Modern Transportation 2024, 13(1), 9292.
https://doi.org/10.18686/mt9292

Article

Polish road type-specific trend models for predicting the frequency of traffic accidents

Piotr Gorzelanczyk

Stanislaw Staszic University of Applied Sciences in Pila, 64-920 Pila, Poland; piotr.gorzelanczyk@ans.pila.pl

CITATION

Gorzelanczyk P. Polish road type-specific trend models for predicting the frequency of traffic accidents. Modern Transportation. 2024; 13(1): 9292.
https://doi.org/10.18686/mt9292

ARTICLE INFO

Received: 10 July 2023
Accepted: 15 August 2023
Available online: 29 November 2024

Abstract: Every year many people are fatally injured or killed on the roads. The number is still quite high, even if it is decreasing from year to year. Although there are fewer accidents on the roads since the epidemic, the number is still relatively high. To minimize the number of road accidents, it is important to understand which types of roads have the highest number of collisions and what the accident forecasts are for the coming years. The purpose of this article is to forecast the number of accidents on Polish roads according to their type. The study consisted of two parts. The first included a forecast of the number of road accidents for 2022–2031, based on an analysis of annual data from police statistics on the number of road accidents in Poland in 2000–2021. The second part of the study focused on monthly data from 2000–2021. In this case, a forecast was also set for the period from January 2022 to December 2023. The results of the study indicate that even after analyzing annual statistics, the number of incidents can be expected to stabilize in the coming years. This is mainly due to the expansion of expressways, especially highways, and the increase in traffic volume on Polish roads. It should be noted that the current epidemic is distorting the results.

Keywords: road traffic accident; forecasting; trend models; road

1. Introduction

Road accidents are situations that result in both damage to property and injury or death to other drivers. According to the WHO, 1.3 million people die each year in car accidents. Road accidents cause a loss of 3% of GDP in a typical country around the world. Children and adolescents between the ages of 5 and 29 are most often killed in road accidents [1]. The UN General Assembly wants to see a 50% reduction in road fatalities and injuries by 2030.

The size of a road traffic collision is a factor in assessing its severity. In order for responsible authorities to develop road safety legislation with the intention of preventing accidents, minimising injuries, fatalities and property damage, it is crucial to quantify the severity of the accident [2,3]. Before implementing countermeasures to prevent and minimise accident severity, it is crucial to identify the critical elements that influence accident severity [4]. A multi-node Deep Neutral Network (DNN) architecture was developed by Yang et al. to predict various degrees of injury, fatality and property loss. This enables a complete and accurate assessment of the severity of road accidents [5].

Accident data come from a variety of sources. Typically, government officials use relevant government agencies to obtain and evaluate them. Numerous sources, including hospital records, insurance company databases and police reports, are used to collect data. As a result, the transport industry is carrying out increasing analyses of data related to road accidents [6].

Currently, the most important source of information for analysing and forecasting traffic incidents is intelligent transport systems. GPS devices mounted on moving vehicles can be used to analyse this data [7]. Roadside microwave vehicle detection systems can continuously capture information about moving vehicles, such as speed, traffic volume and vehicle type [8]. In addition, a lot of traffic data can be collected for a limited period of time using a number plate recognition system [9]. Social media is another possible source of traffic and accident information, although the accuracy of reports may be insufficient due to the inexperience of reporters [10].

Working with a variety of data sources that need to be properly challenged for accident data to have any value. Analytical results can be made more precise by integrating multiple data sources and combining different road accident data [11].

In order to determine the severity of the problem and establish the link between road users and accidents, Vilaca et al. [12] conducted a statistical analysis. The results of the study include raising the bar for road safety rules and implementing more traffic precautions.

Based on the number of road accidents, which serves as a barometer to study the causes of accidents, Bak et al. [13] conducted a statistical analysis of road safety in a selected area of Poland. The study examined the safety variables of accident causes using multivariate statistical analysis.

The type of road problem to be solved determines the source of accident data to be used for the analysis. The accuracy of accident prediction and elimination is increased by combining statistical models with additional data from actual driving or other information collected from intelligent traffic systems [14].

Many accidents probability forecasting techniques can be found in the literature. The most popular methodologies for accident frequency forecasting [15,16] have the disadvantage that they do not allow assessment of prediction accuracy based on previous forecasts and the frequent residual component of autocorrelation [17]. While Sunny et al. [18] used the Holt-Winters exponential smoothing approach, Procházka et al. [19] used a multi-seasonal model. One disadvantage of this model is the inability to account for exogenous variables [20].

Traffic accident frequency has been predicted using the Al-Madani [21] and Monedero et al. [22] curve fitting regression models for fatality analysis, as well as a vector autoregressive model, which has the disadvantage of requiring multiple observations of variables to accurately estimate their parameters [23]. Assuming that the series are already stationary, they require only an order of autoregression [24] and some simple linear combinations [25].

