Forecasting the Number of Road Accidents in Poland by Road Type

by Piotr Gorzelańczyk and Jen Sim Ho

Cite this Article

Article

Forecasting the Number of Road Accidents in Poland by Road Type

Piotr Gorzelańczyk [ID]1,* and **Jen Sim Ho** [ID]2

1   Stanislaw Staszic State University of Applied Sciences in Pila, Podchorazych 10 Street, 64-920 Pila, Poland
2   Malaysian Institute of Road Safety Research, Lot 125-135, Jalan TKS1, Taman Kajang Sentral, 43000 Kajang, Selangor, Malaysia
*   For correspondence: piotr.gorzelanczyk@ans.pila.pl

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. To achieve this, monthly accident data for Poland from the Police's statistics for the years 2007–2021 were analyzed, and a prediction for the years 2022–2024 was created. As is evident, either the number of accidents is rising or it is stabilizing. This is mostly caused by the rise in automobile traffic. 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 in the nation's growing number of freeways. It should be remembered that the current epidemic distorts the findings. Selected time series models were used in the investigation in Statistica.

1. Introduction

Road accidents are situations that result in both property damage and injury or death to other road users. The World Health Organization (WHO) estimates that 1.3 million individuals lose their lives in traffic accidents each year. Road accidents cost the gross domestic product (GDP) of the majority of nations throughout the world roughly 3%. For kids and young adults aged 5 to 29 years old, traffic accidents constitute the main cause of mortality [1]. By 2030, the United Nations (UN) General Assembly wants to cut in half the number of people killed and injured in traffic accidents.

A factor in judging the severity of a traffic collision is its scope. For the appropriate authorities to develop road safety rules aimed at preventing accidents and minimizing injuries, fatalities, and property damage, it is vital to predict accident severity [2,3]. Adopting countermeasures to prevent and reduce accident severity requires the identification of crucial elements that influence accident severity [4]. A multi-node Deep Neutral Network (DNN) structure is put out by Yang et al. [5] for forecasting various degrees of damage, demise, and property loss. It allows for a thorough and precise investigation of the seriousness of road accidents.

The data on accidents comes from several sources. Most frequently, they are gathered and evaluated by government officials using the appropriate government agencies. Police reports, insurance company databases, and hospital records are all used for data collection. The transportation industry is then processing traffic accident data in part on a wider scale [6].

At the moment, the most significant source of information for the analysis and prediction of traffic accidents is intelligent transportation systems. Vehicle-mounted GPS systems can be used to process this data [7]. Roadside microwave vehicle detection systems may continually capture information about moving vehicles, such as speed, amount of traffic, and vehicle type [8]. Additionally, it is feasible to gather a significant amount of traffic data over a supervised time [9] using a license plate recognition system. Social media is another possible source of data for gathering information on traffic and accidents, although their accuracy may not be sufficient owing to the inexperience of reporters [10].

Work with several data sources, which need to be correctly challenged, in order to make accident data useful. The accuracy of analytical findings can be improved by combining diverse data sources and merging heterogeneous traffic accident data [11].

A statistical analysis was undertaken by Vilaca et al. [12] to evaluate the seriousness and determine the connection between traffic participants and accidents. A suggestion to raise the bar for traffic safety regulations and implement more traffic safety measures is the study's outcome.

Based on the number of traffic accidents—a measure of the research of accident causes—Bak et al. [13] conducted a statistical analysis of traffic safety in a chosen Polish area. In order to investigate the safety elements of people who cause accidents, the study employed multivariate statistical analysis.

The type of traffic issue being tackled determines the source of accident data to be used for analysis. The accuracy of accident prediction and accident abolition are improved by combining statistical models with additional natural driving data or other data received through intelligent traffic systems [14].

The literature has a number of techniques for predicting the frequency of accidents. The most popular techniques for predicting the frequency of accidents are time series approaches [15,16], which have the drawback of not being able to evaluate the forecast's accuracy based on previous predictions and the frequent residual component of autocorrelation [17]. While Sunny & Nithya [18] employed the Holt-Winters exponential smoothing approach, Procházka et al. [19] used a multi-seasonality model for forecasting. One of its drawbacks is that exogenous variables cannot be added to the model [20].

The vector autoregressive model, which has the drawback of requiring many observations of variables to accurately estimate their parameters [21], as well as the curve-fitting regression models of Al-Madani [22] and Monederoa et al. [23] for analyzing the number of fatalities, have all been used to predict the frequency of traffic accidents. In turn, these merely need a few straightforward linear connections [24] and an order of autoregression [25] (assuming that the series is already stationary).

Random Forest regression was used by Biswas et al. [26] to forecast the frequency of traffic accidents. In this situation, smaller groups are preferred over bigger ones [27], the approach and peak prediction are unstable [28], and the data contain groups of correlated characteristics of equal importance to the original data. For the given forecasting problem, Chudy-Laskowska & Pisula [29] employed an autoregressive quadratic trend model, a univariate periodic trend model, and an exponential smoothing model. The problem at hand may alternatively be predicted using a moving average model, but this approach has the drawbacks of poor prediction accuracy, loss of data within a sequence, and inability to take trends and seasonal impacts into consideration [30].

