Journal of the Society of Automotive Engineers Malaysia
Volume 9, Issue 1, pp 29-41, January 2025
e-ISSN 2550-2239

Optimization of Driver Behaviour Profiling Using K-means Clustering Algorithm with Environmental Context

M. H. Danial*, Z. Z. Abidin, H. M. Yusof and N. A. Asyqin

Centre for Unmanned Technologies (CUTe), Kulliyah of Engineering, International Islamic University Malaysia (IIUM), Selangor, Malaysia

*Corresponding author: mhariz.hasbullah@live.iium.edu.my

ORIGINAL ARTICLE *Open Access*

Article History:

Received
27 Mar 2025

Accepted
7 Apr 2025

Available online
30 Apr 2025

ABSTRACT – *Advanced Telematic systems have transformed modern vehicles into centralized data collection systems. Contemporary vehicles nowadays are equipped with sensors to collect data and transmit it to the Electronic Control Unit (ECU) of the vehicle. The data gathered from the sensors can provide insight into the driving behaviour and driving patterns. However, the high dimensionality and complexity of the vehicular data hinder its effective utilization for improving road safety; the data requires an advanced method for in-depth analysis. This study applies the K-means clustering algorithm to analyze and categorize driving behaviour patterns, and driving profiling was conducted based on the characteristics of each cluster. To account for external influences, environmental contexts through weather conditions were integrated into the analysis. The findings reveal that K-means effectively identifies six behavioural clusters; two clusters were identified as good driving behaviour, while four clusters exhibit bad driving behaviour. Comparative analysis showed that environmental factors significantly influence driving styles, with more cautious behaviour observed under rainy conditions. The findings highlight the effectiveness of K-means clustering in profiling driving behaviour and emphasize the importance of considering environmental factors for accurate risk assessment. This research provides crucial information to the driver's awareness, gives insight into policymakers and law enforcement, and thus improves road safety.*

KEYWORDS: K-means, telematics systems, driving behavior, clustering algorithms

1. INTRODUCTION

Telematic has emerged as a pivotal technology in the automotive industry, transforming how vehicles operate, communicate, and integrate into broader transportation networks. A Telematic Control Unit (TCU) in a vehicle allows data from a variety of sources within a connected vehicle to be collected (Gekker & Hind, 2020). This technology allows important data to be collected and analyzed, understanding information behind the data collection, potentially improving the safety of the road. The majority of driving accidents are due to human errors; these errors are from reckless and undisciplined driving behaviors and have always been the leading contributors to all sorts of incidents across the globe (Ghaffarpasand et al., 2022). More than 90% of traffic accidents are due to drivers' behavior and their own personal interests. Thus, studying drivers' behavior is essential to reduce traffic accidents (Liu et al., 2020).

A telematic system with an intelligent vehicular telematic platform was proposed, which allows real-time monitoring of vehicular information such as vehicle engine speed, oxygen levels, speed per hour, and water temperature. The concept works by connecting the Controller Area Network (CAN) bus in the vehicle with an On-Board Diagnostic (OBD) bridge construction to receive information. The CAN bus can be used as internal (Local Area Networks) LANs communication in a vehicle, hence outside communication is unable to connect unless OBD is used to directly connect to the Local Area Networks (LANs) communication. OBD provides significant information because of the diagnostic mode to analyze malfunctions in a vehicle (Chen et al., 2016). Most modern cars come with OBD technology, where information about the car can be obtained, such as speed, engine speed, acceleration, and deceleration to be analyzed (Pereira et al., 2016).

A study by Liu et al. (2020) obtains drivers' data from the OBD terminal and combines it with x-axis and y-axis acceleration changes and behavior duration of the vehicle's three-axis acceleration sensor to identify abnormal driving behavior, establishes a hierarchical driving behavior indicator system, and a judgment matrix. It uses threshold standards as references to detect abnormal driving behavior such as rapid acceleration, rapid deceleration, and sharp turns. By analyzing the data from the drivers, bad driving behavior can be detected and raise awareness in the driver. From past studies, different approaches were made to determine driving behavior, such as conducting a data driving simulation or collecting data through GPS to obtain speed and acceleration to evaluate driving behavior (Wu, 2004) Existing methods for detecting aggressive driving often rely on subjective definitions and manually set thresholds based on specific driving parameters (Júnior et al., 2017). For example, the threshold used for driving behavior detection is for safe acceleration, and deceleration lies approximately within a range of ±0.3 g (3m/s2), sudden acceleration and deceleration lie within a range of ±0.5 g (5m/s2), and gradual lane changes produce an average g-force less than ±0.1 g. (approximately 1m/s2), and unsafe lane changes have a g force over ±0.5 g (5m/s2) (Fazeen et al., 2012). Although the threshold method has been widely adopted as a driver profiling system, this approach remains subjective towards the driving environment. A comparison between rule-based and pattern-matching algorithms was made. The results reveal that pattern-matching algorithms give better performance than rule-based algorithms (Saiprasert et al., 2013).

