CEE498 DS - Data Science for CEE - Fall 2022

Authors

- **Amirthavarshini Muraleetharan**
  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ○ [amirthavarshini246](#) · 🐦 [NIL](#)
  Department of CEE, University of Illinois Urbana Champaign

- **Thomas Ngare**
  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ○ [thomasNg](#) · 🐦 [johndoe](#)
  Department of CEE, University of Illinois Urbana Champaign

- **Kapil Shah**
  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ○ [kapilrs2](#) · 🐦 [shkapil12295](#)
  Department of CEE, University of Illinois Urbana Champaign

Abstract

ABCD

"A Machine Learning Based Approach, For Predicting Road Closure Events, Given Data of the US Road Construction and Closure" Amirthavarshini Muraleetharan1, Kapil Shah2, Thomas Ngare Matiki3 A nationwide dataset of road construction and closure events, including data from all 49 US states is chosen for the project. The roadwork included in this dataset's construction events ranges from minor paving repairs to significant undertakings that might take months to complete. Several APIs that provide streaming traffic incident (or event) data are used to collect the data between January 2016 and December 2021. These APIs transmit traffic information gathered by several organizations, including the US and state departments of transportation, law enforcement organizations, traffic cameras, and traffic sensors embedded in the road networks. The number of construction and shutdown records in this dataset currently stands at roughly 6.2 million. In general, this dataset can be used for a wide range of applications, including the prediction of short- and long-term road construction, the prediction of road closures, the study of the life cycle of road construction, the development of insights to help city planners choose construction sites wisely with the most negligible negative impact on traffic flow, and the investigation of the influence of precipitation or other environmental stimuli on the need for road work. The dataset is being updated on an annual basis. The data will be obtained from US Road Construction and Closures (2016 - 2021), Kaggle, and it is available in CSV format. Presently, the dataset contains 6,170,627 observations comprising of features like Construction severity, Latitude and longitude, Precipitation, Traffic signal and many such taking a total of 47 columns. Table 1 elaborates the specifics of this data set.

Using this dataset, a machine learning model will be developed to predict road closure events, given inputs of pertinent features as mentioned in the below Table 1. measurements from pertinent features that will be determined in this study. The developed algorithm will be tested on about 20% of the total samples, and validated with another 20% of the total samples, to be made suitable for real-time applications. The proposed algorithm will have the potential to estimate the likelihood that every road segment satisfying certain requirement will be is closed or open, such that it can be employed in mobile maps (like Google maps). Whenever the likelihood of a closure event exceeds a predetermined threshold, customer can be notified to find the efficient route. For mapping applications, this algorithm can it will automatically update maps further add to its real time monitoring.

To achieve this goal, data wrangling will be performed. The essential data frames for the study will be extracted from the original dataset followed by Exploratory Analysis. Exploratory analysis will enable us to drive insights by forming a pattern for better visualization and exploration. be performed to make sense of the data and visualize them for better exploration. Based on the analysis, a classification model will be developed to predict the class or category for the data or draw a conclusion to the input data given for training. The developed model will be tested for various conditions.

Table 1: Description of Undertaking Dataset Features Description ID Unique identifier of construction record Severity Shows the severity of the construction Start and End Time Shows the start time of construction End Time Shows the end time of construction Latitude and Longitude Shows the GPS coordinates Distance The length of the road extent affected by the construction Street Details Shows the street number, name and right/left side in address field Address Details Shows the city, county, state, country and zip code in address field Time zone Shows time zone based on the location of the construction event Weather Shows the time stamp of weather observation record Temperature, Wind, Humidity, and Pressure Shows the temperature, wind chill, humidity, and pressure Visibility Shows visibility Wind Direction and Speed Shows wind conditions Precipitation and Weather condition Shows precipitation and weather condition Amenity An annotation which indicates presence of amenity in a

nearby location Bump and Crossing Annotations which indicate presence of speed bump or hump and crossings Give way, Junction, railway Annotations which indicate presence of give way, junction and railway Exit, Roundabout, Station, Stop Annotation which indicates presence of no exit, railway, roundabout, and station Traffic Details Annotations which indicate traffic calming, signal, turning loop Light Details Annotations which indicate sunrise, sunset, civil twilight, nautical twilight, astronomical twilight

Once successfully tested, this algorithm can further be developed (given the time permits) to choose construction sites which has less impact of traffic flow, thereby guiding city planners. Also, a study can be undertaken to understand the impact of precipitation on the need of road work.

References:

1. Karimi Monsefi, Amin, Sobhan Moosavi, and Rajiv Ramnath. "Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data." , 2022

2. US Road Construction and Closures (2016 - 2021) | Kaggle

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:

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

An equation too long to fit within page:

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

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

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

References

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