Original Research Article

Forecasting the number of road accidents in Poland depending on weather conditions and COVID-19

Piotr Gorzelanczyk

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

ABSTRACT

COVID-19 fundamentally changed the way that people travel by road in Poland and throughout the world. The lack of mobility at the period, especially during the beginning of the epidemic, had a significant impact on the number of traffic accidents. The goals of this study are to forecast the number of accidents on the basis of weather in Poland and to assess how the COVID-19 epidemic has affected that number. For this objective, annual statistics on weather-related traffic accidents were acquired and evaluated. Based on previous information from police records, the number of traffic accidents was also forecasted for pandemic and non-pandemic variants in order to assess the impact of the pandemic. The number of traffic accidents in Poland was predicted using specific time series models and exponential models in relation to the weather. There is a decrease in the number of traffic accidents during the pandemic. Traffic accidents were on average 22% fewer in 2020 than they were in 2019, and by 2021, the difference was over 24%. When it snows or hails, this is extremely clear. This time period mostly saw the outbreak of the pandemic. The majority of traffic accidents occur when the weather is good. When the weather is bad, drivers are more cautious on the road. Prediction of traffic accidents is important for future planning and measures. The problem of estimating the number of traffic accidents, however, is not one that academics are particularly fond of.

Keywords: road accident; pandemic; Poland; forecasting; weather conditions

1. Literature review

Unfortunate events such as traffic accidents also cause property damage in addition to harming or killing other drivers. According to the WHO, 1.3 million people die in automobile accidents every year. The majority of countries in the globe attribute roughly 3% of their GDP to road accidents. The leading cause of death for children and young adults between the ages of 5 and 29 is vehicular accidents, according to the WHO[1]. The UN General Assembly hopes to see a 50% decrease in traffic-related fatalities and injuries by 2030.

The scope of a traffic collision is a consideration in determining its seriousness. Competent authorities must take into account the severity of the incidents when designing transport safety rules to prevent accidents, reduce injuries, fatalities, and property losses[2,3]. Identifying the critical factors that affect accident severity is required before taking action to prevent and minimize the severity of accidents[4]. Yang et al. [5] propose a DNN (Deep Neutral Network) multi-carbon architecture to predict different levels of injury, death, and property loss severity. It makes it possible to accurately and thoroughly assess how serious a road event is.

ARTICLE INFO

Received: 3 July 2023 | Accepted: 21 August 2023 | Available online: 4 December 2023

CITATION

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

COPYRIGHT

Copyright © 2023 by author(s). Sustainable Social Development is published by Asia Pacific Academy of Science Pte. Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), permitting distribution and reproduction in any medium, provided the original work is cited.
The size of a traffic accident is a consideration when determining its severity. Competent authorities must take into account the severity of the incidents when formulating transport safety laws in order to prevent accidents, reduce injuries, fatalities, and property losses\(^2,3\). It is important to recognize the key factors that affect accident severity before taking steps to prevent and minimize them. Yang et al.\(^5\) propose a multi-carbon DNN (Deep Neutral Network) architecture to forecast different levels of injury, death, and property loss severity. It makes it possible to evaluate road events’ seriousness in-depth and precisely\(^5\).

Intelligent transportation systems are currently the most major source of data for the analysis and prediction of traffic accidents. The data can be processed as a result of the widespread use of GPS devices in automobiles\(^6\). Microwave vehicle detection systems placed at roadside can continuously gather vehicle data (speed, traffic volume, vehicle type, etc.), according to Khaliq et al.\(^7\). Additionally, the Vehicle License Plate Recognition system can be used to collect a lot of traffic data over a monitored period of time\(^8\). Because of the inexperience of the reporters, social media may not be as effective as other sources of data when it comes to gathering information about traffic and accidents\(^9\).

For accident data to be useful, it is necessary to correctly examine a variety of data sources. By merging different data sources and diverse traffic accident data, the accuracy of the study’s findings is increased\(^10\).

A statistical analysis was undertaken by Vilaca et al.\(^11\) to assess the seriousness of accidents and establish their connection to other road users. A recommendation to increase the standard for driving safety and establish new transport safety policies is the study’s output.

Bąk et al.\(^12\) undertook a statistical examination of road safety in a specific Polish region based on the volume of traffic accidents and the speed at which their causes were identified. A thorough statistical analysis was used in the poll to examine how those who cause accidents see safety.

