CS210 - Data Management for Data Science
Project Proposal
November 2, 2025
By: Kaajal Yadav & Rhea Yadav

RoadAware: *Driver Drowsiness Detection & Safety Alert Management System*

One of the major causes of car crashes is because of driver-drowsiness. Statistics proved by the AAA Foundation for Traffic Safety 2002's survey by the National Highway Traffic Safety Administration found that thirty-seven percent of drivers reported falling asleep while driving. This challenge is a tremendous problem as it can cause crashes within seconds. While technology may be advancing and efforts to driver alertness and safety are apparent in today's society, technology can not completely replace the role of a responsible driver. An article within the National Library of Medicine states that in the United States the majority of crashes that are dealt with driver-drowsiness involve head-on collisions and road departures. It is clearly evident that there needs to be an improvement in alert management and driver safety.

RoadAware is a driver drowsiness detection and safety alert management system that intends to analyze data and develop a system that will decrease driver drowsiness. RoadAware consists of three components: a driver drowsiness detection machine learning model that image classifies driver's faces, a safety alert management system, and data analysis of drowsiness events. The project relates to the lectures discussed because it will utilize Python/SQL to achieve data exploration, visualization and analysis, file processing, data frame operations, the development of a database and management system that will establish various relationship schemas, and predictive modeling.

The project is important because it tackles a challenging problem that is present in society. Oftentimes, drivers are not aware of the fact that they are distracted, drowsy, or unattentive while driving. We are excited about the project because we hope to utilize the skills we have learned in the course to develop a meaningful system that will contribute to the improvement of driver safety. Some existing issues in current data management practices include inconsistencies in datasets, biases or unrepresentative datasets of a training model's population, and scattered data that is across various systems and applications. There are prior related works on the topic of driver drowsiness. A few instances of prior works revolve around drowsiness detection using respiration in thermal imaging, fatigue detection based on yawning in thermal images, driver drowsiness detection using face, eye features, and a convolutional neural network. Such prior work is relevant to our project because it showcases how various researchers and data scientists have used various features to create prediction models.

<u>Planning:</u>

- The kind of data that we will be using is from a collection of datasets from Kaggle, the Driver Monitoring dataset, and online datasets (such as Universe Roboflow) that have a sufficient amount of sets to utilize. We will be able to get the data by downloading the dataset. We will then be putting the data into an integrated development environment, such as VSCode for data exploration and analysis. The data will be numerical and/or categorical and will contain the data that will be used throughout the project.

- We will primarily be using predefined datasets, though we plan to include simulated data in order to extract additional details for modeling a relationship. For instance, the simulated data can be used for analyzing the relationship between driver drowsiness and different quantitative/qualitative variables. These can include blinking rate, heart rate, time spent driving, and sudden decreases in speed.
- The models and techniques that we are planning to develop will be useful in supporting our data visualizations. We will develop a machine learning model that will be an image classification tool to analyze drivers. The tools and technologies we may utilize includes Pytorch, Tensorflow, or Google Teachable Machine to classify a driver into three categories: alert, drowsy, or distracted. There would be certain conditions when a driver would be considered to be classified as a category, for example the driver is alert when their hands are on wheel and their eye contact is on the road.
- For the algorithms involved in this project, we will be creating an alert system via a simulation. If a drowsiness event is detected a record will be generated triggering a simulated alert through console output. There will be a certain threshold, for example, if 3 drowsiness detections in 5 minutes the alert will be classified as "critical."

Implementation Steps:

- For this project, we will ensure that we take the necessary steps in order to achieve data analysis, modeling, and data management. The first step would ideally be to create a database in SQL in order to store quantities such as driver age, height, average time spent driving, etc. Once we are able to store values in the database we will then begin on implementing different data visualizations. We will be able to present data in graphs, charts, and plots to better convey our data. Once the visualizations are created, we will begin working on our machine learning model. It is important that we include a large data set to ensure high accuracy. Lastly, we will implement the feature of our project to add more complexity, which is the alert management system. This will be created by utilizing tables to track records and a counter variable to store when the driver is considered as "critical." We will be evaluating our method by performing several steps to ensure that the model and algorithms are accurate. This can be done by performing multiple trials to ensure that the precision of results are matching accurately.

Testing & Measuring Success:

- We will train our image classification machine learning model and later test it. We will be choosing representation and diverse data sets. Through this we aim to achieve high performance, with high training and testing accuracies. Furthermore, we will measure success by creating diagrams and analyzing the data. Success will be measured by the model performance, consistency and accuracy.

- It is important for our model and algorithms to be accurate. This can be performed by noting down the results that we receive. Then from there, we will iteratively improve our results to receive higher accuracies.

Sources:

- https://pmc.ncbi.nlm.nih.gov/articles/PMC11725507/
- https://aaafoundation.org/prevalence-impact-drowsy-driving/
- https://pmc.ncbi.nlm.nih.gov/articles/PMC8914892/
- https://universe.roboflow.com/augmented-startups/drowsiness-detection-cntmz
- https://dmd.vicomtech.org/#Download