In order to ensure that the process is stationary, Prochozka & Camaj [31] employed the GARMA (generalized autoregressive moving average) approach, which places restrictions on the parameter space. Forecasting frequently uses the ARMA (autoregressive moving average) model for a stationary process or the ARIMA (autoregressive integrated moving average) or SARIMA (seasonal autoregressive integrated moving average) model for a non-stationary process [18,19,32,33]. These models provide the analyzed models a great deal of flexibility, but this has the drawback of requiring more research knowledge from the researcher than, say, regression analysis [30]. The linearity of the ARIMA model is another drawback [34].

ANOVA (analysis of variance) was employed by Chudy-Laskowska & Pisula in their study [35] to forecast the frequency of traffic accidents. This method's drawback is that it includes extra presumptions, particularly the presumption of sphericity, the violation of which might result in incorrect findings [36]. The frequency of road accidents is also anticipated using artificial neural network (ANN) algorithms. The drawbacks of ANNs include the requirement for prior knowledge in this area [35,37], the dependence of the final result on the network's initial conditions, and the inability to interpret results in a conventional manner because ANNs are typically referred to as "black boxes", where input data is entered and the model outputs results without knowledge of the analysis [38].

Kumar et al.'s [39] usage of the Hadoop modelfkumar was a novel prediction technique. This method's drawback is that it cannot handle tiny data sets [40]. The Garch model was employed by Karlaftis & Vlahogianni [33] for prediction. This method's intricate model and complex form are a drawback [41,42]. McIlroy and his team employed the ADF (augmented Dickey–Fuller) test [43], which has the drawback of having insufficient power to detect autocorrelation of the random component [44].

Data-mining approaches, which often have the drawback of having enormous collections of generic descriptions [45], have also been employed for predicting [46]. The mixture of models put out by Sebego et al. [47] is another example of a combination of models. Bloomfield [48] also suggests parametric models.

Taking into account the above analysis, the author made forecasts of the number of road accidents in Poland depending on road types. Selected time series models and exponential models were used to forecast the number of accidents. Forecasting of the number of road accidents using factors influencing this value and the use of other forecasting methods (e.g., neural networks, linear regression) can be found in the author's other work in this area [49–61]. In their study, the authors used the police's statics data, not taking into account other factors affecting the occurrence of a traffic accident.

2. Materials and Methods

Every year there is an increase in the number of new cars on Polish roads. Currently, there are nearly 750 vehicles per 1000 inhabitants in Poland. This translates into an increase or stabilization of the number of accidents on the roads (Figure 1). Poland covers an area of 312,705 km² and has the following road categories:

- motorway,
- expressway,
- 2 one-way carriageway,
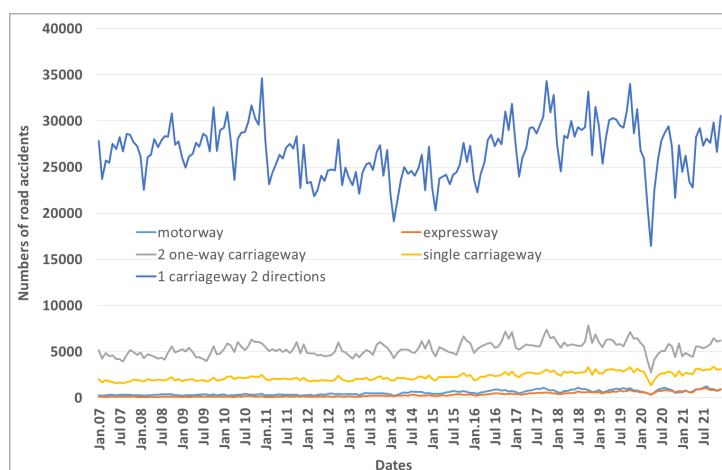- single carriageway,
- 1 carriageway 2 directions.

Figure 1. Number of accidents in Poland in 2007–2021 by road type [62].

The Kruskall-Wallis test was used to examine how the number of accidents on different types of roads changed over time. The test statistic has a value of 830 and a test probability of $p = 0.000$. The result shows that we do not accept the equality of the mean level of traffic accidents (Figure 2).
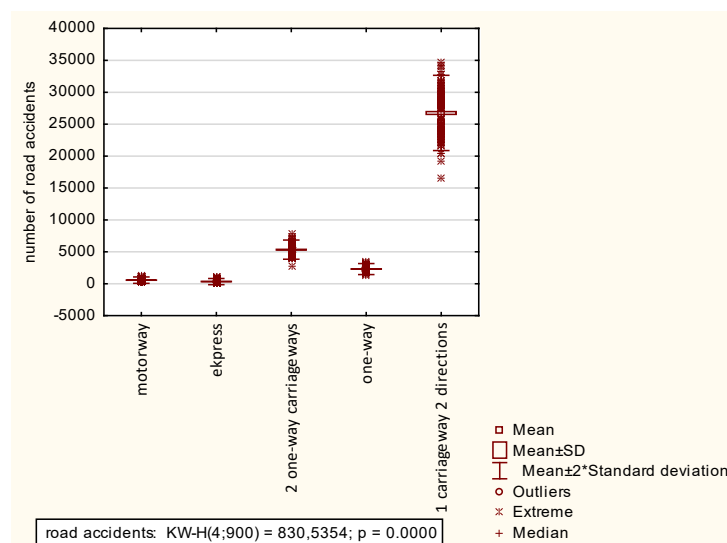

