



Urban Traffic Congestion Analysis Using a Hybrid Machine Learning Model: A Case Study of Nasiriyah



Sarah Abdulkareem Hanoon* 

Department of Surveying and Geomatics, College of Engineering, University of Thi-Qar, 64001 Nasiriyah, Iraq

* Correspondence: Sarah Abdulkareem Hanoon (sarah.abdulkareem@utq.edu.iq)

Received: 10-02-2025

Revised: 12-08-2025

Accepted: 01-30-2026

Citation: S. A. Hanoon, "Urban traffic congestion analysis using a hybrid machine learning model: A case study of Nasiriyah," *Int. J. Transp. Dev. Integr.*, vol. 10, no. 1, pp. 104–116, 2026. <https://doi.org/10.56578/ijtdi100108>.



© 2026 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: In many Iraqi cities, urban traffic congestion is still a problem, and the use of sophisticated analytical methods is constrained by the lack of trustworthy data. The present research combines both supervised learning and clustering algorithms to create a data-driven model to classify traffic levels in Nasiriyah. An example of K-means clustering was used to derive a categorical congestion-level variable by using field data that were collected on sixteen different districts to explain the underlying traffic patterns. The ability of three classification algorithms—J48 decision tree, Naive Bayes, and random forest (RF) to differentiate between low, medium, and high congestion circumstances was then assessed. With an accuracy of 81.25% and a kappa value of 0.70, the J48 model outperformed the other classifiers on the short dataset and had the most consistent performance. The findings also suggest that the lightweight hybrid strategy can provide authoritative congestion information in data-limited settings and, therefore, present a useful tool to support planning and traffic management decisions in fast-growing cities.

Keywords: Urban transportation systems; Smart cities; Traffic density analysis; K-means; J48; Naive Bayes; Machine learning; Sustainability

1 Introduction

The idea of urban transformation is becoming more popular because it represents the Sustainable Development Goals, and cities lie at the centre of the processes of accelerating the change towards sustainability and resiliency on the local and global scales. Among other related issues in the cities, there has been a swift technological change in the recent past with the urban development underway in the city significantly changing the land utilization [1, 2]. The findings also suggest that the lightweight hybrid strategy can provide reliable congestion information in data-limited settings and therefore serve as a useful tool to support planning and traffic management decisions in fast-growing cities [3]. Such problems are particularly widespread in developing cities, as they do not have the digital information and technological innovations that they need to control the traffic more efficiently [4]. Thus, artificial intelligence has emerged as a modern technical instrument for urban planning as a result of this shift. The evaluation of large volumes of traffic data and the accurate prediction of congestion patterns can improve the management of transportation systems. A careful data processing with the help of machine learning algorithms, the abilities and effectiveness of which have been verified in practice, has contributed to the introduction of intelligent solutions in a variety of situations [5, 6].

This study contributes to addressing a documented gap in applied intelligent traffic analysis within Iraqi cities, despite global developments in analytical techniques and the substantial absence of practical studies in Iraqi cities that need to be addressed. It offers a useful model and framework for using machine learning algorithms to examine actual data from Nasiriyah, which may be used in other locations with comparable urban and demographic features.

2 Literature Review

Following the changes in urban transportation planning, particularly after the development of AI tools and their use in transportation planning and intelligent traffic management, studies have examined the prediction of traffic congestion, a model that integrates different datasets and spatial distribution factors on transportation planning and optimization using a set of algorithms called clustering algorithms, which rely on the integration of traffic and spatial data. These AI methods have been widely used in the management of intelligent traffic systems [7].

Connecting these contemporary models to the traditional theoretical framework of traffic has become crucial due to the growing number of studies that use artificial intelligence techniques to study traffic congestion. In this regard, the fundamental theory in understanding the behavior of traffic flows, the dynamics of the congestion formation, and the processes of the instability are presented by the pioneering works of Daganzo [8] and Helbing [9]. By making reference to this work, the current study's theoretical foundation is strengthened and contemporary classification techniques are connected to the core ideas of traffic models [8, 9].

