


Understanding Fatal Crash Reporting Patterns in Bangladeshi Online Media using Text Mining

Subasish Das¹

Transportation Research Record
1–12
© National Academy of Sciences:
Transportation Research Board 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03611981211014200
journals.sagepub.com/home/trr


Abstract

Traffic crashes are a major public health concern. In 2016, traffic crashes resulted in over 1.35 million deaths worldwide. In Bangladesh alone, the number of reported traffic fatalities was 2,376 in 2016. However, the World Health Organization estimated that the true number of traffic fatalities in Bangladesh ranges between 20,730 and 29,177. Editorial traffic crash reports in Bangladesh, and the number of crashes that are reported, vary widely among different media outlets. This study employed a Google News Alert to collect fatal crash reports from online English daily newspapers. The current study compiled a database of 419 fatal crash-related reports over a six-month period (November 2018–April 2019). The reports contain a total of 81,019 words. The results of this study reveal that online news coverage of traffic fatalities tends to vary from news agency to news agency. Furthermore, these reports do not usually cover key contributing factors of crash occurrences; the geometric features of crash occurrence sites are rarely reported. The findings demonstrate the importance of deciphering media coverage to develop potential safety risk measures in Bangladesh. The current findings provide strong support for the need for guidelines to help media outlets adequately document fatal crash reports.

The staggering number of fatalities and injuries resulting from traffic crashes remains a key issue worldwide. In Bangladesh, the information included in editorial reports about traffic crashes varies widely. *New Age (1)*, a local English newspaper, reports that “7,221 [were] killed and 15,466 [were] injured in 2018 road crashes.” Another news outlet, the *Daily Star (2)*, disclosed the same values. Both of these values were cited from the Passenger Welfare Association of Bangladesh. However, *The Independent (3)*, another English-language newspaper, reported that the number of deaths related to traffic crashes in 2016 was only 4,439, and the source given was Nirapad Sharak Chai (Nischa). Because of the difference in the estimates from the World Health Organization (WHO) and editorial-reported crashes, an investigation into the real value is necessary.

In Asian countries, journalists and mass media outlets can play a powerful role in raising road safety awareness. By publicizing preventive messages, news outlets can advocate safer roads and systems, promote safer behavior, and increase people’s understanding and knowledge of the magnitude of roadway fatalities. However, journalists typically only write about safety and traffic-related issues when a fatal crash is reported. Reporters have little time to explore different angles for their stories

or to comprehend the cause of the crash because of the urgency to release articles before other news outlets. The lack of time not only results in unknown information about the consequences for the families, communities, and society at large, but traffic-related news remains limited to the what-who-where-when-why to inform readers about the real scenarios and contributing factors. By delivering a thorough analysis of the associations and consequences of road crashes, journalists can effectively convert traffic crash stories into roadway safety stories.

Media studies offer significant insight into determining road safety measures in Asian countries. In addition to the determination of decision criteria in improving safety, media coverage also shapes political agendas and policies on the outcomes of road safety. This study aims to answer two research questions: 1) RQ1: What are the yearly death tolls from traffic crashes in Bangladesh? 2) RQ2: What are the insights from reports of fatal crashes in online media? To answer the first question, this study explored established newspaper reporting on yearly death

¹Texas A&M Transportation Institute, Bryan, TX

Corresponding Author:

Subasish Das, s-das@tti.tamu.edu

tolls from traffic questions. To answer the second question, this study investigated Bangladeshi news reports on fatal and severe crashes gathered from the Google News Alert setting system. Six months (November 2018–April 2019) of Google News Alert data were collected using a wide range of keywords (some of the keywords are Bangladesh, Bangladeshi, crash, crashes, collision, collisions, accident, accidents, motor vehicle, road death, roadway fatal, road fatal, highway fatal). This study conducted several natural language processing (NLP) tasks to gain insights from online fatal crash reporting.

Literature Review

When forming risk perceptions about public health issues such as road safety, media coverage on fatal road crashes can play a crucial role (4–10). Providing detailed information about any road crashes and portraying the benefits of preventive measures can help others to evaluate their own injury risk, which leads to effective policy-making for the advancement of road safety (4, 6, 7). Both print and online news portals are now easily accessible and more affordable with the help of technological improvements. Detailed media coverage on road crash risks, their root causes, and reported consequences can bring significant social change (4–7) and be an exceptional source of injury surveillance systems for any country (6, 9, 11–13).

In many developing Asian countries, fatal road crashes are often under-reported because of the lack of a valid vital registry system of fatal and non-fatal injuries (9, 13–16). Because road fatalities are under-reported, the true extent of the road safety problem and its devastating effects are not understood accurately as the road toll is rapidly growing every year (12, 13, 16). These developing countries rely heavily on hospital records, police reports, death certificates, post-mortem reports, and newspapers for crash estimates. (12–16). Newspaper reports on fatal road crashes were identified as the least sustainable in acceptability, timeliness, simplicity, and validity (12). When the degree of the injury risk in traffic is imperative for public attention and preventive measure implementation, discrepancies in crash estimates from print media create a biased perception (6).

The comprehension of the nature of media content is crucial to educate others on fatal road crash prevention and implement safety policies (4–7). However, print media in developing countries lose the opportunity for public education by failing to include injury-mitigating measures. For example, 201 road crash articles from Singaporean newspapers were analyzed and revealed that only 8% of the articles mentioned road safety messages, but 74.1% of the articles assigned blame to a particular road user (7). This method of reporting influences the

reader's risk perception, characterizing road fatalities as unforeseen and inevitable events (i.e., biased or wrong perception).