Figure 2. Comparison of the average number of road accidents in Poland by road type from 2007 to 2021.

3. Selected Statistical Methods

In the article, the authors used various static methods. Among them can be mentioned the Brownian, Holt, and Winters methods. The Brownian method is classified as an exponential smoothing method. It is most often used in the case of a time series in which there is no trend, i.e., the series used does not show a development trend, and its fluctuations are due to random factors that occur when forecasting, for example, the number of traffic accidents. The model for the change in the projected number of traffic accidents ($Plwd$) in the analyzed time series takes the form:

$$Plwd_t^* = \alpha Plwd_{t-1} + (1 - \alpha)Plwd_{t-1}^* \tag{1}$$

where:

$Plwd_{t-1}^*$—The forecast values of the number of traffic accidents for the optimal value of the smoothing parameter α;

t—subsequent periods $(1, 2, ..., n + 1)$;

α—constant value of the process smoothing parameter taking a value in the range:

$$\alpha \in \langle 0,1 \rangle \tag{2}$$

Holt Method

Another of the exponential equalization methods presented is the Holt method. In this method, it is assumed that the number of traffic accidents does not fluctuate over time, but only there is a trend and random fluctuations that can change over time. The $Plwd$ change model in the case under consideration takes the form:

$$F_t = \alpha Plwd_t + (1 - \alpha)(F_{t-1} + S_{t-1}) \tag{3}$$

$$S_t = \beta(F_t + F_{t-1}) + (1 - \beta)S_{t-1} \tag{4}$$

where:

F_t—the smoothed value of the forecast variable at time t,

S_t—the smoothed value of the trend increment at time t,

α—constant value of the process smoothing parameter taking a value in the range:

$$\alpha \in \langle 0,1 \rangle \tag{5}$$

β—constant value of the process smoothing parameter taking a value in the range:

$$\beta \in \langle 0,1 \rangle \tag{6}$$

Based on the above relationships, the forecast equation at time $t > k$ takes the form:

$$Plwd(t) = F_k + (t - n)S_k \tag{7}$$

where:

$Plwd(t)$—the forecast value of the number of traffic accidents at time t,

F_k—the smoothed value of the forecast variable at time k,

S_k—the value of the trend increment at the moment k,

N—the number of words in the time series.

The initial value of F_1 is assumed to be the first value of the forecast variable $Plwd$. While the initial value of S_1 is assumed to be the difference in the value of the forecast variable $Plwd_2 - Plwd_1$. The values of α and β coefficient are determined so that their value determines the smallest error of expired forecasts. Expired forecasts are determined from the following formula:

$$Plwd(t) = F_t + S_t, 1 \leq t \leq N \tag{8}$$

Winters Method

Another of the proposed methods for determining the number of traffic accidents over a certain period is the Winters method. This method is used for time series in which there is a development trend and seasonal and random fluctuations.

In the method in question, the following versions of the model are distinguished:

- Additive—occurs when the effect of a season is constant over time.

In this case, the smoothed value of the forecast variable F_t at time t, after removing seasonal data, is determined from the formula:

$$F_t = \alpha(Plwd_t - G_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1}) \tag{9}$$

On the other hand, trend increment analysis at time $t-1$, takes the form of:

$$S_t = \beta(F_t + F_{t-1}) + (1-\beta)S_{t-1} \tag{10}$$

The parameter G_t, the seasonality index at time t, is determined by the formula:

$$G_t = \gamma(y_t + F_t) + (1-\gamma)G_{t-r} \tag{11}$$

where:

r—length of the seasonal cycle, taking the value $1 \leq r \leq k$

α, β, γ—model parameters with values in the interval [0,1], are determined by minimizing the forecast error.

- Multiplicative—occurs if in a time series, the relative amplitude of seasonal changes is constant over time, and seasonal fluctuations are of proportional trend.

