Technical Sciences, 2024, 27, 175–192

UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Yearbook peer-reviewed scientific journal

ISSN 1505-4675
e-ISSN 2083-4527

TECHNICAL SCIENCES

Homepage: https://czasopisma.uwm.edu.pl/index.php/ts/

DOI: https://doi.org/10.31648/ts.10436

# FORECASTING THE NUMBER OF ROAD ACCIDENTS IN POLAND USING TREND MODELS DEPENDING ON THE DAYS OF THE WEEK

***Piotr Gorzelanczyk***

ORCID: 0000-0001-9662-400X
Department of Transportation
Stanislaw Staszic State University of Applied Sciences in Pila

Received 06 August 2024, accepted 25 September 2024, available online 2 October 2024.

Keywords: traffic accident, forecasting, trend models, days of the week.

Abstract

Every year a very large number of people die on the roads. Although the number decreases year by year, it remains high. The pandemic has reduced the number of road accidents, but the value is still very high. For this reason, it is necessary to know on which days the highest number of traffic accidents occur, and to know the forecast of accidents by day of the week for the coming years, so that we can do everything possible to minimize the number of traffic accidents. The purpose of the article is to make a forecast of the number of road accidents in Poland according to the day of the week. The research was divided into two parts. The first was the analysis of annual data from the Police statistics on the number of road accidents in Poland in 2000-2021, and on this basis the forecast of the number of road accidents for 2022-2031 was determined. The second part of the research, dealt with monthly data from 2000-2021. Again, the analyzed forecast for the period January 2022 – December 2023 was determined. The results of the study indicate that we can still expect a decline in the number of accidents in the coming years, which is particularly evident when analyzing annual data. It is worth noting that the prevailing pandemic distorts the results obtained. The research was conducted in MS Excel, using selected trend models.

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

## Introduction

Road accidents are incidents that cause not only injury or death to road users, but also property damage. According to the WHO, about 1.3 million people die each year as a result of road accidents. For most countries around the world, the economic cost of road accidents accounts for about 3% of their GDP. Road accidents are the leading cause of death for minors and young people aged 5-29 (*Global status on road safety* 2018) The UN General Assembly has set an ambitious goal of halving road deaths and injuries by 2030.

The extent of a traffic accident is an attribute to determine its severity. Predicting accident severity is important for relevant authorities when designing transportation safety policies to eliminate accidents, reduce injuries, fatalities and property losses (TAMBOURATZIS et al. 2014, ZHU et al. 2019). Identification of critical factors affecting accident severity is a prerequisite for countermeasures to eliminate and mitigate accident severity (ARTEAGA et al. 2020). YANG et al. (2022) propose a multi-node DNN (Deep Neutral Network) framework for predicting different levels of injury severity, death and property loss. This allows for a comprehensive and accurate analysis of the severity of traffic accidents.

There are several sources of accident data. Most commonly, they are collected and analyzed by government authorities through relevant government agencies. Data collection is done through police reports, insurance databases or hospital records. Partial information about traffic accidents is then processed for the transportation sector on a larger scale (GORZELANCZYK et al. 2020, 2022).

Intelligent transportation systems are currently the most important source of data related to road accident analysis and prediction. This data can be processed through the use of GPS devices in vehicles (CHEN 2017). Microwave vehicle detection systems on roadsides can continuously record vehicle data (speed, traffic volume, vehicle type, etc.) (KHALIQ et al. 2019). A vehicle license plate recognition system can also collect large amounts of traffic data over a monitored period (RAJPUT et al. 2015). Another source of data for obtaining traffic and accident information can be social media, but its relevance may be insufficient due to the incompetence of those reporting (ZHENG et al. 2018).

To ensure the relevance of accident data, it is necessary to work with several data sources, which must be properly confronted. Combining different data sources by consolidating heterogeneous Mahout Data Mining data on traffic accidents helps increase the accuracy of analysis results (ABDULLAH, EMAM 2015).

A statistical study to assess the severity, finding the relationship between road accidents and road users was conducted by VILACA et al. (2017). The result of the study is a proposal to improve road safety standards and adopt other policies related to transportation safety.

BAK et al. (2019) conducted a statistical study of road safety in a selected region of Poland based on the number of road accidents, the pace of learning about the causes of road accidents. The study used multivariate statistical analysis to examine the safety aspect of those responsible for accidents.

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

Various methods for forecasting the number of accidents can be found in the literature. The most commonly used methods for forecasting the number of traffic accidents are time series methods (HELGASON 2016, LAVRENZ et al. 2018), which have the disadvantage of not being able to assess the quality of the forecast based on expired forecasts and frequent autocorrelation of the residual component (*Forecasting based on time series* 2022). PROCHÁZKA et al. (2017) used a multiple seasonality model for forecasting, and SUNNY et al. (2018) the Holt-Winters exponential smoothing method. Its limitations include the inability to introduce exogenous variables into the model (DUDEK 2013, SZMUKSTA-ZAWADZKA, ZAWADZKI 2009).

A vector autoregression model has also been used to forecast the number of traffic accidents, the drawback of which is the need to have a large number of observations of variables in order to correctly estimate their parameters (WOJCIK 2014), as well as the autoregression models of MONEDEROA et al. (2021) for analyzing the number of fatalities and AL-MADANI (2018), curve regression models. These, in turn, require only simple linear relationships (MAMCZUR 2022), and an order of autoregression (assuming the series is already stationary) (PILATOWSKA 2012).

BISWAS et al. (2019) used Random Forest regression to predict the number of traffic accidents. In this case, the data contain groups of correlated features of similar significance to the original data, smaller groups are favored over larger ones (*Random forest* 2022), and there is instability in the method and pin prediction (FIJOREK et al. 2010). CHUDY-LASKOWSKA and PISULA (2014) used an autoregressive quadratic trend model, a single-argument periodic trend model and an exponential equalization model for the forecasting issue under consideration. A moving average model can also be used to forecast the issue at hand, which has the disadvantages of low forecast accuracy, loss of data in the sequence and lack of consideration of trends and seasonal effects (KASHPRUK 2010). PROCHOZKA and CAMEJ (2017) used the GARMA method, in which certain constraints are imposed in the parameter space to guarantee the stationarity of the process. Very often, the ARMA model for a stationary process or ARIMA or SARIMA for a non-stationary process is used for forecasting (SUNNY et al. 2018, PROCHAZKA, CAMAJ 2017, DUTTA et al. 2020, KARLAFTIS, VLAHOGIANNI 2009). There is great flexibility in the models discussed, but this is also a disadvantage, as good model identification requires more experience from the researcher than, for example, regression analysis (LOBEJKO 2015). Another disadvantage is the linear nature of the ARIMA model (DUDEK 2013).

