Insights for Improving Road Safety : Focusing on Vehicle Accidents in Daegu Metropolitan City

Mee Qi Siow, Yang Sok Kim, Mi Jin Noh,
Choong Kwon Lee, Sang Ill Moon, Jae Ho Shin

Abstract

Road accidents not only caused loss of human lives but also costed 3% of gross domestic product in most of the countries. The road accidents pose significant challenges to public safety and urban transportation management. There is a need to identify the high-risk area of accidents along with the critical day of week and vulnerable time period in order to implement effective preventive measures and optimizing the resource allocation. We collected 5,012 accident data from 대구교통종합정보. This study identified the high-risk locations, days of week, and time periods for accidents in Daegu and estimated the conditional probabilities of accidents occurring based on combinations of location, day of the week, and time period. The result is visualized in the form of dashboard in Tableau. This study holds substantial practical significance for urban planners, transportation authorities, and policymakers in Daegu to strategically allocate resources for traffic management, law enforcement, and targeted safety campaigns.

Keywords: Road accidents | Road safety | Conditional Probability | Location | Day of week| Time period

I. INTRODUCTION

According to the fact sheets released by the World Health Organization in June 2022, the casualties caused by road traffic crashes were approximately 1.3 million every year [1]. Centers for Disease Control and Prevention [2] have listed the factor of accidents as one of the leading causes of death. Besides causing loss of human lives, macroeconomically, these accidents cost most countries 3% of their gross domestic product [3]. Road accidents pose significant challenges to public safety and urban transportation management, and understanding the factors contributing to accidents and identifying the areas where accidents happen frequently is crucial for effective preventive measures and optimizing resource allocation.

As one of the largest cities in South Korea, the Daegu Metropolitan City reported that vehicles registered in Daegu increased by 1.15% in 2021 compared to the average number of vehicles registered in the last five years [3]. Meanwhile, according to Traffic Accident Analysis System (TAAS), there is a total of 12,133 accidents happened in Daegu, which is the third city that has the most accidents happened [4].

Previously, studies related to traffic crashes have been conducted to identify factors that caused road accidents [5,6,7] and factors that led to accident severity [8,9,10,11]. Studies that conducted to investigate traffic accidents happened in Daegu mostly related to pedestrian

* This work was supported by Keimyung University.

Manuscript : 2023.10.06
Revised : 2023.11.08
Confirmation of Publication: 2023.11.22
Corresponding Author: Yang Sok Kim, e-mail: yangsok.kim@kmu.ac.kr

accidents [12,13], risk factors [14]. However, these analyses mostly did not look at the connections between high-risk location, day of week, and time period with accidents, causing the related authorities not effectively to apply practical measures for the prevention of accidents.

This study is proposed to address the concern above by providing insights into the relationships between location, day of week, and time period with accidents in Daegu. These will be done by identifying the high-risk locations, days of the week, and time periods for accidents in Daegu and estimating the conditional probabilities of accidents occurring based on combinations of location, day of the week, and time period. A dashboard is used to visualize the high-risk locations and information related to accidents in Daegu. The identification of rush hours during weekdays and visualization of vulnerable location in dashboard are expected to contribute to guiding the traffic management strategies, accident prevention efforts, and resource allocation in high-risk areas and time periods.

II. RELATED WORK

Preceding studies have extensively explored various facets of road accidents, with a focus on identifying causative factors [5-7], assessing accident severity [8-11], and proposing policies for accident reduction [15]. These investigations have encompassed a range of approaches and considerations.

Studies examining accident severity typically categorize accidents as either fatal or non-fatal [8], those involving fatality and injury versus property damage only [10], or establish a hierarchy of severity levels based on factors like property damage and injuries or fatalities [11]. Researchers have employed statistical techniques such as logistic regression [8,9,11] and artificial neural networks [10] to analyze factors effect on accident severity. Key factors consistently found to significantly influence accident severity include location [8,9], the type or cause of the accident [8,10,11], and environmental factors [9,11].

In contrast, research aimed at identifying the factors contributing to the occurrence of road accidents has taken diverse approaches. Study of Hosseinian and Gilani highlighted the significance of environmental factors (such as season, day of the week, time of day, and weather), driver characteristics (age and gender), and road conditions [5]. Meanwhile, Bucsuházy et al. delved into human factors and determined that age, gender, road familiarity, risk behavior tendencies, and driving habits play roles in accident occurrence [6]. Additionally, a study that gathered input from police officers and the public suggested the possibility of underreported causes, emphasizing driver distraction, drug and alcohol impairment, and uncorrected or defective eyesight as under-recognized issues in need of greater law enforcement attention [7].