For the multiplicative model, the variables F_t, S_t and G_t take the form, respectively:

$$F_t = \alpha \frac{Plwd_t}{G_{t-r}} + (1-\alpha)(F_{t-1} + S_{t-1}) \tag{12}$$

$$S_t = \beta(F_t + F_{t-1}) + (1-\beta)S_{t-1} \tag{13}$$

$$G_t = \gamma \frac{Plwd_t}{F_t} + (1-\gamma)G_{t-r} \tag{14}$$

The forecast $Plwd(t)$ at time t is determined from the following relationships:

- For a version of the additive model:

$$Plwd_t^P = F_k + S_k(t-k) + G_{t-r}, t > N \tag{15}$$

- For a version of the multiplicative model:

$$Plwd_t^P = [F_k + S_k(t-k)]G_{t-r}, t > N \tag{16}$$

4. Forecasting the Number of Road Accidents

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.

Measures of analytical forecasting perfection were calculated using the errors of forecasts that had expired, which were calculated using Equations (17)–(21):

- ME—mean error.

$$ME = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_p) \tag{17}$$

- MAE—mean average error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - Y_p| \tag{18}$$

- MPE—mean percentage error.

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - Y_p}{Y_i} \qquad (19)$$

- MAPE—mean absolute percentage error.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - Y_p}{Y_i} \right| \qquad (20)$$

- MSE—mean square error.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - Y_p)^2 \qquad (21)$$

where:

n—the length of the forecast horizon,

Y—observed value of road accidents,

Y_p—the forecasted value of road accidents.

The mean absolute percentage error was reduced to compare the number of accidents during a pandemic.

Using information from the Polish police from 2007 to 2021, the number of accidents on the different types of roads studied was predicted. A number of different methods can be used for forecasting. In this study, the authors use a fast exponential smoothing method.

Figures 3–7 show the results of the road forecasts. The different forecasting techniques used in the study are denoted by M1, M2, ..., M15. The forecasting methods used in the study are shown in Table 1.
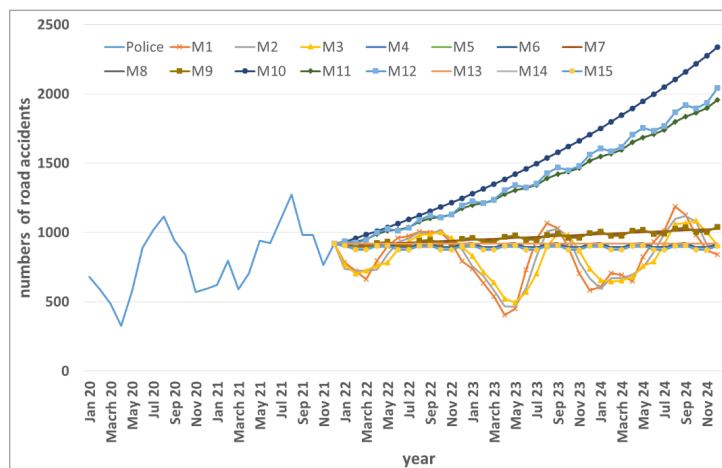

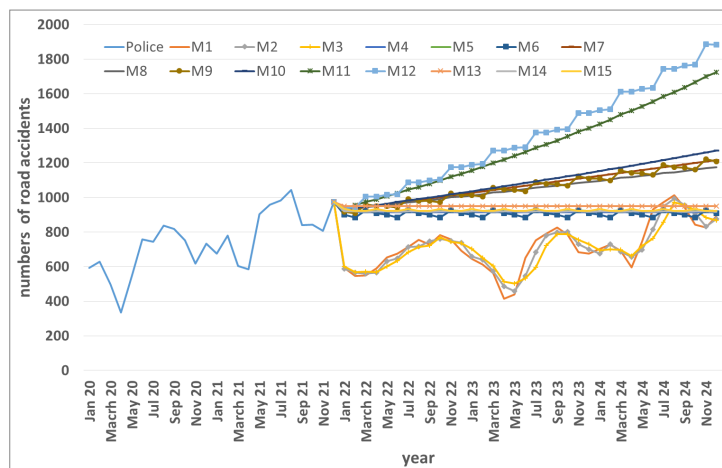**Figure 3.** Forecasting the number of road accidents on the motorway between 2022 and 2024.

