Clustering Algorithms to Analyse Smart City Traffic Data

Praveena Kumari M K[1], Manjaiah D H[2], Ashwini K M[3]*
MCA Dept. NMAM Institute of Technology, Nitte (Deemed to be University), Nitte, India[1, 3]
Department of Computer Science, Mangalore University, Mangalore, India[2]

***Abstract*—Urban transportation systems encounter significant challenges in extracting meaningful traffic patterns from extensive historical datasets, a critical aspect of smart city initiatives. This paper addresses the challenge of analyzing and understanding these patterns by employing various clustering techniques on hourly urban traffic flow data. The principal aim is to develop a model that can effectively analyze temporal patterns in urban traffic, uncovering underlying trends and factors influencing traffic flow, which are essential for optimizing smart city infrastructure. To achieve this, we applied DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture clustering techniques to the traffic dataset of Aarhus, Denmark's second-largest city. The performance of these clustering methods was evaluated using the Silhouette Score and Dunn Index, with DBSCAN emerging as the most effective algorithm in terms of cluster quality and computational efficiency. The study also compares the training times of the algorithms, revealing that DBSCAN, K-Means, and Gaussian Mixture offer faster training times, while Affinity Propagation and Mean Shift are more computationally intensive. The results demonstrate that DBSCAN not only provides superior clustering performance but also operates efficiently, making it an ideal choice for analyzing urban traffic patterns in large datasets. This research emphasizes the importance of selecting appropriate clustering techniques for effective traffic analysis and management within smart city frameworks, thereby contributing to more efficient urban planning and infrastructure development.**

Keywords*—*Clustering; smart city; traffic; analyze

I. Introduction

In the current era of rapid urbanization and technological advancement, cities are increasingly adopting "smart city" initiatives aimed at improving the quality of urban life through data-driven decision-making. A smart city leverages technology, particularly networked sensors and data analytics, to optimize urban services, including traffic management. The massive influx of data generated from various sources—such as government records, online platforms, and IoT devices—is a vital resource for transforming urban environments into innovation hubs. However, this wealth of data presents challenges in extraction, analysis, and application, particularly in traffic management, where traditional methods struggle to keep pace with the growing complexity.

Transportation systems are essential to economic stability and social development, yet they also contribute to urban challenges such as congestion and air pollution. Effective traffic management, a cornerstone of smart city initiatives, requires a deep understanding of traffic patterns and behaviors. Data mining techniques, particularly clustering algorithms, offer powerful tools for uncovering these patterns, facilitating traffic forecasting, and enabling informed decision-making.

This paper focuses on applying clustering algorithms to analyze urban traffic data, specifically within the context of smart city development. By employing techniques such as DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture, we aim to uncover meaningful insights into traffic flow patterns and identify factors influencing congestion. The chosen methods are evaluated on their clustering performance and computational efficiency, as these factors are crucial in real-time traffic management systems.

Previous research, such as the work by Pattanaik, Singh, Gupta, and Singh (2016) [1] proposed a real-time traffic congestion estimation system for urban roads, incorporating K-Means clustering and other clustering techniques. Their approach involves collecting and analyzing real-time traffic data to classify and estimate congestion levels accurately. The system dynamically categorizes traffic patterns to provide timely insights and support effective traffic management. This research highlights the practical application of clustering methods in enhancing urban traffic control and decision-making.

The remainder of this paper is structured as follows: Section II provides a literature review, identifying the state-of-the-art in traffic analysis using clustering techniques. Section III outlines the methodology, detailing the data collection process and the specific algorithms used. Section IV provides the performance measures used to analyze the clustering algorithms Section V presents the results and discussion, comparing the performance of the clustering methods. Finally, Section VI concludes with a summary of findings, implications for smart city initiatives, and suggestions for future research.

II. Literature Review

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file “MSW_A4_format”. The management of urban traffic is a critical component of smart city initiatives, where data-driven strategies are employed to enhance the efficiency and sustainability of urban transportation systems. Clustering algorithms have gained significant attention for their ability to analyze and interpret large-scale traffic data. This section reviews key studies that have applied clustering techniques to traffic data, highlighting the methods used, the contexts in which they were applied, and the findings that contribute to the understanding of urban traffic management.