2.1 Artificial Intelligence and Machine Learning in Traffic Congestion Analysis

High accuracy is essential in AI-based classification systems in traffic management situations. Zhao et al.'s study [10] combined K-means algorithms with self-organizing maps (SOMs) to increase and improve traffic classification accuracy. Therefore, the predictive accuracy of the classification model is boosted by the combination of various artificial intelligence procedures. The hybrid structure and its resulting analytical products have seen a major improvement in the correct identification of primary traffic congestion clusters, a factor that has informed urban planning products, especially in the developing metropolitan setting.

Prior research on urban dynamics indicates that traffic congestion is a complex, dynamic process with a set of parameters that are determined by the relationship between the speed of vehicles, their density, and flow. Nagatani [11] showed that traffic bottlenecks can be self-generated even without external obstacles, making understanding traffic behavior an important foundation before applying modern analysis and artificial intelligence models.

The K-means algorithm is one of the oldest statistical clustering methods. It later became one of the most widely used algorithms for pattern analysis and data classification in engineering and urban fields [12].

Later studies developed the K-means classification algorithm, as demonstrated by Hu et al. [13], by improving it by adding hidden data patterns. This is achieved by combining algorithms, such as decision trees, to produce highly accurate classification results. This enables the analysis of multiple datasets and efficient planning solutions.

Sun et al. [14] also combined a type of GRU recurrent neural network with the K-means algorithm to first classify traffic into traffic levels and then predict traffic flow by identifying the closest pattern. This model was compared with other traditional models and found to achieve significantly higher accuracy than these models, such as SVR and random forest (RF).

Gupta et al.'s study [15] merged K-means and supervised learning algorithms such as decision trees, RF, and SVR. When the hybrid model was used, not only did the RMSE decrease significantly, but also the MAE, which proves the effectiveness of the model in embodying nonlinear relationships inherent in traffic data.

Studies at the Arab and Iraqi levels are still scarce since the models that have been built sometimes rely on simulations or incomplete data rather than real-world deployments of artificial intelligence algorithms. Therefore, there is a need for applied studies based on real data, such as this study, which seeks to fill this gap through a realistic analysis of traffic in the city of Nasiriyah.

This study is a natural extension of these efforts, but it is unique in its integrated, realistic application to an Iraqi urban environment using real data and advanced classification tools.

2.2 Development of Urban Transportation Systems

The urban transportation system is an interconnected system influenced by a range of factors, including population, economy, vehicle traffic, environment, demand and supply of transportation and travel, and congestion levels. Analysis of this system is divided into quantitative analyses, such as population, economy, and vehicle traffic, which in turn affect the transportation system and the surrounding environment. On the other hand, congestion results from the interaction between supply and demand for transportation, which in turn affects the economy [16–18]. The transportation system has evolved, coupled with the need to collect diverse data, poses a major challenge for developing an Integrated urban development model that supports new smart transportation systems [19].

Traffic congestion is a social and environmental challenge that directly and indirectly impacts the national economy [20]. Increasing traffic demand outpaces existing available roads, complicating traffic management. To avoid this in smart city development, accurate and robust predictive transportation approaches and decision-making tools that integrate the predictive nature of traffic must be developed [21, 22]. According to a study by Sarhan et al. [23], decision-makers can better understand urban dynamics and effectively handle urban challenges by combining spatial artificial intelligence techniques with urban data to analyze spatial patterns and predict urban transformations. The results show higher accuracy than traditional methods.

Although autonomous tolls, transit, satellites, and geographic information systems (GIS) are now included in the scope of smart transportation systems, they were initially created to address traffic issues on city roadways. In order to manage transportation networks and make them error-free, AI has emerged as a potent technology [24, 25].

Studies on intelligent traffic analysis are still few and frequently rely on incomplete data or virtual simulations at the Arab and Iraqi levels. For instance, some Iraqi research has concentrated on traffic analysis using conventional

techniques without integrating machine learning algorithms with prediction or classification capabilities. In addition, hybrid models integrating supervisory classification and exploratory analysis are not present in the local literature.

Therefore, this study fills the gap identified by evaluating empirical field data and using advanced models that might become a future framework.

