



FORECASTING THE NUMBER OF ROAD ACCIDENTS IN POLAND USING WEATHER-DEPENDENT TREND MODELS

Piotr Gorzelańczyk

ORCID: 0000-0001-9662-400X

Stanislaw Staszic State University of Applied Sciences in Piła

Received 29 November 2022, accepted 6 February 2023, available online 7 February 2023.

Key words: traffic accident, forecasting, trend models, weather conditions.

Abstract

Every year a very large number of people die on the roads. From year to year, the value decreases, there are still a very high number of them. The pandemic has reduced the number of road accidents, but the value is still very high. For this reason, it is necessary to know under which weather conditions the highest number of road accidents occur, and to know the forecast of accidents according to the prevailing weather conditions for the coming years, in order to be able to do everything possible to minimize the number of road accidents.

The purpose of the article is to make a forecast of the number of road accidents in Poland depending on the prevailing weather conditions. 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 2001-2021, 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 2007-2021. 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.

Introduction

Road traffic accidents are events that cause not only injuries or death to road users, but also damage to property. According to the WHO, approximately 1.3 million people die each year as a result of traffic accidents. Traffic accidents account for around 3% of their GDP for most of the countries in the world. Road traffic accidents are the leading cause of death for minors and young people aged 5-29 (*The Global Status on Road Safety* 2018). The UN General Assembly has set an ambitious goal of halving the number of road deaths and injuries by 2030.

The extent of a traffic accident is an attribute for determining its severity. Predicting the severity of accidents is important for competent authorities when designing transport safety policies to eliminate accidents, reduce injuries, deaths and property losses (TAMBOURATZIS et al. 2014, ZHU et al. 2019). The identification of critical factors that affect the severity of accidents is a precondition for taking countermeasures to eliminate and mitigate the severity of accidents (ARTEAGA et al. 2020). Yang et al. proposes a DNN (Deep Neural Network) multi-carbon framework to predict different levels of severity of injury, death and property loss. It allows a comprehensive and accurate analysis of the severity of traffic accidents (YANG et al. 2022).

There are several sources of accident data. They are mostly collected and analyzed by government authorities through the relevant government agencies. Data collection is carried out through police reports, insurance databases or hospital records. Partial traffic accident information is subsequently processed for the transport sector on a larger scale (GORZELAŃCZYK et al. 2020).

Intelligent transportation systems are currently the most important source of data related to the analysis and prediction of traffic accidents. The data can be processed due to the use of GPS devices in vehicles (CHEN 2017). Microwave vehicle detection systems at roadsides can continuously record vehicle data (speed, traffic volume, vehicle type, etc.) (KHALIQ et al. 2019). The Vehicle License Plate Recognition system also makes it possible to collect large amounts of traffic data over a monitored period (RAJPUT et al. 2015). Another source of data for obtaining traffic and accidents information can be social media, but their relevance may be insufficient due to the incompetence of reporters (ZHENG et al. 2018).

For the relevance of accident data, it is necessary to work with several data sources that need to be confronted correctly. The combination of different data sources by consolidating heterogeneous traffic accident data helps to increase the accuracy of the analysis results (ABDULLAH, EMAM 2016).

A statistical survey aimed at assessing the severity, finding out the connection between traffic accidents and road users was performed by VILACA et al. (2017). The result of the study is a proposal to improve road safety standards and the adoption of other policies related to transport safety.

BAK et al. (2019) conducted a statistical survey of traffic safety in a selected region of Poland based on the number of traffic accidents, the pace of finding out the causes of their occurrence. The survey applied a multidimensional statistical analysis to examine safety aspect of persons responsible for accidents.

The choice of the source of accident data for the analysis depends on the type of traffic problem being addressed. The combination of statistical models with other natural driving data or other data obtained through intelligent transport systems contributes to increasing the accuracy of accident forecasts and contributes to their elimination (CHAND et al. 2021).

Various methods of forecasting the number of accidents can be found in the literature. Most often, time series methods are used for forecasting the number of road traffic accidents (HELGASON 2016, LAVRENZ 2018), the disadvantages of which are the impossibility of assessing the quality of forecast on the basis of expired forecasts and the often-occurring autocorrelation of the residual component (FORECASTING 2022). PROCHÁZKA et al. (2017) used the multiple seasonality model for forecasting and SUNNY et al. (2018) used the Holt-Winters exponential smoothing method. Its limitations include the inability to introduce exogenous variables into the model (DUDEK 2013, SZMUKSTA-ZAWADZKA, ZAWADZKI 2009).

For forecasting the number of road accidents, the vector autoregression model has also been used, whose drawback is the need to have a large number of observations of the variables in order to correctly estimate their parameters (WÓJCIK 2014), as well as the autoregression models of MONEDEROA et al. (2021) for analysing the number of fatalities (MONEDEROA et al. 2021) and AL-MADANI (2018), curve-fitting regression models. These, in turn, require only simple linear relationships (MAMCZUR 2022), and the order of the autoregression (assuming that the series are already stationary) (PIŁATOWSKA 2012).

BISWAS et al. (2019) used Random Forest regression to predict the number of road accidents. In this case, the data contain groups of correlated features with similar significance to the original data, smaller groups are favoured over larger ones (*Las losowy* 2022), and there is instability in the method and spike prediction (FIJOREK et al. 2010). CHUDY-LASKOWSKA and PISULA (2014) used the autoregressive model with quadratic trend, the univariate periodic trend model and the exponential equalization model for the forecasting issue discussed. A moving mean model can also be used for forecasting the discussed issue, the disadvantages of which are low forecast accuracy, loss of data in the sequence, lack of consideration of trends and seasonal effects (KASHPRUK 2010). PROCHOZKA and CAMEJ (2017) used the GARMA method, in which some restrictions 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 (PROCHÁZKA et al. 2017, SUNNY et al. 2018, DUTTA et al. 2020, KARLAFTIS et al. 2009). These models result in very high flexibility of the discussed models, but it is also their disadvantage,

as good model identification requires more experience from the researcher than, for example, regression analysis (ŁOBEJKO et al. 2015). Another disadvantage is the linear nature of the ARIMA model (DUDEK 2013).