A. Clustering Techniques in Traffic Analysis

Clustering, a form of unsupervised learning, has been widely utilized to uncover patterns in traffic data. Various algorithms, such as K-Means, DBSCAN, and Gaussian Mixture Models, have been applied in different contexts to group similar traffic patterns and identify congestion hotspots.

Rouky et al. (2024) [2] investigated traffic congestion in Casablanca using K-Means and DBSCAN clustering methods. Their study focused on identifying congestion patterns and hotspots within the city, contributing to improved traffic management strategies. The findings indicated that clustering could effectively segment traffic data into meaningful clusters, providing insights into daily and seasonal traffic variations. Similarly, Wang et al. (2016) [3] discovered that vehicular traffic levels on metropolitan highways in Shanghai differ dramatically during times of day and night. They discovered that at peak hours, vehicle numbers increase and speeds decrease owing to bottlenecks. Off-peak hours, on the other hand, see a reduction in the amount of traffic and a rise in speed, resulting in a smoother flow of traffic. Their research emphasizes the significance of comprehending these differences for successful traffic control and urban planning in dynamic metropolis contexts such as Shanghai.

Shi et al. (2021) [4] used a density-based moving object clustering technique to determine the spatiotemporal extends of traffic jams. Their approach combines density-based clustering techniques with moving object data to precisely identify crowded zones and their lengths of stay. This technique improves the exactness of bottleneck identification throughout time frames, providing vital knowledge for improving metropolitan traffic control and planning initiatives.

Yang et al. (2017) [5] investigated transportation state fluctuation trends in urban roadways using spectral clustering. Their work used spectral clustering methods to categorize and analyse various traffic situations using spatiotemporal data. The results demonstrated the usefulness of spectral clustering various traffic patterns, such as bottleneck patterns and flow variances across urban regions. This technique offers important conclusions for enhancing metropolitan traffic management techniques and infrastructure design by effectively recording and analyzing complicated traffic behaviors.

Angmo et al. (2021) [6] proposed an enhanced clustering approach that integrates density-based and hierarchical clustering algorithms. Their methodology, which combines spatiotemporal data with advanced clustering techniques, demonstrated increased accuracy and efficiency in identifying relevant locations for traffic management. This study emphasizes the importance of adapting traditional clustering methods to the specific requirements of urban traffic analysis.

Asadi and Regan (2019) [7] presented a method for spatiotemporal clustering of data on traffic through deep-embedded clustering. Their method uses deep learning techniques to embed traffic data in a latent space, allowing for the finding of significant clusters based on geographical and temporal trends. The study sought to increase traffic data analysis and forecast accuracy by utilizing deep-embedded clustering algorithms. This methodological development offers the potential for improving our knowledge about city transportation habits and optimizing roadway safety tactics.

Sfyridis and Agnolucci (2020) [8] created a technique for predicting Annual Average Daily Traffic (AADT) in England and Wales. They used clustering and regression modeling to increase the correctness of AADT predictions. Clustering methods were employed to aggregate road segments that have similar traffic patterns, models and regression were utilised inside these clusters to calculate traffic numbers. The results showed that this comprehensive strategy improved the accuracy of traffic estimates, providing helpful information for planning transportation and construction of infrastructure.

Wang et al. (2020) [9] devised a technique for identifying hotspots in travel security and safety with an emphasis on the regular fluctuation of the flow of traffic and crash data. Their method combines statistical evaluation and geographical information systems (GIS) to detect high-risk regions for vehicular crashes. By taking into account daily traffic changes and collision data, the study improves hotspot identification accuracy, offering useful information for adopting targeted safety measures and enhancing transportation infrastructure.

Taamneh et al. (2017) [10] published research on categorizing road crashes using clustering-based algorithms. They suggested a hybrid technique that combined hierarchical clustering with artificial neural networks (ANNs). Hierarchical clustering was utilised to divide accident data into clusters with comparable features and ANNs were used to categorise and forecast the severity of accidents within each cluster. The findings showed that this combination strategy increased the accuracy of accident categorization and gave useful information for improving road safety measures and tactics.