Driving behavior data can be analyzed using algorithms, statistical analysis, deep learning, or machine learning. By analyzing driving behavior, it can be applied to various situations such as road conditions improvement, safety warning systems, and many more (Anil & Anudev, 2022). Driving analytics adaptations improve other aspects of the transportation system, such as increasing overall security and reducing the usage of vehicle energy and gas emissions. Therefore, by exploring the potential of driver analysis, it could further improve safer and energy-efficient driving style (Syed Ahmad et al., 2022). Identification and categorization of driving behavior is necessary because it improves traffic safety, and this application can be used in intelligent transportation systems (Cai, 2024). Most driving behavior analysis approaches use machine learning to automate driving profiling, mainly supervised and unsupervised algorithms. Studies by Bejani & Ghatee (2020) and Ma et al. (2019) demonstrated the usage of machine learning in driving behavior analysis using supervised algorithms. Although supervised algorithms demonstrated high accuracy in controlled environments, it requires manual labelling of data, which is costly, time-consuming, and often subjective towards types of roads (Azadani & Boukerche, 2022). In contrast, unsupervised learning algorithms, particularly clustering techniques, do not require labeled outputs; instead, it relies on the properties of the data itself (Garkalns et al., 2023). Hence, unsupervised clustering algorithms will be used in our driving behavior analysis.

In this research, K-means is used as our algorithm to detect driving behavior categories. K-Means clustering is an unsupervised learning algorithm used to group an unlabeled dataset into clusters (Du et al., 2023). K-means algorithms have shown satisfactory results in recognizing driving patterns and dividing them into clusters without the need for labelling raw data (Bansal & Singhal, 2017). Limitations of the algorithm also need to be considered, as high-dimensional data with low variance may pose a challenge for clustering algorithms such as K-means, leading to lower accuracy, less meaningful clustering, and high computational complexity.

Aggressive driving behavior studies based on Principal Component Analysis (PCA) and K-means clustering show the effectiveness in recognizing different driving patterns without the use of thresholds. The ability to separate clusters in the K-means shows the relevance of the clustering algorithm to recognize driving behavior patterns. However, the driving pattern is subjectively affected by external factors such as traffic environment, vehicle types, and weather conditions; therefore, it requires further analysis on different datasets (Cai, 2024). Taking these external factors into consideration improves the future progress in driver profiling (Azadani & Boukerche, 2022). Therefore, enhancement in using the K-means models across different types of driving pattern datasets needs to be studied.

In this study, an exploration was conducted to study the driving patterns of the drivers. The aim is to use the K-means algorithm for cluster separation, result analysis, and driving profiling categorization with the use of the different datasets under different external conditions, such as weather conditions.

2. METHODOLOGY

2.1 Scope of Study

This study was conducted in an urban location with a normal two-lane road, and the road conditions are safe for driving. Two parameters were maintained throughout the data collection, which is a fixed route during the collection; this ensures enough data points are collected. The second parameter is to use the same type 1 passenger vehicle during the collection. These parameters are crucial to maintain data reliability and reduce variability. After a certain amounts of data are collected, it is inserted into the K-means algorithm to allow clustering and analysis of the data to be done for driving behavior categorization with a number of clusters set to six. Data collection was performed under two distinct environmental scenarios, namely sunny and rainy conditions, to evaluate the influence of external factors on driving behavior. Figure 1 shows the experimental process flowchart used in this study.

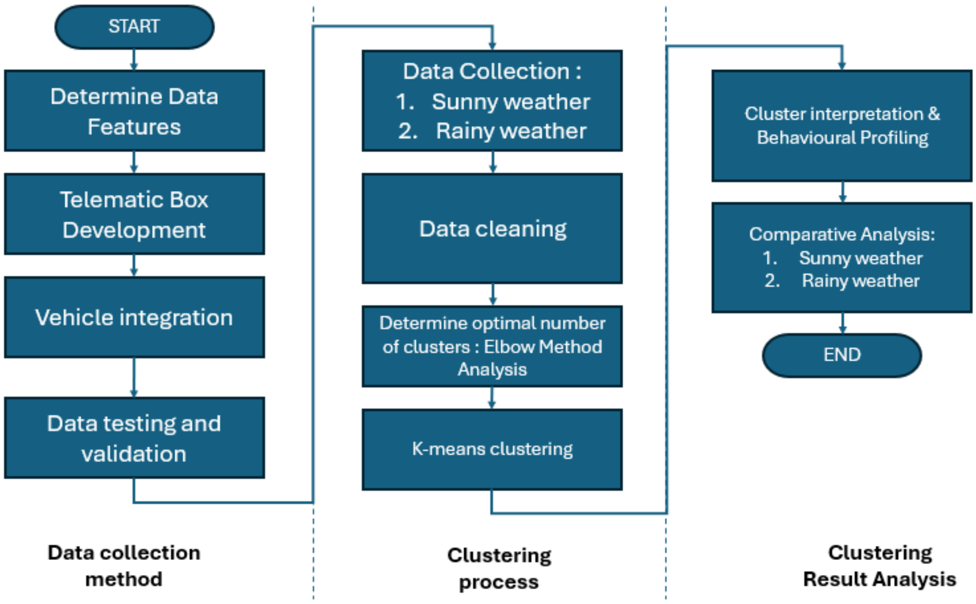

FIGURE 1: Experimental process flowchart

2.2 Experimental Setup

The project aims to collect driving data from drivers, where each driver displays their own driving behavior. A hardware telematic system was developed and implemented in a vehicle for data analysis. The proposed systems helped implement the features and technologies required to achieve our objectives by improving the safety of traffic and vehicles. Various data was collected by the sensors and was sent to the microcontroller of the telematic system to be processed. Then the data was transferred to the K-means algorithm for driving behavior analysis. A fixed route is determined, shown in Figure 2, where the distance of the route is 3.3 km. Figure 2 shows the route that was used during the data collection.