CHUDY-LASKOWSKA and PISULA (2015) in their work used the ANOVA method to forecast the number of road crashes. The disadvantage of this method is the adoption of additional assumptions, especially the assumption of sphericity, the violation of which may lead to erroneous conclusions (GREGORCZYK, SWARCEWICZ 2012). Neural network models are also used to forecast the number of road accidents. The disadvantage of ANN is the need for experience in this field (CHUDY-LASKOWSKA, PISULA 2017, WROBEL 2017) and the dependence of the final solution on the initial conditions of the network, as well as the lack of interpretability in the traditional way since ANN is usually referred to as blackbox where you give input and the model gives output without any knowledge about the analysis (*Techniki zgłębiania danych (data mining)* 2022).

A new prediction method is the use of the Hadoop model by KUMAR et al. (2019). The drawback of this method is its inability to work with small data files (*Top Advantages and Disadvantages of Hadoop 3.* 2022). KARLAFTIS and VLAHOGIANNI (2009) used the Garch model for prediction. The disadvantage of this method is its complex form and complicated model (PERCZAK, FISZEDER 2014, FISZEDER 2009). On the other hand, McIlroy and his team used the ADF test (MCILROY et al. 2019), which has the disadvantage of poor power in the case of 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). The combination of models proposed by SEBEGO et al. (2008) as a combination of different models is also encountered. Parametric models are also proposed in the work of BLOOMFIELD (1973).

Analyzing the above information, trend models were chosen to predict the number of traffic accidents depending on weather conditions.

Number of road accidents

Every year a very large number of people are killed on the roads. From year to year, the value decreases, there are still a very high number of them. Pandemic has reduced the number of road accidents, but the value is still very high. Analyzing the data on the number of road accidents according to the prevailing weather conditions 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 under different weather conditions (Fig. 1, 2).

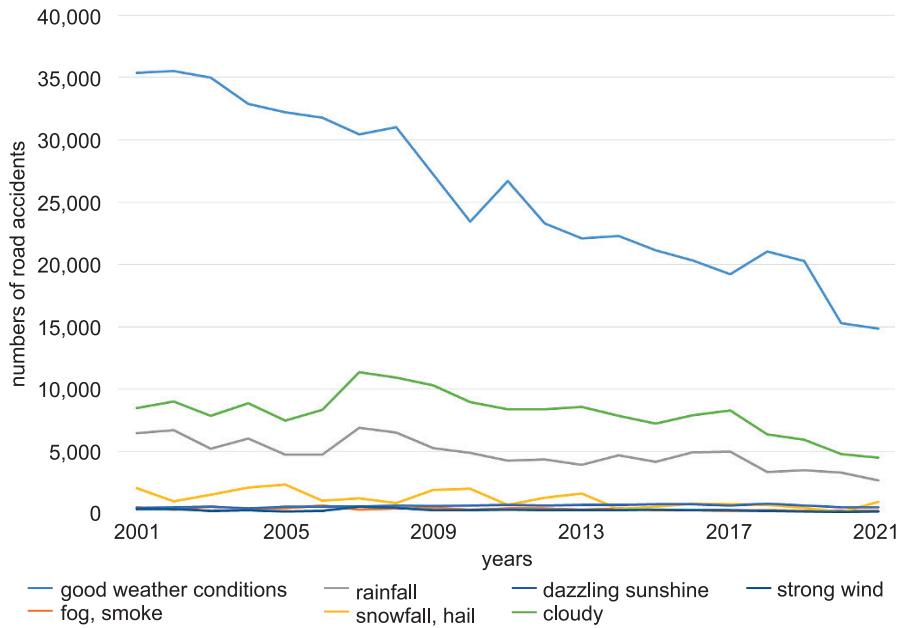


Fig. 1. Number of accidents in Poland from 2001 to 2021 by year
Source: based on data from Statystyka. Portal Polskiej Policji (2022).

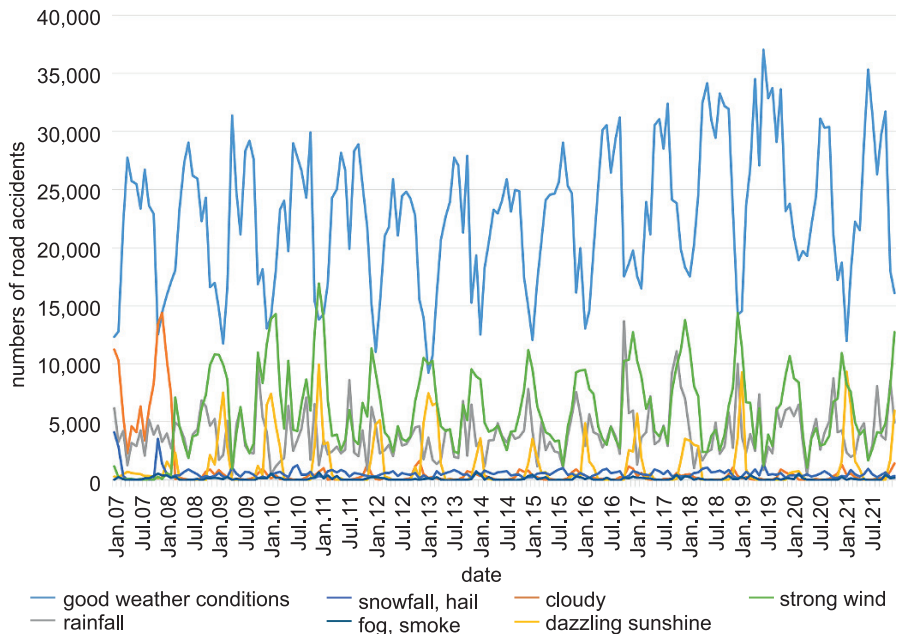


Fig. 2. Number of accidents in Poland from 2007 to 2021 by month
Source: based on data from Statystyka. Portal Polskiej Policji (2022).

For the purpose of this work, it was assumed (KRZYCZKOWSKA 2019):

- good atmospheric conditions are:
 - air temperature $> 3^{\circ}\text{C}$,
 - no precipitation,
 - wind < 5.5 m/s,
 - visibility > 10 km,
 - pressure difference over the day < 8 hPa;
- bad weather conditions (if one of the following factors is met) are:
 - slippery pavement (temperature $< 3^{\circ}\text{C}$ and occurrence of precipitation),
 - heavy rain (temperature $> 0^{\circ}\text{C}$, precipitation > 3 mm),
 - snowstorm (temperature $< 0^{\circ}\text{C}$, precipitation > 3 mm),
 - strong wind (wind > 10 ms/s)
 - dense fog (visibility < 300 m).