A source of accident data is selected for the analysis based on the type of traffic issue being addressed. When statistical models are integrated with extra data from actual driving or other data obtained through intelligent transportation systems, the accuracy of accident forecasts is improved and the number of accidents is decreased\(^13\).

There are numerous methods for predicting the number of accidents in the literature. Time series techniques are typically employed to forecast the number of traffic accidents\(^14,15\). However, these techniques have a number of shortcomings, including the inability to assess a forecast’s accuracy in light of previously issued forecasts and the frequent autocorrelation of the residual component\(^16\). Prochážka et al.\(^17\) multiple seasonality model and Sunny et al.\(^18\) Holt-Winters exponential smoothing method were both used for forecasting. Exogenous variables cannot be included in the model, which is one of its limitations\(^19,20\).

The number of road accidents has also been predicted using the vector autoregression model, which has the drawback of requiring a large number of observations of the variables in order to precisely estimate their parameters\(^21\), as well as the autoregression models of Monederoa et al. for analyzing the number of fatalities\(^22\), and Al-Madani, curve-fitting regression models\(^23\). Since the series are already considered to be stationary, these just require straightforward linear correlations and the order of the autoregression\(^24\).

Biswas et al.\(^25\) used Random Forest regression to predict the frequency of traffic accidents. The linked feature groups in the data have similar relevance to the original data, smaller groups are preferred over larger ones in this case, and the technique and spike prediction are unstable\(^26\). Chudy-Laskowska and Pisula\(^27\) employed the exponential equalization model, the univariate periodic trend model, and the autoregressive model with quadratic trend for the forecasting issue they discussed. It is also possible to forecast the issue at
hand using a moving mean model, however this approach has the disadvantages of having poor forecast accuracy, data loss over time, a lack of trend, and a lack of seasonal consideration\cite{28}.

The GARMA approach, which imposes restrictions on the parameter space to guarantee the process’ stationarity, was used by Prochozka and Camej\cite{29}. Forecasting is typically done using the ARIMA or SARIMA model for non-stationary systems\cite{18,29–31}; the ARMA model for stationary processes. Due to the fact that successful model identification involves more research expertise than, say, regression analysis, these models offer the stated models a significant level of freedom, but this has a downside as well\cite{32}. Another flaw in the ARIMA model is its linearity\cite{33}.

Chudy-Laskowska and Pisula used the ANOVA method in their 2015 study to forecast the frequency of traffic accidents. The disadvantage of this approach is the addition of additional presumptions, especially the presumption of sphericity, whose violation may lead to inaccurate findings. Neural network algorithms are also used to predict the frequency of car accidents. The limitations of ANN include the need for prior expertise in the field\cite{34,35}, the reliance of the final result on the network’s initial conditions, and the lack of conventional interpretability because ANN is frequently referred to as a “blackbox” where you provide input and the model produces results without being aware of the analysis\cite{36}.

The Hadoop model was employed by Kumar et al\cite{37} as a brand-new prediction method. A drawback of this technology is that it cannot process small data files\cite{38}. Karlaftis and Vlahogianni\cite{31} used the Garch model to generate predictions. A disadvantage of the method is its complexity\cite{39,40}. Although the ADF test\cite{41} has the disadvantage of having low power when the random component is autocorrelated\cite{42} when it was used by McIlroy and his team.

Data-mining approaches have also been used by researchers to make predictions\cite{43,44}, although these methods often have the downside of requiring the user to deal with vast quantities of broad descriptions\cite{45}. The model combination offered by Sebego et al\cite{46} as a combination of numerous models is relatively typical. The idea of parametric models is also included in Bloomfield’s\cite{47} work.

The issue of forecasting is taken into consideration by many researchers using a variety of study approaches in light of the aforementioned literature. The problem of estimating the number of traffic accidents, however, is not one that academics are particularly fond of. This is the reason why the author’s research focused on this issue, to forecast Poland’s annual traffic accident rate using weather data. He used a few time series and exponential models.

2. Materials and methods

The purpose of the article is to estimate the number of accidents that will occur on Poland’s roads based on the present weather and to assess how the COVID-19 pandemic will effect that estimate. The article ignores the influence of weather on the frequency of road accidents. The article’s objective is to answer the question, “Under what weather conditions can we anticipate experiencing how many traffic accidents?” without taking into account any external factors that can affect this number.