Random Forest regression was used by Biswas et al. [26] to predict the frequency of traffic accidents. The approach and peak prediction are unstable [27], the data include groups with related characteristics that are as valid as the original data, and smaller groups are preferred over larger groups in this case [28]. For the proposed forecasting problem, Chudy-Laskowska and Pisula [29] used an autoregressive quadratic trend model, a periodic trend model and an exponential smoothing model. The problem at hand can potentially be forecast using a moving average model, but this approach has low forecasting accuracy, loss of data in the sequence and is unable to account for trends and seasonal fluctuations [30].

The GARMA approach, which limits the parameter space, was used by Prochozka and Cameja [19] to ensure process stability. The ARMA model for stationary systems [17,19,31,32] and the ARIMA or SARIMA model for non-stationary phenomena are often used in forecasting. The advantage of these models is that they provide the models under study with considerable flexibility; the disadvantage, however, is that they require more advanced research skills from the researcher than, for example, regression analysis [33]. Another problem is the linearity of the ARIMA model [34].

In their study [35], Chudy-Laskowska and Pisula used ANOVA to predict the frequency of traffic accidents. The disadvantage of this approach is that it makes additional assumptions, in particular the assumption of sphericity, which, if not met, can result in incorrect results. Traffic accident frequency is also predicted using neural networks. Because neural networks are often referred to as "black boxes," in which input data are entered and the model outputs results without awareness of the analysis, they have a number of drawbacks, including the requirement for prior expertise in the field [35,36], the dependence of the final result on the initial conditions of the network, and the inability to interpret the results conventionally [37].

Kumar et al. [38] used the Hadoop model as a state-of-the-art prediction technique. The disadvantage of this strategy is that it cannot handle small data sets [39]. The Garch model was used by Karlaftis and Vlahogianni [32] for prediction. The complicated model and complex form of this strategy pose problems [40,41]. The use of the ADF test by McIlroy and his team [42], however, has the disadvantage that its power is insufficient to detect autocorrelation of the random component [43].

To make predictions, the authors of the publication also used data mining algorithms, which usually struggle with multiple broad descriptions [44]. Another example of model combinations is the Sebego et al. [45] model set. Bloomfield's work [46] also suggests parametric models.

After analyzing the above-mentioned data, trend models were selected to predict the frequency of traffic accidents based on the nature of roads.

2. Materials and methods

Every year many people are fatally injured or killed on the roads. The number is still quite high, even if it is decreasing from year to year. Although there are fewer accidents on the roads since the epidemic, the number is still relatively high. Analyzing the data on the number of road accidents by road type on an annual and monthly basis, it can be seen that there is a clear oscillation with continuous stabilization. Poland still has a significantly higher number of accidents than in other European Union countries. In order to take preventive measures to reduce this value, it is important to know the accident rate forecast for the coming years for each road type (**Figures 1** and **2**).

Statistics from the Polish Police were used for the study. The study consisted of two parts. The first included a forecast of the number of road accidents for 2022–2031, based on an analysis of annual data from police statistics on the number of road accidents in Poland in 2000–2021. The second part of the study focused on monthly data from 2000–2021. In this case, a forecast was also set for the period from January 2022 to December 2023.

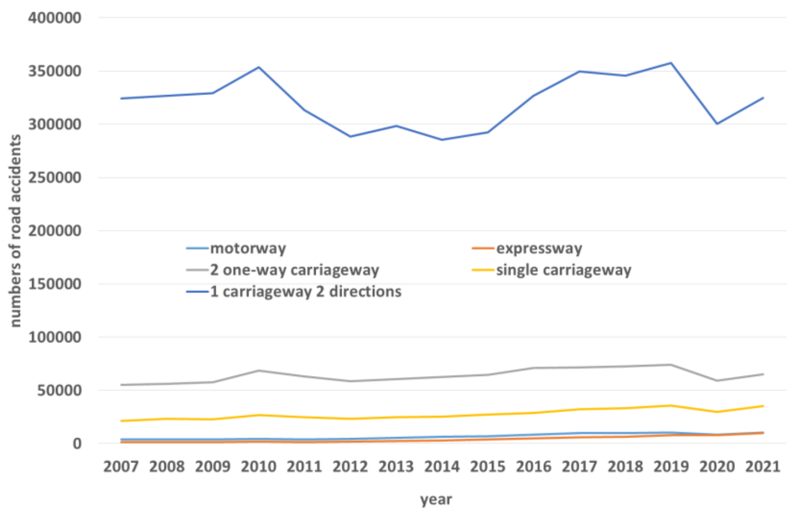

Figure 1. Number of accidents in Poland between 2007 and 2021 by year [47].

