

**CSE523 - Machine Learning**

**Project 11: Identify abnormal driving behavior using spatio-temporal analysis**

**Weekly Report after Mid-Sem Presentation**

**Group: Titans**

**Team Members**

|  |  |
| --- | --- |
| Enrollment Number | Name |
| AU2140043 | Bhavya Khakhar |
| AU2140154 | Harsh Loriya |
| AU2140170 | Krishna Patel |
| AU2140187 | Priyam Shah |

**Previous Approach:**

In the previous approach, the dataset which was provided consisted of mixed files which were normal and abnormal . The trajectory dataset consists of spatial coordinates which are left, top, width and height in each frame number. We separated the files in two different folders and named them normal and abnormal respectively. Now, in each we labeled every row of a file. Then we extracted the features using the Ant colony algorithm and classified the features using Logistic regression. In this approach our model was not able to recognize the pattern of a large data as we were just labeling a row of a file.

**Started work on the Revised Approach.**

In the new approach, we fundamentally changed our way of dealing with the dataset. In the previous approach, we were labelling each cell; instead, in the revised approach, we are labelling each file. The dataset that we have been provided is for three different roads. Each file of the dataset contains a new scenario. So what we have done is, we have separated all files into two folders, that is normal and abnormal. We have renamed them; if it is normal, then “normal\_10\_1”, and for the abnormal file, it is “abnormal\_10\_1”. Why this format, you might wonder? So as to make things easy, this format itself gives us the file name and the label of the file in just one argument.

After labeling all the files, we have stored them in the corresponding arrays. Then, we performed the train test split and randomly selected the file names. Further in the training step, we chose the file name, opened that file, and trained the model for its whole data at once. We performed the same for the test part.

**Problems:**

In this approach, there is a significant data imbalance. Because of that, we are facing some challenges in the training of the model. We are trying new strategies to tackle them without deleting the dataset because we might end up removing essential parts. We are planning to use a feature selection method mainly.