Acun and Gol (2021) [11] discovered that data levels on large-scale traffic networks follow different patterns, which may be efficiently analyzed by employing ARIMA and K-means clustering. By using K-means to combine comparable traffic flow patterns and ARIMA models inside these clusters for prediction, they discovered that this hybrid strategy considerably increases traffic volume estimates. Their findings imply that taking into account both temporal trends and geographical clustering results in more accurate traffic management and prediction on large road systems.

Zou, X., & Chung, E. (2024) [12] proposed a novel traffic prediction approach that combines clustering and deep transfer learning to address the challenge of limited data availability. The study utilizes clustering techniques to group similar traffic patterns and then applies deep transfer learning to enhance prediction accuracy across different datasets. This method significantly improves traffic prediction performance, particularly in scenarios where data is sparse, demonstrating its potential in urban traffic management.

Nguyen, T. T., et al. (2019) [13] proposed a method for feature extraction and clustering analysis to study highway congestion. By utilizing clustering techniques, the study identifies and classifies different congestion patterns based on key traffic features extracted from highway data. The approach focuses on understanding the underlying causes of congestion and its temporal variations. The findings provide valuable insights for designing more effective traffic management strategies on highways, helping to mitigate congestion and improve traffic flow efficiency. The research highlights the importance of data-driven approaches in addressing complex transportation challenges.

B. Smart Cities and Traffic Management

The concept of smart cities involves the use of information and communication technologies (ICT) to optimize urban services, including transportation. Traffic management in smart cities relies heavily on the ability to analyze vast amounts of data generated by sensors, GPS devices, and social media platforms.

Xu et al. (2020) [12] suggested an approach for anticipating traffic jams in Shanghai based on multiperiod hotspot clustering. To forecast bottleneck trends, they use hotspot clustering methods applied across different periods. The work improves overcrowding forecasting accuracy by analyzing past traffic data and using clustering algorithms. This methodological innovation promotes proactive traffic management tactics, hence improving public transportation and the construction of infrastructure in Shanghai.

C. Comparative Analysis of Clustering Algorithms

Several studies have conducted comparative analyses of clustering algorithms to determine their effectiveness in various traffic management scenarios.

Chen et al. (2022) [14] created an automobile flow forecasting system employing a Graph Attention Network (GAT) and spatial-temporal clustering. Their technique makes use of GATs to capture geographical and temporal connections in traffic data, which improves traffic flow prediction precision. The investigation also uses spatial-temporal clustering to organize traffic patterns, which helps manage complicated transportation patterns. The findings show that this combination of approaches increases the accuracy of traffic flow estimates, resulting in improved administration and planning in smart transportation networks.

D. Gaps and Future Directions

While the literature demonstrates the effectiveness of clustering techniques in traffic analysis, several gaps remain. Most studies have focused on traditional algorithms, with limited exploration of novel methods that could offer improved performance or new insights. Moreover, there is a need for more comprehensive evaluations of these techniques within the smart city framework, considering factors such as scalability, adaptability to dynamic data, and integration with other smart city systems.

Additionally, the literature often lacks a critical comparison of how these clustering methods perform in different urban environments, which can vary significantly in terms of traffic patterns and data availability. Future research should focus on developing and testing new algorithms that can address these challenges and contribute to more robust and adaptive traffic management solutions in smart cities.

III. Study Area and Methodology

A. Dataset Description

This research used traffic information from the city of Aarhus to analyze traffic habits in an urban region. Aarhus is Denmark's second-biggest city and main cultural center. Due to its northern latitude, Aarhus experiences significant variations in daylight hours between summer and winter. Various clustering approaches are used to examine data from February 2014 to June 2014 to identify transportation trends. Sensors were set at two nearby places to collect data, tallying the amount of automobiles moving by during a certain period. Every route's report contains a traffic count taken by the sensor on a certain day and time. Table I describes the attributes in the dataset.