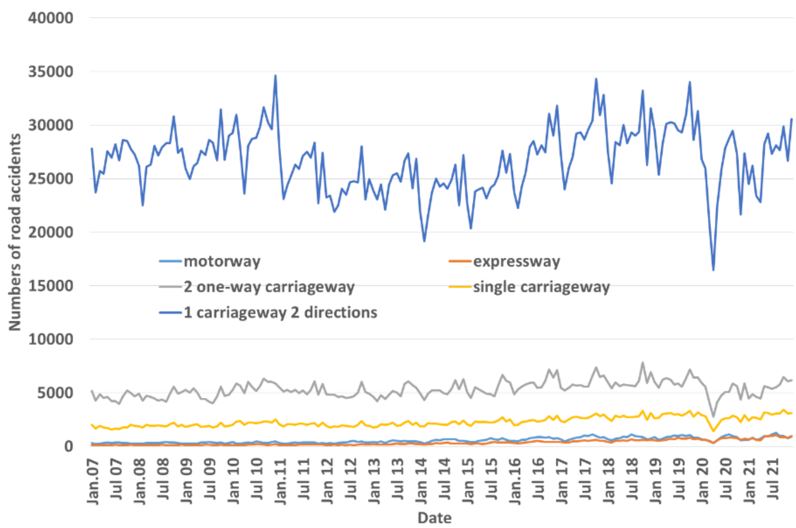

Figure 2. Number of accidents in Poland between 2007 and 2021 by month [47].

3. Forecasting the number of road accidents

For the examined road categories, the number of traffic accidents was projected using the trend models listed below:

- exponential
- linear
- logarithmic,
- polynomial of degree 2,
- polynomial of degree 3,
- polynomial of degree 4,
- polynomial of degree 5,
- polynomial of degree 6,
- power.

The first step was to find the mathematical equation for the trend models, which were analyzed for monthly and annual data. As can be observed, the R-square coefficient, which measures how well the model fits the annual data, is often good or adequate, but is usually bad or unsatisfactory for the monthly data. This is mainly due to the fact that the number of accidents on the different types of roads studied varies seasonally (**Tables 1–5**).

Table 1. Trend models for the motorway.

Data/model	Annual data	Monthly data
exponential	$y = 3047 \times 10^{0.089x}$ $R^2 = 0.9133$	$y = 260.5 \times 10^{0.0074x}$ $R^2 = 0.7402$
linear	$y = 572.53x + 2144.6$ $R^2 = 0.8933$	$y = 4.0266x + 196$ $R^2 = 0.6971$
logarithmic	$y = 2920.5\ln(x) + 1292.9$ $R^2 = 0.7107$	$y = 181.66\ln(x) - 204.82$ $R^2 = 0.4727$
polynomial of 2nd degree	$y = 9.7312x^2 + 416.83x + 2585.8$ $R^2 = 0.8971$	$y = 0.007x^2 + 2.7549x + 234.58$ $R^2 = 0.7017$
polynomial of 3rd degree	$y = -9.979x^3 + 249.23x^2 - 1165.8x + 5028.6$ $R^2 = 0.9527$	$y = -0.0004x^3 + 0.1248x^2 - 5.7933x + 365.29$ $R^2 = 0.738$
polynomial of 4th degree	$y = -0.4601x^4 + 4.745x^3 + 94.296x^2 - 571.62x + 4417.1$ $R^2 = 0.9542$	$y = -1 \times 10^{-6}x^4 - 2 \times 10{-}05x^3 + 0.0767x^2 - 3.8457x + 347.23$ $R^2 = 0.7386$
polynomial of 5th degree	$y = 0.4297x^5 - 17.65x^4 + 253.76x^3 - 1481.4x^2 + 3534.9x + 1244.4$ $R^2 = 0.9715$	$y = 1 \times 10^{-7}x^5 - 7 \times 10^{-5}x^4 + 0.0109x^3 - 0.6676x^2 + 15.613x + 225.27$ $R^2 = 0.7568$
polynomial of 6th degree	$y = 0.0969x^6 - 4.2198x^5 + 68.251x^4 - 511.29x^3 + 1873.7x^2 - 3005.6x + 5340.2$ $R^2 = 0.9819$	$y = 2 \times 10^{-9}x^6 - 1 \times 10^{-6}x^5 + 0.0002x^4 - 0.0162x^3 + 0.5651x^2 - 7.0734x + 328.47$ $R^2 = 0.7671$
potentiometric	$y = 2606.2x^{0.467}$ $R^2 = 0.7681$	$y = 118.69x^{0.345}$ $R^2 = 0.5401$

Table 2. Trend models for the expressway.