Chudy-Laskowska and Pisula (2015) in their work used ANOVA to predict the number of traffic collisions. The disadvantage of this method is that additional assumptions are made, especially the assumption of sphericity, the violation of which can lead to erroneous conclusions (*Road safety assessment handbook* 2022). Neural network models are also used to forecast the number of traffic accidents. The disadvantages of ANNs are the need for experience in the field (Chudy-Laskowska, Pisula 2015) and the dependence of the final solution on the initial conditions of the network, as well as the inability to interpret in the traditional way, as ANNs are usually referred to as a black box in which the user provides the input and the model provides the output without knowledge of the analysis (Data mining techniques 2022).

A new prediction method is the use of the Hadoop model by Kumar et al. (2019). The disadvantage of this method is its inability to work with small data files (Top Advantages and Disadvantages of Hadoop 3 DataFlair, 2022). Karlaftis and Vlahogianni (2009) used the Garch model for prediction. The disadvantages of this method are its complex form and complicated model (Perczak, Fiszeder 2014, Fiszeder 2009). On the other hand, McIlroy et al. (2019) and his team used the ADF test, which has the disadvantage of poor power for autocorrelation of the random component (Muck 2022).

Authors of publications (Shetty et al. 2017, Li et al. 2017) have also used Data-Mining techniques for forecasting, which usually have the disadvantage of huge sets of general descriptions (Marcinkowska 2015). One also encounters the combination of models proposed by Sebego et al. (2008) as a combination of different models. Parametric models are also proposed in the work of Bloomfield (1973).

Analyzing the above information, trend models were selected to forecast the number of traffic accidents on each day of the week. For this purpose, based on statistical data from the Municipal Police Station, various trend models were selected for which forecasts of the number of traffic accidents were made, and then the best one was selected for which the forecast error was the smallest. The study was conducted in order to find out the answer to how the number of traffic accidents will change in the coming years, and on which days there will be the most accidents. The results obtained will form the basis for carrying out preventive measures to reduce the number of traffic accidents. In the paper (Jurkovic et al. 2022), the impact of the COVID-19 pandemic was taken into account.

## Number of road accidents

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

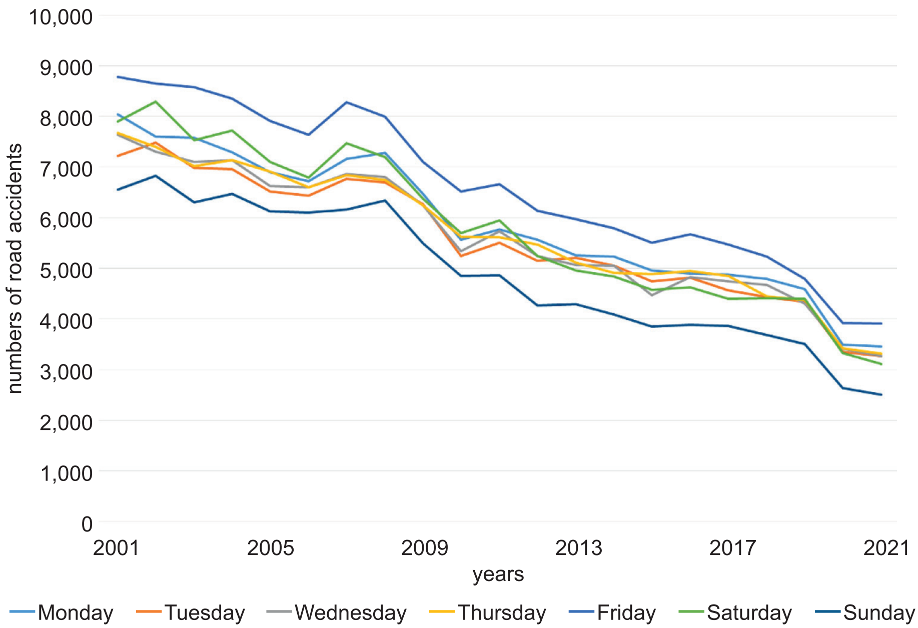


Fig. 1. Number of accidents in Poland from 2001 to 2021 by year
Source: based on Statistic Road Accident (2022).

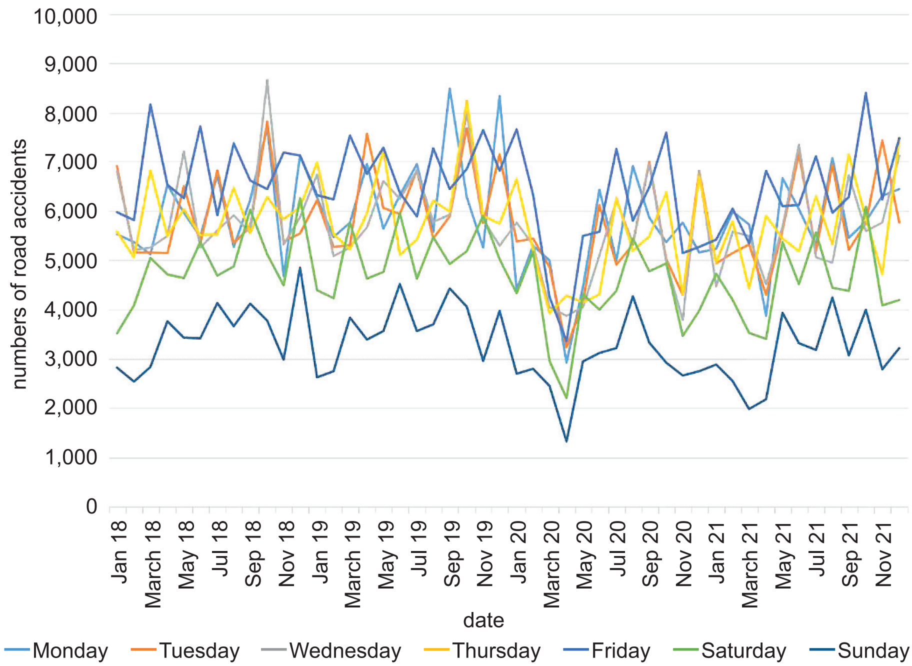


Fig. 2. Number of accidents in Poland in 2018-2021 by month
Source: based on Statistic Road Accident (2022).