TABLE I. Attributes in the Dataset

Attributes	Description
avgMeasuredTime	Shows the duration in seconds of the sensor collecting data.
avgSpeed	This signifies the mean pace of the cars moving within the reported period.
extID	Signifies the distinct identification allocated for every route.
medianMeasuredTime	Like avgMeasuredTime
TIMESTAMP	Represents the beginning time measurement of transportation on a route.
vehicleCount	Quantity of cars traveling across each pair of observations.
_id	A distinctive number, issued for every traffic estimate.
REPORT_ID	Denotes the distinct number of every observation point.

B. Data Preprocessing

The traffic data is processed to enable traffic pattern analytics. Routes with inadequate data are populated using filling missing values techniques. Preprocessing guarantees that the study's final data include comprehensive traffic counts for each road on a given day. Following numerous evaluations of quality, all processed records are grouped into 1-hour intervals. For transportation statistics, we just require an automobile count at periods throughout the day. Therefore, data such as the extID, avgSpeed, avgMeasuredTime and medianMeasuredTime, and other particulars are unneeded. We have trimmed our data collection to contain just the fields important for extracting patterns. Each entry for a certain roadway now has only two fields: date and automobile frequency.

A few rows in the dataset lack accurate timestamps or interval bounds. A defective timestamp indicates that what was observed has changed in time. To resolve this, automobile counts are proportionally adjusted for inadequate timestamps or intervals. The adjusted values are subsequently used for a variety of performance assessments. To eliminate rows with wrong values, the following types of verification procedures are used:

1) Poor-quality detectors may record excessively elevated transport volumes, affecting approximation. Rows with traffic counts exceeding 120 are eliminated from data collection.

2) If nothing is achieved for three hours consecutively in a day, the day is declared worthless. If over thirty percent of the available days for an area are unreliable, it is removed from further investigation.

C. Models

Clustering methods are used when there are not any classes to forecast but instead, instances need to be classified into natural groupings. For the investigation of streets, we employed the DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixtures method to cluster days of the week. Each method takes an alternate strategy to the problem of identifying natural groupings in data. We selected 372 streets and clustered them over all days in the dataset. Fig. 1 shows the data source collection location. Fig. 2 is the proposed methodology of the suggested approach to analyze the traffic pattern.

Fig. 1. Data source location.

DBSCAN (Density-Based Clustering of Applications with Noise) is a clustering technique that detects clusters and outliers in a dataset using density. The algorithm has two parameters: epsilon(ε), which determines the radius for neighborhood searches, and MinPts, which specifies the minimal of points needed to build dense regions. Core points have at least MinPts neighbours within ε, whereas border points are close to a core point but do not have enough neighbors to be core points themselves. The noise points do not belong to any cluster.

DBSCAN performs well for arbitrary clusters and can withstand noise, but its efficiency is significantly dependent on its selection of ε and MinPts.

K-Means is a common clustering technique that divides a dataset into K separate, not interconnected groups. The technique initiates K centroids and allocates every data point to the closest centroid, resulting in K clusters. It then continuously updates the centroids by taking the average of every point in each cluster and reassigning them depending on the new centroids. This procedure repeats until the centroids stop changing appreciably. K-Means is effective and easy to construct, but the number of clusters (K) must be identified in advance. It works with spherical collections but struggles with clusters of different shapes and densities.

Affinity Propagation is a clustering approach that detects exemplars (representative points) inside data by delivering messages between data points. Unlike, K-Means, the total number of groups does not have to be preset. The algorithm operates by repeatedly transferring two sorts of messages: “responsibility”, which shows how well-suited the data point is to serve as an exemplar for another point, and “availability” which represents a point’s appropriateness for assignment to an exemplar. These notifications will be maintained until convergence. Affinity propagation may detect clusters of variable sizes and does not presume an ideal structure for the clusters, making it a useful tool for a variety of grouping jobs. However, it can be extremely costly for huge datasets.

Mean Shift is a nonparametric clustering approach that seeks to identify groups by repeatedly moving data points to the densest region of the feature space. The technique begins by constructing a window (or kernel) around each data point and then computes the average of the points within that window. Each data point is then relocated to the mean, and the procedure is continued until convergence, which causes points to cluster around the density function’s modes (peaks). Mean Shift, unlike K-Means, doesn’t need a pre-specified amount of clusters and may discover clusters of any form. However, its performance is significantly reliant on the selection of the bandwidth option, which controls the window size.