Data/model	Annual data	Monthly data
exponential	$y = 944.33 \times 10^{0.1555x}$ $R^2 = 0.9577$	$y = 82.685 \times 10^{0.013x}$ $R^2 = 0.9108$
linear	$y = 607.82x - 728.04$ $R^2 = 0.8937$	$y = 4.2531x - 40.358$ $R^2 = 0.8315$
logarithmic	$y = 2910.3\ln(x) - 1278.4$ $R^2 = 0.6264$	$y = 178.48\ln(x) - 407.3$ $R^2 = 0.4879$
polynomial of 2nd degree	$y = 51.037x^2 - 208.77x + 1585.6$ $R^2 = 0.9866$	$y = 0.0304x^2 - 1.2467x + 126.47$ $R^2 = 0.9232$
polynomial of 3rd degree	$y = -1.6854x^3 + 91.487x^2 - 476.08x + 1998.2$ $R^2 = 0.988$	$y = -5 \times 10^{-5}x^3 + 0.0434x^2 - 2.1932x + 140.94$ $R^2 = 0.9237$
polynomial of 4th degree	$y = -0.3519x^4 + 9.5764x^3 - 27.013x^2 - 21.583x + 1530.5$ $R^2 = 0.9888$	$y = -9 \times 10^{-7}x^4 + 0.0003x^3 + 0.0065x^2 - 0.6986x + 127.08$ $R^2 = 0.924$

Table 2. (*Continued*).

Data/model	Annual data	Monthly data
polynomial of 5th degree	$y = 0.1509x^5 − 6.3891x^4 + 97.031x^3 − 580.42x^2 + 1420.6x + 416.26$ $R^2 = 0.9907$	$y = 6 \times 10^{−8}x^5 − 3 \times 10^{−5}x^4 + 0.0044x^3 − 0.275x^2 + 6.6604x + 80.957$ $R^2 = 0.9268$
polynomial of 6th degree	$y = 0.0527x^6 − 2.3783x^5 + 40.339x^4 − 319.14x^3 + 1244.7x^2 − 2137.3x + 2644.3$ $R^2 = 0.9934$	$y = 1 \times 10^{−9}x^6 − 6 \times 10^{−7}x^5 + 0.0001x^4 − 0.0095x^3 + 0.3577x^2 − 4.9841x + 133.93$ $R^2 = 0.9297$
potentiometric	$y = 753.57x^{0.7903}$ $R^2 = 0.756$	$y = 22.585x^{0.5879}$ $R^2 = 0.6182$

Table 3. Trend models for road 1 carriageway 2 directions.

Data/model	Annual data	Monthly data
exponential	$y = 316{,}826 \times 10^{0.0013x}$ $R^2 = 0.0064$	$y = 26{,}237 \times 10^{0.0001x}$ $R^2 = 0.0047$
linear	$y = 471.7x + 317{,}334$ $R^2 = 0.0077$	$y = 4.9581x + 26{,}310$ $R^2 = 0.0077$
logarithmic	$y = −462.6\ln(x) + 321{,}968$ $R^2 = 0.0002$	$y = 55.766\ln(x) + 26{,}524$ $R^2 = 0.0003$
polynomial of 2nd degree	$y = 447.11x^2 − 6682.1x + 337{,}604$ $R^2 = 0.1094$	$y = 0.2673x^2 − 43.414x + 27778$ $R^2 = 0.0559$
polynomial of 3rd degree	$y = −100.58x^3 + 2861.1x^2 – 22{,}634x + 362{,}226$ $R^2 = 0.1809$	$y = −0.0036x^3 + 1.2366x^2 − 113.79x + 28{,}854$ $R^2 = 0.0739$
polynomial of 4th degree	$y = −61.577x^4 + 1869.9x^3 – 17{,}873x^2 + 56{,}888x + 280{,}395$ $R^2 = 0.5364$	$y = −0.0002x^4 + 0.0764x^3 − 8.091x^2 + 263.82x + 25{,}351$ $R^2 = 0.2152$
polynomial of 5th degree	$y = 6.4565x^5 − 319.84x^4 + 5611.1x^3 – 41{,}547x^2 + 118{,}585x + 232{,}728$ $R^2 = 0.5858$	$y = 3 \times 10^{−6}x^5 − 0.0014x^4 + 0.2628x^3 − 20.791x^2 + 595.88x + 23{,}270$ $R^2 = 0.2538$
polynomial of 6th degree	$y = 4.1643x^6 − 193.43x^5 + 3373.1x^4 − 27279x^3 + 102{,}695x^2 – 162{,}601x + 408{,}810$ $R^2 = 0.83$	$y = 1 \times 10^{−7}x^6 − 5 \times 10^{−5}x^5 + 0.0102x^4 − 0.8591x^3 + 30.331x^2 − 345.04x + 27{,}550$ $R^2 = 0.3831$
potentiometric	$y = 321{,}797x^{−0.003}$ $R^2 = 0.0007$	$y = 26{,}594x^{−9} \times 10^{−6}$ $R^2 = 6 \times 10^{−9}$

Table 4. Trend models for a one-way road.

Data/model	Annual data	Monthly data
exponential	$y = 20{,}957 \times 10^{0.0332x}$ $R^2 = 0.8151$	$y = 1765.9 \times 10^{0.0028x}$ $R^2 = 0.6167$
linear	$y = 929.4x + 20{,}250$ $R^2 = 0.7975$	$y = 6.5603x + 1713.4$ $R^2 = 0.6239$
logarithmic	$y = 4769.5\ln(x) + 18{,}814$ $R^2 = 0.6422$	$y = 299.17\ln(x) + 1046.9$ $R^2 = 0.4323$
polynomial of 2nd degree	$y = 35.149x^2 + 367.02x + 21{,}843$ $R^2 = 0.8143$	$y = 0.0231x^2 + 2.3759x + 1840.3$ $R^2 = 0.6406$

Table 4. (*Continued*).