Forecasting the number of traffic accidents

The following trend models were used in forecasting the number of traffic accidents for the analyzed weather conditions:

- 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. As can be seen, the *R*-square coefficient, which is a measure of the quality of the model fit for annual data in most cases there is a good or satisfactory fit, while for monthly data there is a poor and unsatisfactory fit. This is mainly due to the seasonality of the number of traffic accidents in the weather conditions analyzed, with the least number of accidents during fog and the most during good conditions (Tab. 1-7). Model parameters were determined using the least squares method.

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

Table 1

Trend models during good weather conditions		
Data/model	Annual data	Monthly data
Exponential	$y = 39,361e^{-0,041x}$	$y = 3,627.8e^{4E-05x}$
	$R^2 = 0.9379$	$R^2 = 0.0565$
Linear	$y = -1,033.1x + 37,145$	$y = 0.9753x - 17,991$
	$R^2 = 0.9528$	$R^2 = 0.0649$
Logarithmic	$y = -7,316 \ln(x) + 41,592$	$y = 40,532 \ln(x) - 408,478$
	$R^2 = 0.8298$	$R^2 = 0.0643$
Polynomial of 2 nd degree	$y = 7.5551x^2 - 1,199.3x + 37,782$	$y = 0.0002x^2 - 14.815x + 311,602$
	$R^2 = 0.9543$	$R^2 = 0.0698$
Polynomial of 3 rd degree	$y = 0.7577x^3 - 17.448x^2 - 974.1x + 37,322$	$y = -2E-07x^3 + 0.0223x^2 - 937.39x + 1E+07$
	$R^2 = 0.9547$	$R^2 = 0.078$
Polynomial of 4 th degree	$y = -0.7724x^4 + 34.742x^3 - 505.92x^2 + 1,548x + 33,972$	$y = -2E-10x^4 + 4E-05x^3 - 2.4024x^2 + 66,559x - 7E+08$
	$R^2 = 0.9664$	$R^2 = 0.1048$
Polynomial of 5 th degree	$y = -0.0836x^5 + 3.828x^4 - 56.383x^3 + 274.62x^2 - 1,148.4x + 36,592$	$y = 3E-14x^5 - 7E-09x^4 + 0.0006x^3 - 25.167x^2 + 541,739x - 5E+09$
	$R^2 = 0.9699$	$R^2 = 0.1058$
Polynomial of 6 th degree	$y = -0.0008x^6 - 0.0282x^5 + 2.4271x^4 - 39.453x^3 + 175.29x^2 - 896.32x + 36,399$	$y = 6E-17x^6 - 1E-11x^5 + 2E-06x^4 - 0.086x^3 + 2,686.4x^2 - 4E+07x + 3E+11$
	$R^2 = 0.9699$	$R^2 = 0.1121$
Potentiometric	$y = 46,046x^{-0.283}$	$y = 0.0001x^{1.789}$
	$R^2 = 0.7625$	$R^2 = 0.056$

Table 2

Trend models during fog, smoke		
Data/model	Annual data	Monthly data
1	2	3
Exponential	$y = 579.63e^{-0.037x}$	$y = 220.69e^{-3E-05x}$
	$R^2 = 0.6162$	$R^2 = 0.0014$
Linear	$y = -14.24x + 557.12$	$y = -0.004x + 293.07$
	$R^2 = 0.5515$	$R^2 = 0.0021$
Logarithmic	$y = -89.69 \ln(x) + 594.29$	$y = -168.1 \ln(x) + 1,915.5$
	$R^2 = 0.3799$	$R^2 = 0.0022$
Polynomial of 2 nd degree	$y = -0.7155x^2 + 1.5018x + 496.77$	$y = 2E-06x^2 - 0.1653x + 3,659.4$
	$R^2 = 0.5921$	$R^2 = 0.0032$
Polynomial of 3 rd degree	$y = 0.0733x^3 - 3.1345x^2 + 23.287x + 452.27$	$y = -1E-09x^3 + 0.0002x^2 - 6.5121x + 91,953$
	$R^2 = 0.6039$	$R^2 = 0.0039$

cont. Table 2

1	2	3
Polynomial of 4 th degree	$y = 0.0089x^4 - 0.3187x^3 + 2.4996x^2 - 5.8036x + 490.9$ $R^2 = 0.6086$	$y = -4E-13x^4 + 6E-08x^3 - 0.004x^2 + 108.79x - 1E+06$ $R^2 = 0.0041$
Polynomial of 5 th degree	$y = 0.0013x^5 - 0.0626x^4 + 1.0984x^3 - 9.6382x^2 + 36.128x + 450.16$ $R^2 = 0.6113$	$y = 2E-15x^5 - 4E-10x^4 + 3E-05x^3 - 1.2797x^2 + 26,738x - 2E+08$ $R^2 = 0.0098$
Polynomial of 6 th degree	$y = 0.0004x^6 - 0.0223x^5 + 0.534x^4 - 6.1118x^3 + 32.663x^2 - 71.229x + 532.65$ $R^2 = 0.6164$	$y = 8E-19x^6 - 2E-13x^5 + 2E-08x^4 - 0.0012x^3 + 36.143x^2 - 598,092x + 4E+09$ $R^2 = 0.0122$
Potentiometric	$y = 634.51x^{-0.232}$ $R^2 = 0.4126$	$y = 1E+07x^{-1.142}$ $R^2 = 0.0015$