By describing the use of preventive measures against road fatalities, the news media have the opportunity to educate the masses. For example, journalists can emphasize safe practices (e.g., use of seat belts, airbags, child seats) and condemn unsafe practices (e.g., use of cell phones, alcohol consumption, speeding while driving) in their news reports (4, 7, 8). However, less than 10% of articles mention protective measures in their articles (7, 8). Reporters are likely to sensationalize and over-emphasize traffic crash stories, rarely mentioning the possible causes of the crash (6, 7). News reports often include information on the vehicle type involved in the crash but rarely report on the characteristics of the people injured and vehicles damaged (8, 10). An analysis conducted on Ghanaian newspapers reveals that only 22% of articles cover pedestrian injuries, which account for 50% of the traffic fatalities in Ghana (8). In many developed countries such as the United States, road fatalities are reported in newspapers as isolated events, concealing the problem with the nature of public health (4–6).

In developing countries, research has been conducted to determine the statistical analysis of newspaper articles on fatal crashes (7–10, 11). Several studies (12–16) used data mining and other analytical approaches to analyze road crash data in Bangladesh. However, the purpose of this paper is to determine the reporting pattern of road crashes from the unstructured textual data in Bangladeshi newspapers using NLP to interpret its influence on public risk perception and policymaking. It is essential to obtain useful knowledge hidden in the media reports on traffic crashes. Mining media coverage reports about traffic crashes in Bangladesh can provide a vital early warning of potential challenges.

Methodology

Media Reports on Yearly Traffic Fatality Counts

This section encompasses a comprehensive review of the media coverage of roadway crashes in Bangladesh in the format of “featured news story” or “editorial.” Some of the popular local news media are *New Age*, the *Daily Star*, *The Independent*, and *Dhaka Tribune*.

One particular concern related to roadway crashes is the number of children that die each year. The *Daily Star* (17) published an article on November 28, 2018 entitled, “Children’s Death on Roads: A danger not even acknowledged.” The article reported that, although roadway crashes are the second most common cause of death for children in Bangladesh, there are no official policies to help prevent crashes or educate children about

traffic safety. A Bangladesh child rights group, the Shishu Adhikar Forum, reported that at least 79 children were injured and 549 killed in the first 10 months of 2018. These statistics reflect a sharp increase from 2017, when 357 children were killed during the year. Young children in urban areas and towns are particularly vulnerable to being harmed in intersection crashes; the Accident Research Institute (ARI) combated this vulnerability by creating a booklet for children about roadway safety that will be distributed in classrooms across Bangladesh.

Another article published by the *Daily Star* (2), entitled “Road Crashes: 7,221 killed last year,” was published on January 25, 2019. This article listed the main causes of roadway crashes in Bangladesh, citing faulty roads and unfit vehicles, as well as road users that are unskilled, reckless, and generally unaware of traffic safety. The title of that article conflicts with the information given in another article by the *Daily Star* (18) published on January 30, 2019, entitled, “Too unsafe to commute.” This article stated that 4,439 people were killed and 7,425 were injured in a total of 3,103 road crash incidents in 2017, and it cited Nirapad Sarak Chai as its source for these statistics. Nirapad Sarak Chai compiled its data by monitoring online news portals, TV channels, and six national daily news agencies. However, the organization said that many crashes go unreported and, thus, were not included in their report. It published 10 recommendations to educate the public and to help prevent crashes including: massive awareness programs in the media; the inclusion of traffic crash issues in textbooks; stricter traffic law enforcement; and better training for drivers. The organization, Bangladesh Jatri Kalyan Samity published the Annual Road Accident Report, stating that at least 16,193 people were injured and 7,397 killed in a total of 4,979 roadway crashes in 2017. Official police reports contained statistics that conflicted with both the Nirapad Sarak Chai and the Jatri Kalyan Samity, stating that only 2,265 people were killed and 1,659 injured in a total of 2,237 crashes during the first 10 months of 2017.

According to the Motor Vehicle Ordinance of 1983, vehicles must undergo a mandatory fitness check every year. Unfit vehicles are a major cause of traffic crashes; vehicles that are not certified can pose a serious risk to the general public. As of 2018, approximately 3.8 million vehicles are registered with the Bangladesh Road Transport Authority (BRTA); of these vehicles, 2.4 million are motorcycles and, therefore, do not require annual fitness certificates.

An article by the *Daily Star* (19) was published on March 23, 2019, with the following headline: “No fitness docs, yet running.” It revealed that there are 71,218 vehicles on the road in Bangladesh that do not been properly

certified in 10 years (since January 2009). The ARI declared that this alarming number of unfit vehicles indicates the dysfunction of the law enforcement system in regard to traffic safety.

New Age (20) published an article on November 30, 2018 entitled, “Underage transport workers a serious threat to road safety.” The article exposed the common practice of underage boys working as drivers or assistants for public transportation in major cities, breaking Bangladesh’s transport and labor laws. The Motor Vehicles Ordinance of 1983 forbids any person under 20 years of age to work as a professional driver and operate any motor vehicle in any public place. Furthermore, young men are typically immature and inexperienced, which can lead to the violation of road safety laws and consequently endanger people’s lives. Additionally, the child laborers working as drivers and assistants are also placed in danger. The Bangladeshi government issued an order on March 13, 2013, under the Bangladesh Labor Law of 2006, that identified 38 activities that it judged to be hazardous for children and outlawed them. These activities included working at automobile workshops and working as helpers for trucks, tempos (four-wheel vehicles), or busses. Despite these laws, 80% of human hauler drivers in Dhaka are between the ages of 13 and 18, according to the Passenger Welfare Association of Bangladesh.

New Age (21) published another article on January 2, 2019, with the headline, “Road safety in Bangladesh remains as elusive as ever.” The article declared that roadway deaths were as pervasive as ever, primarily because of reckless driving by untrained, unlicensed, and underage drivers. Drivers throughout Bangladesh continue to partake in reckless driving, overtaking, and the use of unfit vehicles. The inadequacy of public transportation and pedestrian facilities has further exacerbated the issue. Despite increasing concern for road safety, the BRTA relaxed the licensing rules for drivers in August 2018 to promote the “public interest.” Currently, there are 3.5 million motorized vehicles registered through the BRTA; a total of 1.9 million driving licenses were issued in the first seven months of 2018. On September 20, 2018, the government passed the Road Transport Act to replace the Motor Vehicles Ordinance of 1983 in response to student protests.