Figure 4. Forecasting the number of road traffic accidents on the ecompass road from 2022 to 2024.
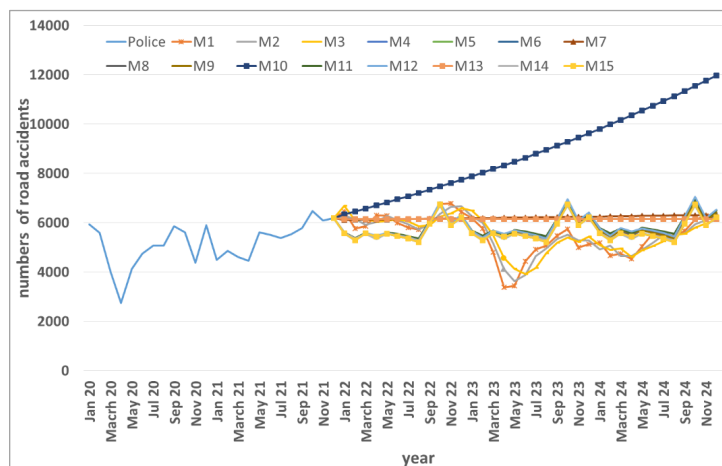
Figure 5. Forecasting the number of road accidents on a 2 carriageway road between 2022 and 2024.
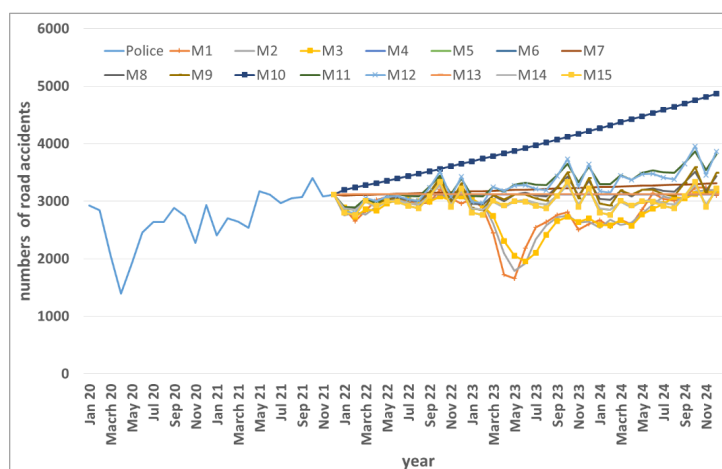

Figure 6. Forecasting the number of road accidents on a one-way road between 2022 and 2024.
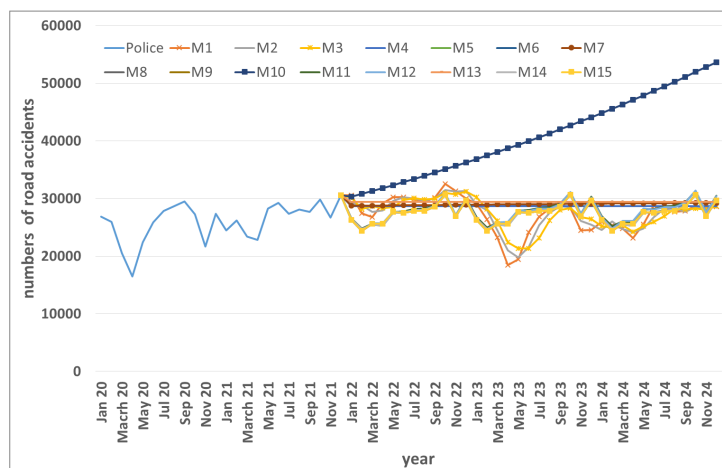
Figure 7. Forecasting the number of road accidents on a 1 carriageway 2 direction road between 2022 and 2024.

Table 1. Forecasting methods.

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

Table 1. (*Continued*)

M6	exponential smoothing no trend seasonal component: multiplicative
M7	exponential smoothing linear trend seasonal component: none HOLT
M8	exponential smoothing linear trend seasonal component: additive
M9	exponential smoothing linear trend seasonal component: multiplicative WINTERS
M10	exponential smoothing exponential seasonal component: none
M11	exponential smoothing exponential seasonal component: additive
M12	exponential smoothing exponential seasonal component: multiplicative
M13	exponential smoothing fading trend seasonal component: none
M14	exponential smoothing fading trend seasonal component: additive
M15	exponential smoothing fading trend seasonal component: multiplicative

The aforementioned information leads to the conclusion that not all procedures utilized in the situation under review are successful. To determine the best forecasting method, the value of mean absolute percentage error MAPE (20) was used for which the analyzed value was the smallest. The best forecasting techniques for each road were determined to be the following:

- motorway—M6;
- expressway—M4;
- 2 one-way carriageway—M10;
- single carriageway—M4;
- 1 carriageway 2 directions—M10.

The data acquired allow it to be determined that the manner of analysis relies on the kind of route being examined. The exponential smoothing without trend and exponential smoothing approaches consistently produced the least MAPE error. Based on this, a prediction of the number of accidents on the roads according to the examined road types was created, which is depicted in Figure 8, and the forecast errors achieved are reported in Table 2 and Figure 9. 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, which should be highlighted. A maximum error value of 1% illustrates the selection of an efficient forecasting method.
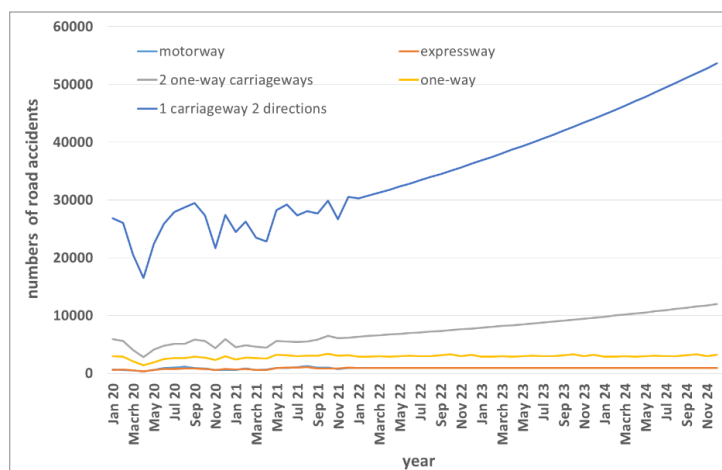

Figure 8. Optimal projected number of road accidents according to the analyzed road types in the period 2022–2024.