Table 3

Trend models during rainfall

Data/model	Annual data	Monthly data
Exponential	$y = 6,785.4e^{-0.033x}$ $R^2 = 0.6669$	$y = 587.07e^{4E-05x}$ $R^2 = 0.013$
Linear	$y = -153.24x + 6,533.4$ $R^2 = 0.6471$	$y = 0.1828x - 3669.6$ $R^2 = 0.0184$
Logarithmic	$y = -1,071 \ln(x) + 7,162.9$ $R^2 = 0.5492$	$y = 7,578 \ln(x) - 76,659$ $R^2 = 0.0181$
Polynomial of 2 nd degree	$y = -2.9002x^2 - 89.436x + 6,288.8$ $R^2 = 0.6539$	$y = 5E-05x^2 - 4.2208x + 88,247$ $R^2 = 0.0215$
Polynomial of 3 rd degree	$y = -0.7386x^3 + 21.473x^2 - 308.94x + 6,737.2$ $R^2 = 0.6661$	$y = -7E-08x^3 + 0.0082x^2 - 346.49x + 5E+06$ $R^2 = 0.0305$
Polynomial of 4 th degree	$y = -0.0587x^4 + 1.8455x^3 - 15.668x^2 - 117.17x + 6,482.5$ $R^2 = 0.6681$	$y = -7E-12x^4 + 1E-06x^3 - 0.0671x^2 + 1,750.1x - 2E+07$ $R^2 = 0.0308$
Polynomial of 5 th degree	$y = -0.0461x^5 + 2.4794x^4 - 48.43x^3 + 414.97x^2 - 1,604.9x + 7,928.1$ $R^2 = 0.7018$	$y = 5E-14x^5 - 1E-08x^4 + 0.0009x^3 - 38.121x^2 + 796,065x - 7E+09$ $R^2 = 0.0516$
Polynomial of 6 th degree	$y = 0.0074x^6 - 0.5352x^5 + 14.841x^4 - 197.82x^3 + 1,291.4x^2 - 3,829.2x + 9,637.2$ $R^2 = 0.7239$	$y = 2E-17x^6 - 4E-12x^5 + 4E-07x^4 - 0.0239x^3 + 740.27x^2 - 1E+07x + 8E+10$ $R^2 = 0.0558$
Potentiometric	$y = 7,648.6x^{-0.225}$ $R^2 = 0.5287$	$y = 3E-05x^{-1.7424}$ $R^2 = 0.0127$

Table 4

Trend models during snowfall, hail		
Data/model	Annual data	Monthly data
Exponential	$y = 2,199.9e^{-0.072x}$	$y = 3.0473e^{0.0001x}$
	$R^2 = 0.4569$	$R^2 = 0.0406$
Linear	$y = -69.145x + 1,948.3$	$y = -0.0329x + 1,943.3$
	$R^2 = 0.4551$	$R^2 = 0.0123$
Logarithmic	$y = -467.5 \ln(x) + 2,197.9$	$y = -1,394 \ln(x) + 15,404$
	$R^2 = 0.3612$	$R^2 = 0.0126$
Polynomial of 2 nd degree	$y = -1.1547x^2 - 43.742x + 1,850.9$	$y = 2E-05x^2 - 1.6199x + 35,069$
	$R^2 = 0.4588$	$R^2 = 0.0204$
Polynomial of 3 rd degree	$y = 0.4937x^3 - 17.448x^2 + 103x + 1,551.1$	$y = -3E-08x^3 + 0.0038x^2 - 159.64x + 2E+06$
	$R^2 = 0.4776$	$R^2 = 0.0603$
Polynomial of 4 th degree	$y = 0.0859x^4 - 3.2849x^3 + 36.863x^2 - 177.43x + 1,923.6$	$y = 1E-11x^4 - 2E-06x^3 + 0.1568x^2 - 4,419.4x + 5E+07$
	$R^2 = 0.4929$	$R^2 = 0.078$
Polynomial of 5 th degree	$y = 0.0037x^5 - 0.116x^4 + 0.7143x^3 + 2.6081x^2 - 59.091x + 1,808.6$	$y = -1E-14x^5 + 2E-09x^4 - 0.0002x^3 + 8.08x^2 - 169,803x + 1E+09$
	$R^2 = 0.4936$	$R^2 = 0.0966$
Polynomial of 6 th degree	$y = 0.0023x^6 - 0.1501x^5 + 3.7718x^4 - 46.27x^3 + 278.26x^2 - 758.66x + 2,346.1$	$y = 1E-17x^6 - 3E-12x^5 + 3E-07x^4 - 0.0164x^3 + 515.46x^2 - 9E+06x + 6E+10$
	$R^2 = 0.5012$	$R^2 = 0.1333$
Potentiometric	$y = 2,734.1x^{-0.465}$	$y = 2E-21x^{5.0502}$
	$R^2 = 0.3348$	$R^2 = 0.0424$

Table 5

Trend models during sun glare		
Data/model	Annual data	Monthly data
1	2	3
Exponential	$y = 557.28e^{0.0123x}$	$y = 17,865e^{-1E-04x}$
	$R^2 = 0.2053$	$R^2 = 0.0052$
Linear	$y = 7.8065x + 560.32$	$y = -0.0295x + 2,396.6$
	$R^2 = 0.2137$	$R^2 = 0.0005$
Logarithmic	$y = 75.638 \ln(x) + 482.74$	$y = -1,211 \ln(x) + 14,048$
	$R^2 = 0.3483$	$R^2 = 0.0005$
Polynomial of 2 nd degree	$y = -1.9886x^2 + 51.556x + 392.61$	$y = -2E-05x^2 + 1.3429x - 26,249$
	$R^2 = 0.6176$	$R^2 = 0.0008$
Polynomial of 3 rd degree	$y = -0.2308x^3 + 5.6286x^2 - 17.045x + 532.77$	$y = 1E-07x^3 - 0.0134x^2 + 560.79x - 8E+06$
	$R^2 = 0.7687$	$R^2 = 0.0264$

cont. Table 5

1	2	3
Polynomial of 4 th degree	$y = -0.016x^4 + 0.4714x^3 - 4.4649x^2 + 35.071x + 463.55$ $R^2 = 0.7882$	$y = -2E-11x^4 + 4E-06x^3 - 0.2607x^2 + 7,443.6x - 8E+07$ $R^2 = 0.0288$
Polynomial of 5 th degree	$y = -0.0005x^5 + 0.0133x^4 - 0.1078x^3 + 0.4966x^2 + 17.931x + 480.2$ $R^2 = 0.7888$	$y = 1E-14x^5 - 3E-09x^4 + 0.0003x^3 - 10.767x^2 + 226,753x - 2E+09$ $R^2 = 0.0305$
Polynomial of 6 th degree	$y = -7E-05x^6 + 0.0043x^5 - 0.1091x^4 + 1.3711x^3 - 8.1799x^2 + 39.951x + 463.28$ $R^2 = 0.7891$	$y = 2E-17x^6 - 6E-12x^5 + 6E-07x^4 - 0.0342x^3 + 1,069.9x^2 - 2E+07x + 1E+11$ $R^2 = 0.039$
Potentiometric	$y = 490.1x^{0.1219}$ $R^2 = 0.352$	$y = 2E+26x^{-5.246}$ $R^2 = 0.0054$