On January 16, 2019, *Dhaka Tribune* (22) published as article with the headline, “2018 saw a surge in road accidents.” Despite massive nationwide road safety movements in 2018, the report indicated that the country had continued to witness many traffic crashes, with total fatalities reaching 4,580. Bangladesh also experienced a significant increase of 19.58% in traffic crashes in 2018 compared with 2017. There were 3,472 fatalities in 2017 in comparison with the 4,317 fatalities in 2018. In

Table 1. Number of Fatalities in 2018 by the Media Sources

Reporting news agency	2018 fatalities	Source
<i>New Age</i> (01/02/2019)	1,853	Bangladesh Road Transport Authority (BRTA)
<i>Dhaka Tribune</i> (01/16/2019)	4,580	National Committee to Protect Shipping, Roads and Railways (NCPSRR)
<i>Daily Star</i> (01/25/2019)	7,221	Bangladesh Jatri Kalyan Samity
<i>Daily Star</i> (01/30/2019)	4,439	Nirapad Sarak Chai

Table 2. Comparison between Local and International News on a Very Serious Fatal Crash Event in Bangladesh

Attributes	Example from local news	Example from international news
Context	<ul style="list-style-type: none"> - Friday - Chouddagram upazila in Cumilla - 13 brick kiln workers were killed in their sleep in a labor shed - Additional injury: two - Coal-laden 10-ton truck overturned - Driver driving illegally with medium driving license - The truck driver and his helper escaped from the scene 	<ul style="list-style-type: none"> - Friday - Eastern Bangladesh (52 mi southeast of Dhaka, the capital) - 13 dead, 2 injured - Crashed into workers' huts at a brickfield - A truck loaded with coal overturned
When and how	<ul style="list-style-type: none"> - Around 6:30 a.m. - Kazi and Co Brick Kiln at Narayanpur village under Gholpasha union in Chouddagram - Truck overturned as the driver lost control 	<ul style="list-style-type: none"> - Chauddogram in Cumilla district - Driver lost control of the truck
Fault	<ul style="list-style-type: none"> - Unload the truck at a wrong point - Illegal permit 	<ul style="list-style-type: none"> - Lax enforcement of traffic laws, reckless driving and bad roads (not the specific crash, generalized view on traffic crashes in Bangladesh)
Compensation	<ul style="list-style-type: none"> - Compensation provided (Tk 20,000) to the families of each victim - Owner provided Tk 10,000 each 	<ul style="list-style-type: none"> - Not mentioned

their latest study, The National Committee to Protect Shipping, Roads, and Railways (NCPSRR), a non-government organization, published their figures. These show that there were 2,998 traffic crashes resulting in 3,412 fatalities—including 470 women and 453 children—and 8,572 injured individuals in 2016. NCPSRR identified six key contributing factors leading to traffic crash fatalities such as reckless driving, the use of overloaded vehicles, an increase in the number of motorbikes and three-wheelers on highways, gross violations of traffic rules in heavily populated areas, unskilled drivers, and distracted driving.

On January 29, 2019, *The Independent* (23) published a report with the headline, “No let-up in road crashes.” According to BRTA statistics, 2.6 million vehicles are registered in the country. However, the agency has only issued 1.7 million driving licenses, indicating that there are almost one million drivers without licenses. Five years ago, 1.4 million vehicles were registered, and 998,000 licenses issued. This report focused mainly on fitness certification issues and related safety concerns. Table 1 lists the number of traffic fatalities reported in 2018 by different news media.

Comparison between National and International News Stories

One of the core objectives of the paper is to improve crash reporting in conventional news outlets. One example is shown to explain how a serious fatal crash (several fatalities) is reported in Bangladeshi online news outlets by comparing the reports with international coverage. International newspapers do not usually report a regular fatal crash unless it is serious. This study analyzed the contexts described in a local newspaper (24) and an international reporting agency (25) to investigate their reporting trends. Table 2 compares how local and international news present the context, when and how, fault, and compensation for a fatal crash event. Overall, the local news examples were more detailed and lengthier than international news examples. The international example did not mention the brick kiln manager at all in addressing the fault of the crash, while the local news example focused heavily on his involvement. Furthermore, the context of the crash was also more thorough in providing specific details about the victims of the crash as well as the driver. Moreover, in Asian countries, compensation for each fatality is normally provided. According to the local

news sources, the monetary compensation was 30,000 Bangladeshi Taka (US\$365 per person). On the other hand, in the United States, the cost of a fatal crash is \$10,243,517, including loss of quality of life (26).

Online Reporting on Traffic Fatalities

Traffic crash data is dependent on police-reported crash statistics in Bangladesh. Hospitals and other emergency services do not manage separate databases for national statistics. By setting a Google News Alert, traffic crash reports were gathered from online daily English-language newspapers from November 2018 to April 2019. For the purposes of data collection, a wide range of keywords were used (some of the keywords are Bangladesh, Bangladeshi, crash, crashes, collision, collisions, accident, accidents, motor vehicle, road death, roadway fatal, road fatal, highway fatal). Note that using additional keywords (e.g., adding district names in place of Bangladesh) can produce additional crash reports. Future studies could explore this option. In a span of six months, this search strategy generated 419 fatal crash reports. Most of the articles were pulled from the following primary news agency (mostly local newspapers) in Bangladesh: UNB, *Dhaka Tribune*, *Daily Star*, *bdnews24*, *Prothom Alo*, *The Independent*, *Asian Age*, *New Age*, *Bangla Tribune*, *Bangladesh Post*. Every article was manually entered into the data set, which contained the following seven variables: name of the paper (i.e., 'Paper'); headlines of the articles (i.e., 'Headline'); date of the article published (i.e., 'Date'); article narratives (i.e., 'Narrative'); place of crash occurrence (i.e., 'District'); number of injured people (i.e., 'Injury'); and number of fatalities (i.e., 'Fatal').

