**Report:1**

**Date: 01-03-2025**

**Group-1**

**Team members:**

| **Aarsha shah** | **AU2240163** |
| --- | --- |
| **Hetav shah** | **AU2240253** |
| **Ved savalia** | **AU2240242** |
| **Parshwa shah** | **AU2240249** |
| **Dev gorakhiya** | **AU2240087** |

**Project Title: Hard stop and momentary stop using vehicle trajectory dataset**

* A hard stop is one in which a vehicle brakes rapidly and comes to a stop within a short distance due to sudden braking for emergencies, instant traffic slowdowns, near misses, or fast reactions to traffic lights. A momentary stop is a transient stop before the vehicle recommences travel, typical in stop-and-go traffic, at intersections, yield signs, or in heavy traffic.
* Since stop types are not labeled in real-world data, unsupervised clustering methods can help identify patterns.
* There are numerous unsupervised clustering methods we have been looking for, such as k-means clustering, density-based clustering, and hierarchical clustering. Of the three, we are going to utilize the DBSCAN algorithm.
* DBSCAN Algorithm
  + DBSCAN is a clustering algorithm that clusters points by density. In contrast to K-Means, it does not need to know the number of clusters beforehand and can identify clusters of any shape as well as indicate noise (outliers).
  + DBSCAN groups points using two key parameters:
    - Epsilon (ε): The maximum distance within which points are considered neighbors.
    - MinPts: The minimum number of points required to form a dense region (cluster).
  + DBSCAN determines neighbors within a radius ε by initially selecting a random, previously unvisited place. A point is designated as a core point and initiates a new cluster if it possesses a minimum of MinPts neighbors. Alternatively, categorize it as noise succinctly. In a single traversal of the remaining new core locations, incorporate directly accessible neighbors to enlarge the cluster. This continues until there are no other points to include. The procedure proceeds to the next unvisited spot and thereafter backtracks. Any remaining points that do not belong to a cluster are ultimately classified as noise.
  + DBSCAN struggles with non-uniform density since a fixed ε may not work well for data sets with regions of varying densities, and it is not ideal for very high-dimensional data since the distance measure becomes less meaningful.
  + In order to use the DBSCAN algorithm, we must select parameters such as features, epsilon, and MinPts from the dataset provided.

**Future work:**

1. Pre-processing of the dataset given by finding the velocity and average velocity of each vehicle’s trajectory.
2. Researching more about different unsupervised clustering algorithms.

**Research papers for unsupervised clustering used in real life:**

The Application of Unsupervised Clustering Methods to Alzheimer’s Disease

<https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2019.00031/full>

Investigate the Correlation of Breast Cancer Dataset using Different Clustering Technique

<https://arxiv.org/abs/2109.01538>

<https://www.youtube.com/watch?v=XwzbFzqhF1Y&t=1s>

<https://www.youtube.com/watch?v=eIdnKMiLeWM&t=2s>