## Forecasting the number of traffic accidents

Various trend methods can be found in the literature, which can be described by the following formulas (1-6):

– linear function:

$$Y_t = \propto_0 + a_1 t \tag{1}$$

– parabolic function:

$$Y_t = \propto_0 + a_1 t + a_2 t^2 \tag{2}$$

– power function:

$$Y_t = \propto_0 \cdot t^{a_1} \tag{3}$$

– exponenional function:

$$Y_t = \propto_0 \cdot \propto_1^t \tag{4}$$

– hyperbolic function:

$$Y_t = \propto_0 + \propto_1 \cdot \frac{1}{t} \tag{5}$$

– logistic function:

$$Y_t = \frac{\propto}{1 + \beta e^{-rt}} \tag{6}$$

where:

$Y_t$ – forecasting the number of traffic accidents over time $t$,
$t$ – time variable ($t$ = 1, 2, …, $n$),
$\propto_0$, $\propto_1$, $\beta$ – parameters of the linear trend function.

Among the aforementioned functions, the following trend models were used to forecast the number of traffic accidents for each day of the week:
– exponential,
– linear,
– logarithmic,
– polynomial of 2nd degree,
– polynomial of 3rd degree,
– polynomial of 4th degree,
– polynomial of 5th degree,
– polynomial of 6th degree,
– potentiometric.

In the first step, for the analyzed trend models, the mathematical formula of the analyzed data on an annual and monthly basis was determined. We can see that the analyzed $R$-square coefficient (a measure of the quality of model fit) for annual data there is a good or satisfactory fit, while for monthly data there is a poor and unsatisfactory fit. This is primarily due to the seasonality of the number of traffic accidents by day of the week, with the least number of accidents on Sunday and the most on Friday (Table 1-7).

Table 1

Trend models for Monday

| Data / model | Annual data | Monthly data |
|---|---|---|
| 1 | 2 | 3 |
| Exponential | $y = 8{,}653.1e^{-0.038x}$ | $y = 1{,}642.3e^{3E\text{-}05x}$ |
| | $R^2 = 0.9189$ | $R^2 = 0.0666$ |
| Linear | $y = -211.59x + 8{,}206.1$ | $y = 0.1638x - 1{,}409.6$ |
| | $R^2 = 0.9448$ | $R^2 = 0.0733$ |
| Logarithmic | $y = -1{,}481\ln(x) + 9{,}078.4$ | $y = 6{,}793.2\ln(x) - 66{,}842$ |
| | $R^2 = 0.8034$ | $R^2 = 0.0723$ |

cont. Table 1

| 1 | 2 | 3 |
|---|---|---|
| Polynomial of 2nd degree | $y = -1.3271x^2 - 182.39x + 8{,}094.1$ | $y = 5E\text{-}05x^2 - 3.6954x + 79{,}144$ |
| | $R^2 = 0.9459$ | $R^2 = 0.085$ |
| Polynomial of 3rd degree | $y = -0.1948x^3 + 5.1003x^2 - 240.28x + 8{,}212.4$ | $y = -2E\text{-}08x^3 + 0.0023x^2 - 99.515x + 1E{+}06$ |
| | $R^2 = 0.9465$ | $R^2 = 0.0885$ |
| Polynomial of 4th degree | $y = -0.1011x^4 + 4.2541x^3 - 58.844x^2 + 89.891x + 7{,}773.9$ | $y = -5E\text{-}11x^4 + 9E\text{-}06x^3 - 0.5408x^2 + 15{,}020x - 2E{+}08$ |
| | $R^2 = 0.9513$ | $R^2 = 0.1425$ |
| Polynomial of 5th degree | $y = -0.0308x^5 + 1.5931x^4 - 29.304x^3 + 228.6x^2 - 903.12x + 8{,}738.8$ | $y = 2E\text{-}14x^5 - 5E\text{-}09x^4 + 0.0004x^3 - 16.707x^2 + 352{,}472x - 3E{+}09$ |
| | $R^2 = 0.9627$ | $R^2 = 0.1612$ |
| Polynomial of 6th degree | $y = 0.0026x^6 - 0.2045x^5 + 5.9831x^4 - 82.358x^3 + 539.86x^2 - 1{,}693.1x + 9{,}345.8$ | $y = 2^{E}\text{-}17x^6 - 5^{E}\text{-}12x^5 + 6^{E}\text{-}07x^4 - 0.0309x^3 + 962.35x^2 - 2^{E}{+}07x + 1^{E}{+}11$ |
| | $R^2 = 0.9649$ | $R^2 = 0.1942$ |
| Potentiometric | $y = 9{,}906x^{-0.254}$ | $y = 0.0206x^{1.172}$ |
| | $R^2 = 0.7279$ | $R^2 = 0.0656$ |