3 Research Objectives

The purpose of this study is to:

1. Identify traffic clusters based on traffic indicators and spatial characteristics using the K-means algorithm in the city of Nasiriyah.
2. Develop a model based on the J48 decision tree to accurately classify traffic conditions and identify congestion levels (low, medium, high).
3. Evaluate the performance of the J48, Naive Bayes, and the RF algorithms with references to 10-fold cross-validation in order to select the most appropriate model to be used in the urban environment.
4. Providing an AI-based analysis tool to the city planners to help in the planning process, contributing ideas on proactive solutions to traffic, and setting priorities of interventions.
5. Offer a methodology framework, which could be used in related urban settings regarding traffic and spatial characteristics, in an effort to enable the dissemination of results and the expansion of its use.

4 Significance of the Study

The deficiency of AI-driven transport research within Iraqi cities is apparent, especially in the cities that use credible data and confirmed analytical frameworks. Such a gap creates the necessity to conduct additional research in this direction.

As transport systems are becoming more complex, Analytical methods are required that provide finer and data-oriented information as compared to the conventional monitoring techniques.

The quality of decision-making in complex urban scenarios is enhanced by machine learning algorithms' capacity to handle data that fluctuates across time and space.

Through proactive traffic analysis, the study provides insights that may support efforts to enhance traffic efficiency and align with broader smart-city development trends.

5 Methodology

The research employed a hybrid technique consisting of three interconnected phases. A device with an Intel Core i7 processor, 16GB RAM, Windows 10, Python software (with Scikit-learn and Pandas libraries), and Weka 3.9.6 software for classification and cross-validation (10-fold cross-validation) was used to implement each model. Using the K-Means technique, an exploratory study was carried out to identify natural groupings in the data. Before the K-means method was applied, the Min-Max Normalization process was applied to all continuous variables in order to reduce the impact of different measurement units and provide higher accuracy in cluster identification. The aim variable (traffic level) was created in the second phase using the characteristics of each cluster. The third stage of supervised classification, or supervised learning, employed three models: J48, Naive Bayes, and RF. Comparing performance and identifying the best model was the goal. Figure 1 provides an overview of the methodology's interconnected phases.



Figure 1. Traffic density patterns in Nasiriyah

5.1 Case Study: Nasiriyah City

Nasiriyah is in the southern Iraqi Dhi Qar Governorate, and it has a population of approximately 500,000. The place is a transportation centre of the region and an agricultural centre due to its position along the Euphrates River. It was selected because of its high rate of urbanization and traffic congestion, which is mainly experienced during rush hour in the morning and evening, as a large number of modes of transport are not available. The use of smart technologies in transportation planning will be adopted in the future to meet the challenges of increased mobility requirements. The city reflects several mobility challenges commonly observed in developing urban areas.

Additionally, it lacks structured traffic data, which makes it a suitable setting for evaluating how well AI algorithms support sustainable urban planning. Sixteen of Nasiriyah’s most crowded and vibrant areas provided data for this study. Along with the monthly number of accidents, the number of vehicles that passed through these neighborhoods each hour was also counted. For every solution, rush hours were also calculated, as Table 1 illustrates. Through field research, artificial intelligence models, and pattern analysis were used to analyze the available data and generate insights that support transportation planning efforts and the dynamic adjustment of traffic lights.

