

A Machine Learning Based Approach, For Predicting Road Closure Events, Given Data of the US Road Construction and Closure

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-jakt@d0e36a2](#) on September 18, 2022.

Authors

- **Amirthavarshini Muraleetharan**

 [677-010-487](#) ·  [amirthavarshini246](#)

Department of CEE, University of Illinois Urbana Champaign

- **Thomas Ngare**

 [652-601-317](#) ·  [thomasNg](#)

Department of CEE, University of Illinois Urbana Champaign

- **Kapil Shah**

 [668-376-620](#) ·  [kapilrs2](#)

Department of CEE, University of Illinois Urbana Champaign

Abstract

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

Features	Description
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 from Kaggle