Classification of Drivers Based on Driving Patterns

Raiyan Diwan

Deeprajsinh Gohil

Vats Patel

Tejas Pansuriya

*Abstract*— **This report provides an unsupervised learning strategy for driver classification based on driving behaviours. Driving characteristics like velocity, warning signs, and crash anticipation are studied. Frame number, coordinates, velocities, warning signal distances, reception anticipation, and collision kinds are in the dataset. The analysis classifies drivers by driving behaviour using KMeans clustering and KNN algorithms. Results show that the suggested method classifies drivers and identifies driving habits.**

Keywords—Unsupervised Learning, Driver Classification, Driving Patterns, KMeans Clustering, K Nearest Neighbors.

# Introduction

To improve road safety and construct intelligent transportation systems, analyse driver behaviour. Driving trends can expose driver behaviour for driver assistance systems, insurance risk assessment, and traffic management. Without tagged training data, unsupervised learning can classify drivers by conduct. A dataset of driving parameters is used to categorise drivers using unsupervised learning methods.

# Methodology

## Data Preprocessing:

## Time-series data were collected from CSV files containing information on vehicle dynamics such as velocity and wind parameters. Preprocessing involved feature extraction, risk calculation, handling missing values, and interpolation to ensure data consistency. This phase ensured the integrity and readiness of the dataset for analysis.

## Resampling and Standardisation:

## To ensure fair comparison across drivers, time-series data were resampled using cubic spline interpolation and standardized to eliminate scale differences. This step enhanced the dataset's uniformity and suitability for clustering analysis.

## Feature Engineering:

## Feature engineering involved transforming raw data into informative features to enhance clustering analysis. Techniques such as calculating risk factors from velocity and wind parameters were employed. Resampling and standardization ensured consistent feature representation across drivers, improving the dataset's informativeness and facilitating accurate clustering.

## Clustering Analysis

Time-series clustering algorithms, including Euclidean k-means and DTW k-means, were utilized to group similar driving patterns based on vehicle dynamics and risk factors. Model evaluation metrics were used to select the optimal clustering model, ensuring robust and interpretable results.

## Exploratory Data Analysis (EDA):

Descriptive statistics and visualization techniques were employed to gain insights into data distribution, trends, and relationships. EDA provided a comprehensive understanding of the dataset's characteristics, guiding subsequent analysis steps effectively.

## Classification

K Nearest Neighbors (KNN) algorithm is utilized to classify drivers based on their driving patterns. Both dynamic time warping (DTW) and Euclidean distance metrics are explored for classification.

# Results

The results reveal distinct clusters of driving behaviors, indicating varying patterns in velocity, warning signal distances, and collision types among drivers. The elbow method suggests an optimal number of clusters for K-means clustering.

Visualization using PCA demonstrates separability among clustered data points. Classification accuracy rates using different algorithms are reported, with KNN achieving notable performance in driver classification. Time series-based classifiers also show promising results, particularly when considering the temporal nature of driving patterns

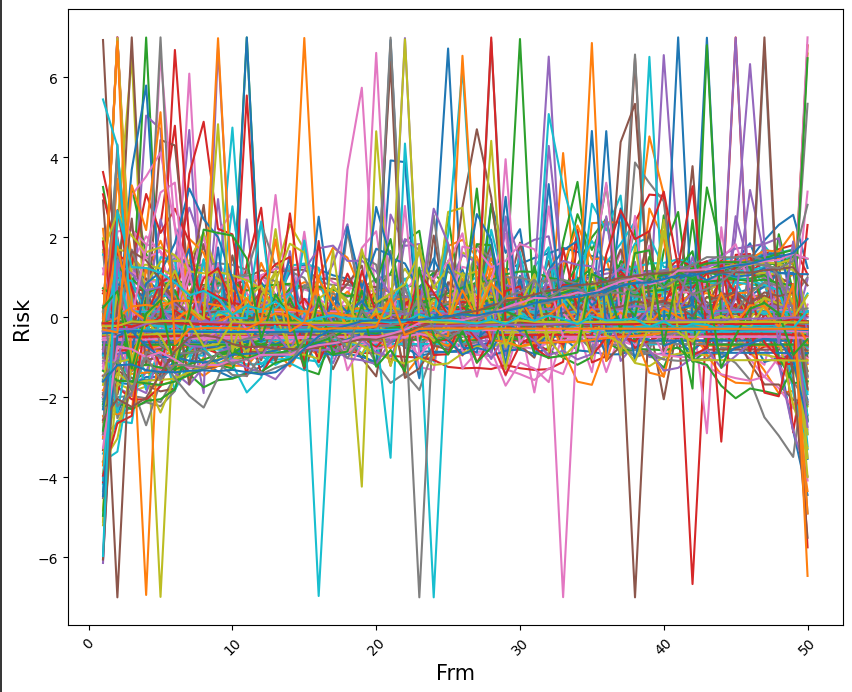
. 

Fig 1 – risks for each drivers

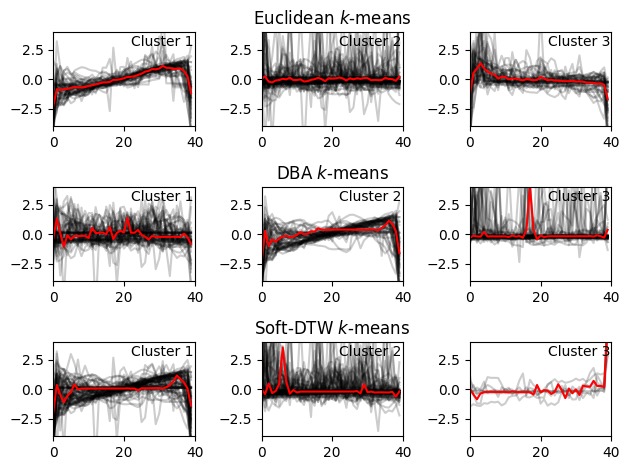


Fig 2 – for 3 clusters

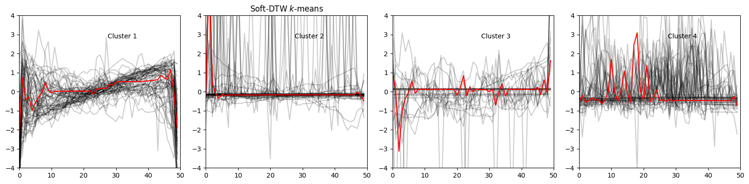
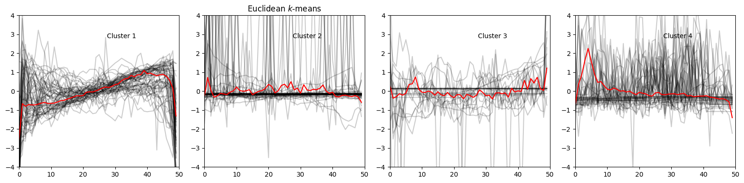


Fig 3 – for 4 clusters

# Discussions

The results highlight the importance of unsupervised learning techniques in driver classification based on driving patterns. Clustering analysis provides valuable insights into the diversity of driving behaviours present in the dataset. Additionally, classification accuracy reflects the ability of the proposed methodology to generalize and classify unseen driver data effectively.

# Conclusion

##### The project successfully clustered drivers based on their behavior using time-series data analysis techniques. Feature engineering, including resampling and standardization, enhanced the dataset's informativeness and facilitated accurate clustering. The insights gained from cluster analysis have practical implications for developing effective road safety measures and personalized driving assistance programs.

##### References

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T.
4. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
5. . Elissa, “Title of paper if known,” unpublished.
6. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
7. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
8. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.