In Bangladesh, crash reporting patterns vary over the different number of news agencies. For the same incidents, the number of casualties may vary from paper to paper, which is indicative of the reporters' inadequate crash report training. Well-known newspapers in Bangladesh (i.e., *Prothom Alo*, *The Independent*, *Daily Star*) only cover vehicle crashes with a high number of casualties and severe injuries. Newspaper articles provide information on the road user type (i.e., age, gender), the number of fatalities, the type of vehicle, and injury severity citing police reports of the incident. However, many news articles lack crucial information about the incidents, such as road geometry, road surface condition, road class, location type, collision type, weather conditions, light conditions, and others. These sets of information can help determine the key causes of road crashes. Figure 1 shows a visual representation of this information.

Table 3 lists the number of crash reports by the top 10 districts. The districts with the greatest number of reports are Dhaka (83), Chittagong (41), and Comilla (28). The

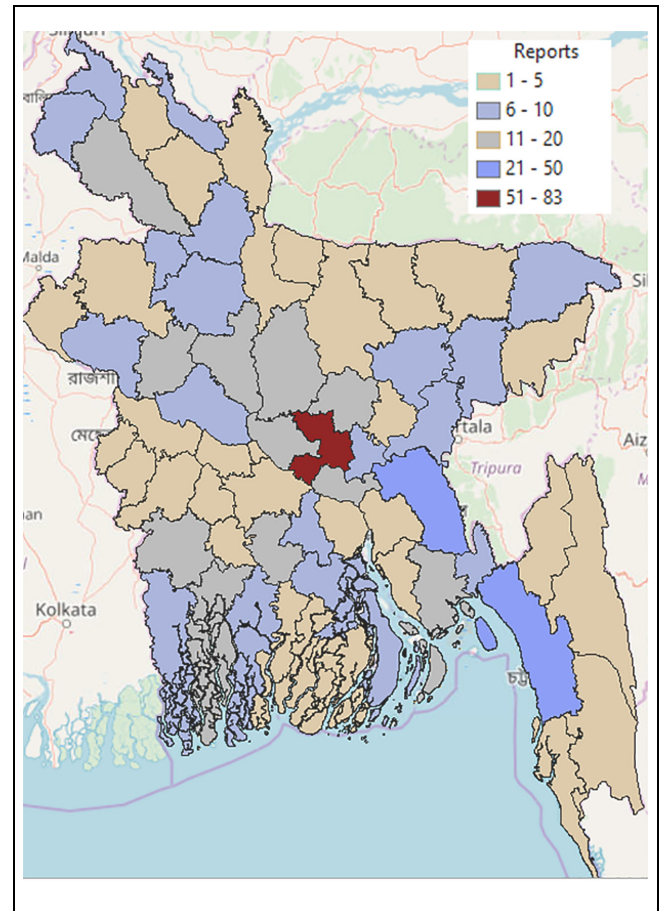


Figure 1. Number of crash reports by districts (November 2018–April 2019).

districts with the smallest number are Dinajpur (14) and Jessore (15). By area, Dhaka and Gopalganj are the top two districts with the highest number of reported fatal crashes. By population, Gopalganj and Khulna are the top two districts with the highest number of reported fatal crashes.

Figure 2 is a heat map that illustrates the number of crash reports by date for November 2018 to April 2019. White and lighter yellow squares represent dates that had fewer crash reports, while dark orange to red squares represent dates that had a greater number of reports. It does not show any significant pattern; dates that had the highest amount of crash reports appear to be random.

Table 4 lists the number of crash reports published by the top 10 online news agencies. UNB had the most reports (154 reports), followed by the *Dhaka Tribune* (71 reports) and the *Daily Star* (45 reports). The *Bangladesh Post* had the fewest crash reports (6 reports). Table 4 also lists the minimum, maximum, mean, and standard deviation for the number of words used in the reports for each news agency. *New Age* had the highest mean (352.62), followed by the *Daily Star* (338.31). UNB had

Table 3. Number of Fatal Crash Reports by Top 10 Districts

Districts	Crash report counts (November 2018–April 2019)	Area (km ²)	Population estimate (2016)
Dhaka	83	1,464	13,142,000
Chittagong	41	5,283	8,440,000
Comilla	28	3,085	6,046,000
Gazipur	20	1,800	3,809,000
Tangail	20	3,414	3,923,000
Gopalganj	18	1,490	1,277,000
Khulna	18	4,394	2,528,000
Sirajgonj	18	2,498	3,455,000
Jessore	15	2,567	3,029,000
Dinajpur	14	3,438	3,268,000

Source: <https://www.citypopulation.de/Bangladesh-Mun.html>.

