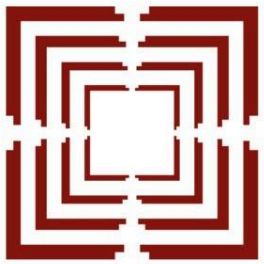

Ahmedabad University

CSE623 - Machine Learning Theory and Practice
Section - 1

Weekly Report 5
Identify Hard stop and momentary stop using the vehicle trajectory dataset

---

Team Name: **The Overfitters**

Name	Enrolment Number
Jinil Savaj	AU2240159
Jay Raval	AU2240151
Meet Suthar	AU2240198
Karan Prajapati	AU2240161
Vishesh Bhatia	AU2240027

1. Introduction

In our work toward improving vehicle stop detection with trajectory data sets, our first effort was to use the **Random Forest** algorithm. This choice was made to see how well it could distinguish hard stops from momentary stops. For future work, we intend to build upon our work by integrating and comparing other methods, specifically density-based cluster methods like **DBSCAN**. Our progress with Random Forest and our intention to investigate DBSCAN and other cluster-based methods are chronicled here.

2. Random Forest for Stop Detection

a. Motivation

Random Forest, being an ensemble learning approach, builds many decision trees and combines their outputs to provide a more accurate and robust prediction. We used it to classify stops and non-stops from features extracted in vehicle trajectory data.

b. Implementation

i. **Feature Selection:** Derived speed fluctuations, acceleration trends, and temporal variations in car motion.
ii. **Model Training:** Trained the model using labeled trajectory data to differentiate between hard stops and momentary stops.
iii. **Evaluation Metrics:** Evaluated performance based on accuracy, precision, recall, and F1-score.

c. Results and Observations

i. The model handled well the recognition of stops with good accuracy.
ii. However, classification mistakes were observed and noted.
iii. Computational performance was adequate but could be enhanced by feature selection optimization.

3. Future Work: Exploring DBSCAN and Other Methods

a. Exploring Density-Based Clustering

One of the alternative methods we seek to explore is **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise) for stop identification in vehicle trajectory data. DBSCAN's density-based clustering strategy can identify stops effectively without the need for labeled data.

b. Comparative Analysis Plan

In order to enhance our results, we will undertake a comparative analysis of **Random Forest, DBSCAN, K-Nearest Neighbors (KNN), and Gaussian Mixture Models (GMM)**. Major parameters to be evaluated are:

- **Accuracy & Precision:** Measuring the performance of classification under varying stop situations.
- **Computational Efficiency:** Assessing processing time and scalability.
- **Robustness:** Detecting parameter tuning and noise sensitivity.

c. Implementation Plan for DBSCAN and Other Models

i. Preprocessing the Data:

- Normalize trajectory data to have a uniform scale.
- Extract important features like speed changes, acceleration, and stop duration.
- Eliminate noise and outliers that could impact clustering performance.

ii. Implementing DBSCAN:

- Choose suitable hyperparameters, especially epsilon (ε) and minimum points (MinPts), for best clustering performance.
- Run DBSCAN on the preprocessed dataset to detect stop clusters.
- Verify clustering results against ground truth stop data.

iii. Comparing with Other Models:

- Run KNN, GMM, and Random Forest on the same dataset.
- Tune the hyperparameters of each model to achieve optimal detection accuracy.
- Compare outcomes based on standardized metrics including precision, recall, F1-score, and computation time.

iv. Evaluating and Refining Models:

- Examine differences in performance and improve feature selection.
- Maximize computational efficiency, especially for real-time environments.
- Test final models on an independent data set for the purpose of robustness evaluation.

4. Conclusion

Our work until now has proven Random Forest effective in vehicle stop detection. We will extend our methodology with **DBSCAN** and **other clustering algorithms** from here onwards. With a systematic comparison, we would like to ascertain the best algorithm for accurate and efficient computational stop detection. The experimental setup plan laid out above will dictate how we carry out experiments and verify the models strictly.