Table 2. Forecast errors.

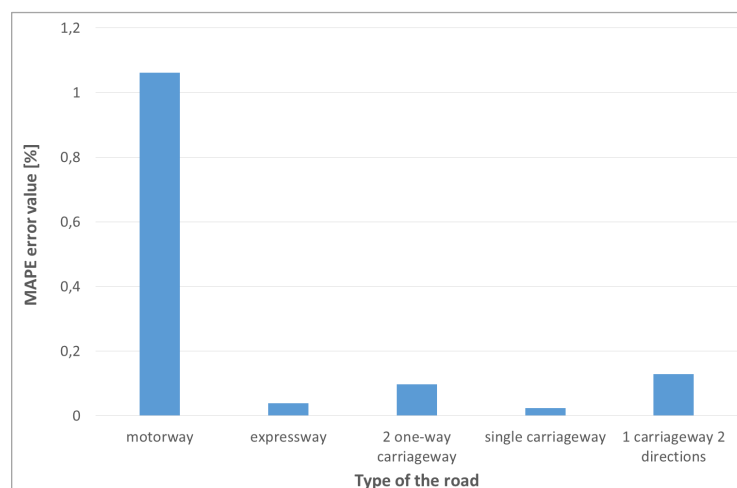| Road Type/Forecast Error	ME	MPE	MSE	MAPE [%]	MAE [%]
motorway	3.973694	79.10152	1,895,329	1.062161	14.55316
expressway	7.895166	44.23115	748,560	0.038562	13.61563
2 one-way carriageway	36.11964	492.0262	73,097,240	0.097067	9.37681
single carriageway	10.51187	118.3365	5,788,476	0.023715	5.318702
1 carriageway 2 directions	102.9622	2119.331	1,306,117,717	0.129038	8.167838

Figure 9. Forecast errors.

5. 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 before 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, which should be highlighted. The selection of an efficient forecasting technique may be demonstrated by an error value of no more than 1.06%.

In addition, based on the study, the largest projected increase in traffic accidents is on highways. By the end of 2024, it could reach more than 50,000, in which case we can expect a large increase. For the rest of the analyzed roads (expressway, 2 one-way carriageways, 1 one-way carriageway, and 1 two-way carriageway), we can expect the number of accidents to stabilize over the analyzed period. A minimal increase is observed for 2 one-way roadways. In this case, the number of accidents could reach more than 10,000 by the end of 2024.

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

The authors intend to consider further variables affecting accident rates in Poland in their future study. Examples of these can be the amount of traffic, the day of the week, or the age of the accident's perpetrator.

Data Availability

The research was made on the basis of publicly available Statistics of the Polish Police.

Author Contributions

Methodology: P.G., & J.S.H.; Project administration: P.G., Resources: P.G.; Software: P.G.; Supervision: P.G., & J.S.H.; Validation: P.G.; Visualization: P.G., & J.S.H.; Writing – original draft: P.G.; Writing – review & editing: P.G., & J.S.H.

Conflicts of Interest

The authors have no conflict of interest to declare.