2.3 Data Sources

Data features selected for this study have eight input features: speed (km/h), engine load (%), three-axis acceleration (m/s2), and three-axis angular velocity (°/s). The data for this study was collected using the OBD-II UART board by SparkFun and the MPU-6050 sensor, both of which are essential tools for capturing the necessary data from the vehicle for driving behavior data. A microcontroller, ESP8266, is used to process the data from the sensors and saved to a local machine.

The OBD-II UART Board by SparkFun is a device designed to interface with a vehicle's On-Board Diagnostics (OBD-II) system. It retrieves real-time vehicle parameters required for this study, such as speed and engine load. Speed, measured in kilometers per hour (km/h), provides critical information on the velocity of the vehicle, while engine load indicates the percentage of engine capacity being utilized. These parameters are vital for profiling driving behavior, as they directly reflect how the vehicle is being operated in different conditions. The board sends requests for data by sending the Parameter Identification (PID) to the vehicle computer to request the information. However, PID standards may differ depending on the car manufacturer, therefore further research is needed to identify what PID the vehicle is. Figure 3 shows the SparkFun OBD-II UART.

The MPU-6050 is an acceleration sensor that captures both acceleration and gyroscopic data along the X, Y, and Z axes. The acceleration data format is in "g" which is the gravitational unit where 1 g refers to 9.8 $m/s2$ thus, acceleration data provides insights into linear motion. Meanwhile, gyroscopic data captures rotational movements of the vehicle by giving the data format in degrees. Derived metrics such as acceleration magnitude, jerk, and angular velocity are calculated to better represent driving behavior. These metrics are crucial for detecting sudden changes, such as harsh braking, rapid acceleration, sharp turns, or erratic movements, which are often indicative of aggressive or unsafe driving styles. The sensor's high sensitivity and versatility make it a critical component of the data collection system. Figure 4 shows the MPU-6050 accelerometer sensor.

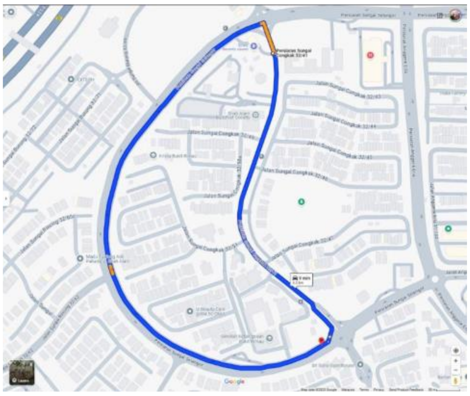

FIGURE 2: Single-lane road route

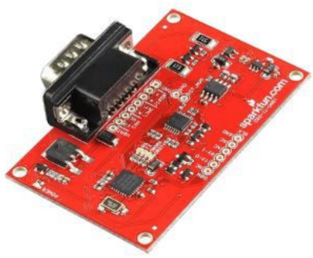

FIGURE 3: SparkFun OBD-II UART

FIGURE 4: MPU-6050 accelerometer sensor

2.4 Hardware Telematic System

The telematic box used for this research is an integrated device that is housed in a compact electrical box containing components necessary to collect data. The components consist of the main microcontroller, which is ESP8266 microcontroller to process data from the OBD UART board and MPU-6050 sensors. ESP8266 translates serial communication from these external modules to readable data and sends data requests to the vehicle's ECU, hence supporting real-time data streaming during vehicle operation. The SparkFun OBD-II UART is used to retrieve vehicle parameters such as speed and engine load through the OBD-II port of the vehicle. This board is necessary as a gateway between the ESP8266 microcontroller and the vehicle's ECU, as certain requirements are needed before requesting the data. Figure 5 shows the telematic box hardware setup.

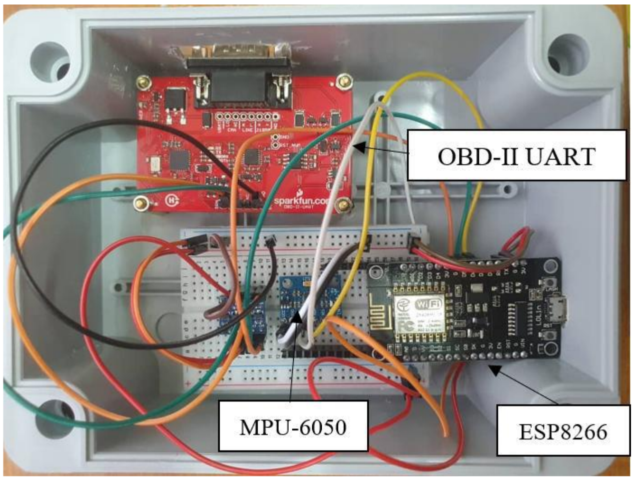

FIGURE 5: Telematic box hardware

2.5 Integration with Vehicle

The telematic box is connected to the vehicle via the OBD-II port and positioned securely within the vehicle cabin. The setup includes vehicle compatibility to retrieve vehicular data through the telematic box. The telematic box is compatible with any vehicle equipped with an OBD-II port, a standard feature in most modern cars. For this study, the test vehicle used was a 2022 Proton Saga MC2; this car model has an OBD-II port and provides reliable OBD-II data access. After integration was successful, data collection was able to proceed. Figure 6 below shows the car model used for the setup and the OBD port location.

FIGURE 6: Vehicle (left), OBD port location (right)

2.6 Clustering Algorithm Process

2.6.1 K-means

After data collection, K-means algorithms were employed to cluster driving behaviors. The K-means clustering algorithm was implemented using Python in the JupyterLab environment. The scikit-learn library was utilized to configure and execute the clustering process.