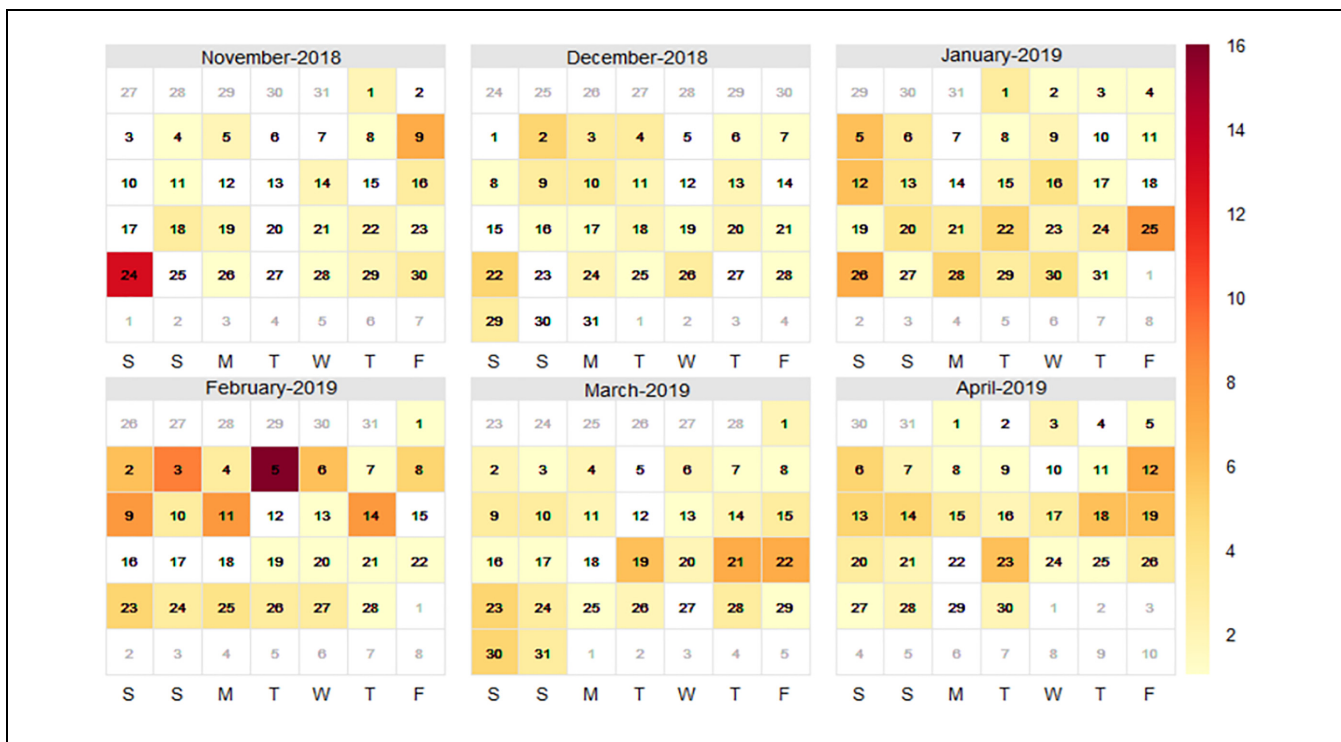


Figure 2. Number of crash reports by date.

the smallest mean (120.83); additionally, its minimum number of words used in a report was only 10 words.

Text Mining on Online News Reports

This study aims to investigate the underlying reasons why the discrepancies of fatality counts arise and how different text mining methods could help understand the key association factors and associated needs for better journalism. Text mining has been gaining more and more attention in several industrial fields (e.g., aerospace, automotive, railway, biomedicine, manufacturing,

customer services) as a result of its ability to identify insightful information from a complex set of unstructured textual contents. NLP and text mining in combination with different machine learning algorithms can be used as excellent tools to collect, analyze, and extract interesting trends, patterns, and knowledge from big data text corpora (plural of corpus; corpus indicates the collection of documents [for example, one report can be considered as a document] based on a theme or group) such as media report databases. This study developed a framework for recognizing various interesting patterns and anomalies in the data using text mining pipelines.

Table 4. Crash Reports by Top 10 News Media

Reporting news agency	Reports	Number of words used in a report			
		Minimum	Maximum	Mean	Standard deviation
UNB	154	10	520	120.83	78.52
Dhaka Tribune	71	69	1,015	294.86	175.46
Daily Star	45	60	1,126	338.31	252.01
bdnews24	32	35	411	132.09	73.82
Prothom Alo	23	61	865	158.52	178.24
The Independent	23	68	1,086	281.57	299.17
Asian Age	21	34	235	108.43	54.26
New Age	13	75	765	352.62	223.33
Bangla Tribune	11	93	265	156.64	64.4
Bangladesh Post	6	67	243	127	78.21

Term Frequency Inverse Document Frequency (TF-IDF): the parameter TF-IDF has been widely used to differentiate between documents by estimating how relevant their contents are to a set of terms in a search string. It combines two different weighting parameters: TF and IDF. In TF-IDF, terms are viewed as having different levels of importance; some terms are weighted more while others are weighted less (27). TF can be denoted as $tf(t, d)$, which indicates the number of occurrences of the term (t) in the document (d). The parameter IDF of a term (t) can be defined as:

$$idf(t, D) = \log \frac{|D|}{df(t, D)} \quad (1)$$

where $|D|$ is the total number of documents in a corpus D , and $df(t, D)$ is the number of documents that contain term t :

$$df(t, D) = |\{d \in D : t \in d\}| \quad (2)$$

The parameter IDF provides additional weight to terms that are found in few documents to reduce the weight of terms that frequently appear in all documents. The TF-IDF score for a term t is the product of the TF and the IDF scores, so for a string containing the set of terms q , the TF-IDF score of documents d in corpus D :

$$TF - IDF(q, D) = \sum_{t \in q} tf(t, d) \times idf(t, D) \quad (3)$$

Figure 3a illustrates the top 12 keywords generated from a month-based corpus. The keywords are sorted based on the TF-IDF values. The keywords associated with locations, such as Jhenaidah and Rajendrapur, indicate that the presence of these locations is unique compared with other months. The two top word pairs generated in the February corpus are 5,514 roads and 7,221 people. As indicated in the February reports, the

news agencies generate editorial reports on the overall crash incidents in the last year.

Rapid automatic keyword extraction (RAKE): a domain-independent, unsupervised, and language-independent method for obtaining keywords from individual documents is known as RAKE (28). By defining its text into a set of candidate keywords, RAKE begins the keyword extraction on a document. First, the document text is categorized into an array of words by the specified word delimiters. This array is then separated into sequences of contiguous words at stop word positions and phrase delimiters. Words placed within a sequence are assigned the same position in the text and are then considered a candidate keyword. Extracted keywords do not contain interior stop words because RAKE splits candidate keywords by stop words. Because of its ability to extract highly specific terminology, a strong interest in identifying keywords that contain interior stop words such as the axis of evil was expressed. RAKE searches for pairs of keywords that are attached one another in the same order and at least twice in the same document to find the keywords. A combination of those keywords and their interior stop words creates a new candidate keyword. The sum of its member keyword scores is the score for the new keyword.