Furthermore, accidents often occur due to drivers' behavior influenced by environmental factors [5,16]. For instance, accidents resulting in injuries tend to happen more frequently on weekdays and during rush hours in both the morning and

evening [16]. Another study applied data mining techniques to identify and compare characteristics of high-frequency and low-frequency accident locations, offering insights into road accident patterns [17]. Study of Goniewicz et al. addressed the root causes of underscored the importance of enforcing traffic regulations, improving infrastructure, and ensuring vehicle roadworthiness as ongoing issues contributing to road accidents, despite the multifaceted nature of these incidents [15]. This study also advocated for strategies and programs aimed at reducing the risk of exposure to accidents, preventing accidents, and enhancing post-accident medical care to improve overall road safety.

While these studies have provided valuable insights, it's essential to note that they often pertain to specific locations, and some factors may be beyond the control of transportation authorities or policymakers. Consequently, this study proposes to investigate the relationships between location, day of the week, and time periods in the context of road accidents in Daegu, with the aim of offering more specific insights to inform road safety initiatives for accident reduction.

III. DATASET AND PREPROCESSING

This study used the accident data collected from 대구교통종합정보, which recorded information on accidents that occurred in Daegu City [18]. The provided information includes details such as the start date and time of the accident, end date and time of the accident, location of the accident, and condition of the accident. To collect this data, Selenium, a web browser automation tool that enables users to automate website interactions through programming [19], was employed. It's worth noting that, in accordance with Article 24-2 of the Copyright Act (regarding the free use of public works), Daegu Metropolitan City either holds full copyright or has obtained the consent of the right holder for the works. As a result, these works are labeled with the "Korean Open Government License (KOGL)," which allows for free use without the need for separate permission [18].

The data collected for this study, encompassing 5012 accident entries from January 1, 2020, to November 21, 2022, falls under the provisions of the KOGL, which means that it can be freely used without the necessity of additional permissions.

Before conducting this study, Excel was used to extract the Day of the Week, Time Period, and Road Name from the data collected. The road names with wrong spelling are corrected. The table below shows the road names with modifications.

Table 1. Road names modifications

Initial Road Name	Modified Road Name
돈부도	동부로
유니버시아도로	유니버시아드로
파겨로	파계로
호곡로	호국로

After obtaining all the road names from the raw data, geopy is applied to obtain the coordinates of the roads. Geopy is a Python client for several popular geocoding web services that enable people to locate the coordinates of addresses, cities, countries, and landmarks using third-party geocoders and other data

sources [20]. The coordinates are needed in the visualization of the high-risk location in the dashboard. After that, all the accident entries that happened on the highway and with missing coordinates are removed from this study. Hence, there are a total of 3,057 accident entries used in this study.

IV. PROPOSED METHOD

Probability denotes the likelihood of the occurrence of an event [21] and conditional probability assesses the probability under a specified condition [21]. Conditional probability disclosed the uncertainties about one event on the assumption that the specific condition occurs [22] with two important assumptions, which are (i) the attributes are independent of each other and do not affect each other performance, and (ii) all the features are given equal importance [23]. Conditional probability has been widely used in various researcher, such as proving the of non-randomness in occurrence of distant metastasis for lung cancer [24], assessing the potential of Tsunami [25], and prediction of corporate failure [26].

In this study, the conditional probability of an accident happening on a particular day and time according to the location is estimated using the data collected. The estimation is calculated by the multiplication of the probability of X day given the accident happened at location A, the probability of Y time given the accident happened at location A, and the probability of the accident happened at location A, divided by the product of the probability of accident happened on X day, and probability of accident happened on Y time. Hence, the probability of an accident happening at location A, given that the day of the week is X and the time period is Y is estimated for this study.

$$P(Location\ A\ accident\ |X\ day, Y\ time) = \frac{P(X\ day\ |Location\ A\ accident) * P(Y\ time\ |Location\ A\ accident) * P(Location\ A\ accident)}{P(Accident\ on\ X\ day) * P(Accident\ on\ Y\ time)} \quad (1)$$

Meanwhile, data normalization is a process of transforming data into a consistent and comparable format, which can improve data quality, analysis, and integration [27]. Min-Max Normalization is a normalization method that produces a comparable value by performing linear transformations of the original data [28]. Min-Max Normalization helps to preserve the relationships among the original data values, and the bounded range helps to reduce the standard deviations, which can suppress the effect of outliers [27].

$$Min - Max\ Normalization = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Since the result of the estimated probabilities is very small in value, this study normalized the probabilities using Min-Max Normalization. The probabilities are normalized by comparing data across all locations (global normalization) and according to locations (relative normalization). This aims to make a better comparison across the probabilities.