2.6.2 Elbow Method

The Elbow method is used to determine the optimal number of clusters (k) by evaluating the within-cluster sum of squares (WCSS), also known as inertia (Anil & Anudev, 2022). This metric represents the total squared distance between data points and the centroid of the assigned cluster center for different values of k. As k increases, the WCSS decreases, indicating the clusters are more compact. However, after a certain number of k, the reduction in WCSS will be minor. A WCSS visualization plot figure is constructed against k values to determine at which an 'elbow point' will form to identify the optimum number of clusters.

2.6.3 Principal Component Analysis

Principal Component Analysis (PCA) is used to visualize high-dimensional clustering results in a lower-dimensional space, such as 2-dimensional or 3-dimensional, which are easier to analyze (Li et al., 2023). High-dimensional data used in this study has 8 input features: speed (km/h), engine load (%), three- axis acceleration ($m/s2$), and three-axis angular velocity (°/s), which are fed into the K-means algorithm. As these features reside in an eight-dimensional space, PCA was applied to reduce the dimensionality to two principal components for effective visualization of the cluster distributions.

3. RESULTS AND DISCUSSION

This chapter presents the results of the clustering analysis performed on driving behavior data collected under different weather conditions: Rainy and Sunny. The objective is to profile driving behaviors based on longitudinal acceleration and lateral acceleration features, and to observe behavioral adaptations influenced by weather conditions. K-means clustering was applied separately to each dataset, and cluster characteristics were analyzed to identify bad and good driving patterns. A comparative analysis was conducted to highlight differences between rainy and sunny driving behaviors.

3.1 Elbow Method Analysis

Figure 7 and Figure 8 show the Elbow method analysis applied to the dataset for Rainy and Sunny conditions, respectively. From the figures shown, the number of clusters used in clustering on the x-axis and on the y- axis represents the compactness of clusters (inertia). The lower the inertia difference value, the higher the chance of the elbow point occurring. As seen from the graph, cluster k was tested up to 10 clusters. The slope of the data plot decreased gradually at different rates as clusters increased, eventually forming an elbow-like shape. From clusters 1 to 2, the inertia difference is high, indicating lower compact clusters. As the clusters increase, the decrease in inertia becomes lesser showing that adding more clusters gives diminishing returns. Six clusters were chosen for clustering because beyond cluster 6 does not significantly reduce the inertia and complexity of the model also increased.

3.2 K-means Clustering Algorithm Analysis

3.2.1 Elbow Method Analysis

Figures 9 and 10 show the 2-dimensional (2D) PC, K-means clustering visualization with k = 6 of the Dataset Rainy and Sunny, respectively. The figure shows the clustering result with 6 clusters, whereby each cluster is color-coded to distinguish the separation. Each cluster represents different driving patterns, where K-means effectively separates the different driving patterns for analysis. The clustering result is represented by Principal Component Analysis (PCA) pattern to reduce the high-dimensional feature space into 2 principal components for readable analysis. Cluster 0 is represented by purple, cluster 1 is dark blue, cluster 2 is light blue, cluster 3 is light green, cluster 4 is green, and cluster 5 is yellow. The PCA-based clustering visualization effectively illustrates six different driving patterns. Figures 11 and 12 illustrate the visualization in 3-dimensional (3D) PCA components.

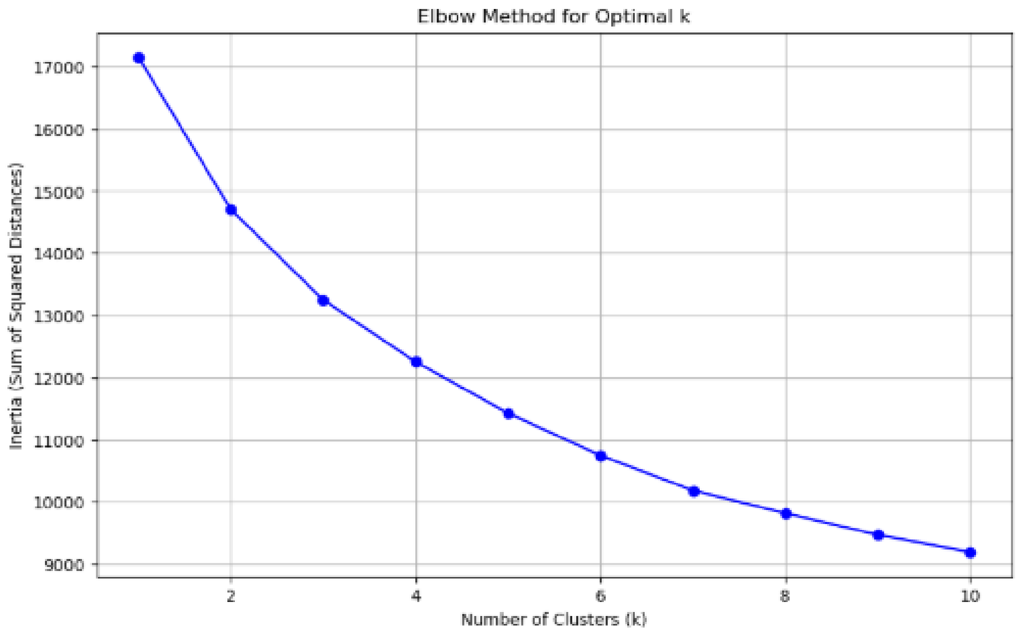

FIGURE 7: Elbow method analysis (Dataset Rainy)

.