Table 2

Trend models for Tuesday

| Data / model | Annual data | Monthly data |
|---|---|---|
| Exponential | $y = 8{,}141.1e^{-0.037x}$ | $y = 1{,}471.4e^{3E\text{-}05x}$ |
| | $R^2 = 0.9155$ | $R^2 = 0.086$ |
| Linear | $y = -195.65x + 7{,}721.6$ | $y = 0.172x - 1{,}863.8$ |
| | $R^2 = 0.946$ | $R^2 = 0.0936$ |
| Logarithmic | $y = -1{,}345\ln(x) + 8{,}475.6$ | $y = 7{,}136.4\ln(x) - 70{,}607$ |
| | $R^2 = 0.776$ | $R^2 = 0.0924$ |
| Polynomial of 2nd degree | $y = -2.4557x^2 - 141.63x + 7{,}514.5$ | $y = 4E\text{-}05x^2 - 3.512x + 75{,}033$ |
| | $R^2 = 0.9503$ | $R^2 = 0.1059$ |
| Polynomial of 3rd degree | $y = -0.1333x^3 + 1.9422x^2 - 181.23x + 7{,}595.4$ | $y = -2E\text{-}08x^3 + 0.0021x^2 - 88.547x + 1E{+}06$ |
| | $R^2 = 0.9507$ | $R^2 = 0.1091$ |
| Polynomial of 4th degree | $y = -0.1206x^4 + 5.1717x^3 - 74.308x^2 + 212.47x + 7{,}072.5$ | $y = -4E\text{-}11x^4 + 7E\text{-}06x^3 - 0.4358x^2 + 12{,}100x - 1E{+}08$ |
| | $R^2 = 0.9585$ | $R^2 = 0.1497$ |
| Polynomial of 5th degree | $y = -0.0167x^5 + 0.7956x^4 - 12.976x^3 + 81.137x^2 - 324.53x + 7{,}594.3$ | $y = 2E\text{-}14x^5 - 4E\text{-}09x^4 + 0.0003x^3 - 13.984x^2 + 294{,}904x - 2E{+}09$ |
| | $R^2 = 0.9624$ | $R^2 = 0.1649$ |
| Polynomial of 6th degree | $y = 0.0014x^6 - 0.1086x^5 + 3.1195x^4 - 41.061x^3 + 245.91x^2 - 742.7x + 7{,}915.6$ | $y = 3^{E}\text{-}17x^6 - 6^{E}\text{-}12x^5 + 7^{E}\text{-}07x^4 - 0.0363x^3 + 1{,}134.7x^2 - 2^{E}{+}07x + 1^{E}{+}11$ |
| | $R^2 = 0.9631$ | $R^2 = 0.2174$ |
| Potentiometric | $y = 9{,}221.8x^{-0.245}$ | $y = 0.0076x^{1.2632}$ |
| | $R^2 = 0.7039$ | $R^2 = 0.0849$ |

Table 3

Trend models for Wednesday

| Data / model | Annual data | Monthly data |
|---|---|---|
| Exponential | $y = 8{,}300.1e^{-0.038x}$ | $y = 1{,}732.9e^{3E\text{-}05x}$ |
| | $R^2 = 0.9155$ | $R^2 = 0.0643$ |
| Linear | $y = -202.41x + 7{,}866.7$ | $y = 0.1514x - 972.78$ |
| | $R^2 = 0.9456$ | $R^2 = 0.0692$ |
| Logarithmic | $y = -1{,}408\ln(x) + 8{,}682.9$ | $y = 6{,}278.3\ln(x) - 61{,}444$ |
| | $R^2 = 0.7945$ | $R^2 = 0.0682$ |
| Polynomial of 2nd degree | $y = -1.6298x^2 - 166.55x + 7{,}729.3$ | $y = 4E\text{-}05x^2 - 3.5491x + 76{,}270$ |
| | $R^2 = 0.9474$ | $R^2 = 0.081$ |
| Polynomial of 3rd degree | $y = -0.1139x^3 + 2.1295x^2 - 200.41x + 7{,}798.4$ | $y = -3E\text{-}08x^3 + 0.0033x^2 - 140.89x + 2E{+}06$ |
| | $R^2 = 0.9476$ | $R^2 = 0.089$ |
| Polynomial of 4th degree | $y = -0.1174x^4 + 5.0496x^3 - 72.087x^2 + 182.79x + 7{,}289.5$ | $y = -4E\text{-}11x^4 + 7E\text{-}06x^3 - 0.4615x^2 + 12{,}797x - 1E{+}08$ |
| | $R^2 = 0.9545$ | $R^2 = 0.1327$ |
| Polynomial of 5th degree | $y = -0{,}0312x^5 + 1.599x^4 - 28.948x^3 + 219.12x^2 - 823.21x + 8{,}267$ | $y = 2E\text{-}14x^5 - 4E\text{-}09x^4 + 0.0004x^3 - 15.311x^2 + 322{,}766x - 3E{+}09$ |
| | $R^2 = 0.9674$ | $R^2 = 0.1501$ |
| Polynomial of 6th degree | $y = 0.0023x^6 - 0.1841x^5 + 5.4642x^4 - 75.659x^3 + 493.17x^2 - 1{,}518.7x + 8{,}801.4$ | $y = 3^{E}\text{-}17x^6 - 8^{E}\text{-}12x^5 + 9^{E}\text{-}07x^4 - 0.0489x^3 + 1{,}527x^2 - 3^{E}{+}07x + 2^{E}{+}11$ |
| | $R^2 = 0.9692$ | $R^2 = 0.2403$ |
| Potentiometric | $y = 9{,}476.2x^{-0.253}$ | $y = 0.0414x^{1.105}$ |
| | $R^2 = 0.7183$ | $R^2 = 0.0633$ |

Table 4

Trend models for Thursday

| Data / model | Annual data | Monthly data |
|---|---|---|
| 1 | 2 | 3 |
| Exponential | $y = 8{,}361.2e^{-0.037x}$ | $y = 1{,}401.3e^{3E\text{-}05x}$ |
| | $R^2 = 0.9272$ | $R^2 = 0.0943$ |
| Linear | $y = -202.52x + 7{,}918.4$ | $y = 0.1772x - 2{,}057$ |
| | $R^2 = 0.9617$ | $R^2 = 0.0955$ |
| Logarithmic | $y = -1{,}403\ln(x) + 8{,}723.1$ | $y = 7{,}365.5\ln(x) - 73{,}019$ |
| | $R^2 = 0.8017$ | $R^2 = 0.0947$ |
| Polynomial of 2nd degree | $y = -2.2541x^2 - 152.93x + 7{,}728.3$ | $y = 3E\text{-}05x^2 - 2.5639x + 55{,}159$ |
| | $R^2 = 0.9652$ | $R^2 = 0.1021$ |
| Polynomial of 3rd degree | $y = -0.1912x^3 + 4.0562x^2 - 209.77x + 7{,}844.5$ | $y = -3E\text{-}08x^3 + 0.0036x^2 - 151.93x + 2E{+}06$ |
| | $R^2 = 0.9659$ | $R^2 = 0.1116$ |

