

Predicting Road Accident Risk in the City of San Pablo, Laguna: A Predictive Model Using Time Series Forecasting Analysis with Multiplicative Model

Reymar V. Manaloto*, Ronnel A. Dela Cruz

College of Computer Studies, Laguna State Polytechnic University, San Pablo Campus, Philippines

Received July 19, 2020; Revised August 27, 2020; Accepted September 29, 2020

Cite This Paper in the following Citation Styles

(a): [1] Reymar V. Manaloto, Ronnel A. dela Cruz, "Predicting Road Accident Risk in the City of San Pablo, Laguna: A Predictive Model Using Time Series Forecasting Analysis with Multiplicative Model," *Computer Science and Information Technology*, Vol. 8, No. 4, pp. 75 - 80, 2020. DOI: 10.13189/csit.2020.080401.

(b): Reymar V. Manaloto, Ronnel A. dela Cruz (2020). *Predicting Road Accident Risk in the City of San Pablo, Laguna: A Predictive Model Using Time Series Forecasting Analysis with Multiplicative Model*. *Computer Science and Information Technology*, 8(4), 75 - 80. DOI: 10.13189/csit.2020.080401.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This research used the mechanisms in time series analysis such as secular trend, irregular fluctuation, cyclical and seasonal patterns to conclude forecast using the multiplicative model. In the conduct of this study, time, day and location are the main focuses in predicting road accident patterns. By analyzing road accident patterns based on time series forecasting using a multiplicative model, this research will provide suggestions for the government to take effective measures to reduce accident impacts and improve traffic safety. This research analyzed the road accident data in San Pablo City, Laguna Philippines from the year 2014 – 2016 and forecast the possible prevalence of road accidents and its pattern. A total of 1229 road accidents were included in this study. As recorded based on the cumulative frequency, Barangay San Francisco has the greatest number of road accidents with an average of 101 cases per year. It can be attributed to the fact that the road is considered the busiest because it is the only gateway of provincial travelers from Metro Manila to Southern Provinces. While for the monthly pattern prediction, April is the most risky in road accidents with possible 33 cases in a year with 13.08% mean absolute error or 86.92% accuracy, probably because most of the community is in summer vacation. And in terms of the daily pattern, Sunday is the crucial day in terms of road accidents with 44 possible cases with 10.29% mean absolute error or 89.71% accuracy, and the majority of the possible road accident arises between 6:00 pm to 9:00 pm

with 45 possible cases with 13.21% mean absolute error or 86.79% accuracy.

Keywords Road Accidents, Predictive Model, Time Series, Multiplicative Model, Forecasting

I. Introduction

Based on the survey conducted by the World Health Organization last 2017, road accident is one of the top leading causes of deaths in the world. Tragically, as many as 1.3 million deaths are a result of road accidents every year. It is the leading cause of accidental death. Road accidents often lead to death as the impact of a collision can cause severe injuries such as head trauma (brain swelling, hemorrhages, and blood clots), external and/or internal bleeding, or a broken neck. Being one of the most significant causes of deaths, a more effective and efficient accident management is needed. [1] A Predictive Model using time series forecasting Analysis with the Multiplicative Model will be used in predicting road accident risk in the City of San Pablo, Laguna.

The multiplicative Model assumes that as the data increase, so does the seasonal pattern. Most time series plots exhibit such a pattern. In this model, the trend and seasonal components are multiplied and then added to the

error component.

Predictions are made for new data when the actual outcome may not be known until some future date. The future is being predicted, but all prior observations are almost always treated equally. Time series analysis involves developing models that best capture or describe an observed time series to understand the underlying causes. This often involves making assumptions about the form of the data and decomposing the time series into constitution components. Time Forecasting involves taking models to fit historical data and using them to predict future observations. The skill of a time series forecasting model is determined by its performance at predicting the future. This is often at the expense of being able to explain why a specific prediction was made, confidence intervals, and an even better understanding of the underlying causes behind the problem.

The researcher will be focusing on the data gathered from the San Pablo City Police Station from the year 2015 up to 2017. San Pablo City Police Station is located in the southern tip of the Province of Laguna in the island of Luzon at geographical coordinates, 14 degrees 4' north latitude and 121 degrees 19' east longitude.

San Pablo City Police Station is operating on manual procedures when they are managing records by simply filing it in folders and forms. They are also manually drawing the map in the old school application Paint. With the invention of new technologies come a variety of new ways for law enforcement agencies to reduce the number of road accidents in San Pablo City, Laguna.