Table 6

Trend models when it was cloudy

Data/model	Annual data	Monthly data
Exponential	$y = 10,429e^{-0.025x}$ $R^2 = 0.453$	$y = 8E+07e^{-3E-04x}$ $R^2 = 0.0994$
Linear	$y = -178.12x + 10,060$ $R^2 = 0.4164$	$y = -0.5751x + 24,912$ $R^2 = 0.164$
Logarithmic	$y = -900.3 \ln(x) + 10,046$ $R^2 = 0.1847$	$y = -24,466 \ln(x) + 261,193$ $R^2 = 0.1702$
Polynomial of 2 nd degree	$y = -28.854x^2 + 456.67x + 7,626.2$ $R^2 = 0.7347$	$y = 0.0004x^2 - 37.462x + 794,863$ $R^2 = 0.3571$
Polynomial of 3 rd degree	$y = -0.132x^3 - 24.5x^2 + 417.46x + 7,706.3$ $R^2 = 0.7349$	$y = -3E-07x^3 + 0.034x^2 - 1,437.2x + 2E+07$ $R^2 = 0.4939$
Polynomial of 4 th degree	$y = 0.083x^4 - 3.784x^3 + 27.992x^2 + 146.42x + 8,066.3$ $R^2 = 0.7369$	$y = 1E-10x^4 - 2E-05x^3 + 1.5732x^2 - 44,286x + 5E+08$ $R^2 = 0.5726$
Polynomial of 5 th degree	$y = -0.0803x^5 + 4.4978x^4 - 91.233x^3 + 777.04x^2 - 2,441.2x + 10,581$ $R^2 = 0.7854$	$y = -5E-14x^5 + 1E-08x^4 - 0.0008x^3 + 35.426x^2 - 750,906x + 6E+09$ $R^2 = 0.5874$
Polynomial of 6 th degree	$y = 0.0071x^6 - 0.5459x^5 + 16.268x^4 - 233.48x^3 + 1,611.6x^2 - 4,559.2x + 12,208$ $R^2 = 0.7949$	$y = 4E-18x^6 - 1E-12x^5 + 1E-07x^4 - 0.0066x^3 + 216.65x^2 - 4E+06x + 3E+10$ $R^2 = 0.5876$
Potentiometric	$y = 10,465x^{0.13}$ $R^2 = 0.209$	$y = 5E+62x^{-13.07}$ $R^2 = 0.1031$

Table 7

Trend models during high winds		
Data/model	Annual data	Monthly data
Exponential	$y = 375.22e^{-0.022x}$	$y = 0.0396e^{0.0003x}$
	$R^2 = 0.2222$	$R^2 = 0.1357$
Linear	$y = -6.626x + 380.89$	$y = 0.3377x - 8,288.8$
	$R^2 = 0.1862$	$R^2 = 0.0229$
Logarithmic	$y = -41.03 \ln(x) + 396.67$	$y = 1,4507 \ln(x) - 148,530$
	$R^2 = 0.124$	$R^2 = 0.0242$
Polynomial of 2 nd degree	$y = -0.7757x^2 + 10.438x + 315.47$	$y = -0.0004x^2 + 33.469x - 699,845$
	$R^2 = 0.2605$	$R^2 = 0.0858$
Polynomial of 3 rd degree	$y = -0.0652x^3 + 1.3774x^2 - 8.9522x + 355.09$	$y = 2E-07x^3 - 0.0315x^2 + 1331.9x - 2E+07$
	$R^2 = 0.2751$	$R^2 = 0.1333$
Polynomial of 4 th degree	$y = 0.0162x^4 - 0.7787x^3 + 11.633x^2 - 61.904x + 425.42$	$y = -2E-10x^4 + 3E-05x^3 - 2.1338x^2 + 59,853x - 6E+08$
	$R^2 = 0.2994$	$R^2 = 0.1927$
Polynomial of 5 th degree	$y = -0.0047x^5 + 0.2752x^4 - 5.9089x^3 + 55.575x^2 - 213.71x + 572.92$	$y = 9E-14x^5 - 2E-08x^4 + 0.0017x^3 - 70.718x^2 + 1E+06x - 1E+10$
	$R^2 = 0.3533$	$R^2 = 0.2173$
Polynomial of 6 th degree	$y = 0.0011x^6 - 0.0748x^5 + 2.0458x^4 - 27.307x^3 + 181.11x^2 - 532.31x + 817.74$	$y = 6E-17x^6 - 2E-11x^5 + 2E-06x^4 - 0.0919x^3 + 2,858.8x^2 - 5E+07x + 3E+11$
	$R^2 = 0.4232$	$R^2 = 0.2389$
Potentiometric	$y = 395.23x^{0.134}$	$y = 2E-51x^{11.77}$
	$R^2 = 0.1485$	$R^2 = 0.1414$

In the next step, expired forecast errors were determined for the obtained forecasts based on equations (1-5):

- ME – mean error

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

- MAE – mean absolute error

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y_p| [-] \quad (2)$$

- MPE – mean percentage error

$$\text{MPE} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i - Y_p}{Y_i} [\%] \quad (3)$$

- MAPE – mean absolute percentage error

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - Y_p|}{Y_i} [\%] \quad (4)$$

- MSE – mean square error

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_p)^2 [-] \quad (5)$$

where:

- n – number of information about road accidents,
- Y – observed value of road accidents,
- Y_p – forecasted value of road accidents.

To forecast the number of traffic accidents depending on the prevailing weather conditions, 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 in most cases the exponential model was the best fit, while for good weather conditions the linear model was the best fit, and for strong wind conditions the power model was the best fit. For all atmospheric conditions analysed, the maximum MAPE error was 12.3%. However, for monthly data for the exponential model, which also proved to be the best, except for solar glare over cloudy – the linear model, the error ranged from 79% to 28794%. This is a very large value (Tables 8, 9). On this basis, the projected number of accidents for the following years was determined on a monthly and annual basis (Tab. 10, 11, Fig. 3, 4). Based on Table 10, 11 and 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.