Figure 3b illustrates the top 20 keywords generated from the complete corpus. The keywords are sorted based on the RAKE values. In Figure 3b, the terms are illustrated in a bar graph to identify the words that are most commonly used based on the RAKE measures. The term “deputy commissioner” is significantly larger than any other term listed in Figure 3b with a rake of over 4.5. The next highest term on the graph is “bangla nagar” with a rake of a little over 4.0. The terms with the smallest rake are “mawa highway,” “hospital police,” and “highway police.” It is important to note that the reports mostly document the information that is relevant to the police rather than to the crash itself. The words

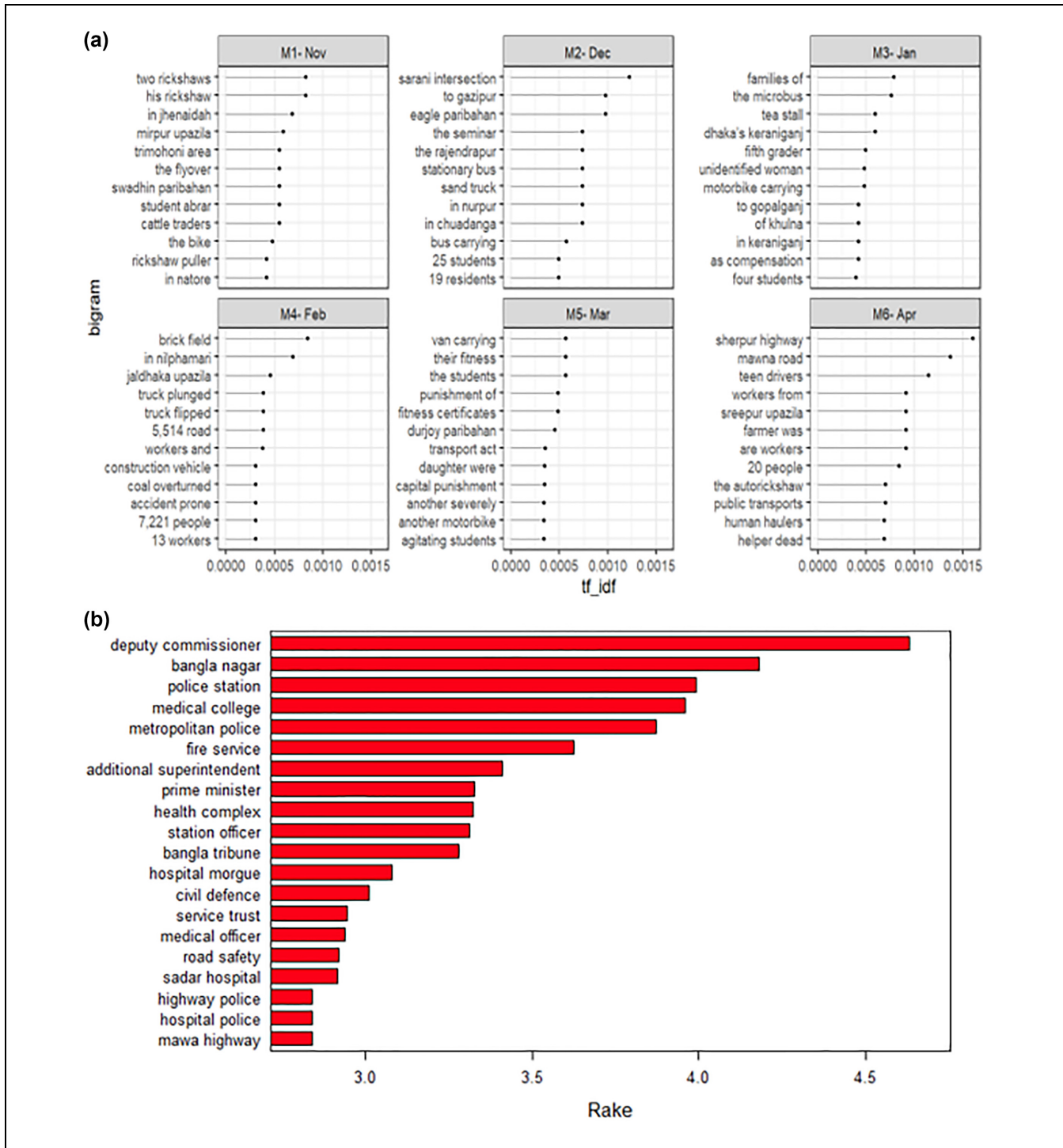


Figure 3. Performance measures: (a) TF-IDF (b) RAKE.
 Note: TF-IDF = term frequency inverse document frequency; RAKE = rapid automatic keyword extraction.

included in Figure 3 indicate that road geometry and environmental factors causing the crashes are rarely reported in online reports.

Word Cloud analysis: word cloud is one of the popular text mining techniques widely used to provide a broader

context of unstructured data. Word clouds generally represent the distribution of the keywords based on the frequencies in the collected textual content. The size of the words indicates the relative frequency of the word in the database. The larger the word size, the larger the