The author will examine how the amount of road accidents is impacted by the weather in the piece that follows. This was accomplished by the yearly collection of collision statistics and the analysis of weather information. The number of road accidents for the pandemic and non-pandemic variations were anticipated and compared based on previous data from police statistics in order to ascertain whether the pandemic affected the number of road accidents in Poland depending on meteorological conditions.

A couple exponential alignment models were used to forecast how many traffic accidents would occur in response to the weather. A weighted moving average of earlier values is used in this case to reduce the
variability of the predicted variable’s studied time series, with the weights chosen in line with the exponential function. During the inquiry, the statistic program was utilized to select the weights for the applicable analyses that worked best.

The anticipated number of traffic accidents under the analyzed weather circumstances was calculated using the weighted average of the current and historical records. The model selection, as well as the model’s optimum parameter and weight values, affect the forecast’s outcomes.

The number of accidents in Poland was predicted using a few time series models with a linear trend. One of the methods presented is the Brownian approach, which is known as an exponential smoothing technique. It is most often used when a time series lacks a trend, which means that its variances are brought on by chance events that occur when projecting the number of traffic accidents, and the series being used does not reflect a developing trend.

The weights used in this method are determined according to the right exponential. The model of the change in \( lwd(t) \) over the analyzed time takes the form: (1) For the first moment in time:

\[
Plwd_1^* = lwd_o
\]

(2) To subsequent time periods:

\[
Plwd_t^* = \alpha \ast Plwd_{t-1}^* + (1 - \alpha) \ast Plwd_{t-1}^*
\]

where, \( Plwd_{t-1}^* \) is the forecast value of the number of traffic accidents for the optimal value of the smoothing parameter \( \alpha \), and \( \alpha \) is the constant value of the process smoothing parameter taking a value in the range \( \alpha \in (0, 1) \).

While calculating measures of analytical forecast perfection the following errors of expired forecasts determined from Equations (3)–(7) were used:

- **ME**—Mean error:

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

(3)

- **MAE**—Mean average error:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - Y_p|
\]

(4)

- **MPE**—Mean percentage error:

\[
MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - Y_p}{Y_i}
\]

(5)

- **MAPE**—Mean absolute percentage error:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i - Y_p|}{Y_i}
\]

(6)

- **SSE**—Mean square error:

\[
SSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - Y_p)^2}
\]

(7)

where: \( n \): The length of the forecast horizon; \( Y \): Observed value of road accidents; \( Y_p \): Forecasted value of road accidents.
In order to compare the number of accidents during the pandemic and if the pandemic had not occurred for the weather conditions analyzed, the mean absolute percentage error was minimized. In the research, the author considered other factors affecting the number of traffic accidents such as: speed, traffic volume, vehicle type, alcohol intake status, road shape, rode slope, driver license, gender and etc.

3. Results

The variability of the number of road accidents depending on weather conditions was carried out using the Kruskall-Wallis test. In the analyzed case the value of the test statistic is 133.8816 with probability \( p < 0.05 \).

Based on the results of the test, it can be said that the hypothesis, which states that the average number of traffic accidents is equal regardless of the weather, should be rejected. On this foundation, it is possible to draw the conclusion that the number of accidents studied over time demonstrates a consistent decline in the average level of accidents. Additionally, there is a definite change in the amount of accidents depending on the weather. In good weather, there are more accidents than in bad, and there are less accidents in bad weather. This is the time when people travel more cautiously and carefully (Figure 1).

38 million people call 31,2705 km\(^2\) of Poland home. According to police statistics, fewer accidents occurred on Polish roads during the analyzed period each year. When it rains, snows, or hails, the decline is compounded, making it very clear (Figure 2). The epidemic also decreased the amount of road accidents between 2020 and 2021. During the outbreak, fewer car accidents are reported. Traffic accidents were on average 22\% fewer in 2020 than they were in 2019, and by 2021, the difference was over 24\%. When it snows or hails, this is extremely clear.
This time period mostly saw the outbreak of the pandemic. The majority of traffic accidents occur when the weather is good. When the weather is bad, drivers are more cautious on the road. In comparison to the rest of the European Union, Poland continues to have a high accident rate\cite{48,49}.