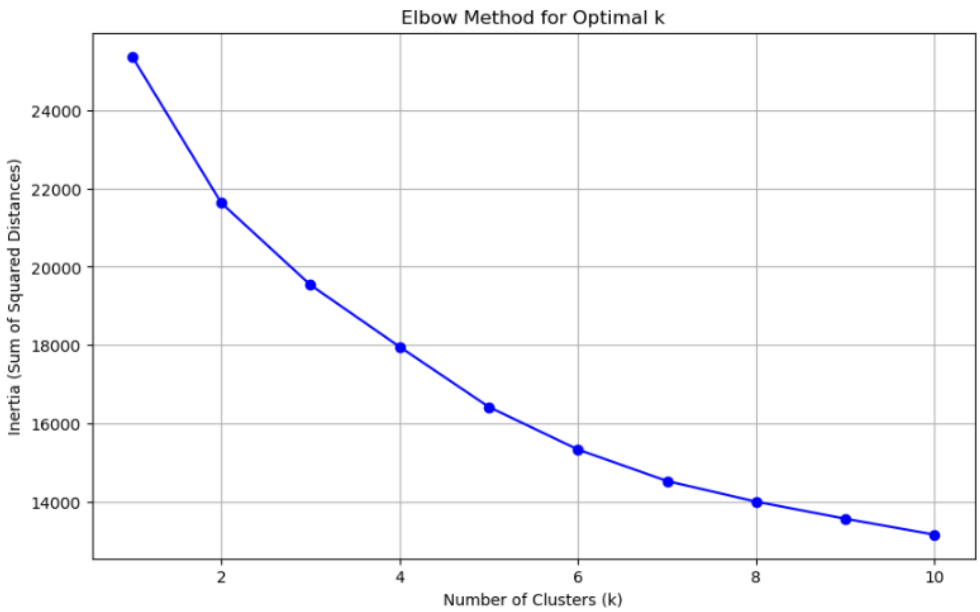

FIGURE 8: Elbow method analysis (Dataset Sunny)

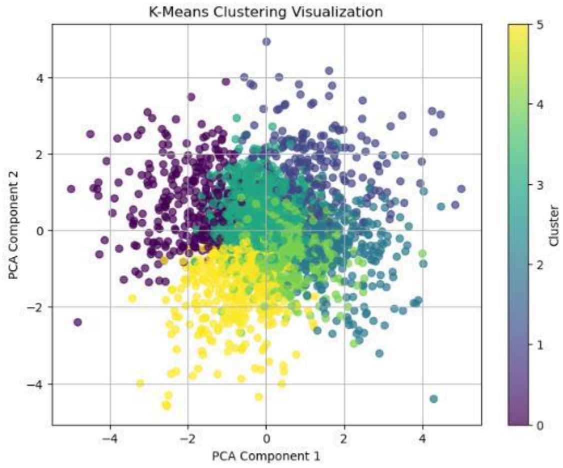

FIGURE 9: 2D PCA visualization clusters (Dataset Rainy)

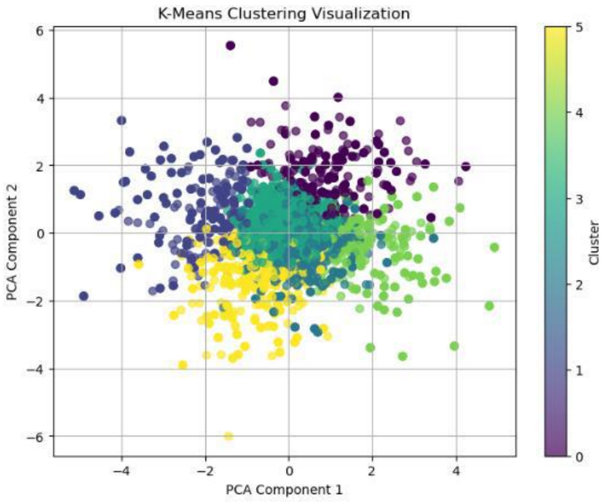

FIGURE 10: 2D PCA visualization clusters (Dataset Sunny)

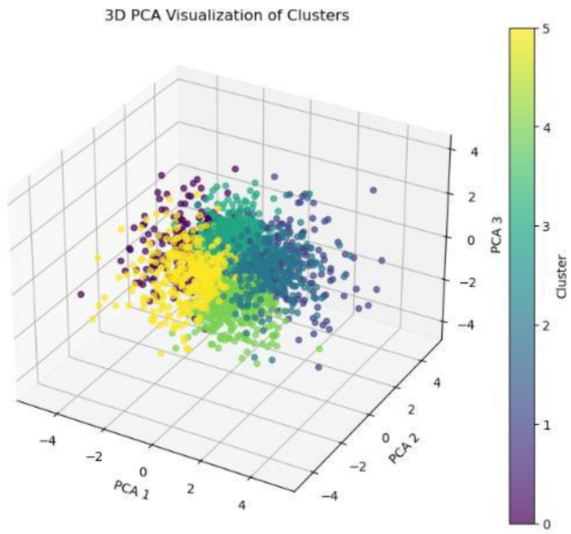

FIGURE 11: 3D PCA visualization clusters (Dataset Rainy)