This research study focuses mainly on predicting road accident risk using time series forecasting Analysis with Multiplicative Model in the City of San Pablo, Laguna based on the data gathered from the San Pablo City, Laguna Police Station.

2. Objectives of the Study

The general objective of the study is to predict road accident risk in San Pablo City, Laguna using Time Series Forecasting Analysis with the Multiplicative Model.

Specifically, the study aims to:

- i To analyze road accidents patterns using the multiplicative model;
- ii To provide suggestions for the government to take effective measures to reduce accident impacts and improve traffic safety; and
- iii To document the findings and further analyze road accident patterns in the City of San Pablo.

3. Review of Related Literature

According this research [6] road accidents and injuries occur because of human fault or vehicle fault or

infrastructure fault or sometimes combinations of these factors. Each of these factors individually or in combination may cause an accident.

Moreover, as cited by [2] one important aspect of road safety (time series) analysis is that road safety cannot be measured unambiguously. There is no unique measure of road safety. Usually, road safety is measured in terms of the amount of 'lack of road safety', for instance, the number of accidents occurring per time unit. Even if the number of accidents is selected as the measure of road safety, it could still be all accidents, injury accidents, serious accidents or fatal accidents, or other types of accidents. But even then, the number of victims per accident may be of interest, as well as the number of fatalities per accident. It should further be considered that influences on road safety may primarily affect certain parts of the road safety process.

Some research analyzed traffic incidents to improved traffic management service "Critical Transport Infrastructure in Urban Areas: Impacts of Traffic Incidents Assessed Using Accessibility Based Network Vulnerability Analysis" [3], and "Analysis of traffic injury severity: An application of non-parametric classification tree techniques" [4], these researches used statistical regression model in analyzing traffic incident. However, this research uses the Time series forecasting model that focuses on historical data to get the seasonal pattern and predict possible risks in the future. Moreover, another research that is same with this study that is conducted [5] the study entitled, A Multiplicative Time Series Model for Air Transport Demand Forecasting, concluded that when the seasonal pattern changes, these will be appropriately projected into the forecast. However, an initially remote forecast will be continually updated and as the lead shortens, greater accuracy will be possible.

4. Methodology

This research study focused mainly on predicting road accident risk using time series forecasting Analysis with Multiplicative Model in the City of San Pablo, Laguna based on the data gathered from the San Pablo City, Laguna Police Station. This model assumes that as the data increase, so does the seasonal pattern. Most time series plots exhibit such a pattern. In this model, the trend and seasonal components are multiplied and then added to the error component. Data that is being used for analysis is from the year 2015 up to the year 2017.

Time Series Forecasting using the Multiplicative Model

In the multiplicative model an original time series is expressed as a product of three components. [7].

- the trend-cycle component (TCt);
- the seasonal effects (St);

- the irregular fluctuations (It);
- the Observed series (Ot),

Therefore, the multiplicative decomposition is expressed by the following equation:

$$\text{Observed series} = \text{Trend} \times \text{Seasonal} \times \text{Irregular}$$

$$O_t = T C_t \times S_t \times I$$

[8] Smoothing: Moving average smoothing

In Moving Average Smoothing, each observation is assigned an equal weight, and each observation is forecasted by using the average of the previous observation(s). Using the time series $X_1, X_2, X_3, \dots, X_t$, this smoothing technique predicts X_{t+k} 5.2 Smoothing Time Series. (N.D.). as follows.: $S_t = \text{Average}(x_{t-k+1}, x_{t-k+2}, \dots, x_t)$, $t = k, k+1, k+2, \dots, N$ where, k is the smoothing parameter.

The absolute mean percentage error

$$\text{MAPE} = \frac{\sum \frac{|A-F|}{A} \times 100}{N}$$

A= Actual, F= Forecast, N= Number of observations, and the vertical bars stand for absolute values.

5. Results and Discussion

This research study focused mainly on predicting road accident risk using time series forecasting Analysis with Multiplicative Model in the City of San Pablo, Laguna based on the data gathered from the San Pablo City, Laguna Police Station. This model assumes that as the data increase, so does the seasonal pattern. Most time series plots exhibit such a pattern. In this model, the trend and seasonal components are multiplied and then added to the error component. Data that is being used for analysis is from the year 2015 up to the year 2017.