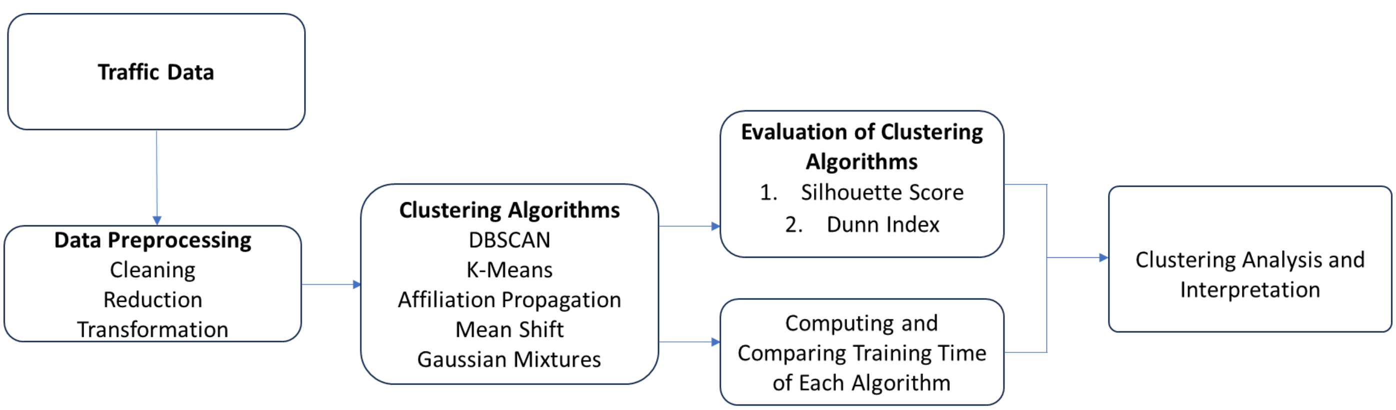

Fig. 2. Methodology of the suggested approach.

Gaussian Mixture Models (GMM) are probabilistic clustering methods that presume data is created by a combination of multiple Gaussian distributions with unidentified variables. Each Gaussian component has a mean and a covariance, and the model is described by the components’ mixture weights, means, and covariances. GMMs employ the

Expectation-Maximization (EM) method to continuously modify these variables to maximize the probability of the data being observed. Unlike K-Means, GMMs may detect the elliptical form of clusters and give soft categorization, which means that each data point has a chance to be assigned to each group. This makes GMMs more adaptable and stronger when modelling real-world data with different cluster shapes and densities. However, they can be sensitive to startup and may require careful adjustment.

IV. Study Area and Methodology

To determine which clustering approach is optimal for our model, we apply the following validity indices.

1) *Silhouette score*
2) *Dunn index*

A. *The Silhouette Score*

The Silhouette Score is a way to evaluate the quality of clusters generated by a clustering algorithm. It determines how comparable a data point is to its group relative to different groups. The score goes from -1 to 1, with a large number indicating that the data points are compatible within the same cluster but not sufficiently matched with neighboring groups. The score for each point is determined and averaged to produce an overall evaluation. A higher average Silhouette Score indicates a more clearly defined and suitable grouping structure. This method is useful for determining the ideal number of clusters and evaluating the effectiveness of various clustering algorithms. The Silhouette Score for a single data point is given by the formula:

$$s(i) = \frac{z(j)-y(j)}{max(y(j),z(j))} \quad (1)$$

where:

- y(j) is the average distance between j and all other points in the same cluster (also known as intra-cluster distance).
- z(j) is the minimum average distance from j to all points in the nearest cluster that j is not a part of (also known as the nearest cluster distance).

The Silhouette Score for the entire dataset is the mean Silhouette Score of all individual data points. The formula is:

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) \quad (2)$$

where n is the total number of data points. The overall score indicates the clustering quality, with values closer to 1 indicating better clustering.