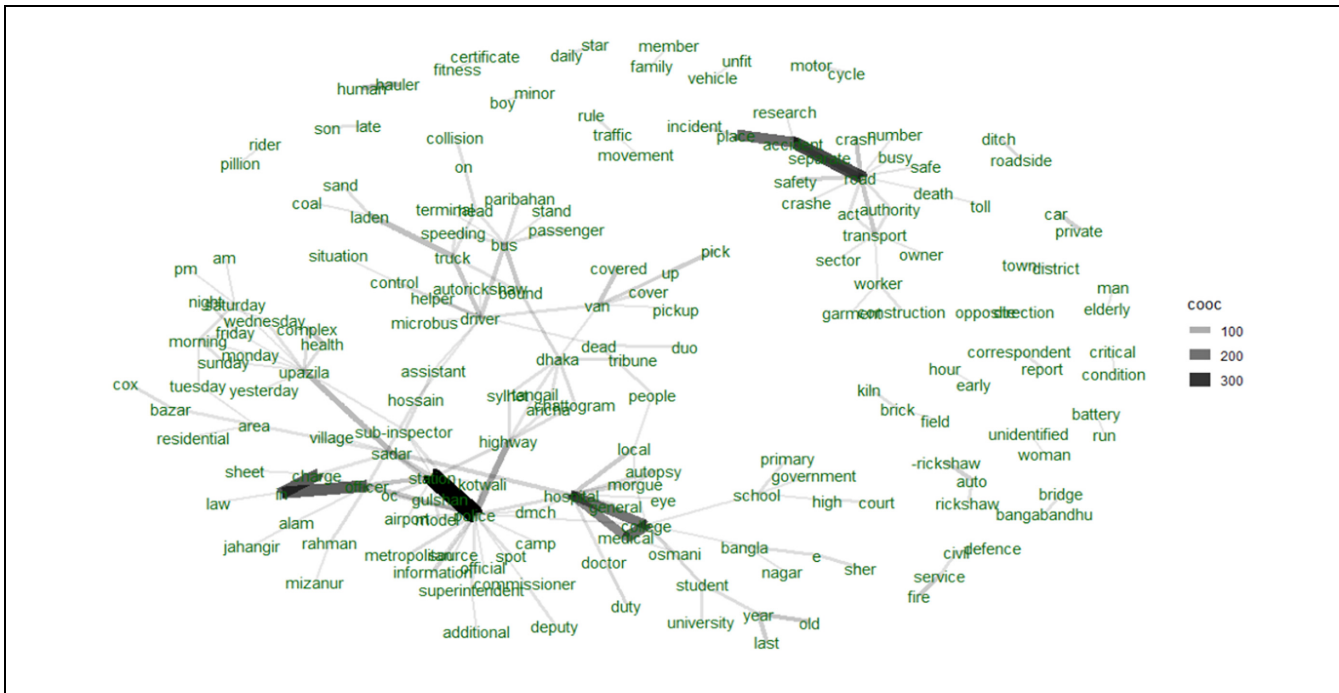


Figure 5. Co-occurrence of the words.

road safety. It is the eighth leading cause of death for all age groups surpassing HIV/AIDS, tuberculosis, and diarrheal diseases (30).” Based on reports from established print media, traffic fatalities in Bangladesh ranges from 1,853 to 7,221.

Police data is the single data source that provides relevant data on road accidents in Bangladesh, and it is processed and made available by the ARI. The recording process is activated when a police case is lodged on the accident. This may not always happen if the affected parties are reluctant to become embroiled in police procedures, are otherwise disinclined, or if the police themselves fail to record the case. Thus, there is an in-built element of under-reporting for those whose magnitude and variation are difficult to assess. The traffic crash statistics from different sources underscore that road deaths are not an inevitable concomitant of development, but they can be addressed and minimized through judicious and timely action.

In recent years, many studies have applied text mining to identify the hidden trends in crash reports (31–35). This study applied several NLP techniques to identify the reporting patterns of crashes from online news outlets by examining six months of accident reports in Bangladesh. The findings show that there is a systemic loss of key information in the crash reporting in Bangladesh. In many cases, the reports do not cover any specific information about roadway geometry,

environmental factors, and other key contributors. After performing text mining on 82,000 words generated in these reports, the research team identified several key attributes such as nighttime crashes, large-truck- or bus-related crashes, autorickshaw-related crashes, and teen drivers. Furthermore, this study does not address vehicle conditions. One of the key contributions of this project is the development of a framework that analyzes news outlet information and deciphers the key information through a text mining pipeline. This framework can be adopted by other users to extend the scope of the current study. The findings strongly support the idea that there is a need for guideline development to educate the journalists so that they can adequately document crash reports in the media.