cont. Table 4

| 1 | 2 | 3 |
|---|---|---|
| Polynomial of 4th degree | $y = -0.1071x^4 + 4.5232x^3 – 63.706x^2 + 140.11x + 7,379.7$ | $y = -4E\text{-}11x^4 + 7E\text{-}06x^3 – 0.4254x^2 + 11,790x – 1E{+}08$ |
| | $R^2 = 0.9718$ | $R^2 = 0.1492$ |
| Polynomial of 5th degree | $y = -0.0254x^5 + 1.2898x^4 – 23.149x^3 + 173.32x^2 – 678.71x + 8,175.4$ | $y = 2E\text{-}14x^5 – 4E\text{-}09x^4 + 0.0003x^3 – 13.34x^2 + 281,365x – 2E{+}09$ |
| | $R^2 = 0.9804$ | $R^2 = 0.1624$ |
| Polynomial of 6th degree | $y = 0.0033x^6 – 0.2441x^5 + 6.8169x^4 – 89.943x^3 + 565.19x^2 – 1,673.3x + 8,939.6$ | $y = 4^E\text{-}17x^6 – 9^E\text{-}12x^5 + 9^E\text{-}07x^4 – 0.0527x^3 + 1,645.7x^2 – 3^E{+}07x + 2^E{+}11$ |
| | $R^2 = 0.9842$ | $R^2 = 0.2678$ |
| Potentio-metric | $y = 9,514.7x^{-0.25}$ | $y = 0.0043x^{1.3182}$ |
| | $R^2 = 0.7207$ | $R^2 = 0.0934$ |

Table 5

Trend models for Friday

| Data / model | Annual data | Monthly data |
|---|---|---|
| Exponential | $y = 9,846e^{-0.039x}$ | $y = 2,049.1e^{3E\text{-}05x}$ |
| | $R^2 = 0.9354$ | $R^2 = 0.0669$ |
| Linear | $y = -244.36x + 9,300.5$ | $y = 0.1652x – 841.85$ |
| | $R^2 = 0.9603$ | $R^2 = 0.0758$ |
| Logarithmic | $y = -1,682\ln(x) + 10,246$ | $y = 6,849,9\ln(x) – 66,819$ |
| | $R^2 = 0.7897$ | $R^2 = 0.0748$ |
| Polynomial of 2nd degree | $y = -2.6135x^2 – 186.86x + 9,080$ | $y = 5E\text{-}05x^2 – 3.745x + 80,777$ |
| | $R^2 = 0.9635$ | $R^2 = 0.088$ |
| Polynomial of 3rd degree | $y = 0.0574x^3 – 4.5084x^2 – 169.79x + 9,045.2$ | $y = -4E\text{-}08x^3 + 0.0047x^2 – 198.19x + 3E{+}06$ |
| | $R^2 = 0.9636$ | $R^2 = 0.1028$ |
| Polynomial of 4th degree | $y = -0.1317x^4 + 5.8522x^3 – 87.798x^2 + 260.26x + 8,474$ | $y = -4E\text{-}11x^4 + 7E\text{-}06x^3 – 0.4646x^2 + 12,865x – 1E{+}08$ |
| | $R^2 = 0.9697$ | $R^2 = 0.1438$ |
| Polynomial of 5th degree | $y = -0.0199x^5 + 0.9609x^4 – 15.79x^3 + 97.577x^2 – 380.14x + 9,096.2$ | $y = 3E\text{-}14x^5 – 5E\text{-}09x^4 + 0.0004x^3 – 18.838x^2 + 396,389x – 3E{+}09$ |
| | $R^2 = 0.9733$ | $R^2 = 0.1684$ |
| Polynomial of 6th degree | $y = 0.0032x^6 – 0.2324x^5 + 6.3317x^4 – 80.696x^3 + 478.37x^2 – 1,346.6x + 9,838.9$ | $y = 3^E\text{-}17x^6 – 8^E\text{-}12x^5 + 9^E\text{-}07x^4 – 0.0473x^3 + 1,477.1x^2 – 2^E{+}07x + 2^E{+}11$ |
| | $R^2 = 0.9758$ | $R^2 = 0.2466$ |
| Potentiometric | $y = 11,231x^{-0.258}$ | $y = 0.073x^{1.0635}$ |
| | $R^2 = 0.7208$ | $R^2 = 0.0659$ |

Table 6

Trend models for Saturday

| Data / model | | Annual data | Monthly data |
|---|---|---|---|
| Exponential | | $y = 9{,}132.9e^{-0.045x}$ | $y = 3{,}426e^{6E\text{-}06x}$ |
| | | $R^2 = 0.9358$ | $R^2 = 0.0033$ |
| Linear | | $y = -246.76x + 8{,}519.6$ | $y = 0.0355x + 2{,}992.8$ |
| | | $R^2 = 0.9526$ | $R^2 = 0.0057$ |
| Logarithmic | | $y = -1{,}715\ln(x) + 9{,}511.6$ | $y = 1{,}423.3\ln(x) - 10{,}666$ |
| | | $R^2 = 0.7991$ | $R^2 = 0.0053$ |
| Polynomial of 2nd degree | | $y = -0.57x^2 - 234.22x + 8{,}471.5$ | $y = 6E\text{-}05x^2 - 4.9231x + 106{,}496$ |
| | | $R^2 = 0.9528$ | $R^2 = 0.0378$ |
| Polynomial of 3rd degree | | $y = 0.358x^3 - 12.384x^2 - 127.82x + 8{,}254.1$ | $y = -3E\text{-}08x^3 + 0.0042x^2 - 176.36x + 2E{+}06$ |
| | | $R^2 = 0.9544$ | $R^2 = 0.0566$ |
| Polynomial of 4th degree | | $y = -0.1359x^4 + 6.3362x^3 - 98.311x^2 + 315.84x + 7{,}664.8$ | $y = -4E\text{-}11x^4 + 7E\text{-}06x^3 - 0.4473x^2 + 12{,}391x - 1E{+}08$ |
| | | $R^2 = 0.9607$ | $R^2 = 0.1187$ |
| Polynomial of 5th degree | | $y = -0.0308x^5 + 1.5562x^4 - 27.18x^3 + 188.78x^2 - 675.93x + 8{,}628.5$ | $y = 1E\text{-}14x^5 - 3E\text{-}09x^4 + 0.0002x^3 - 9.3644x^2 + 19{,}8521x - 2E{+}09$ |
| | | $R^2 = 0.9692$ | $R^2 = 0.1282$ |
| Polynomial of 6th degree | | $y = -0.0009x^6 + 0.0284x^5 + 0.0598x^4 - 9.0966x^3 + 82.681x^2 - 406.67x + 8{,}421.6$ | $y = 2^E\text{-}17x^6 - 4^E\text{-}12x^5 + 4^E\text{-}07x^4 - 0.0244x^3 + 761.72x^2 - 1^E{+}07x + 9^E{+}10$ |
| | | $R^2 = 0.9694$ | $R^2 = 0.1622$ |
| Potentiometric | | $y = 10{,}674x^{-0.299}$ | $y = 342.06x^{0.2404}$ |
| | | $R^2 = 0.7313$ | $R^2 = 0.0029$ |