Table 1. Peak hours for residential areas

Region	Street Width (m)	Traffic Lights	Nearby Facilities	Land Use	Intersections	Parking Lot	Average Speed (km/h)	Road Type	Public Services	Vehicles per Hour	Monthly Accidents	Rush Hour Time (min)
Al-Shohada District	12	2	3	Mixed	4	Yes	30	Main	Yes	1200	4	45
Al-Zahraa District	10	1	2	Residential	3	No	35	Secondary	No	900	2	35
Al-Mualimeen District	14	3	4	Commercial	5	Yes	25	Main	Yes	1500	6	60
Baghdad Street	12	2	3	Mixed	4	Yes	29	Main	Yes	1300	5	50
Al-Shumoukh District	12	0	1	Residential	2	No	40	Secondary	No	700	1	30
Economists District	10	1	2	Mixed	3	No	34	Secondary	No	1000	3	40
Al-Askari District	11	1	2	Residential	3	No	33	Secondary	Yes	1400	5	55
Housing Area	9	0	1	Residential	2	No	36	Secondary	No	800	2	30
Al-Risala District	10	2	3	Mixed	4	Yes	28	Main	Yes	950	2	38
Al-Tadhiyah District	11	1	2	Residential	3	No	32	Secondary	Yes	1100	3	42
Al-Ghadeer District	12	2	3	Mixed	4	Yes	26	Main	Yes	600	1	25
Local Administration Street	10	1	2	Mixed	3	No	34	Secondary	No	1250	4	48
Al-Aqtisadien District	14	2	5	Mixed	4	Yes	30	Main	Yes	1500	5	65
UR District	12	1	1	Residential	2	No	40	Secondary	No	700	1	30
Al-Haboobi	14	3	4	Commercial	5	Yes	25	Main	Yes	1500	5	55
Al-Iskan ALqadeem	10	2	2	Residential	4	Yes	3	Main	Yes	700	2	30

Source: Field Survey, 2025

5.2 Exploratory Analysis Using the K-Means Algorithm (Phase 1)

The K-means algorithm was applied as an unsupervised clustering method to analyze the traffic and field-collected data. The algorithm was used to process traffic data collected from field observations and official reports.

The number of cars per hour, the number of accidents per month, and the length of rush hour in minutes were the three primary traffic indicators that were gathered in the field and through official reports.

The analysis combined several spatial indicators, which are related to the distribution of traffic.

The clustering ($K = 3$) approach was used to group the areas on the basis of traffic patterns.

As shown in Table 2, the K-means results identified three distinct clusters. In addition to supporting the reasoning behind the creation of the categorical variable (traffic level), which was subsequently employed in supervisory classification models, the results provide an initial understanding of underlying traffic patterns.

The Elbow Method has been used to determine the best number of clusters, based on variation in within-cluster dispersion.

These measures are crucial quantitative instruments that allow understanding the traffic density and the accompanying effect on the urban infrastructure. Figure 2 provides an illustration of these indications.

The scatter plot shows how Nasiriyah’s metropolitan districts are categorized according to the amount of traffic and the length of rush hour. Using the K-means clustering algorithm, each hue denotes a unique cluster of congestion levels.

Table 2. Statistical properties of each K-means cluster

Area	Vehicles per Hour	Monthly Accidents	Rush Hour Time (min)	Cluster
Al-Shohada District	1200	4	45	2
Al-Zahraa District	900	2	35	1
Al-Mualimeen District	1500	6	60	0
Baghdad Street	1300	5	50	0
Al-Shumoukh District	700	1	30	1
Economists District	1000	3	40	2
Al-Askari District	1400	5	55	0
Housing Area	800	2	30	1
Al-Risala District	950	2	38	2
Al-Tadhivah District	1100	3	42	2
Al-Ghadeer District	600	1	25	1
Local Administration Street	1250	4	48	2
ALAqtisadien	1500	5	65	0
UR District	700	1	30	1
Al-Haboobi	1500	5	55	0
Al-Iskan ALqadeem	700	2	30	1

Source: Field Survey, 2025. Units: vehicles/h, min

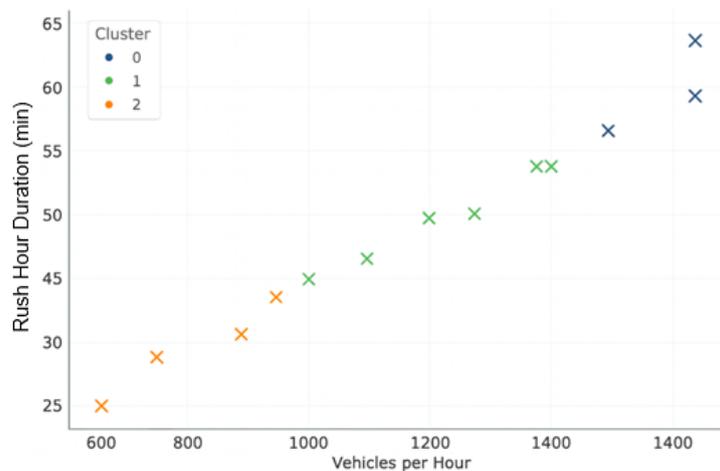


Figure 2. Urban traffic clusters in Nasiriyah

5.3 Phase Two: Developing the Target Variable (Traffic Level)

A categorical traffic-level variable was created for each urban region using K-means clustering results supported by descriptive statistics. This variable was later used as the target in supervised classification models. The variable was developed using two key indicators that reflect congestion characteristics: rush hour (minutes) and the quantity of cars per hour.