Figure 2. Number of road accidents in Poland depending on weather conditions.

Road accidents have a seasonal component, according to a review of the data on their frequency. This enables subsequent investigations to forecast the amount of traffic accidents using the presumptive time series models. Polish police statistics from 2001 to 2021 were used to predict the frequency of accidents based on meteorological conditions. Weather conditions were divided into the following categories:

- Good weather conditions;
- Fog, Smoke;
- Precipitation;
- Snowfall, Hail;
- Dazzling sun;
- Cloudy;
- Strong wind.

In order to investigate the impact of the pandemic on the number of road accidents under different weather conditions, the study was divided into two time frames:

- 2001–2021 (considering the occurrence of pandemic).
- 2001–2019 (considering that there were no pandemic).

The study assumes that the onset of the pandemic is in 2020. The individual forecasting methods used in the study are coded $M_1$, $M_2$, ......, $M_n$. The individual forecasting techniques used in the study are as follows:

- 2-point moving average method;
- 3-point moving average method;
- 4-point moving average method;
- Exponential smoothing method without trend seasonal component: None;
- Exponential smoothing method no trend seasonal component: Additive;
- Exponential smoothing method no trend seasonal component: Multiplicative;
- Exponential smoothing of the seasonal component of a linear trend: None, HOLT;
- Exponential smoothing of the seasonal component of a linear trend: Additive;
- Exponential smoothing of the seasonal component of a linear trend: Multiplicative WINTERS;
- Exponential smoothing exponential seasonal component: None;
Exponential smoothing exponential seasonal component: Additive;
Exponential smoothing exponential seasonal component: Multiplicative;
Exponential smoothing seasonal component of trend distribution: None, exponential smoothing;
Seasonal component of trend distribution exponentially smoothed: Additive;
Exponential smoothing seasonal component of trend distribution: Multiplicative.

Forecasting methods for which the average percentage error was the smallest were selected for further analyses. Forecasting methods for the following weather conditions were obtained as the best:

Pandemic occurrence:
- Good weather conditions—Exponentially smoothed trend vanishing seasonal component: Multiplicative;
- Fog, smoke—Exponentially smoothed vanishing trend seasonal component: Additive;
- Rainfall—Seasonal component of declining trend smoothed exponentially: Additive;
- Snowfall, hail—Seasonal component of the fading trend smoothed exponentially: Multiplicative;
- Glare—Exponential smoothing without trend seasonal component: Additive;
- Cloudy weather—Exponential smoothing with declining trend seasonal component: Multiplicative;
- Strong wind—Exponentially smoothed seasonal component with vanishing trend: Additive.

No pandemic:
- Good weather—Exponential smoothing with no trend seasonal component: Additive;
- Fog, smoke—Exponentially smoothed seasonal component with vanishing trend: Additive;
- Rainfall—Exponential smoothing with no trend seasonal component: Additive;
- Snowfall, hail—Exponentially smoothed seasonal component with vanishing trend: Multiplicative;
- Glare—Exponential smoothing with no trend seasonal component: Multiplicative;
- Cloudiness—Exponential smoothing with decreasing trend seasonal component: Multiplicative;
- Strong wind—Exponential smoothing with decreasing trend seasonal component: Additive.

It is possible to conclude that the pandemic decreased the number of traffic accidents by an average of 23% based on the findings obtained after accounting for the pandemic. Due to the lack of police information on the number of traffic accidents by weather conditions and months, the year 2020 was assumed to be the beginning of the pandemic during the study. For this comparison, the actual number of road accidents reported by the police (black line) and the predicted number of road accidents in 2020 and 2021 for various weather conditions were used. Figures 3–9 show the results of the forecasting process, and Tables 1 and 2 show the mistakes. As we can see, the pandemic influenced the number of road accidents and reduced its projected number.
Figure 3. Comparison of the number of road accidents in good weather conditions with and without pandemic.

Figure 4. Comparison of the number of road accidents when there is fog, smoke with and without pandemic.

Figure 5. Comparison of the number of road accidents when rainfall occurs with and without pandemic.
Figure 6. Comparison of the number of road accidents if snow, hail with and without pandemic.

Figure 7. Comparison of the number of road accidents if there is dazzling sun with and without pandemic.