V. RESULT

The accident frequency according to the day of the week and time period is

assessed using the collected data. The figure below shows the frequency of accidents according to the day of the week. Based on our result, accidents that happened during weekdays are more frequent than on weekends in Daegu. Particularly, accidents that happened on Tuesday and Friday show a slightly high frequency compared to the other days.

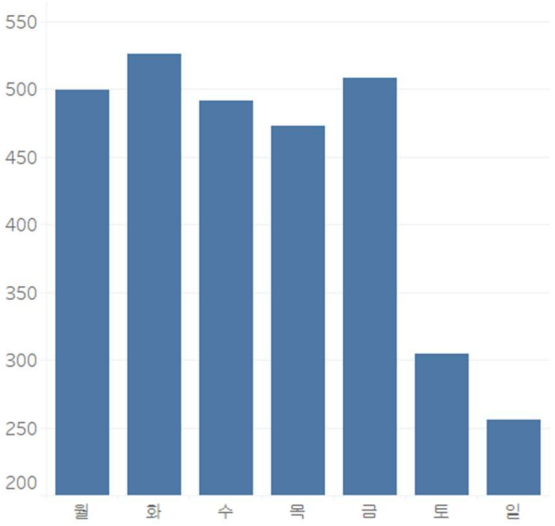

Fig.1. Accident frequency according to day of week

Meanwhile, Figure 2 shows the frequency of accidents according to the time period. According to Figure 2, the number of accidents surged between the time interval from 6 to 9 in the morning. To investigate the effect of workday commuting hours on the accident occurrence, Figure 3 shows the proportion of days of the week in the frequency of accidents that happened in each time period. We can confirm that the critical day and time period for accidents in Daegu is during the weekday morning rush hours when people go to work.

Fig.2. Accident frequency according to time period

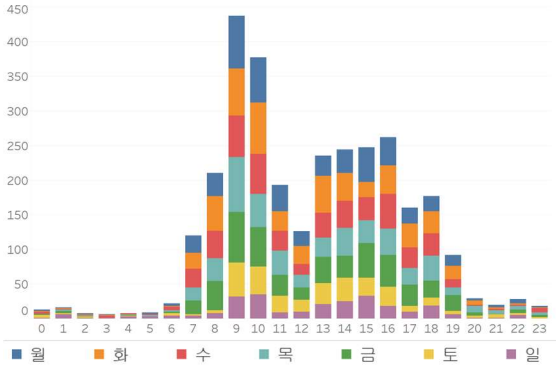

Fig.3. Accident frequency according to time period and day of week

Next, the conditional probability of an accident and the globally and relatively normalized probabilities are estimated. The probabilities were further plotted according to the location, as shown in Figure 4.

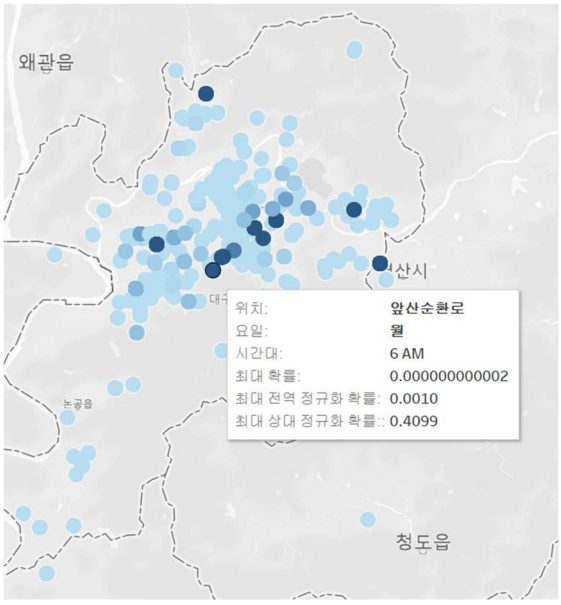

Fig.4. Visualization of high-risk locations

The figure below shows the dashboard with the complete information of accident data in Daegu. With this information, the relevant authorities can identify locations and times with a high probability of accidents. Thus, relevant authorities will be able to apply better prevention.