The following guidelines were used to determine three levels of congestion (low, medium, and high), which were defined according to the following rules:

- (1) A district is considered to have low congestion if the rush hour lasts less than thirty minutes.
- (2) When the rush hour lasts more than thirty minutes:
 - a. If the number of vehicles is less than 1,300 per hour, the district is categorized as having medium congestion.
 - b. If there are more than 1,300 vehicles per hour, the district is said to have substantial congestion.

The trends identified in the exploratory phase are represented in this simple and interpretable classification model. This variable served as the target used in the supervised models, such as J48, Naive Bayes, and RF, to train the classifiers. These criteria were used to classify field data from sixteen Nasiriyah neighborhoods according to the length of peak traffic and the number of cars per hour. Table 3 presents the classification results and the corresponding congestion levels for each area based on the specified criteria.

Table 3. Traffic level classification results for urban districts in Nasiriyah

Region	Vehicles per Hour	Monthly Accidents	Rush Hour Time (min)	Traffic Level
Al-Shohada District	1200	4	45	Medium
Al-Zahraa District	900	2	35	Medium
Al-Mualimeen District	1500	6	60	High
Baghdad Street	1300	5	50	Medium
Al-Shumoukh District	700	1	30	Low
Economists District	1000	3	40	Medium
Al-Askari District	1400	5	55	High
Housing Area	800	2	30	Low
Al-Risala District	950	2	38	Medium
Al-Tadhiyah District	1100	3	42	Medium
Al-Ghadeer District	600	1	25	Low
Local Administration Street	1250	4	48	Medium
ALAqtisadien	1500	5	65	High
UR District	700	1	30	Low
Al-Haboobi	1500	5	55	High
Al-Iskan ALqadeem	700	2	30	Low

5.4 Phase Three: Applying Supervised Classification Models

Following the specification of the target classification variable (traffic level) based on the exploratory study, three supervised machine learning algorithms were used to classify urban areas of Nasiriyah into three congestion levels: low, medium, and high.

In this stage, the models' accuracy and ability to distinguish between the three congestion levels will be evaluated.

The models were implemented using Weka version 3.9.6, employing the 10-fold cross-validation technique, which divides the data into ten parts, using nine for training and one for testing, repeated over ten rounds to provide a robust evaluation of model performance, as shown in Figure 3.

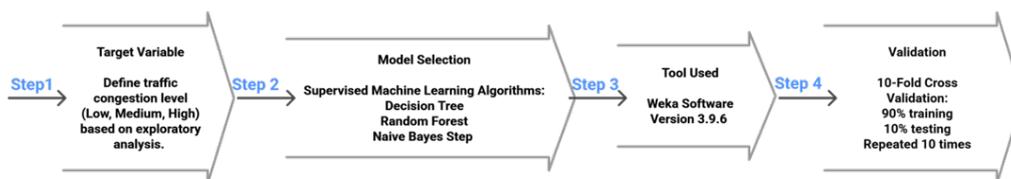


Figure 3. Workflow for traffic congestion classification

5.4.1 Algorithms used

1. Decision Tree (J48): A C4.5-based algorithm, configured with a pruning level ($C = 0.25$) and a minimum number of instances per leaf ($M = 2$) to control model complexity;
2. RF: Implemented with the default settings (100 trees); and
3. Naive Bayes: Implemented using the default settings because of its probabilistic structure and computational efficiency.

5.4.2 Data structure used

The dataset consisted of 16 urban areas within Nasiriyah City, with five main characteristics (number of vehicles per hour, peak traffic duration in minutes, land use type, etc.), and the target variable (traffic level).