Table 8
Summary of errors for annual data

Error/weather conditions	Good weather conditions	Fog, smoke	Rainfall	Snowfall, hail	Dazzling sunshine	Cloudy	Strong wind
Best model	linear	exponential	exponential	exponential	potegal	exponential	exponential
ME	0.48	4.89	33.09	93.82	5.47	87.63	10.81
MPE	1,072.08	60.87	546.81	392.58	61.25	1,154.20	61.61
Sum of squares	40,680,995.31	137,979.40	10,359,303.92	4,968,270.79	145,974.13	38,104,078.44	152,810.86
MSE	1,937,190.25	6,570.45	493,300.19	236,584.32	6,951.15	1,814,479.93	7,276.71
MAPE [%]	0.33	2.01	1.28	12.33	0.92	1.58	2.49
MAE [%]	4.59	14.31	11.34	43.11	9.90	14.61	19.45

Table 9
Summary of errors for monthly data

Error/weather conditions	Good weather conditions	Fog, smoke	Rainfall	Snowfall, hail	Dazzling sunshine	Cloudy	Strong wind
Best model	exponential	exponential	exponential	exponential	linear	linear	exponential
ME	19,140.63	93.36	3,382.59	564.23	1,229.30	23,992.02	5,830.50
MPE	19,140.63	146.71	3,387.78	564.23	2,173.27	23,992.02	5,830.50
Sum of squares	72,538,421,905.45	4,894,311.41	2,876,780,693.66	97,038,950.94	1,046,654,730.96	10,450,947,6326.06	8,366,291,597.63
MSE	402,991,232.81	27,190.62	15,982,114.96	539,105.28	5,814,748.51	580,608,201.81	46,479,397.76
MAPE [%]	82.66	457.22	78.81	97.91	22,279.68	28,794.07	99.99
MAE [%]	82.66	469.42	80.08	97.91	22,295.39	28,794.07	99.99

Table 10

Forecast number of road accidents

Specification	Year	Good weather conditions	Fog, smoke	Rainfall	Snowfall, hail	Dazzling sunshine	Cloudy	Strong wind
Data from the police	2001	35,348	523	6,490	2,078	479	8,488	391
	2002	35,503	373	6,719	1,028	536	9,000	399
	2003	34,990	596	5,212	1,544	611	7,872	253
	2004	32,858	445	6,037	2,102	456	8,890	281
	2005	32,214	454	4,735	2,376	586	7,510	225
	2006	31,755	680	4,780	1,059	591	8,372	250
	2007	30,441	343	6,889	1,272	603	11,349	577
	2008	31,003	466	6,524	873	695	10,931	495
	2009	27,226	505	5,276	1,952	635	10,303	279
	2010	23,452	371	4,878	2,017	669	8,973	292
	2011	26,710	448	4,254	743	756	8,384	329
	2012	23,290	460	4,354	1,323	705	8,392	290
	2013	22,102	343	3,929	1,643	720	8,595	294
	2014	22,268	468	4,715	399	724	7,859	320
	2015	21,162	319	4,162	579	798	7,262	336
	2016	20,349	315	4,969	837	762	7,931	304
	2017	19,205	259	5,020	768	698	8,312	320
	2018	21,055	281	3,348	730	825	6,390	236
	2019	20,281	254	3,504	505	679	5,978	219
	2020	15,316	244	3,309	160	522	4,815	173
	2021	14,881	263	2,698	953	520	4,500	205
Forecast of the number of road accidents	2022	14,417	257	3,283	451	714	6,017	231
	2023	13,384	247	3,176	420	718	5,868	226
	2024	12,351	239	3,073	391	722	5,724	221
	2025	11,318	230	2,974	364	726	5,582	216
	2026	10,284	221	2,877	338	729	5,444	212
	2027	9,251	213	2,784	315	732	5,310	207
	2028	8,218	206	2,693	293	736	5,179	203
	2029	7,185	198	2,606	273	739	5,051	198
	2030	6,152	191	2,521	254	742	4,926	194
	2031	5,119	184	2,439	236	745	4,805	190

Table 11

Forecast number of road accidents

Specification	Data	Good weather conditions	Fog, smoke	Rainfall	Snowfall, hail	Dazzling sunshine	Cloudy	Strong wind
1	2	3	4	5	6	7	8	9
Data from the police	Jan 20	20,948.00	106.00	5,500.00	425.00	712.00	473.00	8,779.00
	Feb 20	18,942.00	538.00	6,506.00	407.00	765.00	68.00	8,395.00
	March 20	19,695.00	161.00	3,157.00	380.00	253.00	42.00	3,876.00
	Apr 20	19,268.00	33.00	397.00	492.00	12.00	9.00	1,058.00
	May 20	21,893.00	28.00	3,708.00	306.00	36.00	54.00	3,568.00
	Jun 20	24,470.00	15.00	5,245.00	428.00	9.00	81.00	4,489.00
	Jul 20	31,116.00	13.00	2,616.00	566.00	6.00	46.00	2,970.00
	Aug 20	30,353.00	33.00	4,064.00	759.00	1.00	47.00	3,114.00
	Sep 20	30,408.00	23.00	5,068.00	535.00	9.00	103.00	3,825.00
	Oct 20	21,091.00	68.00	8,777.00	188.00	21.00	346.00	6,737.00
	Nov 20	17,216.00	67.00	4,115.00	181.00	343.00	575.00	6,997.00
	Dec 20	18,709.00	272.00	4,253.00	344.00	1,649.00	1356.00	10,951.00
	Jan 21	11,945.00	61.00	2,376.00	310.00	9,316.00	447.00	8,228.00
	Feb 21	17,840.00	147.00	2,011.00	859.00	6,376.00	552.00	7,593.00
	March 21	22,210.00	93.00	2,204.00	512.00	1,652.00	228.00	4,987.00
	Apr 21	21,496.00	62.00	4,295.00	319.00	916.00	87.00	3,903.00
	May 21	28,950.00	83.00	4,880.00	535.00	34.00	40.00	4,339.00
	Jun 21	35,311.00	15.00	1,724.00	967.00	14.00	28.00	1,685.00
	Jul 21	31,534.00	8.00	2,998.00	491.00	13.00	51.00	2,668.00
	Aug 21	26,311.00	26.00	8,110.00	261.00	25.00	124.00	4,143.00
	Sep 21	29,750.00	67.00	3,794.00	480.00	6.00	151.00	4,087.00
	Oct 21	31,742.00	473.00	3,495.00	689.00	26.00	225.00	4,894.00
	Nov 21	17,989.00	151.00	8,554.00	235.00	1,736.00	756.00	7,976.00
	Dec 21	16,085.00	217.00	4,925.00	371.00	5,996.00	1,439.00	12,733.00
Forecast of the number of road accidents	Jan 22	3,654.16	219.49	591.34	3.10	2,391.26	24,807.91	0.04
	Feb 22	3,654.31	219.49	591.36	3.10	2,391.23	24,807.33	0.04
	March 22	3,654.45	219.48	591.38	3.10	2,391.20	24,806.76	0.04
	Apr 22	3,654.60	219.48	591.41	3.10	2,391.17	24,806.18	0.04
	May 22	3,654.75	219.47	591.43	3.10	2,391.14	24,805.61	0.04
	Jun 22	3,654.89	219.46	591.45	3.10	2,391.11	24,805.03	0.04
	Jul 22	3,655.04	219.46	591.48	3.11	2,391.08	24,804.46	0.04
	Aug 22	3,655.18	219.45	591.50	3.11	2,391.05	24,803.88	0.04

