A Machine Learning Model for Generating Traffic Data on California Highway System

This project will develop a Machine Learning (ML) model to predict statewide traffic volumes and speeds.

Traffic Operations

MAY 2024

Project Title: A Machine Learning Model for Generating Traffic Data on California Highway System

Task Number: 4413

Start Date: June 1, 2024

Completion Date: TBD

Task Manager:
Charina Guarino
Transportation Engineer (Civil)
charina.guarino@dot.ca.gov

WHAT IS THE NEED?

The California Department of Transportation (Caltrans) is looking to increase the accuracy and reliability of the traffic data collected statewide. The primary methods of traffic data collection by Caltrans are the use of Intelligent Transportation System (ITS) elements such as Traffic Census and Performance Measurement System (PeMS) stations. The data collected through these methods are limited by the sparsity of covered area, unreliable sensor functionality and unreliable quality. Caltrans needs accurate traffic data to perform essential functions such as traffic flow optimization, infrastructure development, safety enhancement and emergency response. There is a need to improve and ensure Caltrans traffic data in robust and reliable using existing traffic data sources. A Machine Learning (ML) model would be developed to fill in the gap in Caltrans traffic data. This approach has the potential to eliminate the need to procure additional traffic sensors and other data sources.

WHAT ARE WE DOING?

In this research, the Advanced Highway Maintenance and Construction Technology (AHMCT) Research Center at the University of California (UC) Davis will conduct a literature review on a mixture of data integration techniques including Ensemble Learning, Deep Learning, Time Series Analysis for Temporal data, Geospatial Analysis, Feature Engineering, Transfer Learning, Outlier detection, and Multi-modal models to determine which technique(s) would be the most appropriate for Caltrans needs. AHMCT will also conduct data preprocessing, exploratory data analysis (EDA) and data integration using traffic data provided by Caltrans. AHMCT will propose features to include in the development of ML model for forecasting traffic flow and speed.

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

This will be accomplished in partnership with the Division of Traffic Operation's Business intelligence & Automation team to understand their needs. The results will be compiled in a final report, along with the ML model and technical documentation for model training and evaluation.

WHAT IS OUR GOAL?

he goal of this research project is to develop a ML model with the ability to forecast and impute traffic volume and speed information, filling in the data gap of current Caltrans traffic data sources.

WHAT IS THE BENEFIT?

The results of this project will help improve the reliability of traffic data available to Caltrans without the procurement of additional data collection equipment and software. The ML model will bridge the data gap across the State Highway System as Caltrans works to meet its Strategic Management Plan.

WHAT IS THE PROGRESS TO DATE?

As of May 2024, contract was sent to Caltrans Legal team for review. Legal has requested an IT consultation to determine if project will involve Generative Artificial Intelligence (GenAI) as a result of new Governor's policy on GenAI.