Table 7

Trend models for Sunday

| Data / model | | Annual data | Monthly data |
|---|---|---|---|
| 1 | | 2 | 3 |
| Exponential | | $y = 9{,}132.9e^{-0.045x}$ | $y = 4{,}181.4e^{-6E\text{-}06x}$ |
| | | $R^2 = 0.9358$ | $R^2 = 0.0025$ |
| Linear | | $y = -246.76x + 8{,}519.6$ | $y = -0.011x + 3{,}757.7$ |
| | | $R^2 = 0.9526$ | $R^2 = 0.0008$ |
| Logarithmic | | $y = -1{,}715\ln(x) + 9{,}511.6$ | $y = -478.6\ln(x) + 8{,}389$ |
| | | $R^2 = 0.7991$ | $R^2 = 0.0009$ |
| Polynomial of 2nd degree | | $y = -0.57x^2 - 234.22x + 8{,}471.5$ | $y = 2E\text{-}05x^2 - 1.4855x + 34{,}536$ |
| | | $R^2 = 0.9528$ | $R^2 = 0.0049$ |
| Polynomial of 3rd degree | | $y = 0.358x^3 - 12.384x^2 - 127.82x + 8254.1$ | $y = -3E\text{-}08x^3 + 0.0033x^2 - 139.3x + 2E{+}06$ |
| | | $R^2 = 0.9544$ | $R^2 = 0.0224$ |

cont. Table 7

| 1 | 2 | 3 |
|---|---|---|
| Polynomial of 4th degree | $y = -0.1359x^4 + 6.3362x^3 - 98.311x^2 + 315.84x + 7{,}664.8$ | $y = -3E\text{-}11x^4 + 6E\text{-}06x^3 - 0.3619x^2 + 10{,}027x - 1E{+}08$ |
| | $R^2 = 0.9607$ | $R^2 = 0.081$ |
| Polynomial of 5th degree | $y = -0.0308x^5 + 1.5562x^4 - 27.18x^3 + 188.78x^2 - 675.93x + 8{,}628.5$ | $y = 2E\text{-}14x^5 - 4E\text{-}09x^4 + 0.0004x^3 - 14.971x^2 + 31{,}4962x - 3E{+}09$ |
| | $R^2 = 0.9692$ | $R^2 = 0.1176$ |
| Polynomial of 6th degree | $y = -0.0009x^6 + 0.0284x^5 + 0.0598x^4 - 9.0966x^3 + 82.681x^2 - 406.67x + 8{,}421.6$ | $y = 1^{E}\text{-}17x^6 - 3^{E}\text{-}12x^5 + 4^{E}\text{-}07x^4 - 0.0198x^3 + 615.72x^2 - 1^{E}{+}07x + 7^{E}{+}10$ |
| | $R^2 = 0.9694$ | $R^2 = 0.1504$ |
| Potentiometric | $y = 10{,}674x^{-0.299}$ | $y = 51{,}681x^{-0.26}$ |
| | $R^2 = 0.7313$ | $R^2 = 0.0026$ |

Next, using data from Tables 1-7, the projected number of traffic accidents was determined. For annual data it was 2022-2031, while for monthly data it was the period from January 2022 to December 2023. The forecast in this case was based on a weighted average of current and historical data of series values. The result of the forecast using this method, depends on the choice of the model and its fit.

Then for the obtained forecasts, the errors of the expired forecasts were determined based on equations (7-11):

– ME – mean error

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right) \tag{7}$$

– MAE – mean everage error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - Y_p\right| \tag{8}$$

– MPE – mean percentage error

$$\mathrm{MPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - Y_p}{Y_i}\cdot 100\% \tag{9}$$

– MAPE – mean absolute percentage error

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - Y_p\right|}{Y_i}\cdot 100\% \tag{10}$$

– MSE – mean square error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right)^2 \tag{11}$$

where:

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

To forecast the number of traffic accidents by day of the week, trend models were selected for which the mean percentage error and mean absolute percentage error were the smallest. On this basis, it was found that for annual data, the exponential model was the best fit, for which the maximum MAPE error was 0.69% on all analyzed days of the week. In contrast, for monthly data, for the exponential model, which also proved to be the best, the error ranged from 21% to 73%. This is a very large value (Tabs. 8, 9). On this basis, the projected number of accidents for the following years was determined on a monthly and annual basis (Figs. 3, 4). Based on Figures 3 and 4, we can expect a further decrease in the number of traffic accidents in the following years. Note that the pandemic has caused significant changes in the forecasts. As can be seen in Figure 4, the trend models do not take into account the seasonality present in traffic accidents and should not be used in the case under analysis.