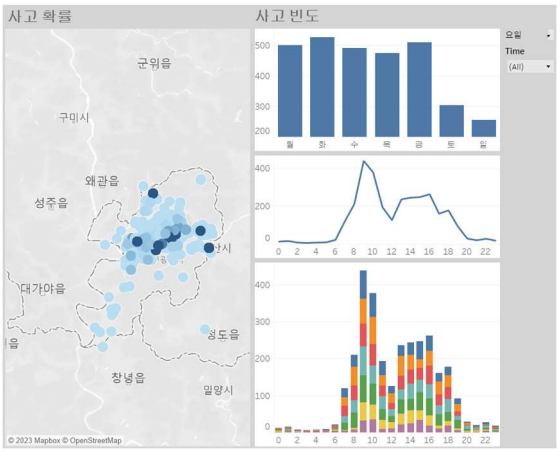

Fig.5. Dashboard presentation

VI. CONCLUSION

This study shed some light on the relationship of high-risk locations, days of week, and time periods for accidents in Daegu by using data collected from the Daegu Transportation Comprehensive Information System. According to the obtained result, accident happens the most during the weekday morning rush hours when people go to work. The visualized result holds substantial practical significance for urban planners, transportation authorities, or policymakers in Daegu. The identification of high-risk locations, critical days of the week, and vulnerable time periods for accidents enables the stakeholders to strategically allocate resources for traffic management, law enforcement, and targeted safety campaigns, hence reducing the occurrence of road accidents. However, this study had limited information about the accident site, future study is proposed to include information of road conditions and weather conditions. We believe that this information is significant for a more accurate traffic management.

REFERENCES

[1] Road traffic injuries (2022). https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (Accessed Mar., 20, 2023).

[2] Leading Causes of Death (2023). https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm (Accessed Mar., 20, 2023).

[3] 2022 Transportation Census Report (2023). https://car.daegu.go.kr/board/trafficData/14 (Accessed Aug., 12, 2023).

[4] Accidents According To Metropolitan Cities (2023). https://taas.koroad.or.kr/sta/acs/exs/typical.do?menuId=WEB_KMP_OVT_UAS_TAT (Accessed Aug., 12, 2023).

[5] S.M. Hosseinian, and V.N.M. Gilani, "Analysis of factors affecting urban road accidents in rasht metropolis," *Eng Transactions*, vol. 1, pp. 1-4, 2020.

[6] K. Bucsuházy, E. Matuchová, R. Zůvala, P. Moravcová, M. Kostíková, and R. Mikulec, "Human factors contributing to the road traffic accident occurrence," *Transportation Research Procedia*, vol. 45, pp. 555-561, Sep. 2020.

[7] J.J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accident Analysis & Prevention,* vol. 115, pp. 11-24, Jun. 2018.