References

1. World Health Organization (WHO). (Ed.). (2018). *The Global status on road safety 2018*. https://www.who.int/publications/i/item/9789241565684 (accessed 10 February 2022).
2. 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. https://doi.org/10.2478/jaiscr-2014-0023
3. 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
4. Arteaga, C., Paz, A., & Park, J. (2020). Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Safety Science*, *132*, 104988. https://doi.org/10.1016/j.ssci.2020.104988
5. 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
6. 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. https://doi.org/10.20858/sjsutst.2020.107.3
7. Chen, C. (2017). Analysis and forecast of traffic accident big data. *ITM Web of Conferences*, *12*, 04029. https://doi.org/10.1051/itmconf/20171204029
8. Khaliq, K. A., Chughtai, O., Shahwani, A., Qayyum, A., & Pannek, J. (2019). Road Accidents Detection, Data Collection and Data Analysis Using V2X Communication and Edge/Cloud Computing. *Electronics*, *8*, 896. https://doi.org/10.3390/electronics8080896
9. Rajput, H., Som, T., & Kar, S. (2015). An automated vehicle license plate recognition system. *Computer*, *48*(8), 56–61. https://doi.org/10.1109/MC.2015.244
10. Zheng, Z., Wang, C., Wang, P., Xiong, Y., Zhang, F., & Lv, Y. (2018). Framework for fusing traffic information from social and physical transportation data. *PLOS ONE*, *13*(8), e0201531. https://doi.org/10.1371/journal.pone.0201531
11. Abdullah, E., & Emam, A. (7–9 December 2015). *Traffic accidents analyzer using big data*. 2015 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA. https://doi.org/10.1109/CSCI.2015.187
12. 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
13. 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
14. 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
15. Helgason, A. F. (2016). Fractional integration methods and short Time series: evidence from a simulation study. *Political Analysis*, *24*(1), 59–68.
16. Lavrenz, S. M., Vlahogianni, E. I., 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
17. Kowalski, L. (2022). *Forecasting based on time series* (in Polish). http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf (accessed 10 February 2022).
18. Sunny, C. M., Nithya, S., Sinshi, K. S., Vinodini, V., Lakshmi, A., Anjana, S., et al. (7–9 August 2018). *Forecasting of Road Accident in Kerala: A Case Study*. 2018 International Conference on Data Science and Engineering (ICDSE), Kochi, India. https://doi.org/10.1109/ICDSE.2018.8527825
19. Procházka, J., Flimmel, S., Čamaj, M., & Bašta, M. (30 August–3 September 2017). *Modelling the Number of Road Accidents*. 20-th AMSE. Applications of Mathematics and Statistics in Economics. International Scientific Conference, Szklarska Poręba, Poland. https://doi.org/10.15611/amse.2017.20.29
20. Szmuksta-Zawadzka, M., & Zawadzki, J. (2009). Forecasting on the basis of Holt-Winters models for complete and incomplete data (in Polish). *Ekonometria*, *24*(38), 85–99.
21. Wójcik, A. (2014). A*utoregressive vector models as a response to the critique of multi-equation structural econometric models*. University of Economics in Katowice.
22. Al-Madani, H. M. N. (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
23. 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
24. Mamczur, M. (31 August 2020). *Machine learning How does linear regression work? And is it worth using?* https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac (accessed 10 February 2022).
25. Piłatowska, M. (2012). The choice of the order of autoregression depending on the parameters of the generating model (in Polish). *Econometrics*, *4*(38), 16–35.
26. Biswas, A. A., Mia, M. J., & Majumder, A. (6–8 July 2019). *Forecasting the Number of Road Accidents and Casualties using Random Forest Regression in the Context of Bangladesh*. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India. https://doi.org/10.1109/ICCCNT45670.2019.8944500
27. Wikipedia. (2022). *Random forest*. https://en.wikipedia.org/wiki/Random_forest (accessed 10 February 2022).
28. Fijorek, K., Mróz, K., Niedziela, K., & Fijorek, D. (2010). Forecasting electricity prices on the day-ahead market using data mining methods (in Polish). *Energy Market*.
29. Chudy-Laskowska, K., & Pisula, T. (2014). Forecast of the number of road accidents in Poland (in Polish). *Logistics*, *6*, 2710–2722.
30. Łobejko, S. (2015). *Time series analysis and forecasting with SAS*. Main business school in Warsaw.
31. Prochazka, J., & Camaj, M. (2017). Modelling the number of road accidents of uninsured drivers and their severity. In *Proceedings of International Academic Conferences 5408040*, International Institute of Social and Economic Sciences.
32. Dutta, B., Barman, M. P., & Patowary, A. N. (2020). Application of Arima model for forecasting road accident deaths in India. *International Journal of Agricultural and Statistical Sciences*, *16*(2), 607–615.
33. Karlaftis, M. G., & Vlahogianni, E. I. (2009). Memory properties and fractional integration in transportation time-series. *Transportation Research Part C: Emerging Technologies*, *17*(4), 444–453. https://doi.org/10.1016/j.trc.2009.03.001

34. Dudek, G. (2013). Forecasting Time Series with Multiple Seasonal Cycles Using Neural Networks with Local Learning. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, & J. M. Zurada (Eds.), *Artificial Intelligence and Soft Computing. ICAISC 2013. Lecture Notes in Computer Science* (Vol. 7894). Springer. https://doi.org/10.1007/978-3-642-38658-9_5

35. Chudy-Laskowska, K., & Pisula, T. (2015). Forecasting the number of road accidents in Subcarpathia (in Polish). *Logistics*, *4*, 2782–2796.

36. Gregorczyk, A., & Swarcewicz, M. (2012). Analiza wariancji w układzie powtarzanych pomiarów do określenia efektów czynników wpływających na pozostałości linuronu w glebie (in Polish). *Polish Journal of Agronomy*, *11*, 15–20.

37. Wrobel, M. S. (2017). *Application of neural fuzzy systems in chemistry* [PhD Thesis]. University of Silesia in Katowice.

38. StatSoft. (2022). *Data Mining Techniques* (in Polish). https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html (accessed 10 February 2022).

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

40. DataFlair. (2022). *Top Advantages and Disadvantages of Hadoop 3*. https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop (accessed 10 February 2022).

41. Perczak, G., & Fiszeder, P. (2014). *GARCH model - using additional information on minimum and maximum prices* (in Polish). Bank and Credit.