Figure 8. Comparison of number of road accidents when it is cloudy with and without pandemic.
Figure 9. Comparison of the number of road accidents when there is strong wind with and without pandemic.

Table 1. Measurement errors.

<table>
<thead>
<tr>
<th>Errors</th>
<th>Good weather conditions</th>
<th>Fog, smoke</th>
<th>Rainfall</th>
<th>Snowfall, hail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With pandemic</td>
<td>No pandemic</td>
<td>With pandemic</td>
<td>No pandemic</td>
</tr>
<tr>
<td></td>
<td>1.48 × 10^3</td>
<td>3.29 × 10^2</td>
<td>1.93 × 10^3</td>
<td>2.69 × 10^2</td>
</tr>
<tr>
<td></td>
<td>4.62 × 10^3</td>
<td>1.72 × 10^3</td>
<td>1.35 × 10^3</td>
<td>1.17 × 10^3</td>
</tr>
<tr>
<td></td>
<td>1.70 × 10^6</td>
<td>6.87 × 10^6</td>
<td>4.85 × 10^6</td>
<td>6.37 × 10^6</td>
</tr>
<tr>
<td></td>
<td>2.43 × 10^0</td>
<td>1.75 × 10^0</td>
<td>2.96 × 10^0</td>
<td>1.37 × 10^0</td>
</tr>
<tr>
<td></td>
<td>1.53 × 10^1</td>
<td>6.39 × 10^1</td>
<td>3.21 × 10^1</td>
<td>2.87 × 10^1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Errors</th>
<th>Good weather conditions</th>
<th>Fog, smoke</th>
<th>Rainfall</th>
<th>Snowfall, hail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With pandemic</td>
<td>No pandemic</td>
<td>With pandemic</td>
<td>No pandemic</td>
</tr>
<tr>
<td></td>
<td>5.90 × 10^6</td>
<td>9.26 × 10^6</td>
<td>4.20 × 10^5</td>
<td>2.64 × 10^2</td>
</tr>
<tr>
<td></td>
<td>6.22 × 10^0</td>
<td>5.98 × 10^1</td>
<td>1.51 × 10^3</td>
<td>1.28 × 10^1</td>
</tr>
<tr>
<td></td>
<td>6.25 × 10^3</td>
<td>4.83 × 10^3</td>
<td>1.18 × 10^7</td>
<td>6.17 × 10^6</td>
</tr>
<tr>
<td></td>
<td>2.02 × 10^1</td>
<td>2.44 × 10^1</td>
<td>3.49 × 10^1</td>
<td>2.12 × 10^0</td>
</tr>
<tr>
<td></td>
<td>1.01 × 10^3</td>
<td>9.66 × 10^0</td>
<td>1.85 × 10^1</td>
<td>1.55 × 10^1</td>
</tr>
</tbody>
</table>

4. Conclusion

Based on the investigated climatic conditions, the statistical program statistic was used to estimate the expected number of accidents in Poland. To lower the mean absolute percentage error, the software computed the study’s weights.

According to the data gathered, the epidemic decreased Poland’s annual average number of road accidents by 23%. These ranges may change depending on the meteorological factors taken into account. These ranges change according to the climate, ranging from 2% during rains to 277% during winter, hail, and a decrease in traffic on Polish roads — all of which correspond to the pandemic’s waves.

The study’s findings suggest that it is still realistic to expect that there will be fewer accidents on Polish roads in the future, especially in light of the current COVID-19 pandemic. The measured forecast error levels show how trustworthy the used models are. It is also possible to assert that the number of accidents is significantly influenced by the weather. The prediction methods applied in this study are also very accurate.
The best method for the scenarios under examination was proven to be exponential smoothing with a declining trend, with the average absolute percentage error being typically modest.

The data gathered in the study for estimating the number of road accidents can be used to build new strategies in the future to decrease the number of accidents dependent on weather conditions. Future research by the authors will incorporate other factors influencing Poland’s accident rates. It can be the amount of traffic or the age of the perpetrator.

Conflict of interest

The author declares no conflict of interest.

References


45. Marcinkowska K. Statistical methods and data mining in assessing the occurrence of syncope in the group of narrow QRS complex (AVNRT and AVRT) (Polish) [PhD thesis]. Medical University in Poznan; 2015.