B. *Dunn Index*

The Dunn Index is a tool for evaluating the effectiveness of algorithms for clustering that measures group density and separation. It can be defined as the ratio of the lowest intercluster distance to the maximal intra-cluster distance. A higher Dunn Index implies better grouping since it implies that groups are clearly distinguished and densely packed. Here is the formula:

$$D = \frac{(max_{1\le i<j<k} d(C_i C_j)}{(max_{1<l<k} \delta(C_l)} \quad (3)$$

where $d(C_i, C_j)$ is the distance between the two clusters C_i and C_j, and $\delta(C_l)$ is the diameter of the cluster C_l. The Dunn Index is particularly useful for comparing different clustering results on the same dataset. However, it can be computationally expensive for large datasets due to the need to compute all pairwise distances between cluster.

V. Experiments and Results

Everyday travel patterns from various days are clustered using the DBSCAN, K-Means, Affiliation Propagation, Mean Shift, and Gaussian Mixture algorithms. Records for working and nonworking days were combined for all routes from February to June.

A. *Clustering Analysis*

Clustering was performed on 372 streets, each having enough information to analyze. The number of clusters and related traffic patterns varied according to location. Affinity Propagation clustering divided days into two categories: weekdays and weekends. In DBSCAN, K-Means, Mean Shift, and Gaussian Mixture clustering days were divided into four main groups. A thorough investigation of several roadways was conducted, providing reliable outcomes. The results of DBSCAN, K-Means, Affinity Propagation, Mean Shift and Gaussian Mixture clustering are presented in Fig. 3, Fig. 4, Fig. 5, Fig. 6, and Fig. 7 respectively. Except Affinity Propagation in other clustering techniques each street's days were separated into four groups.

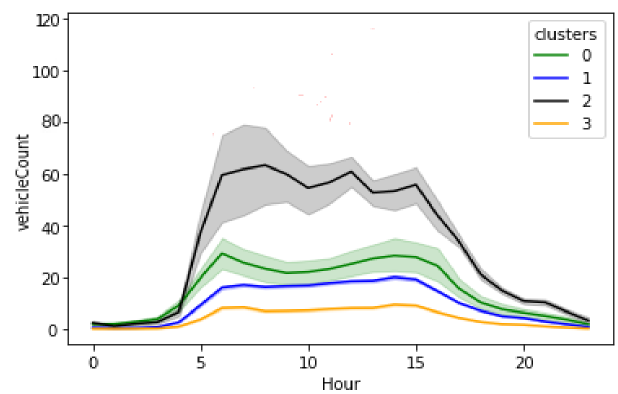

Fig. 3. Traffic pattern representing four clusters using DBSCAN.

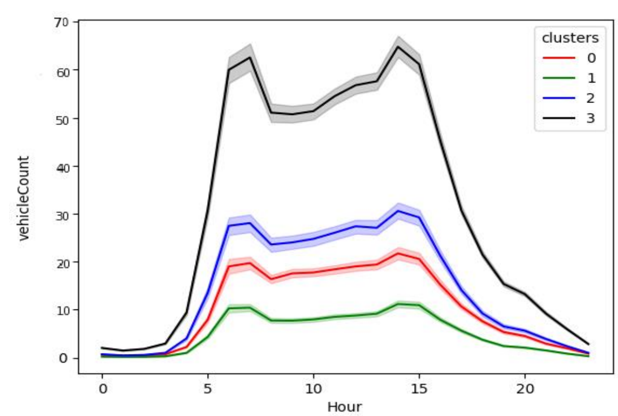

Fig. 4. Traffic pattern representing four clusters using K-Means.

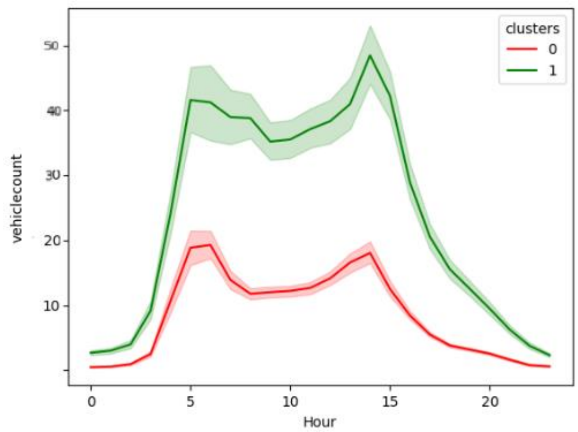

Fig. 5. Traffic pattern representing four clusters using affinity propagation.

