

# DRISI

CALTRANS DIVISION OF RESEARCH,  
INNOVATION AND SYSTEM INFORMATION

# Research Results

Transportation  
Safety and  
Mobility

JUNE 2023

**Project Title:**

Developing a Safety Effectiveness  
Evaluation Tool for California

**Task Number:** 4016

**Start Date:** June 1, 2022

**Completion Date:** March 31, 2023

**Task Manager:**

Jerry Kwong  
Transportation Engineer (Electrical)  
jerry.kwong@dot.ca.gov

## Developing a Safety Effectiveness Evaluation Tool for California

Developing an Excel based tool to predict Crash Modification Factors using the federal CMF clearinghouse data.

### WHAT WAS THE NEED?

Safety is one of the top priorities of Caltrans. Caltrans is in urgent need of a safety tool to evaluate various safety countermeasures and rank and prioritize safety projects. In the planning and development of safety projects, state Departments of Transportation rely on an assessment tool to evaluate the impacts of alternative safety countermeasures. To respond to this need, it is necessary to develop a safety project evaluation tool that employs the state-of-the-practice methods, reflects local conditions, and tailors to Caltrans' needs to quantify the impacts of one or more countermeasures using expected fatal and serious injury reduction.

### WHAT WAS OUR GOAL?

The goal is to develop an easy to use Crash Modification Factors (CMF) prediction tool that can help assess the safety benefits of safety improvement projects in California.

### WHAT DID WE DO?

This project developed a data-driven framework for CMF prediction through mining the Federal Highway Administration (FHWA's) CMF clearinghouse data. The proposed framework handled the heterogeneity of CMF clearinghouse data in a flexible way, including the unstructured descriptions of countermeasures, the high-cardinality categories, missing rates, and noises. This is achieved through introducing the cutting-edge Natural Language Processing (NLP) techniques and the target encoding method. In addition, it combined multiple machine learning methods to make predictions on combined



DRISI provides solutions and  
knowledge that improves  
California's transportation system

countermeasures. The framework was first coded into Python and trained and evaluated against FHWA clearinghouse data. The results show that there is a good consistency between model predictions and the publicly known CMFs. Finally, the Python based tool was converted into a simpler, Excel based CMF prediction tool and the Excel tool was transferred to Caltrans for testing and evaluation.

## WHAT WAS THE OUTCOME?

The project developed both a Python based and an Excel based CMF prediction tool that are capable of predicting CMFs for individual and combined safety countermeasures. The Excel tool was delivered to Caltrans traffic safety for evaluation and feedback.

## WHAT IS THE BENEFIT?

This project demonstrates that the machine learning based, data driven approach is capable to mine the hidden relations between safety countermeasures and CMF embedded in the federal CMF clearinghouse. Supplied with local CMF data, the tool can fine tune its predictions and will be able to help Caltrans evaluate and prioritize its safety improvement projects in the future.

## LEARN MORE

Review the complete report.

### Project Contacts:

Caltrans:

Abdelrahman Beshair, [Abdelrahman.beshair@dot.ca.gov](mailto:Abdelrahman.beshair@dot.ca.gov)

Ray Estahri, [ray.estahri@dot.ca.gov](mailto:ray.estahri@dot.ca.gov)

Jerry Kwong, [jerry.kwong@dot.ca.gov](mailto:jerry.kwong@dot.ca.gov)

UC Davis:

Michael Zhang, [hmzhang@ucdavis.edu](mailto:hmzhang@ucdavis.edu)

## IMAGES

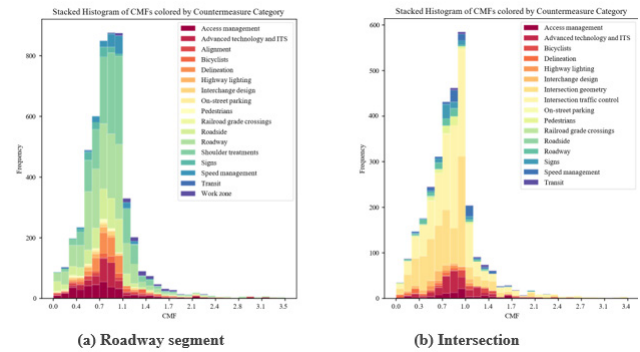


Image 1: The CMF frequency distribution over different countermeasure categories for the roadway segment (a) and intersection (b) types in the FHWA CMF clearinghouse

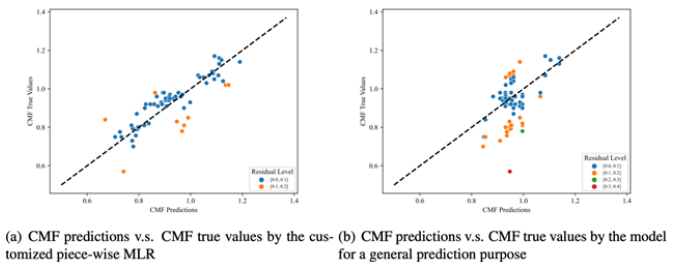


Image 2: The true CMFs vs the predictions from the piecewise MLR model (a) and the general prediction model (b) on the test set for the shoulder width category