cont. Table 11

1	2	3	4	5	6	7	8	9
	Sep 22	3,655.33	219.44	591.53	3.11	2,391.02	24,803.31	0.04
	Oct 22	3,655.48	219.44	591.55	3.11	2,391.00	24,802.73	0.04
	Nov 22	3,655.62	219.43	591.57	3.11	2,390.97	24,802.16	0.04
	Dec 22	3,655.77	219.42	591.60	3.11	2,390.94	24,801.58	0.04
	Jan 23	3,655.92	219.42	591.62	3.11	2,390.91	24,801.01	0.04
	Feb 23	3,656.06	219.41	591.64	3.11	2,390.88	24,800.43	0.04
	March 23	3,656.21	219.40	591.67	3.11	2,390.85	24,799.86	0.04
	Apr 23	3,656.35	219.40	591.69	3.11	2,390.82	24,799.28	0.04
	May 23	3,656.50	219.39	591.71	3.11	2,390.79	24,798.71	0.04
	Jun 23	3,656.65	219.38	591.74	3.11	2,390.76	24,798.13	0.04
	Jul 23	3,656.79	219.38	591.76	3.11	2,390.73	24,797.56	0.04
	Aug 23	3,656.94	219.37	591.79	3.11	2,390.70	24,796.98	0.04
	Sep 23	3,657.09	219.36	591.81	3.11	2,390.67	24,796.40	0.04
	Oct 23	3,657.23	219.36	591.83	3.11	2,390.64	24,795.83	0.04
	Nov 23	3,657.38	219.35	591.86	3.11	2,390.61	24,795.25	0.04
	Dec 23	3,657.52	219.34	591.88	3.11	2,390.58	24,794.68	0.04

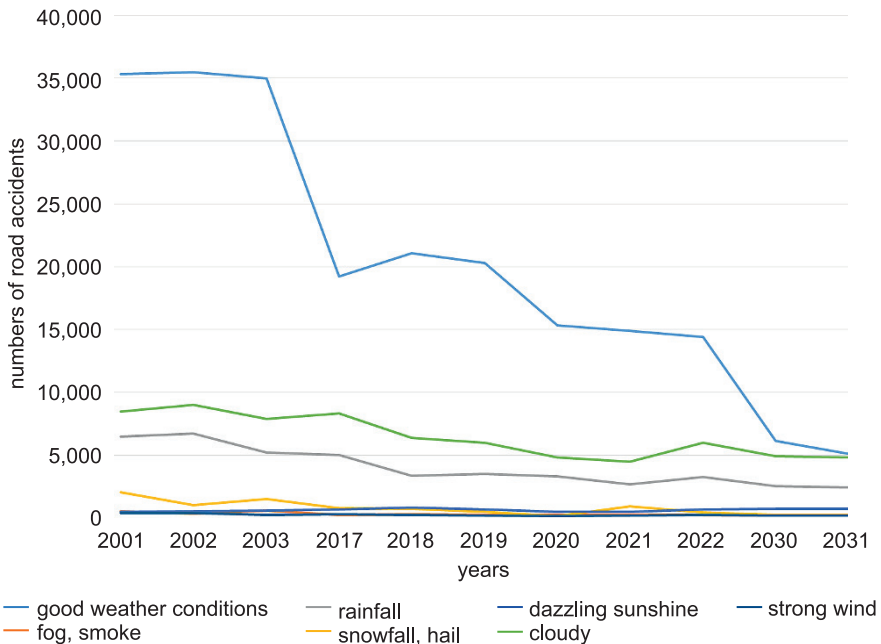


Fig. 3. Forecasted number of road accidents for 2022-2031

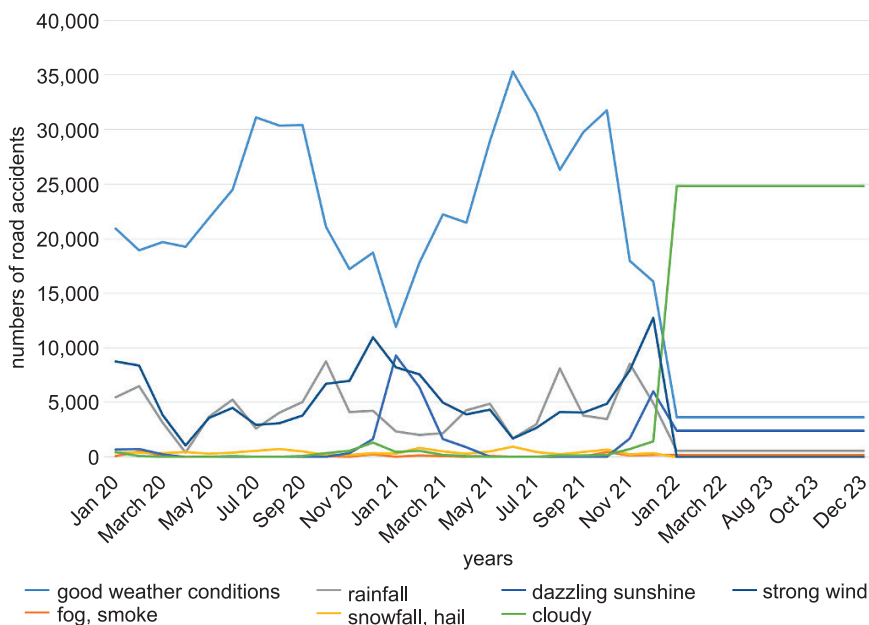


Fig. 4. Forecasted number of road accidents for 2022-2023

Conclusions

Forecasts of the number of accidents in Poland for the analysed weather conditions were determined by selected trend models using Excel. The results show that we can still expect a decrease in the number of traffic accidents in the coming years. It should be noted that the pandemic has skewed the results obtained, and if it continues and traffic restrictions are imposed, the proposed model may not be adequate. The average value of the MAPE for the cases analyzed was 3%, for annual data, may indicate 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 analysed country. These measures may include, for example, the introduction of higher fines for traffic offenses on Polish roads from January 1, 2022.