The current study is a starting point for understanding online media coverage of crash fatality reporting in Bangladesh. The current study has some limitations. First, the sample size is limited to a six-month period. Future studies could collect data for additional years to perform a more robust study. Note that using additional keywords (e.g., adding a district name in place of Bangladesh) can produce additional crash reports to create a more comprehensive database as well. Second, only reports from English online dailies were considered in this study. The limitations of the current study can be minimized by expanding the sample size time frame and including Bengali newspaper reporting.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: Subasish Das; data collection: Subasish Das; analysis and interpretation of results: Subasish Das; draft manuscript preparation: Subasish Das. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. *New Age*. 7,211 Killed, 5,514 Injured in Road Crashes in 2018. *New Age*. <http://www.newagebd.net/article/62886/7221-killed-5514-injured-in-road-crashes-in-2018>. Accessed: July 29, 2019.
2. The *Daily Star*. Road Crashes: 7,211 Killed Last Year. *The Daily Star*. <https://www.thedailystar.net/city/news/7221-people-killed-road-accident-2018-bjks-report-1692772>. Accessed: July 29, 2019.
3. The *Daily Star*. 4,439 Killed in Road Crashes in 2018. *The Daily Star*. <http://www.theindependentbd.com/post/185412>. Accessed: July 29, 2019.
4. Rosales, M., and L. Stallones. Coverage of Motor Vehicle Crashes with Injuries in U.S. Newspapers, 1999–2002. *Journal of Safety Research*, Vol. 39, 2008, pp. 477–482.
5. Ralph, K., E. Iacobucci, C. G. Thigpen, and T. Goddard. Editorial Patterns in Bicyclist and Pedestrian Crash Reporting. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 663–671.
6. Daniels, S., T. Brijs, and D. Keunen. Official Reporting and Newspaper Coverage of Road Crashes: A Case Study. *Safety Science*, Vol. 48, 2010, pp. 1469–1476.
7. Heng, K. W., and A. Vasu. Newspaper Media Reporting of Motor Vehicle Crashes in Singapore: An Opportunity Lost for Injury Prevention. *European Journal of Emergency Medicine*, Vol. 17, 2010, pp. 173–176.
8. Yankson, I. K., E. N. L. Browne, H. Tagbor, P. Donkor, R. Quansah, G. E. Asare, C. N. Mock, and B. E. Ebel. Reporting on Road Traffic Injury: Content Analysis of Injuries and Prevention Opportunities in Ghanaian Newspapers. *Injury Prevention*, Vol. 16, 2010, pp. 194–197.
9. Ghaffar, A. Newspaper Reports as a Source for Injury Data in Developing Countries. *Health Policy and Planning*, Vol. 16, 2001, pp. 322–325.
10. Mogambi, H., and F. Nyakeri. Media Priming of Road Traffic Accidents in Kenya: Praxis, Patterns, and Issues. *SAGE Open*, Vol. 5, 2015, 2158244015606491.
11. Bhalla, K., S. Shahraz, D. Bartels, and J. Abraham. Methods for Developing Country Level Estimates of the Incidence of Deaths and Non-Fatal Injuries from Road Traffic Crashes. *International Journal of Injury Control & Safety Promotion*, Vol. 16, 2009, pp. 239–248.
12. Rahman, F. Potential of using Existing Injury Information for Injury Surveillance at the Local Level in Developing Countries Experiences from Bangladesh. *Public Health*, Vol. 114, 2000, pp. 133–136.
13. Kayani, A., J. J. Fleiter, and M. J. King. Underreporting of Road Crashes in Pakistan and the Role of Fate. *Traffic Injury Prevention*, Vol. 15, 2014, pp. 34–39.
14. Janstrup, K. H., S. Kaplan, T. Hels, J. Lauritsen, and C. G. Prato. Understanding Traffic Crash Under-Reporting: Linking Police and Medical Records to Individual and Crash Characteristics. *Traffic Injury Prevention*, Vol. 17, 2016, pp. 580–584.
15. Rainey, D. Y., and C. W. Runyan. Newspapers: A Source for Injury Surveillance? *American Journal of Public Health*, Vol. 82, 1992, pp. 745–746.
16. Ponnaluri, R. V., and Y. D. Santhi. Road Crash History and Risk Groups in India: Need for New Initiatives and Safety Policies. *Transportation Research Record: Journal of the Transportation Research Board*, 2009. 2114: 64–71.
17. The *Daily Star*. Children’s Death on Roads: A Danger Not Even Acknowledged. *The Daily Star*. <https://www.thedailystar.net/backpage/children-death-on-road-accident-danger-not-much-acknowledged-1666222>. Accessed July 29, 2019.
18. The *Daily Star*. Too Unsafe to Commute. *The Daily Star*. <https://www.thedailystar.net/country/news/4439-road-crash-deaths-2018-ili-as-kanchan-1694449>. Accessed July 29, 2019.
19. The *Daily Star*. No Fitness Docs Yet Running. *The Daily Star*. <https://www.thedailystar.net/frontpage/news/no-fitness-docs-yet-running-1719148>. Accessed July 29, 2019.
20. *New Age*. Underage Transport Workers a Serious Threat to Road Safety. *New Age*. <http://www.newagebd.net/article/57485/underage-transport-workers-a-serious-threat-to-road-safety>. Accessed July 29, 2019.
21. *New Age*. Road Safety in Bangladesh Remains as Elusive as Ever. *New Age*. <http://www.newagebd.net/article/60589/road-safety-in-bangladesh-remains-as-elusive-as-ever>. Accessed July 29, 2019.
22. Dhaka Tribune. 2018 Saw Surge in Road Accidents. *Dhaka Tribune*. <https://www.dhakatribune.com/bangladesh/nation/2019/01/16/2018-saw-surge-in-road-accidents>. Accessed July 29, 2019.
23. The Independent. No Let-Up in Road Crashes. *The Independent*. <http://m.theindependentbd.com/post/185268>. Accessed July 29, 2019.
24. The *Daily Star*. Truck Kills 13 Workers in Their Sleep. *The Daily Star*. <https://www.thedailystar.net/country/news/13-killed-comilla-road-crash-1692745>. Accessed July 29, 2019.
25. AP News. 13 Dead as Truck Crashes into Workers’ Huts in Bangladesh. *AP News*. <https://www.apnews.com/e954802091cb43aebf75ae956dd6a338>. Accessed July 29, 2019.
26. Schneider, H. *Louisiana Traffic Records Data Report 2017*. Louisiana State University, Baton Rouge, LA. http://datareports.lsu.edu/Reports/SpecialReports/2017_Factbook.pdf. Accessed July 29, 2019.
27. Silge, J., and D. Robinson. tidytext: Text Mining and Analysis using Tidy Data Principles in R. *Journal of Open Source Software*, Vol. 1, 2016, p. 37.

28. Cramer, N., S. Rose, D. Engel, and W. Cowley. *Automatic Keyword Extraction from Individual Documents*. John Wiley & Sons, Ltd., 2010.
29. Wijffels, J. *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*. R Package Version 0.8.2, 2019.
30. World Health Organization (WHO). Global Status Report on Road Safety 2018. https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/ Accessed July 2019.
31. Marucci-Wellman, H. R., M. R. Lehto, and H. L. Corns. A Practical Tool for Public Health Surveillance: Semi-Automated Coding of Short Injury Narratives from Large Administrative Databases using Naïve Bayes Algorithms. *Accident Analysis & Prevention*, Vol. 84, 2015, pp. 165–176.
32. Chatterjee, S. A Connectionist Approach for Classifying Accident Narratives. Theses and dissertations. Available from ProQuest, 1998, pp. 1–345.
33. Das, S., A. Mudgal, A. Dutta, and S. R. Geedipally. Vehicle Consumer Complaint Reports Involving Severe Incidents: Mining Large Contingency Tables. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672(32): 72–82.
34. Das, S., M. Le, and B. Dai. Application of Machine Learning Tools in Classifying Pedestrian Crash Types: A Case Study. *Transportation Safety and Environment*, Vol. 2, No. 2, 2020, pp. 106–110.
35. Das, S., A. Dutta, and I. Tsapakis. Motorcycle Crash Causation Study: Exploratory Topic Models from Crash Narrative Reports. *Transportation Research Record: Journal of Transportation Research Board*, 2021.