5.4.3 Evaluation

The models were evaluated using the following indicators: overall accuracy, kappa coefficient, and confusion matrix, and performance indicators for each category were precision, recall, F-measure, and receiver operating characteristic area.

Detailed model results and comparisons are provided in the Results section.

6 Results

This section presents the outputs of analytical models developed using traffic and spatial data from several areas in Nasiriyah City, using an approach that includes unsupervised exploratory analysis (K-means), followed by supervised classification using machine learning algorithms (J48, Naive Bayes, RF). The workflow then concludes with model evaluation using 10-fold cross-validation and standard performance metrics, as shown in Figure 4.

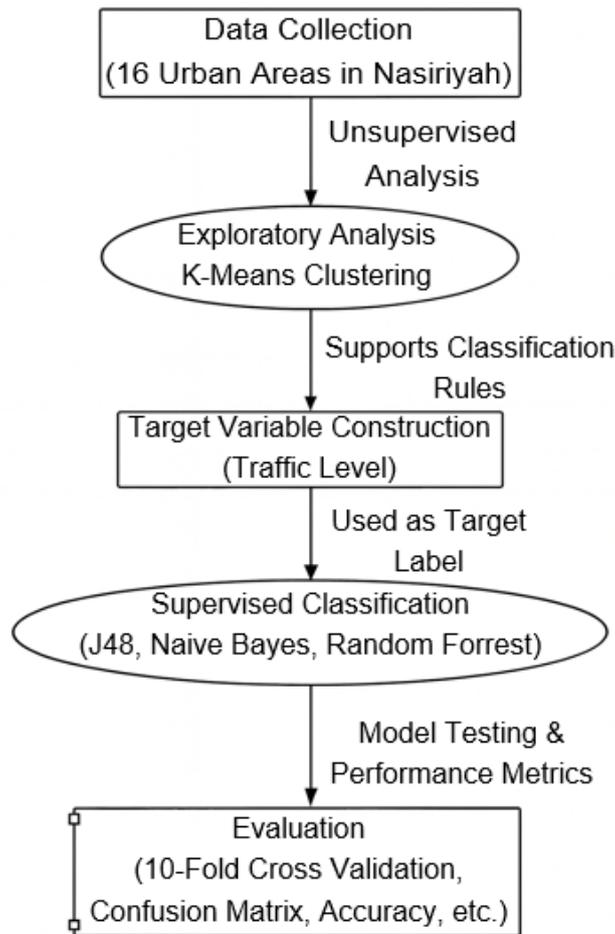


Figure 4. Classification workflow for congestion analysis

6.1 Exploratory Analysis Results: K-Means Clustering Algorithm

The K-means algorithm was applied to cluster 16 urban areas in Nasiriyah City into three groups based on three indicators: the vehicles per hour, accidents per month, and peak-hour duration in minutes, collected in the field. The results showed clear differentiation in traffic density patterns and associated risks, as shown in Table 4.

The K-means clustering analysis identified three unique congestion patterns, which varied in terms of the amount of traffic volume, frequency of the incidences and the duration of peaks. These inconsistencies seem to be associated with the latent spatial features and land-use typologies in the study regions. Areas with intensive activity showed greater magnitudes of congestion, whereas regions with less intensive activity showed longer periods of peak. All these results explain the interaction between the spatial configuration and the observed traffic dynamics.

Cluster 0 has the greatest congestion indicators, Cluster 1 has the least, with Cluster 2 representing the intermediate conditions. These results were used in the creation of the categorical variables that were utilized later.

1. Cluster 0: High traffic, risky areas.

Table 4. K-means results—statistical characteristics of each cluster

Cluster	Number of Areas	Vehicles/Hour	Accidents per Month	Peak Time (min)	Cluster Areas
0	3	1400.00	5.33	55.00	Al-Mualimeen, Baghdad St, Al-Askari
1	4	750.00	1.50	30.00	Al-Zahraa, AlShumoukh, Housing Area, Al-Ghadeer
2	5	1100.00	3.20	42.60	Al-Shohada, Economists, Al-Risala, Al-Tadhivah, Local Admin Street

Source: Field Survey, 2025; Units: vehicles/h, min

The three key locations are the ones that have the highest peak length (55 minutes), the largest monthly accident rate (5.33 incidents), and the highest average traffic volume (1,400 vehicles per hour). The cluster represents the most congested and dangerous areas that require immediate infrastructure and signal control.