Data/model	Annual data	Monthly data
polynomial of 3rd degree	$y = -6.3496x^3 + 187.54x^2 - 640.03x + 23{,}398$ $R^2 = 0.8219$	$y = -0.0002x^3 + 0.0875x^2 - 2.3014x + 1911.8$ $R^2 = 0.6443$
polynomial of 4th degree	$y = -4.1036x^4 + 124.96x^3 - 1194.2x^2 + 4659.4x + 17{,}944$ $R^2 = 0.8642$	$y = -1 \times 10^{-5}x^4 + 0.0044x^3 - 0.4523x^2 + 19.551x + 1709.1$ $R^2 = 0.6662$
polynomial of 5th degree	$y = 0.7362x^5 - 33.553x^4 + 551.57x^3 - 3893.7x^2 + 11{,}695x + 12{,}509$ $R^2 = 0.8813$	$y = 3 \times 10^{-7}x^5 - 0.0001x^4 + 0.0231x^3 - 1.7276x^2 + 52.897x + 1500.1$ $R^2 = 0.6842$
polynomial of 6th degree	$y = 0.3569x^6 - 16.397x^5 + 282.99x^4 - 2267.6x^3 + 8469.9x^2 - 12{,}407x + 27{,}602$ $R^2 = 0.9293$	$y = 1 \times 10^{-8}x^6 - 5 \times 10^{-6}x^5 + 0.001x^4 - 0.0825x^3 + 3.0863x^2 - 35.704x + 1903.2$ $R^2 = 0.7374$
potentiometric	$y = 19787x^{0.1737}$ $R^2 = 0.6819$	$y = 1313.9x^{0.1296}$ $R^2 = 0.4511$

Table 5. Trend models for a 2-lane one-way road.

Data/model	Annual data	Monthly data
exponential	$y = 57{,}354 \times 10^{0.0133x}$ $R^2 = 0.3751$	$y = 4775.2 \times 10^{0.0011x}$ $R^2 = 0.1684$
linear	$y = 845.66x + 57311$ $R^2 = 0.3659$	$y = 6.2031x + 4778.3$ $R^2 = 0.1848$
logarithmic	$y = 5225.1\ln(x) + 54{,}358$ $R^2 = 0.4271$	$y = 343.27\ln(x) + 3893.7$ $R^2 = 0.1886$
polynomial of 2nd degree	$y = -117.46x^2 + 2725.1x + 51{,}986$ $R^2 = 0.47$	$y = -0.0581x^2 + 16.717x + 4459.4$ $R^2 = 0.2198$
polynomial of 3rd degree	$y = -23.502x^3 + 446.59x^2 - 1002.4x + 57{,}739$ $R^2 = 0.5278$	$y = -0.0009x^3 + 0.1969x^2 - 1.7951x + 4742.5$ $R^2 = 0.2389$
polynomial of 4th degree	$y = -8.2774x^4 + 241.37x^3 - 2340.5x^2 + 9687.2x + 46{,}739$ $R^2 = 0.623$	$y = -2 \times 10^{-5}x^4 + 0.0074x^3 - 0.7796x^2 + 37.737x + 4375.8$ $R^2 = 0.2627$
polynomial of 5th degree	$y = 1.1445x^5 - 54.056x^4 + 904.53x^3 - 6536.9x^2 + 20{,}624x + 38{,}290$ $R^2 = 0.646$	$y = 5 \times 10^{-7}x^5 - 0.0003x^4 + 0.0468x^3 - 3.4638x^2 + 107.92x + 3935.9$ $R^2 = 0.2891$
polynomial of 6th degree	$y = 0.957x^6 - 44.793x^5 + 794.66x^4 - 6654.3x^3 + 26{,}612x^2 - 4{,}3998x + 78{,}757$ $R^2 = 0.8371$	$y = 3 \times 10^{-8}x^6 - 1 \times 10^{-5}x^5 + 0.0026x^4 - 0.2342x^3 + 9.3402x^2 - 127.74x + 5007.9$ $R^2 = 0.4137$
potentiometric	$y = 54{,}667x^{0.083}$ $R^2 = 0.4466$	$y = 4042.5x^{0.0637}$ $R^2 = 0.1802$

The information from **Tables 1–5** was then used to calculate the anticipated number of traffic incidents. This covered the period from 2022 to 2031 for the yearly data and January 2022 to December 2023 for the monthly data. In this instance, the projection was based on a weighted average of the value data from the current and previous series. This method's forecasting outcome is based on the model selection and model fit.

The errors of the forecasts that had already expired were calculated for the received forecasts using Equations (1–5) in the next step:

- ME—mean error

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

- MAE—mean everage error

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

- 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

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

- MSE—mean square error

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

where:

n—the length of the forecast horizon,

Y—observed value of road accidents,

Y_p—forecasted value of road accidents.