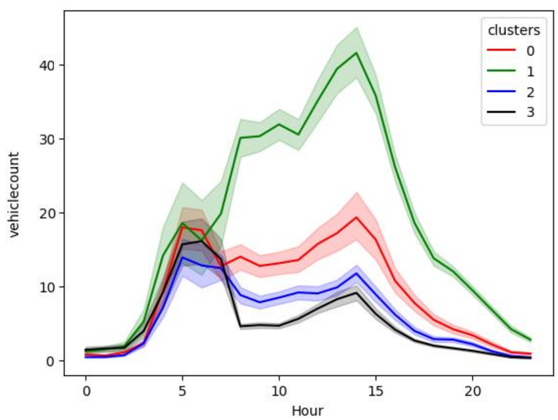

Fig. 6. Traffic pattern representing four clusters using mean shift.

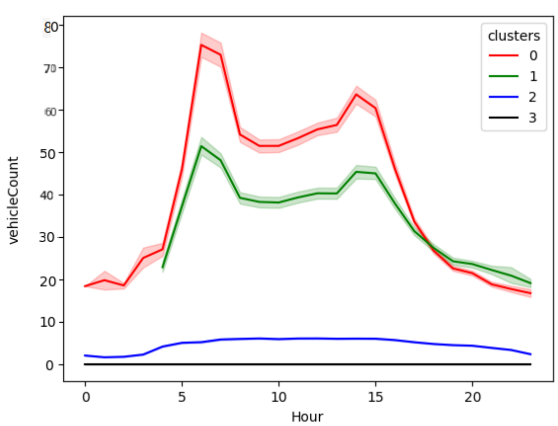

Fig. 7. Traffic pattern representing four clusters using Gaussian Mixture.

The lower level cluster in Fig. 3 to Fig. 7 includes Saturdays, Sundays, and certain weekdays. Weekdays with vacations or wet weather have significantly lower transportation than regular days. This cluster consists of routes with low to medium traffic throughout the day. The other groups include a variety of weekdays.

NearestNeighbors is used to calculate the estimated closeness between points of data and it is utilized as eps in DBSCAN. The _n_neighbors_ in NearestNeighbors is used as the minimum sample point in DBSCAN. The elbow approach determines the ideal number of groups in K-Means clustering. To assess which clustering strategy is best for our model Silhouette Score and Dunn Index are used. The DBSCAN gives better values for the Silhouette coefficient and Dunn Index as seen in the results shown above in Table II.

TABLE II. SILHOUETTE SCORE AND DUNN INDEX OF DIFFERENT CLUSTERING TECHNIQUES

Technique	Silhouette Score	Dunn Index
DBSCAN	0.3664	0.03752
K-Means	0.3119	0.00395
Affinity Propagation	0.0017	0.00123
Mean Shift	0.0042	0.00350
Gaussian Mixture	0.2032	0.00390

We have also compared the training times taken by each clustering algorithm. Fig. 8 depicts the training time for each investigated clustering algorithm. Fig. 9 depicts the training time for each clustering algorithm excluding Affinity Propagation. On the scale of the training timeframes for the Mean Shift and Affinity Propagation algorithms, the training durations of the other methods completely evaporate. Furthermore, Affinity Propagation is eliminated to compare the training durations of the remaining algorithms on a more reasonable scale.

Fig. 8. Training time for each clustering algorithm.

Fig. 9. Training time for each clustering algorithm excluding affinity propagation.

Compared to Affiliation Propagation and Mean Shift, training time is less for DBSCAN, K-Means, and Gaussian Mixture.