2. Cluster 1: Low-density and quiet places.

The cluster has four stable traffic areas with an average of only 750 vehicles per hour, a low number of accidents (1.5), and a short length of the peak (30 minutes). These places have relatively low congestion signs and need to be checked on a regular basis.

3. Cluster 2: Areas of Moderate Congestion.

There are five areas in which the traffic rate (1,100 vehicles per hour), the median accidents (3.2), and the maximum duration are about 42.6 minutes. This cluster represents transitional areas that may experience increased traffic density in the future.

The findings of the clustering can be used to give a clue about which of the areas can be given attention earlier, as summarized in Table 5.

Table 5. Smart intervention priority

Level	Target Cluster	Proposed Priorities
First	Cluster 0 (High)	<ul style="list-style-type: none"> ● Infrastructure improvement <ul style="list-style-type: none"> ● Smart traffic signals ● Alternative routes ● Awareness campaigns
Second	Cluster 2 (Moderate)	<ul style="list-style-type: none"> ● Safety monitoring ● Public transport enhancement ● Intersection management
Third	Cluster 1 (Low)	<ul style="list-style-type: none"> ● Maintain stability ● Monitor future changes

K-means results provided a clear grouping of urban areas based on traffic density indicators. This classification has contributed to revealing underlying patterns in the data, supporting the development of a classification variable (traffic level) and offering insights that may support future planning decisions based on data-driven analysis.

Based on these clustering results, the next step was to create a categorical variable for the level of congestion for use in supervisory models.

6.2 Traffic Level Classification Method

Based on the K-means results, a “traffic level” variable was derived to classify areas into three categories:

- Low: If the peak duration is ≤ 30 minutes;
- Medium: If the peak duration is > 30 minutes and the number of vehicles is $\leq 1,300$;
- High: If the peak duration is > 30 minutes and the number of vehicles is $> 1,300$.

These rules were applied to each area in the sample, as shown in Figure 5.

6.3 Results of Supervisory Classification Models

Three machine learning algorithms were applied to classify areas based on traffic density (traffic level): J48 Decision Tree, Naive Bayes, and RF.

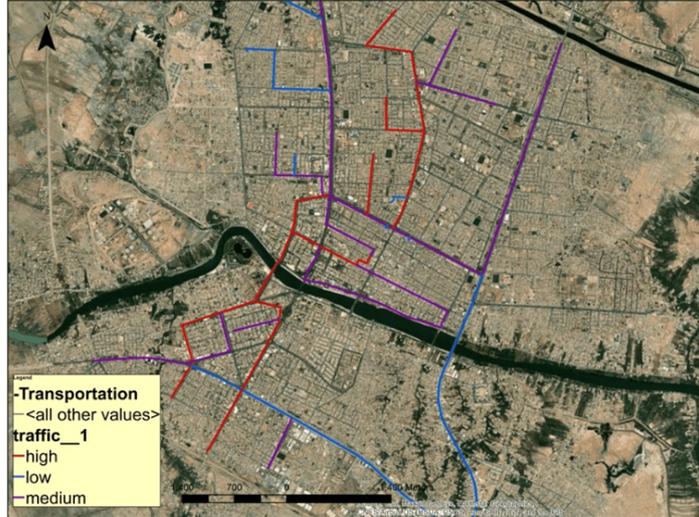


Figure 5. Traffic level classification results

These three algorithms were evaluated using 10-fold cross-validation. Table 6 summarizes the model settings for the algorithms.