In order to predict the frequency of traffic accidents based on the road types studied, trend models were selected for which the mean percentage error and mean absolute percentage error were the smallest. Based on this, it was discovered that the exponential model (2 one-way roadways and 1 two-way roadway) and the third-degree polynomial model (highway, expressway) provided the best fit to the annual data in most situations. However, the exponential model performed best for a road with one lane. The largest MAPE error for all road types analyzed was 0.45%, suggesting a very strong model fit.

However, for the monthly data, a highway, two one-way roads and one two-way road performed better with the exponential model. However, for the expressway, the exponential model and the linear model performed better, respectively. The maximum MAPE error in this analysis was 3%, indicating a very good model fit (**Tables 6** and **7**). On this basis, the predicted monthly and annual number of accidents in the following years was calculated (**Figures 3** and **4**). Based on **Figures 3** and **4**, it can be predicted that the number of accidents on Polish roads will continue to stabilize. This is mainly due to the fact that new roads, especially highways, have begun to appear on the road map of Poland. Note that the epidemic has led to significant adjustments in the forecasts. As shown in **Figure 4**, trend models are not fully adequate for analyzing the number of road accidents, because they do not take into account the seasonality that occurs in road accidents.

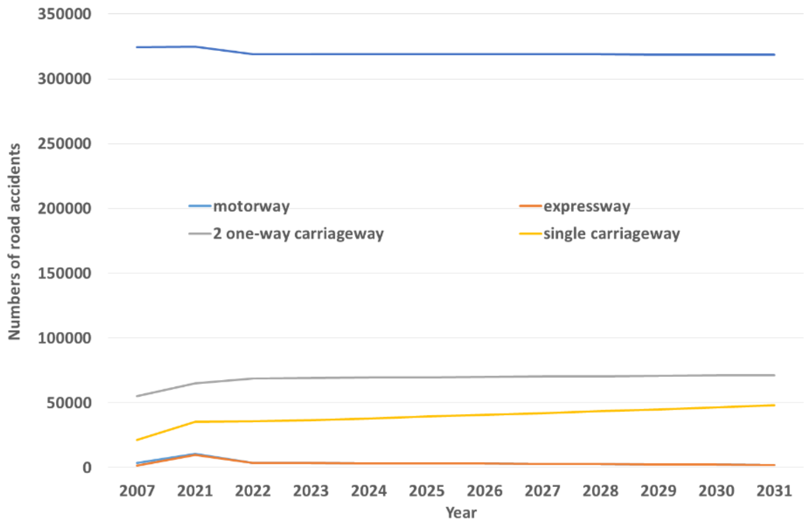

Figure 3. Forecasting number of road accidents for 2022–2031.

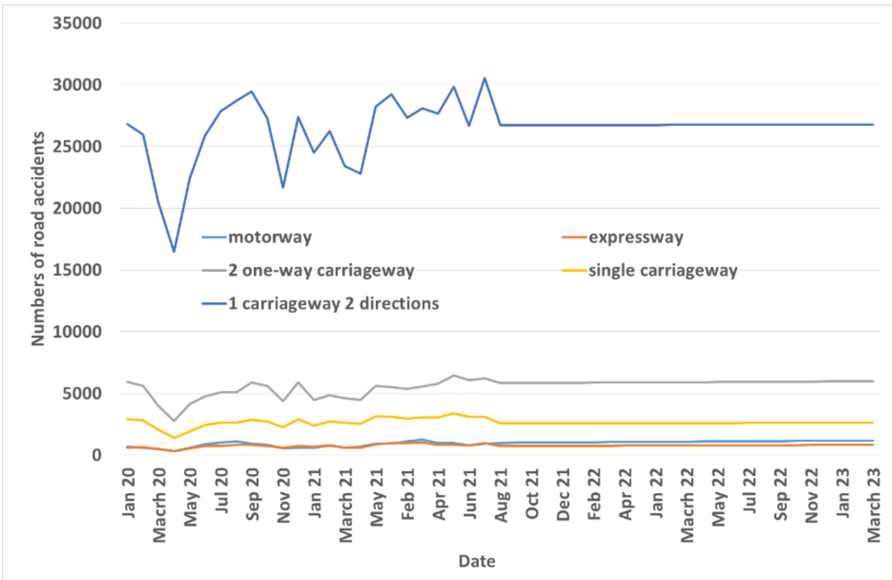

Figure 4. Forecasting number of road accidents for 2022–2023.

Table 6. Summary of errors for annual data.

Error/type of road	motorway	expressway	2 one-way carriageway	one-way carriageway	1 carriageway 2 directions
Best model	3 degree polynomial model	3 degree polynomial model	power model	exponential model	power model
ME	0.506667	0.032	160.6781	70.68791	1100.824
MPE	459.524	219.6597	3892.145	1594.736	19856.25
Sum of squares	4864491	1391739	3.14×10^8	58242425	8.13×10^9
MSE	324299.4	92782.6	20936211	3882828	5.42×10^8
MAPE [%]	0.450046	0.439782	0.242333	0.231562	0.184612
MAE [%]	7.256851	6.662241	5.974687	5.612987	6.223259

Table 7. Summary of errors for monthly data.