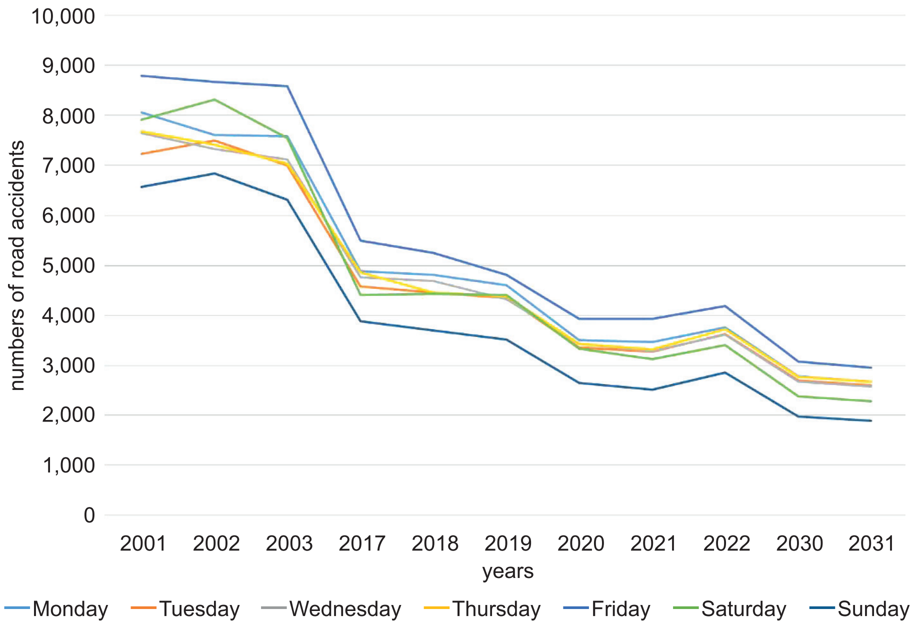


Fig. 3. Projected number of road accidents for 2022-2031

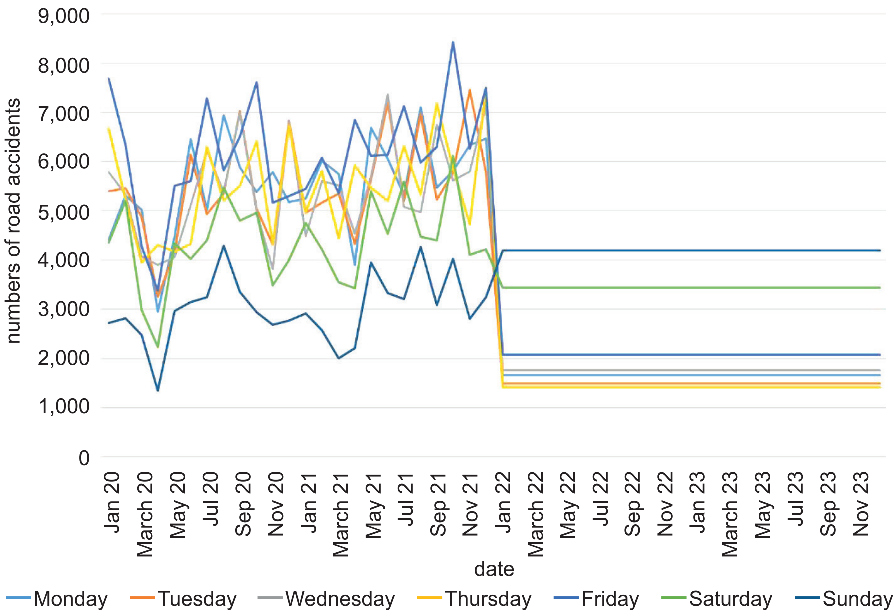


Fig. 4. Projected number of road accidents for 2022-2023

Table 8

Summary of errors for annual data

| Error / day of the week | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| ME | 29.73096 | 13.31825 | 29.9056 | 15.65675 | 20.9265 | 29.05283 | 7.791874 |
| MPE | 296.0938 | 289.0418 | 298.7438 | 252.1597 | 301.3413 | 311.1243 | 296.1362 |
| Suma kwadratów | 2,712,094 | 2,459,100 | 2,536,060 | 2,008,412 | 3,117,205 | 3,345,804 | 3,211,355 |
| MSE | 129,147.3 | 117,100 | 120,764.8 | 95,638.64 | 148,438.3 | 159,324 | 152,921.7 |
| MAPE [%] | 0.20663 | 0.093904 | 0.221561 | 0.613665 | 0.081506 | 0.200802 | 0.694331 |
| MAE [%] | 5.420082 | 5.612984 | 5.75191 | 4.894084 | 4.78237 | 5.613153 | 6.422599 |

Table 9

Summary of errors for monthly data

| Error / day of the week | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| ME | 3,791.339 | 3,849.998 | 3,620.231 | 3,945.472 | 4,009.132 | 1,050.306 | 882.1192 |
| MPE | 3,791.339 | 3,849.998 | 3,620.231 | 3,945.472 | 4,009.132 | 1,078.197 | 935.8334 |
| Suma kwadratów | 2.75E+09 | 2.81E+09 | 2.51E+09 | 2.95E+09 | 3.05E+09 | 2.97E+08 | 2.09E+08 |
| MSE | 15,288,042 | 15,611,732 | 13,933,953 | 16,387,446 | 16,971,339 | 1,652,674 | 1,159,591 |
| MAPE [%] | 68.79714 | 71.53983 | 66.66251 | 73.03121 | 65.27333 | 21.2692 | 31.5986 |
| MAE [%] | 68.79714 | 71.53983 | 66.66251 | 73.03121 | 65.27333 | 22.33479 | 32.7702 |

## Conclusions

Forecast of the number of accidents in Poland for each day of the week determined on the basis of selected Excel method trend models. The results show that we can still expect a decrease in the number of traffic accidents in the coming years. It should be noted that the pandemic has skewed the results obtained, and if the pandemic continues and traffic restrictions are implemented, the proposed model may not be adequate. The error value of at most 0.69%, for annual data, can testify to the choice of an effective forecasting method. As can be seen, trend models fail for forecasting the monthly number of traffic accidents, where there is seasonality. In contrast, for annual data, the results are at a high level. The advantage of trend models is the speed of determining the forecast.

The forecast of the number of traffic accidents obtained in the article, can be used in the future to formulate further measures to minimize the number of accidents in the analyzed country. These measures may include, for example, the introduction of higher fines for traffic offenses on Polish roads from January 1, 2022. In addition, based on the survey, it can be concluded that the most accidents will be on Fridays. For this reason, it is advisable to introduce traffic controls on this day, in order to reduce the number of traffic accidents.