42. Fiszeder, P. (2009). *GARCH class models in empirical financial research*. Scientific Publishers of the Nicolaus Copernicus University.

43. McIlroy, R. C., Plant, K. A., Hoque, M. S., Wu, J., Kokwaro, G. O., Nam, V. H., et al. (2019). Who is responsible for global road safety? A cross-cultural comparison of actor maps. *Accident Analysis & Prevention*, *122*, 8–18. https://doi.org/10.1016/j.aap.2018.09.011

44. Muck, J. (2022). *Econometrics. Modeling of time series. Stationary. Unit root tests. ARDL models. Co-integration* (in Polish). http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja5.pdf (accessed 10 February 2022).

45. Li, L., Shrestha, S., & Hu, G. (7–9 June 2017). *Analysis of road traffic fatal accidents using data mining techniques*. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), London, UK. https://doi.org/10.1109/SERA.2017.7965753

46. Marcinkowska, J. (2015). *Statistical methods and data mining in assessing the occurrence of syncope in the group of narrow-QRS tachycardia (AVNRT and AVRT)* (in Polish). Medical University of Karol Marcinkowski in Poznań. http://www.wbc.poznan.pl/Content/373785/index.pdf (accessed 10 February 2022).

47. Sebego, M., Naumann, R. B., Rudd, R. A., Voetsch, K., Dellinger, A. M., & Ndlovu, C. (2014). The impact of alcohol and road traffic policies on crash rates in Botswana, 2004–2011: a time-series analysis. *Accident Analysis & Prevention*, *70*, 33–39. https://doi.org/10.1016/j.aap.2014.02.017

48. Bloomfield, P. (1973). An exponential model in the spectrum of a scalar time series. *Biometrika*, *60*(2), 217–226. https://doi.org/10.2307/2334533

49. Gorzelanczyk, P. (2023) Application of neural networks to forecast the number of road accidents in provinces in Poland. *Heliyon*, *9*(1), e12767. https://doi.org/10.1016/j.heliyon.2022.e12767

50. Gorzelanczyk, P. (2023). Forecasting the Number of Road Accidents in Polish Provinces Using Trend Models. *Applied Sciences*, *13*, 2898. https://doi.org/10.3390/app13052898

51. Gorzelanczyk, P. (2023). Forecasting the number of road accidents in Poland using weather-dependent trend models. *Technical Sciences*, *26*(26), 57–76. https://doi.org/10.31648/ts.8289

52. Gorzelanczyk, P. (2023). Using neural networks to forecast the number of road accidents in Poland taking into account weather conditions. *Results in Engineering*, *17*, 100981. https://doi.org/10.1016/j.rineng.2023.100981

53. Gorzelanczyk, P. (2023). The use of neural networks to forecast the number of road accidents in Poland. *Scientific Journal of Silesian University of Technology. Series Transport*, *118*, 45–54. https://doi.org/10.20858/sjsutst.2023.118.4

54. 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), 431–435. https://doi.org/10.3311/PPtr.22203

55. Gorzelanczyk, P., & Grobelny, P. (2024). An Estimate of the Number of Accidents on Polish Highways Based on the Kind of Road. *Communications - Scientific Letters of the University of Žilina*, *26*(1), F13–F12. https://doi.org/10.26552/com.C.2024.012

56. Gorzelanczyk, 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 Žilina*, *24*(4), A216–A231. https://doi.org/10.26552/com.C.2022.4.A216-A231

57. Gorzelanczyk, P., Kalina, T., Jurkovič, M., & Mohanty, M. (2023). Forecasting the number of road accidents in Poland by day of the week and the impact of pandemics and pandemic-induced changes. *Cogent Engineering*, *10*(1), 2185956. https://doi.org/10.1080/23311916.2023.2185956

58. Gorzelanczyk, P., & Tylicki, H. (2023). Forecasting the number of road accidents in Poland depending on the day of the week using neural networks. *LOGI - Scientific Journal on Transport and Logistics*, *14*(1), 35–42. https://doi.org/10.2478/logi-2023-0004

59. Gorzelanczyk, P. (2023). Forecasting the number of road accidents in Poland depending on weather conditions and COVID-19. *Sustainable Social Development*, *1*(2), 2221. https://doi.org/10.54517/ssd.v1i2.2221

60. Gorzelanczyk, P. (2023). Impact of information on the number of traffic accidents on the outcome of the forecast. *Technical Sciences*, *26*(26), 219–230. https://doi.org/10.31648/ts.8945

61. Jurkovic, M., Gorzelanczyk, P., Kalina, T., Jaros, J., & Mohanty, M. (2022). Impact of the COVID-19 pandemic on road traffic accident forecasting in Poland and Slovakia. *Open Engineering*, *12*, 578–589. https://doi.org/10.1515/eng-2022-0370

62. Statystyka. (2022). *Statistic Road Accident* (in Polish). https://statystyka.policja.pl (accessed 10 February 2022).