[8] A.S. Al-Ghamdi, "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis &*

Prevention, vol. 34, no. 6, pp. 729-741, Nov. 2002.
[9] L. Eboli, C. Forciniti, and G. Mazzulla, "Factors influencing accident severity: an analysis by road accident type," *Transportation Research Procedia*, vol. 47, pp. 449-456, Jan. 2020.
[10] F. Rezaie Moghaddam, S. Afandizadeh, and M. Ziyadi, "Prediction of accident severity using artificial neural networks," *International Journal of Civil Engineering*, vol. 9, no. 1, pp. 41-48, Mar. 2011.
[11] V. Shankar, F. Mannering, and W. Barfield, "Statistical analysis of accident severity on rural freeways," *Accident Analysis & Prevention*, vol. 28, no. 3, pp. 391-401, May 1996.
[12] Y. Hwang, S. Park, H. Choi, and S. Yoon, "Spatial clustering of pedestrian traffic accidents in Daegu," *Journal of Digital Convergence*, vol. 20, no. 3, pp. 75-83, Mar. 2022.
[13] S.H. Park, and M.K. Bae, "Exploring the determinants of the severity of pedestrian injuries by pedestrian age: a case study of Daegu Metropolitan City, South Korea," *International journal of environmental research and public health*, vol. 17, no. 7, pp. 2358, Mar. 2020.
[14] Y.Y. Kim, K.H. Cho, and Y. Kim, "Analysis of risk factors for traffic accidents in Daegu area," *The Korean Data & Information Science Society*, vol. 31, no. 3, pp. 503-510, May 2020.
[15] K. Goniewicz, M. Goniewicz, W. Pawłowski, and P. Fiedor, "Road accident rates: strategies and programmes for improving road traffic safety," *European Journal of Trauma and Emergency Surgery*, no. 42, pp .433-438, Aug. 2016.
[16] C. Cabrera-Arnau, R. Prieto Curiel, and S.R. Bishop, "Uncovering the behaviour of road accidents in urban areas," *Royal Society Open Science*, vol. 7, no. 4, pp. 191739, Apr. 2020.
[17] S. Kumar, and D. Toshniwal, "A data mining approach to characterize road accident locations," *Journal of Modern Transportation*, vol. 24, pp. 62-72, Feb. 2016.
[18] Accident Information (2022). https://car.daegu.go.kr/metro/outbreak (Accessed Nov., 21, 2022).
[19] B. Zhao, *Encyclopedia of big data*, Springer Cham, pp. 1-3, 2017.
[20] Geopy (2023) https://github.com/geopy/geopy (Accessed Aug., 31,2023).
[21] R. Warner, *Optimizing the Display and Interpretation of Data*, Elsevier, pp. 117-134, 2016.
[22] J.M. Joyce, "The development of subjective Bayesianism," *In Handbook of the History of Logic*, vol. 10, pp. 415-475, Jan. 2011.
[23] Naïve Bayes Classifier (2018). https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c (Accessed: Apr., 2, 2023).
[24] A. Oikawa, H. Takahashi, H. Ishikawa, K. Kurishima, K. Kagohashi, and H. Satoh, "Application of conditional probability analysis to distant metastases from lung cancer," *Oncology Letters*, vol. 3 no. 3, pp. 629-634, Mar. 2012.
[25] K. Orfanogiannaki, and G.A. Papadopoulos, "Conditional probability approach of the assessment of tsunami potential: application in three tsunamigenic regions of the Pacific Ocean," *Pure and Applied Geophysics*, 164, pp. 593-603, Mar. 2007.
[26] L. Lin, and J. Piesse, "Identification of corporate distress in UK industrials: a conditional probability analysis approach," *Applied Financial Economics*, vol. 14, no. 2, pp. 73-82, Jan. 2004.
[27] What are the benefits and drawbacks of different normalization techniques? (2023).

https://www.linkedin.com/advice/0/what-benefits-drawbacks-different-normalization#:~:text=Min%2Dmax%20normalization%20is%20a,magnifies%20the%20effect%20of%20outliers. (Accessed Aug., 2023).

[28] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *International Journal of Informatics and Information Systems,* vol. 4, no. 1, pp. 13-20, Mar. 2021.

Authors

Mee Qi Siow

She received her B.S. degree at the Department of Economics, University of Malaya, Malaysia, in year 2020. She is currently pursuing her Master studies at the Department of Management Information Systems in Keimyung University, South Korea. Her research interests include Big Data Analytics, Text Mining and Data Visualization.

Yang Sok Kim

He has been serving as an Associate Professor at the department of Management Information Systems, Keimyung University, South Korea. He received his Ph.D. from University of Tasmania (UTAS), Australia. His research interests are Machine Learning, Web Search/Mining, Social Network, and Recommenders Systems. He has published papers in Electronic Commerce Research and Applications, Expert Systems with Applications, Mobile Information Systems, International Journal of Human-Computer Studies, Sustainability, and other reputed journals.

Mi Jin Noh

She received her M.S. and Ph.D. degree in Management Information Systems from Kyungpook National University, Korea in 2001 and 2006, respectively. Since 2022, she has been an assistant professor in Department of Business Big Data, Keimyung University, Korea. Her research interests are Big Data Analysis, Smart City, Text Mining and Mobile Services.

Choong Kwon Lee

He is professor in MIS at Keimyung University. His research interests are Big Data, IT Innovation, and Skills and Knowledge for IT Workforce.

Sang Ill Moon

He is studying at the Graduate School of Business Administration in Yeungnam University, South Korea. He is also the manager of the Strategic Planning Department of Moon Chang Co., Ltd. His research interests include Water and Sewage Engineering, Integrated Water Management Systems, and Water Storage Devices.

Jae Ho Shin

He received her B.S. degree at the Department of Management Information Systems in Keimyung University, South Korea, in 2015. He has been serving as the head of the IT Strategy Division of Moon Chang Co., Ltd. His research interests are AIoT, Smart City Infrastructure and Integrated Platform, Water Management Integrated System, and Video Analysis.