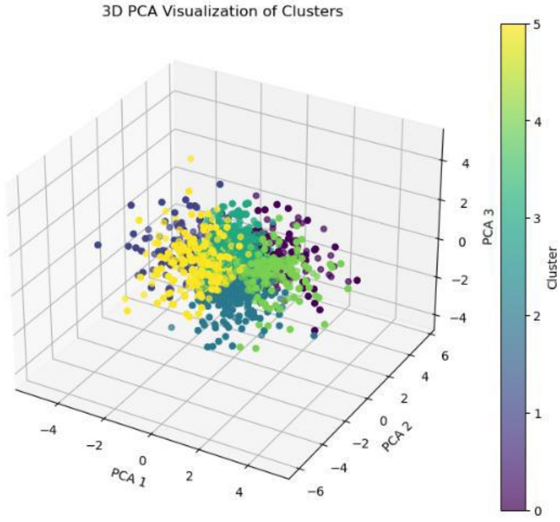

FIGURE 12: 3D PCA visualization clusters (Dataset Sunny)

3.2.2 Driver Behaviour Profiling

The dataset of sunny and rainy conditions was clustered into six groups using the K-means algorithm. Tables 1 and 2 summarize the characteristics of each cluster based on longitudinal acceleration (AccX) and lateral acceleration (AccY), along with behavioral descriptions and risk categorizations. The acceleration data reveals distinct driving behavior patterns categorized into good and bad driving behavior, along with a description of driving patterns. The range of AccX and AccY values for each cluster further supports these classifications, as clusters associated with bad driving exhibit wider acceleration variations, whereas good drivers maintain a more stable and controlled range.

TABLE 1: Driver profile (Dataset Rainy)

K = 6	AccX (m/s^2)	AccY (m/s^2)	Description	Category
Cluster 1	(-1.385) to 2.980	(-2.735) to 2.458	**Harsh** acceleration and **Smooth** braking, **Stable** lane changes	Bad driver
Cluster 2	(-2.378) to 2.383	(-2.337) to 2.403	**Moderate** acceleration and breaking, **Stable** lane changes	Good driver
Cluster 3	(-2.772) to 1.857	(-2.475) to 2.393	**Harsh** braking and **Smooth** acceleration, **Stable** lane changes	Bad driver
Cluster 4	(-1.692) to 1.918	(-0.499) to 2.521	**Smooth** acceleration and braking, **Stable** lane changes	Good driver
Cluster 5	(-2.592) to 2.667	(-2.824) to 0.324	**Moderate** acceleration and braking, **Sharp** lane changes	Bad driver
Cluster 6	(-2.627) to 2.018	(-2.682) to 2.271	**Harsh** braking and **Smooth** acceleration, **Sharp l**ane changes	Bad driver

In Table 1, Cluster 1 shows the AccX values range from -1.385 to 2.980 m/s2, showing the highest magnitude of acceleration, but smooth braking (deceleration), for AccY lane changes range from -2.735 to 2.458 m/s2, showing stable lane changes. In cluster 2, it shows AccX value of -2.378 to 2.383 m/s2, indicating a moderate magnitude value of acceleration and braking as the value is lesser than cluster 1 for acceleration and cluster 3 for braking, for AccY lane changes range from -2.337 to 2.403 m/s2, showing stable lane changes. In cluster 3, the AccX value ranges from -2.772 to 1.857 m/s2, showing the highest magnitude of braking but smooth acceleration, for AccY ranges from -2.475 to 2.393 m/s2, indicating stable lane changes. In cluster 4, the AccX value ranges from -1.692 to 1.918 m/s2, showing smooth acceleration and braking, and the AccY value ranges from -0.499 to 2.521 m/s2, indicating stable lane changes. For cluster 5, AccX shows a value range -2.592 to 2.667 m/s2, showing moderate acceleration and deceleration, for AccY value range -2.824 to 0.324 m/s2, indicating sharp lane changes. Finally, for cluster 6, AccX ranges from -2.627 to 2.018 m/s2, showing moderate acceleration and braking, for AccY ranges from -2.682 to 2.271 m/s2, showing sharp lane changes.

TABLE 2: Driver profile (Dataset Sunny)

K = 6	AccX (m/s^2)	AccY (m/s^2)	Description	Category
Cluster 1	(-2.543) to 3.173	(-2.310) to 2.403	**Moderate** acceleration and braking, **Stable** lane changes	Good driver
Cluster 2	(-1.385) to 3.768	(-2.421) to 2.857	**Harsh** acceleration and **Smooth** braking, **Stable** lane changes	Bad driver
Cluster 3	(-3.260) to 1.795	(-3.821) to (-0.050)	**Harsh** braking and **smooth** acceleration, **Sharp** lane changes	Bad driver
Cluster 4	(-1.870) to 3.119	(-0.891) to 2.862	**Moderate** acceleration and braking, **Stable** lane changes	Good driver
Cluster 5	(-3.396) to 1.857	(-2.541) to 2.392	**Smooth** acceleration and **Harsh** braking, **Stable** lane changes	Bad driver
Cluster 6	(-3.148) to 2.726	(-2.970) to 3.536	**Harsh** braking and **Smooth** acceleration, **Sharp** lane changes	Bad driver