In his 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 road accidents. We can include traffic volume, day of the week or age of the accident perpetrator, among others.

References

- ABDULLAH E., EMAM A. 2016. *Traffic accidents analyzer using big data*. 2015 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, Las Vegas, NV, p. 392–397. <https://doi.org/10.1109/CSCI.2015.187>.
- AL-MADANI H. 2018. *Global road fatality trends estimations based on country-wise microlevel data*. Accident Analysis & Prevention, 111: 297–310. <https://doi.org/10.1016/j.aap.2017.11.035>.
- ARTEAGA C., PAZ A., PARK J. 2020. *Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach*. Safety Science, 132: 104988. <https://doi.org/10.1016/j.ssci.2020.104988>.
- BAK 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*. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), p. 1–5. <https://doi.org/10.1109/ICCCNT45670.2019.8944500>.
- 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>. 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: 2710–2721.
- CHUDY-LASKOWSKA K., PISULA T. 2015. *Prognozowanie liczby wypadków drogowych na Podkarpaciu*. Logistics, 4(2): 2782–2796.
- DUDEK G. 2013a. *Exponential smoothing models for short-term power system load forecasting*. Rynek Energii, 106(3): 14–19.
- DUDEK G. 2013b. *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.
- GORZELAŃCZYK P., PYSZEWSKA D., KALINA T., JURKOVIC M. 2020. *Analysis of road traffic safety in the Pila powiat*. Scientific Journal of Silesian University of Technology, Series Transport, 107: 33–52. <https://doi.org/10.20858/sjsutst.2020.107.3>.
- GREGORCZYK A., SWARCZYK M. 2012. *Analiza wariancji w układzie powtarzanych pomiarów do określenia efektów czynników wpływających na pozostałości linuronu w glebie*. Polish Journal of Agronomy, 11: 15–20.
- HELGASON A. 2016. *Fractional integration methods and short Time series: evidence from a simulation study*. Political Analysis, 24(1): 59–68. Retrieved from <http://www.jstor.org/stable/24573204>.
- KARLAFTIS M., VLAHOGIANNI E. 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>.

- KASHPRUK N. 2010. *Comparative research of statistical models and soft computing for identification of time series and forecasting*. Politechnika Opolska, 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*. Electronics, 8(8): 896. <https://doi.org/10.3390/electronics8080896>.
- KOWALSKI L. 2022. *Prognozowanie na podstawie szeregów czasowych*. Prognozowanie i Symulacje. Strona Lucjana Kowalskiego. Retrieved from <http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf>.
- KRZYCZKOWSKA Z. 2019. *Przy jakiej pogodzie najczęściej dochodzi do wypadków? Wcale nie podczas złych warunków*. Gazeta.pl. Moto.pl. Retrieved from <https://moto.pl/MotoPL/7,88389,25510393,przy-jakiej-pogodzie-najczesciej-dochodzi-do-wypadkow-wcale.html>.
- 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>
- Las losowy (2023). Wikipedia. Wolna encyklopedia. Retrieved from https://pl.wikipedia.org/wiki/Las_losowy.
- LAVRENZ S., VLAHOIANNI E., GKRTZA K., KE Y. 2018. *Time series modeling in traffic safetyresearch*. Accident Analysis & Prevention, 117: 368–380.
- LI L, SHRESTHA S., HU G. 2017. *Analysis of road traffic fatal accidents using data mining techniques*. 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., MASLOWSKA K., WOJDAN R. 2015. *Analiza i prognozowanie szeregów czasowych z programem SAS*. Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa.
- MAMCZUR M. 2020. *Jak działa regresja liniowa? I czy warto ją stosować?* Mirosław Mamczur. Blog o data science, sztucznej inteligencji, uczeniu maszynowym i wizualizacji danych Retrieved from <https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/>.
- MARCINKOWSKA J. 2015. *Metody statystyczne i eksploracji danych (data mining) w ocenie występowania omdleń w grupie częstoskurczu z wąskim zespołem QRS (AVNRT i AVRT)*. Katedra i Zakład Informatyki i Statystyki Uniwersytetu Medycznego im. Karola Marcinkowskiego, 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 factor 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. *Ekonometria. Modelowanie szeregów czasowych. Stacjonarność. Testy pierwiastka jednostkowego. Modele ARDL. Kointegracja*. 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.
- PILATOWSKA 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*. Publishing House of the University of Economics, 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>.
- SEBEGO M., NAUMANN R.B., RUDD R.A., VOETSCH K., DELLINGER A.M., NDLOVU C. 2011. *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.
- Statystyka. Portal Polskiej Policji. 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*. 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. <https://doi.org/10.2478/jaiscr-2014-0023>.
- Techniki zgłębiania danych (data mining)*. 2022. StatSoft Polska. Electronic Statistics Textbooks. Retrieved from https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html.
- The Global Status on Road Safety*. 2018. World Health Organization. Retrieved from <https://www.who.int/publications/i/item/9789241565684>.
- Top Advantages and Disadvantages of Hadoop 3*. 2022. DataFlair. Retrieved from <https://dataflair.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, Katowice.
- WROBEL M.S. 2017. *Application of neural fuzzy systems in chemistry*. PhD thesis. University of Silesia, Katowice.
- YANG Z., ZHANG W., FENG J. 2022. *Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework*. Safety Science, 146: 105522. <https://doi.org/10.1016/j.ssci.2021.105522>.
- 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.