Based on the result acquired in the combined analysis of traffic incidents by location from 2014 to 2016, a total of 1229 occurrences were stated in San Pablo City, Laguna composed of thirty (30) Barangays excluding the other fifty-one (51) Barangays that do not have any traffic incidents from the inclusive years of the study. Table 1 (Traffic Location by Incidents) shows the occurrence of a traffic incident in the Barangay. Barangay San Francisco, San Pablo City, Laguna, got the greatest number of cases with 203 incidents, followed by Bagong Pook, Barangay VI-C, San Pablo City, Laguna with 145 incidents.

Table 1. Traffic Location by Incidents

No.	Location	Incident
1	Barangay San Francisco	203
2	Baong Pook VI-C	145
3	Barangay San Rafael	109
4	Barangay San Nicolas	93
5	Barangay Bagong Bayan II-A	88
6	Barangay San Ignacio	79
7	Barangay II-B	45
8	Barangay Concepcion	37
9	Barangay I-A	35
10	Barangay San Roque	34
11	Barangay Bagong Pook VI-C	31
12	Barangay San Cristobal	30
13	Barangay San Antonio II	28
14	Barangay Del Remedio	26
15	Baramgay San Marcos	25
16	Barangay San Concepcion	25
17	Barangay II – E	23
18	Barangay II – C	24
19	Barangay V – A	20
20	Barangay I – B	16
21	Barangay San Antonio I	15
22	Barangay III – C	14
23	Barangay II – F	14
24	Barangay San Gabriel	13
25	Barangay San Gregorio	11
26	Barangay Sto. Angel	11
27	Barangay Sta Maria Magdalena	9
28	Barangay V – B	9
29	Barangay I – C	9
30	Barangay IV – C	8
	Total	1229

Cumulative Frequency Analysis shows the relationship and related pattern between Time Series Forecasting as shown in Fig.1 and shows a detailed presentation in Table 2. As predicted based on the Multiplicative model it shows that the most risk in traffic incidents is in the month of April with thirty-three (33) possible cases in a year, followed by the month of June with twenty-five (25) possible cases and month of February with the least traffic incident of six (6) possible cases in a year 13.08% mean absolute error or 86.92% accuracy.

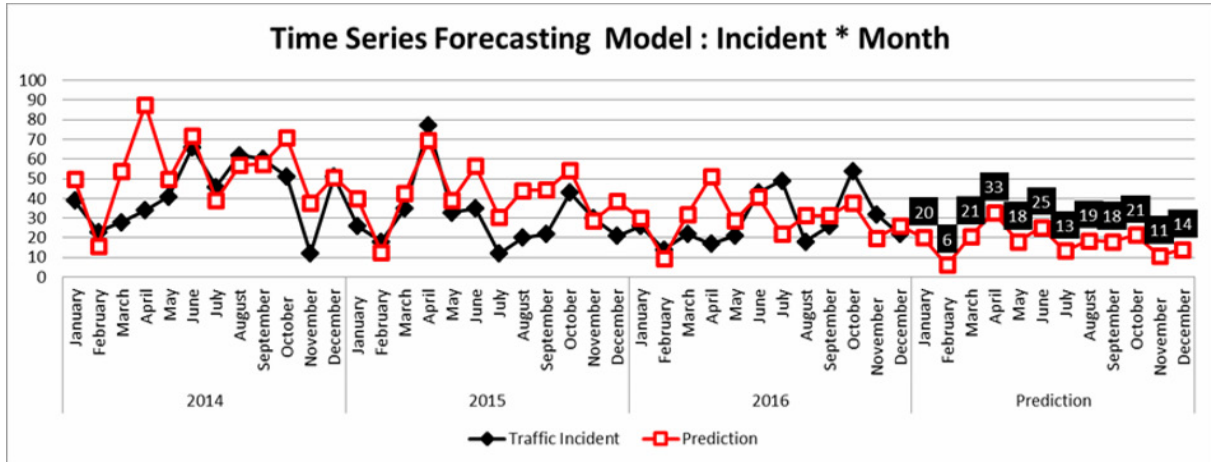


Figure 1. Traffic Incidents by Month

Table 2. Traffic Incidents by Month

Incident by Month (Prediction*)					
Month	2014	2015	2016	Total	P*
January	39	26	26	91	20
February	23	18	14	55	6
March	28	35	22	85	21
April	34	77	17	128	33
May	41	33	21	95	18
June	66	35	43	144	25
July	46	12	49	107	13
August	62	20	18	100	19
September	60	22	26	108	18
October	51	43	54	148	21
November	12	30	32	74	11
December	51	21	22	94	14
TOTAL				1229	

Figure 2 shows the predicted cases based on the Multiplicative model in terms of Incident by Day and actual detailed cases shown in table 3, it shows that the most critical day in traffic incidents happened every Sunday with forty four (44) possible cases in a week, followed by Saturday with thirty eight (38) possible cases and Monday with twenty nine (29) possible cases in a week with 10.29% mean absolute error or 89.71% accuracy.