In Table 2, Cluster 1 shows the AccX values range from (-2.543) to 3.173 m/s2, showing a moderate magnitude of acceleration and braking (deceleration), for AccY lane changes range from (-2.310) to 2.403m/s2, showing stable lane changes. In cluster 2, it shows AccX value of (-1.385) to 3.768 m/s2 indicating the highest magnitude of acceleration but has smooth braking. AccY lane changes range from (-2.421) to 2.857 m/s2, showing stable lane changes. In cluster 3, the AccX value ranges from (-3.260) to 1.795 m/s2, showing harsh braking but smooth acceleration. For AccY, it ranges from (-3.821) to (-0.050) m/s2, indicating sharp lane changes. In cluster 4, the AccX value ranges from (-1.870) to 3.119 m/s2, showing moderate acceleration and braking, and the AccY value ranges of (-0.891) to 2.862 m/s2, indicating stable lane changes. For cluster 5, AccX shows a value range (-3.396) to 1.857 m/s2showing smooth acceleration and harsh braking, for AccY value range (-2.541) to 2.392 m/s2indicating stable lane changes. Finally, for cluster 6, AccX ranges from (-3.148) to 2.726 m/s2, showing smooth acceleration but harsh braking, for AccY range of (-2.970) to 3.536 m/s2, showing sharp lane changes.

By observation of the range value in Tables 1 and 2. The highest and the lowest magnitude of acceleration data AccX and AccY indicate a threshold for driving pattern evaluation. In the Dataset Rainy Condition, Clusters 1, 3, 5, and 6, are categorized as bad drivers because it exhibits the driving pattern of harsh acceleration, harsh braking, and sharp lane changes. Meanwhile, clusters 2 and 4 are categorized as good drivers as they show more stable changes and are below the magnitude threshold value. In the Dataset Sunny Condition, Clusters 2, 3, 5, and 6 are categorized as bad drivers, while Clusters 1 and 4 exhibit good driver behavior patterns.

3.3 Comparative Analysis between the Datasets Rainy and Sunny Conditions

Table 3 summarizes the comparative analysis between the rainy and sunny condition datasets, which revealed distinct behavioural adaptations in response to environmental factors. Under rainy conditions, drivers predominantly exhibited cautious behaviour, characterized by lower longitudinal accelerations (AccX) and reduced lateral movements (AccY). Most clusters identified during rain demonstrated smooth acceleration and braking with stable lane changes, reflecting a general tendency to minimize abrupt maneuvers to maintain vehicle stability on slippery road surfaces. High-risk behaviors under rain were limited and primarily associated with occasional sharp lane changes, as evidenced in Clusters 5 and 6.

TABLE 3: Comparative analysis of driving behavior

Observation	Sunny Conditions	Rainy Conditions
Longitudinal Acceleration (AccX)	Higher acceleration (3.768 m/s^2) and deceleration magnitude (-3.396 m/s^2)	Lower, smooth, and cautious acceleration (2.98 m/s^2) and deceleration (-2.772 m/s^2).
Lateral Acceleration (AccY)	Higher magnitude of lane changes of (-3.82) m/s^2 and 3.536 m/s^2	Lower magnitude of lane changes of (-2.824) m/s^2 and 2.521 m/s^2
Driving Behavioral Adaptation	Increased in more aggressive driving style and maneuvers. Exhibit more confinement and a dynamic driving style.	Fewer aggressive maneuvers and cautions driving style. Exhibit a more cautious and controlled driving style.

In contrast, the sunny condition dataset captured a broader and more dynamic spectrum of driving behaviors. Drivers exhibited significantly higher mean longitudinal accelerations, with some clusters (notably Cluster 1) recording mean AccX values exceeding 1.7 m/s^2, indicative of aggressive acceleration practices. Moreover, lateral movements were more pronounced, particularly in Cluster 5, where mean AccY exceeded 1.3 m/s^2, suggesting sharp lane changes or aggressive turning maneuvers. These patterns reflect increased driver confidence and a reduced perception of risk under favorable weather conditions.

The shift from predominantly low-risk, cautious behaviors in rain to more aggressive and dynamic behaviors in sunny conditions underscores the influence of environmental context on driver decision-making. Dry road conditions appear to encourage higher speeds, greater acceleration, and sharper steering actions, whereas adverse weather fosters more conservative and controlled driving patterns. These findings emphasize the necessity for adaptive driver assistance systems that account for weather-induced behavioural variability to enhance road safety outcomes.

It is important to note that the same magnitude of acceleration does not inherently carry the same behavioral implications across different environmental conditions. For example, an acceleration of approximately 3 m/s^2 may be considered an acceptable and controlled behavior under sunny conditions due to the high friction and stability provided by dry roads. However, under rainy conditions, where road surfaces are significantly more slippery and vehicle grip is compromised, the same level of acceleration could substantially increase the risk of vehicle instability, hydroplaning, or loss of control. Therefore, driving behaviors must be interpreted in context, and acceleration thresholds should be assessed relative to prevailing weather and road surface conditions. This highlights the necessity for dynamic driver profiling models that adapt behavioral risk assessments based on environmental factors.