VI. CONCLUSION AND FUTURE WORK

Various clustering methods have been employed to extract patterns from urban traffic data, including DBSCAN, K-Means, Affinity Propagation, Mean Shift, and Gaussian Mixture, notably DBSCAN, which produced enlightening results. DBSCAN proved to be very successful in its capacity to manage noise and detect groups of arbitrary shape, making it ideal for the complex and variable traffic patterns seen across multiple streets, the clustering procedure showed various traffic patterns, in some categorizing days a weekdays or weekends and many cases into four primary clusters. The rigorous clustering method enabled a thorough understanding of the traffic behavior, allowing for more focused traffic control tactics. DBSCANs exceptional efficiency when handling and analyzing traffic data demonstrates its potential to improve urban traffic systems.

For future studies, it is recommended to use big data technology to handle and analyze large amounts of traffic data collected from diverse sources such as GPS, detectors, and online platforms. This strategy will give more specific information on traffic trends. Furthermore, utilizing clustering algorithms to undertake comparative evaluations of travel habits across multiple cities can aid in identifying common difficulties and best practices, as well as revealing trends and solutions that may be broadly implemented.

REFERENCES

[1] Pattanaik, V., Singh, M., Gupta, P. K., & Singh, S. K. (2016, November). Smart real-time traffic congestion estimation and clustering technique for urban vehicular roads. In 2016 IEEE region 10 conference (TENCON) (pp. 3420-3423). IEEE.

[2] Rouky, N., Bousouf, A., Benmoussa, O., & Fri, M. (2024). A spatiotemporal analysis of traffic congestion patterns using clustering algorithms: A case study of Casablanca. Decision Analytics Journal, 10, 100404.

[3] Wang, X., Qu, X., & Jin, S. (2020). Hotspot identification considering daily variability of traffic flow and crash record: A case study. Journal of Transportation Safety & Security, 12(2), 275-291.

[4] Shi, Y., Wang, D., Tang, J., Deng, M., Liu, H., & Liu, B. (2021). Detecting spatiotemporal extents of traffic congestion: A density-based moving object clustering approach. International Journal of Geographical Information Science, 35(7), 1449-1473.

[5] Yang, S., Wu, J., Qi, G., & Tian, K. (2017). Analysis of traffic state variation patterns for urban road network based on spectral clustering. Advances in Mechanical Engineering, 9(9), 1687814017723790.Stathopoulos, A., & Karlaftis, M. (2001). Temporal and spatial variations of real-time traffic data in urban areas. Transportation Research Record, 1768(1), 135-140.

[6] Angmo, R., Aggarwal, N., Mangat, V., Lal, A., & Kaur, S. (2021). An improved clustering approach for identifying significant locations from spatio-temporal data. Wireless Personal Communications, 121(1), 985-1009.

[7] Asadi, R., & Regan, A. (2019, November). Spatio-temporal clustering of traffic data with deep embedded clustering. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility (pp. 45-52).

[8] Sfyridis, A., & Agnolucci, P. (2020). Annual average daily traffic estimation in England and Wales: An application of clustering and regression modelling. Journal of Transport Geography, 83, 102658.

[9] Wang, W., Shao, F., Du, B., & Lu, G. (2016). Analysis of vehicular traffic flow characteristics on urban expressways based on traffic flow data in Shanghai. Transportation Research Part C: Emerging Technologies, 69, 30-49.

[10] Taamneh, M., Taamneh, S., & Alkheder, S. (2017). Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. International journal of injury control and safety promotion, 24(3), 388-395.

[11] Acun, G., & Gol, E. (2021). Traffic prediction and management using ARIMA and K-means clustering on large-scale traffic networks. Journal of Intelligent Transportation Systems, 25(5), 385-399.

[12] Zou, X., & Chung, E. (2024). Traffic prediction via clustering and deep transfer learning with limited data. Computer-Aided Civil and Infrastructure Engineering.

[13] Nguyen, T. T., Krishnakumari, P., Calvert, S. C., Vu, H. L., & Van Lint, H. (2019). Feature extraction and clustering analysis of highway congestion. Transportation Research Part C: Emerging Technologies, 100, 238-258.

[14] Chen, Y., Li, Z., & Zhang, J. (2022). Automobile flow forecasting with Graph Attention Network and spatial-temporal clustering. Computer-Aided Civil and Infrastructure Engineering, 37(3), 230-245.