Error/type of road	motorway	expressway	2 one-way carriageway	one-way carriageway	1 carriageway 2 directions
Best model	exponential model	linear model	exponential model	power model	exponential model
ME	2306.8818	0.003105556	2306.8818	22.90197369	283.1256091
MPE	2306.8818	80.81473222	2306.8818	248.1568511	2370.930111
Sum of squares	979,361,291.1	1,781,287.143	979,361,291.1	17,965,755.18	1,558,686,454
MSE	5,440,896.062	9896.039681	5,440,896.062	99,809.75102	8,659,369.187
MAPE [%]	3.02852656	1.547122305	0.666652737	0.927991546	0.20568071
MAE [%]	473.8053144	36.28240696	43.76224349	10.77221508	9.068673406

4. Conclusions

Using selected trend models and Excel, the predicted number of accidents in Poland was determined for the routes studied. According to the results, we can still predict that the number of road accidents will stabilize in the coming years. It should be noted that the epidemic has affected the results, and in the case of its maintenance and the introduction of traffic restrictions, the proposed model may not be sufficient. The average maximum error rate of 0.45% for annual data and 3% for monthly data may indicate the use of an effective forecasting technique.

As can be observed, if there is seasonality in the monthly number of traffic accidents, trend models do not perform well. However, the results are at a high level for annual data. The advantage of trend models is how quickly they can produce forecasts.

Future actions to reduce the number of accidents in the country under study can be developed using the road accident forecast developed in this article. These measures could, for example, begin on 1 January 2022, with the imposition of stricter penalties for traffic offenses on Polish roads.

In his future research, the author intends to use more techniques to forecast the number of traffic accidents, as well as a wider range of variables that affect the accident rate in Poland. These could include traffic volume, day of the week or age of the accident victim.

Conflict of interest: The author declares no conflict of interest.

References

1. WHO. The Global status on road safety. Available online: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023 (accessed on 16 October 200).
2. Tambouratzis T, Souliou D, Chalikias M, et al. 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. doi: 10.2478/jaiscr-2014-0023
3. Zhu L, Lu L, Zhang W, et al. Analysis of Accident Severity for Curved Roadways Based on Bayesian Networks. Sustainability. 2019; 11(8): 2223. doi: 10.3390/su11082223
4. Arteaga C, Paz A, Park J. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. Safety Science. 2020; 132: 104988. doi: 10.1016/j.ssci.2020.104988
5. 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. doi: 10.1016/j.ssci.2021.105522