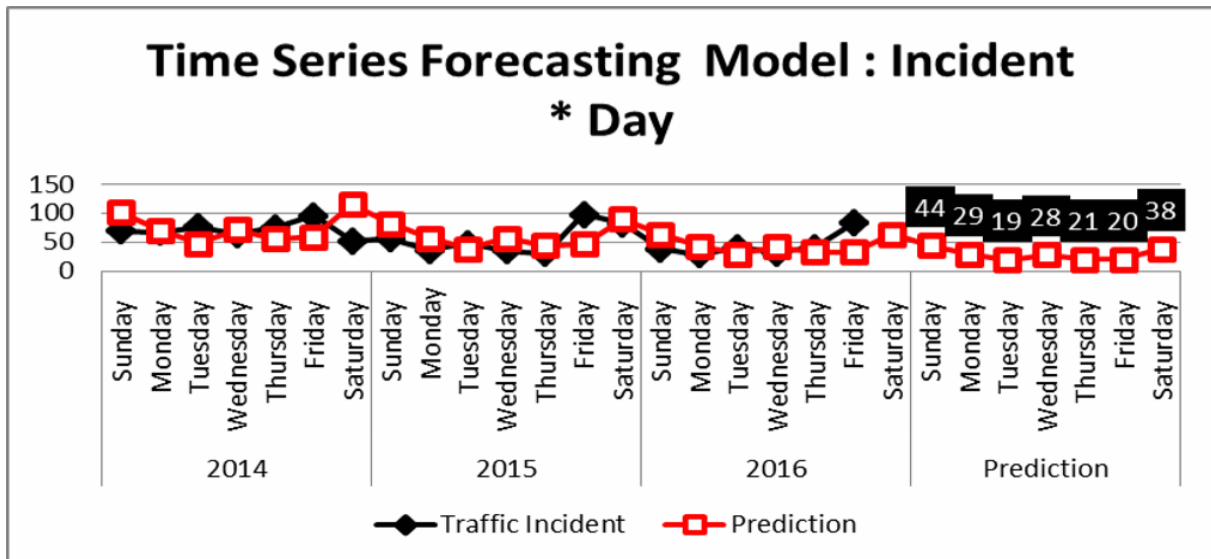


Figure 2. Traffic Incidents by Day

Table 3. Traffic Incidents by Day

Incident by Day (Prediction*)					
Day	2014	2015	2016	Total	P*
Sunday	84	52	82	218	44
Monday	69	56	38	163	29
Tuesday	67	36	28	131	19
Wednesday	76	48	41	165	28
Thursday	64	35	31	130	21
Friday	75	30	41	146	20
Saturday	95	98	83	276	38
TOTAL				1229	

In Figure 3 and Table 4 shows the crucial time that the traffic incident will possibly rise is between 06:00pm to 9:00pm and also this particular time is considered by many as “rush hours”. Moreover the second most possible incident occurrence in terms of time is between 01:00pm to 06:00pm with thirty six (36) possible traffic incidents cases with 13.21% mean absolute error or 86.79% accuracy.

