Sobriety Test

Siddharth Bhattacharya : Model Training and Validation, Data Analysis, Testing
Kanav Talwar : Data Collection, Data Analysis, Report Writing, Testing
Laxman Mohanarajan : Data Collection, Data Analysis, Testing

Overview

In recent years, there has been a significant increase in the number of students who have been involved in cases of intoxication. This is a serious problem that can have a number of negative consequences, including injuries, accidents, and even death. Our project aims to address this problem by developing a new method for detecting intoxication. Our method uses the user's phone's inbuilt accelerometer and gyroscope to conduct a walking test called the Tandem Gait. In this test, the person takes five heel-to-toe steps, turns 180°, and then repeats the process. The data collected by the phone is then used to classify the user's state of intoxication. We believe that our method has the potential to be a valuable tool for preventing intoxication-related injuries and accidents.

Application

This project is designed to assist individuals, specifically college students, who are unsure of their state of intoxication. It does this by providing a means to assess their levels of intoxication.

Features

As mentioned above, using the phone’s accelerometer and gyroscope we classify the person’s gait into three categories, Sober, Tipsy and Drunk. Looking at the amplitude of the peaks from the given live gyroscope and accelerometer data when performing the test, our program classifies the data into one of the three categories using the data collected by us for training our classifier.

Data Collection

All collection of the data was simulated to enact the stages of sobriety. Looking at online resources and videos of inebriated individuals, we overall recorded around 100 samples of the Tandem Gait with 3 different levels of intoxication. The accelerometer data and gyroscope data was calibrated while recording. All data was collected and zipped using the Sensor Logger application on our mobiles, renamed for easier accessibility and stored in a CSV format.

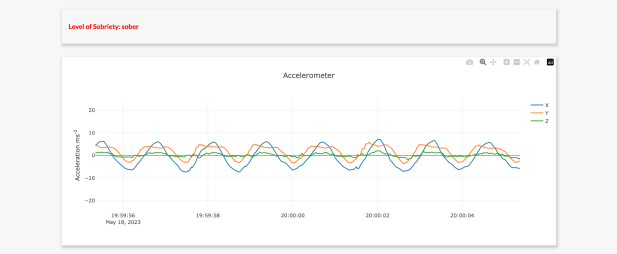
Data Analysis and Model Training

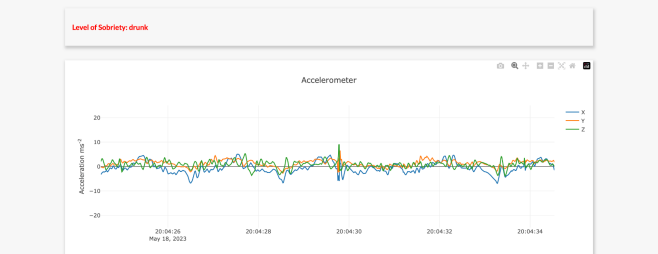
Using inbuilt python functions in NumPy, each accelerometer data point in our recorded data is individually reoriented using a loop, which calls the **reorient()** function and accepts the required arguments. The reoriented data that results is kept in a NumPy array which we use for data classification. A sliding window approach is used for feature extraction and label assignment. A 10-fold cross-validation technique is used to divide the data into train and test datasets, which

then loops through each fold, training the decision tree classifier, making predictions, and evaluating performance using metrics like accuracy, precision, and recall.

Results

During our run of the program these are a few of the results we received:



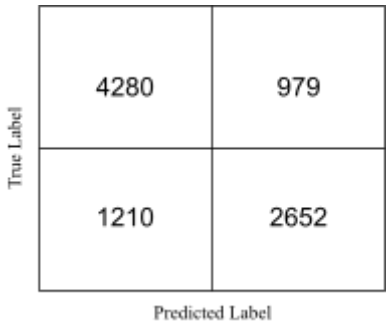


You can see the difference between the accelerometer reading in the graph and what our classifier gives us as an output for it.

While training the program, we were getting poor results with the classifier's accuracy, precision and recall as the data for tipsy and drunk were really similar and in real life can depend from person to person. To improve the accuracy of our classifier, we then tried to run it with there being only two classes instead of three with most of the tipsy data being used as drunk. Here is a table displaying the two results:

	Three-way classification	**Two-way classification**
Average Accuracy	0.5517102676463158	0.7546336061951155
Average Precision	0.5579847610553056	0.7538792879607776
Average Recall	0.5623704727930093	0.7441069003383675

After comparing the above results, we chose to go with the two-way classifier as it gave us more accurate results. The confusion matrix of the calculated two-way classification was:



Learnings from this Project

A few things we can take away from this project:

- How recording real vs simulated data can affect analysis and training models
- Importance of processing data correctly with respect to the project we are working on
- Each person's state of intoxication is different, making data collection complicated from person to person.
- Importance of using the correct sensors which can detect the respective symptoms regarding intoxication.

How can we improve?

We feel that we could have used more sensors in this project, for example used the blood oxygen levels as well as used the heart rate sensor as even these are affected during high levels of intoxication. Our data collection could have been more accurate as well due to it being simulated.