Application of Data Science Technologies to Take Proactive Decisions to Control Road Crashes

Ishan De Silva[1] and Loshaka Perera[2]

Abstract

Traditionally, road traffic crash analysis and accident modeling resorted to regression models and discrete choice models. Many countermeasures have been identified and implemented but still, the number of crashes and severities are increasing every year. Since road traffic crashes occur across space and time, conventional approaches have failed to provide alerts and insights in relation to geospatial regions, enabling proactive prevention measures. Aggregation of other data sources such as real-time weather, traffic flow counts and congestion levels etc. to alert authorities on increased crash risks is another gap that needs attention. The lack of geospatial analysis or visualization on available crash data (e.g., crash hotspots identification) limits road agencies' abilities in prioritizing funds allocation to more impactful improvements. The enforcement authorities also find it difficult to deploy their staff strength to high-risk areas. The latest advancements in programmatic geospatial analysis, interactive map visualizations and open-source software offer a unique opportunity to fill these gaps in a cost-effective way. This paper presents an application of data science and data visualization technologies to analyze road crashes. Popular packages written in Python programming language were used for the analysis. GeoPandas library provided the ability to process GPS locations (latitude and longitude) while Matplotlib was used to generate static maps. Folium library and the underlying Leaflet.js library were applied to generate interactive maps to help visualize crash hot spots. The study developed algorithms to combine GPS location data from crash records with boundary and attributes data from geospatial files to generate road crash density maps by administrative division areas and population. Interactive maps that allow authorities to drill down (or zoom in) to hot spots were also developed. Unlike GUI-driven analysis tools such as ArcGIS or QGis, the programmatic approach developed in this study enables the repeatable application of the analysis and visualization to new and old datasets with minimal effort. The application of existing geospatial analysis tools to road crash data is the key contribution of this study. The findings from the study lay the foundation for a digital system that can become an online platform for road and enforcement agencies to obtain reports and alerts on road crash risks and hot spots. The application was tested using crash data in Sri Lanka and outcomes are presented in this study. Future work such as real-time prediction of crash risk using machine learning technologies and fusion of multiple data sources onto the same platform can bridge the current gaps in crash prevention measures.

Keywords: *road crashes, geospatial analysis, interactive maps, crash prediction*

Author Details;

1. Postgraduate Student, University of Moratuwa, ishan.desilva@gmail.com
2. Senior Lecturer, Department of Civil Engineering, University of Moratuwa. loshakap@uom.lk