**UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN**

**DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING**

**CEE498 DS – Data Science for CEE - Fall 2022**

***Term Project Proposal***

**A Machine Learning Based Approach, For Predicting Road Closure Events, Given Data of 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](https://www.kaggle.com/datasets/sobhanmoosavi/us-road-construction-and-closures), and it is available in CSV format. Presently, the dataset contains 6,170,627 observations/samples and 47 columns/features. The columns of the dataset includes information about the ID, Severity, start time, end time, starting latitude, ending latitude, ending longitude, distance, description, number, street, side, country, state, zip code, country, time zone, airport\_code, weather timestamp, temperature(F),wind\_chill(F), humidity(%), pressure(in), visibility(mi), wind\_direction, wind\_speed(mph), precipitation(in), weather\_condition, amenity, bump, crossing, give\_way, junction, no\_exit, railway, roundabout, station, stop, traffic\_calming, traffic\_signal, turning\_loop, sunrise\_sunset, civil\_twilight, nautical\_twilight, and astronomical\_twilight.

Using this dataset, a machine learning model will be developed to predict road closure events, given 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 is closed or open, such that it can be employed in mobile maps (like Google maps). For mapping applications, it will automatically update maps, whenever the likelihood of a closure event exceeds a predetermined threshold.

To achieve this goal, the essential data frames for the study will be extracted from the original dataset. An exploratory data analysis will 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.

*Reference:*

*1. Karimi Monsefi, Amin, Sobhan Moosavi, and Rajiv Ramnath. “*[*Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data*](https://www.kaggle.com/datasets/sobhanmoosavi/us-road-construction-and-closures)*.”, 2022*

2. [US Road Construction and Closures (2016 - 2021) | Kaggle](https://www.kaggle.com/datasets/sobhanmoosavi/us-road-construction-and-closures)