6. Gorzelańczyk P, Pyszewska D, Kalina T, et al. Analysis of road traffic safety in the Piła poviat. Scientific Journal of Silesian University of Technology Series Transport. 2020; 107: 33–52. doi: 10.20858/sjsutst.2020.107.3
7. Chen C. Analysis and Forecast of Traffic Accident Big Data. In: Proceedings of the ITM Web of Conferences; 5 September 2017. 12, pp. 04029.
8. Khaliq KA, Chughtai O, Shahwani A, et al. Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing. Electronics. 2019; 8(8): 896. doi: 10.3390/electronics8080896
9. Rajput H, Som T, Kar S. An Automated Vehicle License Plate Recognition System. Computer. 2015; 48(8): 56–61. doi: 10.1109/mc.2015.244
10. Zheng Z, Wang C, Wang P, et al. Framework for fusing traffic information from social and physical transportation data. PLOS ONE. 2018; 13(8): e0201531. doi: 10.1371/journal.pone.0201531
11. Abdullah E, Emam A. Traffic Accidents Analyzer Using Big Data. In: Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence (CSCI); 7–9 December 2015; Las Vegas, NV, USA. pp.392–397.
12. 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. doi: 10.1016/j.trpro.2017.12.113
13. Bąk I, Cheba K, Szczecińska B. The statistical analysis of road traffic in cities of Poland. Transportation Research Procedia. 2019; 39: 14–23. doi: 10.1016/j.trpro.2019.06.003
14. Chand A, Jayesh S, Bhasi AB. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. Materials Today: Proceedings. 2021; 47: 5135–5141. doi: 10.1016/j.matpr.2021.05.415
15. Helgason AF. Fractional Integration Methods and Short Time Series: Evidence from a Simulation Study. Political Analysis. 2016; 24(1): 59–68. doi: 10.1093/pan/mpv026
16. Lavrenz SM, Vlahogianni EI, Gkritza K, et al. Time series modeling in traffic safety research. Accident Analysis & Prevention. 2018; 117: 368–380. doi: 10.1016/j.aap.2017.11.030
17. Kowalski L. Forecasting based on time series. Available online: http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf (accessed on 17 May 2024).
18. Sunny CM, Nithya S, Sinshi KS, et al. Forecasting of Road Accident in Kerala: A Case Study. In: Proceedings of the 2018 International Conference on Data Science and Engineering (ICDSE). 1–5 August 2018; Kochi, India.
19. Prochazka J, Camaj M. Modelling the number of road accidents of uninsured drivers and their severity. In: Proceedings of the 32nd International Academic Conference, Geneva; 27 June 2017; Geneva, Switzerland.
20. Dudek G. Exponential smoothing models for short-term power system load forecasting. Energy Market. 2013; 106(3): 14–19.
21. Al-Madani HMN. Global road fatality trends' estimations based on country-wise micro level data. Accident Analysis & Prevention. 2018; 111: 297–310. doi: 10.1016/j.aap.2017.11.035
22. Monedero BD, Gil-Alana LA, Martínez MCV. Road accidents in Spain: Are they persistent? IATSS Research. 2021; 45(3): 317–325. doi: 10.1016/j.iatssr.2021.01.002
23. Wójcik A. Autoregressive vector models as a response to the critique of multi-equation structural econometric models. Publishing house of the University of Economics in Katowice; 2014.
24. Mamczur M. Machine learning How does linear regression work? And is it worth using? (Polish). Uczenie Maszynowe. 2020.
25. Piłatowska M. The choice of the order of autoregression depending on the parameters of the generating model. Econometrics. 2012; 4(38).
26. Biswas AA, Mia MJ, Majumder A. Forecasting the Number of Road Accidents and Casualties using Random Forest Regression in the Context of Bangladesh. In: Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT); 6–8 July 2019; Kanpur, India.
27. Fijorek K, Mróz K, Niedziela K, Fijorek D. Forecasting electricity prices on the day-ahead market using data mining methods. Energy Market. 2010; 12.
28. Trevor H, Robert T, Jerome F. The Elements of Statistical Learning. Wyd. 2. Springer; 2008.
29. Chudy-Laskowska K, Pisula T. Forecast of the number of road accidents in Poland. Logistics. 2014; 6.
30. Kashpruk N. Comparative research of statistical models and soft computing for identification of time series and forecasting. Opole University of Technology; 2010.

31. 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.
32. Karlaftis MG, Vlahogianni EI. Memory properties and fractional integration in transportation time-series. Transportation Research Part C: Emerging Technologies. 2009; 17(4): 444–453. doi: 10.1016/j.trc.2009.03.001
33. Łobejko S. Time series analysis and forecasting with SAS. Main business school in Warsaw; 2015.
34. Dudek G. Forecasting Time Series with Multiple Seasonal Cycles Using Neural Networks with Local Learning. In: Proceedings of the 12th International Conference, ICAISC 2013; 9–13 June 2013; Zakopane, Poland.
35. Chudy-Laskowska K. Pisula T. Forecasting the number of road accidents in Podkarpacie (Polish). Logistics. 2015; 4: 2782–2796.
36. Wrobel MS. Application of neural fuzzy systems in chemistry [PhD thesis]. University of Silesia; 2017.
37. Vijay K. https://www.spiceworks.com/tech/big-data/articles/what-is-data-mining/ (accessed on 13 April 2024).
38. Kumar PS, Viswanadham V, Bharathi B. Analysis of Road Accident. IOP Conference Series: Materials Science and Engineering. 2019; 590(1): 012029. doi: 10.1088/1757-899x/590/1/012029
39. DataFlair Team. Top Advantages and Disadvantages of Hadoop 3. Available online: https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/ (accessed on 12 March 2024).
40. Fiszeder P. GARCH class models in empirical financial research. Scientific Publishers of the Nicolaus Copernicus University; 2009.
41. Perczak G, Fiszeder P. GARCH model—using additional information on minimum and maximum prices. Bank and Credit. 2014.
42. McIlroy RC, Plant KA, Hoque MS, et al. Who is responsible for global road safety? A cross-cultural comparison of Actor Maps. Accident Analysis & Prevention. 2019; 122: 8–18. doi: 10.1016/j.aap.2018.09.011
43. Muck J. Econometrics. Modeling of time series. Stationary. Unit root tests. ARDL models. Co-integration. Available online: http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja5.pdf (accessed on 12 March 2024).
44. Li L, Shrestha S, Hu G. Analysis of road traffic fatal accidents using data mining techniques. In: Proceedings of the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA); 7–9 June 2017; London, United Kingdom. pp.363–370.
45. Sebego M, Naumann RB, Rudd RA, et al. 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. doi: 10.1016/j.aap.2014.02.017
46. Bloomfield P. An exponential model for the spectrum of a scalar time series. Biometrika. 1973; 60(2): 217–226. doi: 10.1093/biomet/60.2.217
47. Symon E, Rzepka P. Available online: https://statystyka.policja.pl/st/ruch-drogowy/76562,Wypadki-drogowe-raporty-roczne.html (accessed on 23 April 2024).