Feng J. Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. Safety Science. 2022;146:105522. https://doi.org/10.1016/j.ssci.2021.105522.

[35] Zheng Z, Wang C, Wang P, Xiong Y, Zhang F, Lv Y. Framework for fusing traffic information from social and physical transportation data. PLoS One. 2018;13(8):e0201531. https://doi.org/10.1371/journal.pone.0201531.

[36] Zhu L, Lu L, Zhang W, Zhao Y, Song M. Analysis of accident severity for curved roadways based on bayesian networks. Sustainability. 2019;11(8):2223. https://doi.org/10.3390/su11082223.