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

## References


ABDULLAH E., EMAM A. 2015. *Traffic accidents analyzer using big data*. International Conference on Computational Science and Computational Intelligence (CSCI). 7-9 Dec. https://doi.org/10.1109/CSCI.2015.187

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

ARTEAGA C., PAZ A., PARK J. 2020. *Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach*. Safety Science, 132, Article 104988.

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

BISWAS A.A., MIA J., MAJUMDER A. 2019. *Forecasting the Number of Road Accidents and Casualties using Random Forest Regression in the Context of Bangladesh*. July 2019 Economics 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

BLOOMFIELD P. 1973. *An exponential model in the spectrum of a scalar time series*. Biometrika, 60: 217-226. Retrieved from https://www.jstor.org/stable/2334533

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

CHEN C. 2017. *Analysis and forecast of traffic accident big data*. ITM Web of Conferences, 12: 04029. https://doi.org/10.1051/itmconf/20171204029

CHUDY-LASKOWSKA K., PISULA T. 2014. *Forecast of the number of road accidents in Poland*. Logistics, 6.

CHUDY-LASKOWSKA K., PISULA T. 2015. *Forecasting the number of traffic accidents in Subcarpathia*. Logistics, 4.

Data mining techniques. 2022. StatSoft. Retrieved from https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html

DUDEK G. 2013. *Exponential smoothing models for short-term power system load forecasting*. Energy Market, 3(106).

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

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.

FIJOREK K., MRÓZ K., NIEDZIELA K., FIJOREK D. 2010. *Forecasting electricity prices on the day-ahead market using data mining methods*. Energy Market, 12.

FISZEDER P. 2009. *GARCH class models in empirical financial research*. Scientific Publishers of the Nicolaus Copernicus University, Toruń.

*Forecasting based on time series*. 2022. PiS-W-5. pis.rezolwenta.eu.org. Retrieved from http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf

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

GORZELAŃCZYK P., JURKOVIČ M., KALINA T., MOHANTY M. 2022. *Forecasting the road accident rate and the impact of the COVID-19 on its frequency in the polish provinces*. Communications – Scientific Letters of the University of Zilina, 24(4). https://doi.org/10.26552/com.C.2022.4.A216-A231

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

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

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(1): 578-589. https://doi.org/10.1515/eng-2022-0370

KARLAFTIS M., VLAHOGIANNI E. 2009. *Memory properties and fractional integration in trans--portation time-series*. Transportation Research. Part C. Emerging Technologies, 17(4): 444-453. https://doi.org/10.1016/j.trc.2009.03.001

KASHPRUK N. 2010. *Comparative research of statistical models and soft computing for identification of time series and forecasting*. Opole University of Technology, Opole.

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*. Electron, 8, https://doi.org/10.3390/electronics8080896

KUMAR S., VISWANADHAM V.,BHARATHI B. 2019. *Analysis of road accident*. IOP Conference Series Materials Science and Engineering, 590(1): 012029. https://doi.org/10.1088/1757-899X/590/1/012029

LAVRENZ S., VLAHOGIANNI E., GKRITZA K., KE Y. 2018. *Time series modeling in traffic safety research*. Accident Analysis & Prevention, 117: 368–380.

LI L, SHRESTHA S., HU G. 2017. *Analysis of road traffic fatal accidents using data mining techniques*. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), p. 363-370. https://doi.org/10.1109/SERA.2017.7965753

ŁOBEJKO S. 2015. *Time series analysis and forecasting with SAS*. Main Business School, Warszawa.

MAMCZUR M. 2022. *Machine learning How does linear regression work? And is it worth using?* Retrieved from https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/

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

MCILROY R.C., PLANT K.A., HOQUE M.S., WU J., KOKWARO G.O., NAM V.H., STANTON N.A. 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

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

MUCK J. 2022. *Econometrics. Modeling of time series. Stationary. Unit root tests. ARDL models. Co-integration.* Retrieved from http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja5.pdf

PERCZAK G., FISZEDER P. 2014. *GARCH model – using additional information on minimum and maximum prices.* Bank and Credit, 2.

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

PROCHAZKA J., CAMAJ M. 2017. *Modelling the number of road accidents of uninsured drivers and their severity.* Proceedings of International Academic Conferences 5408040, International Institute of Social and Economic Sciences.

PROCHÁZKA J., FLIMMEL S., ČAMAJ M., BAŠTA M. 2017. *Modelling the Number of Road Accidents.* Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław. https://doi.org/10.15611/amse.2017.20.29

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

*Random forest.* 2022. Wikipedia. Wolna encyklopedia. Retrieved from https://pl.wikipedia.org/wiki/Las_losowy

*Road safety assessment handbook.* 2022. Retrieved from https://www.iung.pl/PJA/wydane/11/PJA11_3.pdf

SEBEGO M., NAUMANN R.B., RUDD R.A., VOETSCH K., DELLINGER A.M., NDLOVU C. 2008. *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

SHETTY P., SACHIN P.C., KASHYAP V.K., MADI V. 2017. *Analysis of road accidents using data mining techniques.* International Research Journal of Engineering and Technology, 4.

Statistic Road Accident. 2022. Retrieved from https://statystyka.policja.pl/

SUNNY C.M., NITHYA S., SINSHI K.S., VINODINI V.M.D. LAKSHMI A.K.G., ANJANA S., MANOJKUMAR T.K. 2018. *Forecasting of Road Accident in Kerala: A Case Study.* 2018 International Conference on Data Science and Engineering (ICDSE). https://doi.org/10.1109/ICDSE.2018.8527825

SZMUKSTA-ZAWADZKA M., ZAWADZKI J. 2009. *Forecasting on the basis of Holt-Winters models for complete and incomplete data.* Research papers of the Wrocław University of Economics, 38.

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

*Top Advantages and Disadvantages of Hadoop 3 DataFlair.* 2022. Retrieved from https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/

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

WÓJCIK A. 2014. *Autoregressive vector models as a response to the critique of multi-equation structural econometric models.* Publishing house of the University of Economics in Katowice, 193.2014.

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

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. https://doi.org/10.1371/journal.pone.0201531

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.