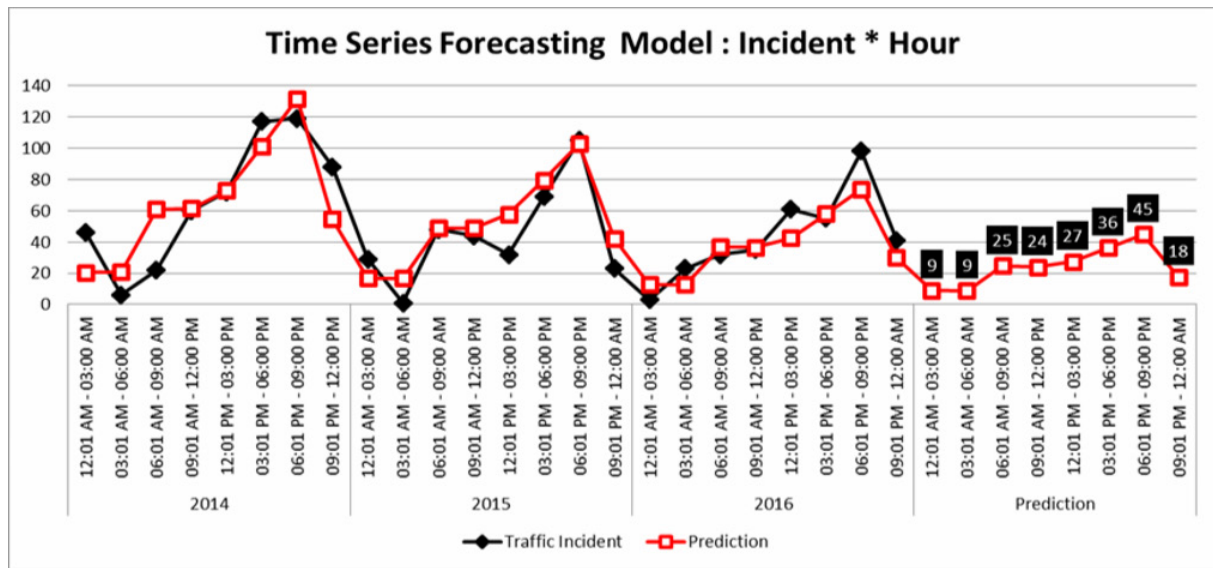


Figure 3. Traffic Incidents by Hour

Table 5. Traffic Incidents by Hour

Incident by Hour (Prediction*)					
Hour	2014	2015	2016	Total	P*
12:01 AM - 03:00 AM	46	29	3	78	9
03:01 AM - 06:00 AM	6	1	23	30	9
06:01 AM - 09:00 AM	22	48	32	102	25
09:01 AM - 12:00 PM	60	44	35	139	24
12:01 PM - 03:00 PM	72	32	61	165	27
03:01 PM - 06:00 PM	117	69	55	241	36
06:01 PM - 09:00 PM	119	105	98	322	45
09:01 PM - 12:00 PM	88	23	41	152	18
TOTAL				1229	

6. Conclusions

After the completion of the project, the following outputs are realized:

- i.) Multiplicative Model analyzed the road accident patterns recorded in San Pablo City Police Station;
- ii.) The study helps the police force in managing and controlling road accidents and aids improvement in traffic safety management and take effective measures to reduce accident impacts and improve traffic safety management;
- iii.) It helps the government to have effective and efficient decision-making in terms of road and traffic management based on the result of the study; and
- iv.) It also helps future researchers whose interest in road accident pattern, to use data set on this study and compare it to other predictive models for better prediction accuracy.

Acknowledgment

The researchers would like to extend their sincerest thanks to all the people who became part of this research. Special thanks are extended to the Philippine National Police – San Pablo City Chapter for providing all the necessary data and information needed for this research, to Laguna State Polytechnic University Administration for all the support, and to the College of Computer Studies for the motivation to complete this research paper.

REFERENCES

- [1] World Health Organization: WHO. "The Top 10 Causes of Death." <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, 24 May 2018, www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
- [2] Commandeur, Jacques JF, et al. "On statistical inference in time series analysis of the evolution of road safety." *Accident Analysis & Prevention* 60 (2013): 424-434.
- [3] Taylor, Michael AP. "Critical transport infrastructure in Urban areas: impacts of traffic incidents assessed using accessibility - based network vulnerability analysis." *Growth and Change* 39.4 (2008): 593-616.
- [4] Chang, Li-Yen, and Hsiu-Wen Wang. "Analysis of traffic injury severity: An application of non-parametric classification tree techniques." *Accident Analysis & Prevention* 38.5 (2006): 1019-1027.
- [5] Awad, Mohammed. "A Multiplicative Time Series Model." <https://www.slideshare.net/>, 29 Nov. 2011, www.slideshare.net/wings_of_wisdom/a-multiplicative-time-series-model.
- [6] Vasavi, S. "Extracting hidden patterns within road accident data using machine learning techniques." *Information and Communication Technology*. Springer, Singapore, 2018. 13-22.
- [7] "Additive Models and Multiplicative Models." <https://support.minitab.com/>, support.minitab.com/en-us/minitab/19/help-and-how-to/statistical-modeling/time-series/supporting-topics/time-series-models/additive-and-multiplicative-models. Accessed 13 Sept. 2020.
- [8] "Statistics Online | STAT ONLINE." PennState: Statistics Online Courses, online.stat.psu.edu/statprogram. Accessed 13 Sept. 2020.