4. CONCLUSION

This study successfully applied K-means clustering to analyze and classify driving behaviors based on driving data of longitudinal acceleration (AccX) and lateral acceleration (AccY). The data was captured on a type 1 vehicle and on a fixed route. This is to ensure data has stable variability. The clustering results revealed six distinct driver profiles, where in Dataset Rainy conditions, Clusters 2 and 4 were identified as good drivers with smooth acceleration, braking, and stable lane changes, while Clusters 1, 3, 5, and 6 exhibited aggressive tendencies such as harsh acceleration, harsh braking, and sharp lane changes, categorizing them as bad drivers. In Dataset Sunny conditions, Clusters 2, 3, 5, and 6 are categorized as bad drivers while Clusters 1 and 4 exhibit good driver behavior patterns. The acceleration variations in bad driver clusters showed significantly wider ranges compared to good driver clusters, supporting the effectiveness of clustering in differentiating between safe and risky driving behaviors. Additionally, different driving datasets with environmental factors (weather conditions) were analyzed using K-means clustering, revealing the shift of driving behavior style and the magnitude of maneuver. These findings highlight the effectiveness of K-means clustering in differentiating between good and bad driving behaviors under different external factors, such as weather conditions, providing valuable insights for applications in road safety analysis, driver monitoring systems, and insurance risk assessments. These findings emphasize the potential of unsupervised machine learning in driver profiling, which can be leveraged for applications such as driver monitoring systems, insurance risk assessments, and intelligent transportation safety programs. However, this study's limitation is the data quality. The dataset can be further improved with higher variations and more data points; this would improve cluster separability and more meaningful behavioral distinctions. Future research can explore real-time clustering applications, multi-sensor data integration, external factor datasets, and adaptive learning models to further enhance driver behavior classification accuracy and robustness.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their detailed comments, which have improved the quality of this paper. The main author is supported by the project of International Islamic University Malaysia (IIUM) with Collaborative Research in Engineering, Science & Technology Center (CREST) and Delloyd R&D (M) Sdn. Bhd. This work is under grant funding (grant ID CREST24-001-0002).

REFERENCES

Anil, A. R., & Anudev, J. (2022). Driver behavior analysis using K-means algorithm. Proc. of the 2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies: Computational Intelligence for Smart Systems, ICICICT 2022, 1555–1559.

Azadani, M. N., & Boukerche, A. (2022). Driving behavior analysis guidelines for intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 23(7), 6027–6045.

Bansal, A., & Singhal, A. (2017). A survey on driver behaviour detection techniques for intelligent transportation systems. IEEE.

Bejani, M. M., & Ghatee, M. (2020). Convolutional Neural Network with adaptive regularization to classify driving styles on smartphones. IEEE Transactions on Intelligent Transportation Systems, 21(2), 543–552.

Cai, G. (2024). Aggressive driving behavior recognition based on K-means clustering. Scholarly Review Journal SR Online: Showcase (Winter 2024/2025).

Chen, L. B., Li, H. Y., Chang, W. J., Tang, J. J., & Li, K. S. M. (2016). An intelligent vehicular telematics platform for vehicle driving safety supporting system. 2015 International Conference on Connected Vehicles and Expo, ICCVE 2015 - Proceedings, 210–211.

Du, X., Kang, X., Gao, Y., & Wang, X. (2023). Driving behavior characterization and traffic emission analysis considering the vehicle trajectory. Frontiers in Psychology, 14.

Fazeen, M., Gozick, B., Dantu, R., Bhukhiya, M., & González, M. C. (2012). Safe driving using mobile phones. IEEE Transactions on Intelligent Transportation Systems, 13(3), 1462-1468.

Garkalns, P., Nikiforova, O., Zabiniako, V., & Kornienko, J. (2023). Analysis of the behavior of company employees as users of information systems or tools, based on employees clustering with K-means algorithm. 64th International Scientific Conference on Information Technology and Management Science of Riga Technical University, ITMS 2023 - Proceedings.

Gekker, A., & Hind, S. (2020). Infrastructural surveillance. New Media and Society, 22(8), 1414–1436.

Ghaffarpasand, O., Burke, M., Osei, L. K., Ursell, H., Chapman, S., & Pope, F. D. (2022). Vehicle telematics for safer, cleaner and more sustainable urban transport: A review. Sustainability, 14(24), 16386.

Júnior, J. F., Carvalho, E., Ferreira, B. V., De Souza, C., Suhara, Y., Pentland, A., & Pessin, G. (2017). Driver behavior profiling: An investigation with different smartphone sensors and machine learning. PLoS ONE, 12(4).

Li, Y., Dong, E., Zhang, Y., Xu, S., & Zan, J. (2023). Driving style clustering study based on HighD dataset. 2023 5th International Conference on Artificial Intelligence and Computer Applications, ICAICA 2023, 380–383.

Liu, T., Yang, G., & Shi, D. (2020). Construction of driving behavior scoring model based on OBD Terminal Data Analysis. Proceedings - 2020 5th International Conference on Information Science, Computer Technology and Transportation, ISCTT 2020, 24–27.

Ma, Y., Li, Z., Li, Y., Li, H., & Malekian, R. (2019). Driving style estimation by fusing multiple driving behaviors: A case study of freeway in China. Cluster Computing, 22, 8259-8269.

Pereira, A., Alves, M., & Macedo, H. (2016). Vehicle driving analysis in regards to fuel consumption using Fuzzy Logic and OBD-II devices. In 2016 8th Euro American conference on telematics and information systems (EATIS) (pp. 1-4). IEEE.

Saiprasert, C., Pholprasit, T., & Pattara-Atikom, W. (2013). Detecting driving events using smartphone. In Proceedings of the 20th ITS World Congress (Vol. 11, No. 11).

Syed Ahmad, S. S., Muhammad, M., & Jawi, Z. M. (2022). Driving analytics - Data science approach based on smartphone vehicle telematic data. Journal of the Society of Automotive Engineers Malaysia, 6(2), 77–84.

Wu, J. (2004). Analysis of taxi drivers' driving behavior based on a driving simulator experiment. Electronic Theses and Dissertations. 4587. University of Central Florida.