Machine Learning-Driven Insights into Driver Behavior and Risk Factors

Donaven Lobo

Project Proposal

ME 8813
Dr. Yan Wang | Dr. Surya Kalidindi
Georgia Institute of Technology
North Avenue, Atlanta, GA 30332

1. Introduction

What is the engineering problem(s) that you plan to study/solve? Why are such problems and issues important?

The focus of this project is to develop a system that accurately assesses driving risk based on behavior patterns. The importance of this problem lies in its potential to proactively identify and mitigate risky driving habits, enhancing road safety for all users. By alerting drivers to potential dangers related to their current behavior, the system aims to encourage safer driving practices. Furthermore, from an auto-insurers perspective, this project supports the creation of a usage-based insurance model. This model assigns risk scores to vehicles and drivers based on their driving habits, promoting safer driving through the incentive of reduced insurance premiums. The significance of this project is the incentivised safe driving practices, promising to enhance insurance practices by rewarding responsible driving with financial benefits.

2. Project Plan

What is the machine learning model(s) that you plan to use to solve the problems/issues? How do you plan to obtain data to train and validate your machine learning model(s)?

Through my research with Ford Motor Company and Ford Insure, I have access to their 1-Hz BlueCruise dataset, which is primarily used for developing their autonomous vehicle algorithms. This dataset contains many low-level vehicle signals, including velocity, lateral and longitudinal acceleration, and visible range, among many others. My aim is to repurpose this data to analyze real driver habits when the vehicle is not in autonomous mode. Although the dataset is limited to Ford's 'BlueZones'—areas restricted to highways and protected roads—I think it could still hopefully provide valuable insights into risky driving behaviors such as late braking, weaving in traffic, tailgating, etc.

One challenge is that the dataset contains raw signals without direct labels for driving behavior, making it an unlabeled dataset. To address this, I plan to employ unsupervised learning algorithms to uncover patterns in the data. Additionally, I intend to develop a simple scoring model to evaluate driving behaviors and generate artificial labels, enabling the use of supervised learning algorithms for further analysis.

The ultimate goal is to create a type of Recurrent Neural Network (RNN) capable of identifying time-sequenced patterns in the data that lead to dangerous situations. By doing so, I hope to develop predictive models that can actively monitor a vehicle's driving behavior and trigger warnings when risky patterns are detected.

Here are a few ML algorithms/methods I anticipate having to use for this project: K-Means Clustering, Principal Component Analysis (PCA), Bayesian Optimization, K-Fold Validation, Support Vector Machines (SVM), Decision Trees/ Random Forests, RNN/ Long Short-Term Memory (LSTM), and whatever else arises.