Crash Risk Prediction Model using Data Science

Authors

- **Lara Diab**

  ⓘD [0000-0001-8489-2015](#) · ◯ [diablara](#)

  Illinois Center for Transportation (ICT); Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign · Funded by The Grainger College of Engineering

- **Renan Santos Maia.**

  ⓘD [0000-0002-0877-4006](#) · ◯ [renanssmaia](#)

  Illinois Center for Transportation (ICT); Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign · Funded by The Grainger College of Engineering

- **Farid Saud**

  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ◯ [fsaudm](#)

  Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign · Funded by The Grainger College of Engineering

- **Johann J Cardenas Huaman**

  ⓘD [0000-0002-4695-7639](#) · ◯ [Johann-Cardenas](#) · 🐦 [transporter_pe](#)

  Illinois Center for Transportation (ICT); Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign · Funded by The Grainger College of Engineering

Abstract

Introduction

Road accidents are responsible for a significant number of injuries reported every year. According to World Health Organization (WHO), approximately 1.3 million people die each year as a result of road traffic crashes (June, 2022). In addition, road traffic crashes cost countries 3% of their gross domestic product (WHO, 2022). Consequently, understanding what influences these accidents on roads is of utmost importance. However, it is not easy to decide which exact conditions lead to these accidents. Different road, climate, vehicle and driver conditions affect the likelihood of a driver to be in a fatal/serious accident.

The ability of predicting in an accurate way the potential occurence of crases is a valuable contribution for road safety. In an approach vastly used in the literature, crash records' data are used for the development of crash prediction models, so that agencies can allocate investments to priority areas of the roadway network. However, given that the budget for infrastructure improvements is limited, adopting countermeasures for all facilities that crashes are potentially occuring is not financially feasible. Therefore, informing drivers about the potential safety risks is a way to proactively compensate the aforementioned limitations. Moreover, with the development of connected and autonomous vehicles, this information can be provided in a more optimized way, contributing for vehicles' route decision, as well as for real-time alerts that can lead drivers to take the necessary precautions to operate more safely.

Project Proposal and Objective

The objective of this work is to use the Illinois Department of Transportation (IDOT) extensive crashes' data to be analyzed and, finally, be used for a crash risk prediction model based on main categorical data that can be real-time updated (such as the weather/lighting/pavement conditions). Ideally, it could be used by navigation systems to allert drivers to adopt more cautious behavior as soon as they enter higher-risk sections.

Description of the Data Set

IDOT has generated datasets with statewide crash locations produced by the Crash Information Section of the Illinois Department of Transportation (IDOT). The accident data has been collected througout the years usign Application Programming Interfaces (APIs) that provided streaming traffic incident data. There are about 300,000 accident records per year in these datasets, and each record contains attributes that include conditions like (among others that are not listed and described because these are not relevant for this study):

1. Time and date (day, month, year)
2. Coordinates (x,y)
3. Type of collision
4. A quantitative description of fatalities and injuries
5. Crash severity classification based on their impact on traffic
6. The road surface condition ("Dry", "Wet", "Snow or slush", "Ice", or "Sand/Dirt/Mud")
7. Road defects ("Debris on roadway", "Rut/Holes", "Unkown", or "No defects")
8. Light conditions (rated in a scale from 1 to 9)
9. Geometric characteristics of the road section
10. Work Zone ("construction", "maintenance", "utility", "unknown", or "N/A")
11. Possible causes of the accident.

The datasets for different years are available for downloan as a .CSV file in the IDOT's website:

https://gis-idot.opendata.arcgis.com/search?groupIds=6d2862031a6d47c7a8c211e38e423e05

References

Batty, S. (n.d.). Illinois Roadway Crash Data. Retrieved September 23, 2022, from https://idot.illinois.gov/transportation-system/safety/Illinois-Roadway-Crash-Data

Illinois Department of Transportation | IDOTAdmin https://gis-idot.opendata.arcgis.com/search?groupIds=6d2862031a6d47c7a8c211e38e423e05

Safarpour, H., Khorasani-Zavareh, D., & Mohammadi, R. (2020). The common road safety approaches: A scoping review and thematic analysis. Chinese Journal of Traumatology, 23(2), 113–121. https://doi.org/10.1016/j.cjtee.2020.02.005

World Health Organization. (n.d.). Road traffic injuries. World Health Organization. Retrieved September 23, 2022, from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Yu, R., L. Han, and H. Zhang. Trajectory Data Based Freeway High-Risk Events Prediction and Its Influencing Factors Analyses. Accident Analysis and Prevention, Vol. 154, No. March, 2021, p.106085. https://doi.org/10.1016/j.aap.2021.106085.

1. **Sci-Hub provides access to nearly all scholarly literature**
   Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
   eLife (2018-03-01) https://doi.org/ckcj
   DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

2. **Reproducibility of computational workflows is automated using continuous analysis**
   Brett K Beaulieu-Jones, Casey S Greene
   Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/
   DOI: 10.1038/nbt.3780 · PMID: 28288103 · PMCID: PMC6103790

3. **Bitcoin for the biological literature.**
   Douglas Heaven
   Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888
   DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

4. **Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**
   cOAlition S
   (2018-09-04) https://www.wikidata.org/wiki/Q56458321

5. **Open access**
   Peter Suber
   MIT Press (2012)
   ISBN: 9780262517638

6. **Open collaborative writing with Manubot**
   Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
   Manubot (2020-05-25) https://greenelab.github.io/meta-review/

7. **Opportunities and obstacles for deep learning in biology and medicine**
   Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, …

Casey S Greene
Journal of The Royal Society Interface (2018-04) https://doi.org/gddkhn
DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

8. **Open collaborative writing with Manubot**
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
PLOS Computational Biology (2019-06-24) https://doi.org/c7np
DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653

This manuscript is a template (aka "rootstock") for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Basic formatting

Bold text

Semi-bold text

<div align="center">Centered text</div>

<div align="right">Right-aligned text</div>

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
   a. Sub-item
   b. Sub-item
      i. Sub-sub-item
3. Ordered list item
   a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Document organization

Document section headings:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 1`'s are recommended to be reserved for the title of the manuscript.

`Heading 2`'s are recommended for broad sections such as *Abstract*, *Methods*, *Conclusion*, etc.

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 1

Figure 2

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 4: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	Jane	John	Alice	Bob
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	Digits 34-66	Digits 67-99	Ref.
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 3: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \tag{2}$$

Special

⚠ **WARNING** *The following features are only supported and intended for* `.html` *and* `.pdf` *exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as* `.docx` *.*

> **LINK STYLED AS A BUTTON**

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the Font Awesome icon set:

✔ ? ★ 🔔 ✖ •••

> 🏳 **Light Grey Banner**
> useful for *general information* - manubot.org

> ℹ **Blue Banner**
> useful for *important information* - manubot.org

> 🚫 **Light Red Banner**
> useful for *warnings* - manubot.org