Table 6. The model settings for the algorithms

Algorithm	Settings	Accuracy	Kappa	Average F-Measure
J48	$C = 0.25, M = 2$	81.25%	0.7073	0.87
Naive Bayes	Default	81.25%	0.7073	0.85
Random Forest (RF)	100 Trees (default)	62.5%	0.306	0.45

6.3.1 J48 decision tree model results

The J48 Decision Tree model achieved the highest accuracy among the three evaluated algorithms. The settings ($C = 0.25$ and $M = 2$) were chosen based on preliminary tests conducted using Grid Search, where these settings provided an acceptable trade-off between complexity and performance. Achieving a classification accuracy of 81.25% and a kappa coefficient of 0.7073, demonstrating a high agreement between predictions and actual data. Additionally, the model showed a balanced distribution of correct classifications across the three classes, as indicated in Tables 7 and 8.

Confusion Matrix:

Table 7. Decision tree J48 model evaluation matrix

Actual/Predicted	Medium	High
Medium	6	1
High	0	3
Low	0	1

Per-Class Metrics:

Table 8. Decision tree J48 model metrics

Class	Precision	Recall	F-Measure
Medium	0.86	0.86	0.86
High	0.75	1.00	0.86
Low	1.00	0.83	0.91

6.3.2 Random Forest Model

The RF model achieved lower performance, with an accuracy of 62.5% and a Kappa coefficient of 0.306, with significant weakness in class discrimination, especially for the high-congestion class, as shown in Table 9.

Confusion Matrix:

Table 9. Random Forest model evaluation matrix

Actual/Predicted	Medium	High	Low
Medium	4	1	2
High	1	1	1
Low	1	2	3

6.3.3 Naive Bayes model

With a marginally different error distribution, Naive Bayes attained the same accuracy (81.25%) as J48. Due to its computational efficiency, this model can perform well when the available data are limited, as shown in Table 10.

Confusion Matrix:

Table 10. Naive Bayes model evaluation matrix

Actual/Predicted	Medium	High	Low
Medium	6	1	0
High	0	3	0
Low	1	0	5

For every category, additional performance metrics such as precision, recall, F1 score, and receiver operating characteristic curve were computed.

The results indicated that the decision-tree model performed comparatively better across these metrics than the other models, which indicates a consistent performance pattern in classifying congestion levels.

The RF model results showed some instances of misclassification, which may be related to the limited number of observations available in the dataset. The samples where misclassification occurred were reviewed and compared with the actual values. The errors were found to be caused by overlap between the characteristics of medium- and high-density areas, in addition to the limited data available for each category. This analysis highlights specific limitations in RF performance, which generally improve with larger and more diverse datasets, which explains its limited effectiveness in this study.

6.3.4 Final comparison

Based on the evaluation results, the findings indicate that the J48 model performed most consistently within this dataset, given its balance between accuracy and interpretability in this specific context, which may support its potential use in exploratory urban analysis, as shown in Table 11.

Table 11. Final comparison of the algorithms (J48, Naive Bayes, Random Forest)

Algorithm	Accuracy	Kappa	Average F-Measure	Low
J48	81.25%	0.7073	0.87	0
Naive Bayes	81.25%	0.7073	0.85	0
Random Forest (RF)	62.5%	0.306	0.45	5

7 Discussion

This study used a hybrid approach that combines supervised classification and exploratory clustering to analyze traffic density patterns in Nasiriyah. The findings of the analysis, which included supervisory modeling using three classification algorithms and unsupervised clustering using K-means, highlighted several patterns that can be examined from both technical and urban perspectives.

7.1 Cluster Analysis and Its Spatial Implications

The application of the K-means technique revealed a clear variation in traffic characteristics across the urban areas of Nasiriyah, indicating the presence of distinguishable cluster patterns within the traffic data. Within the first cluster (Cluster 0), the most noteworthy localities that demonstrated high traffic and accident rates are “Al-Mualimeen” and Baghdad Street, which are also marked by high commercial activity. On the other hand, the second cluster (Cluster 1) comprised districts that exhibited lower congestion levels, such as “Al-Zahraa” and “Housing Area”. These findings highlight the importance of cluster analysis as a starting point for policy development and indicate a correlation among the land-use patterns, traffic infrastructure, and the resultant traffic situation.

7.2 Effectiveness of Classification Models

J48 algorithm showed the best classification of 81.25%, which is better than that of Naive Bayes and RF. The advantage of J48 is that it can produce a straightforward decision tree using simple logical rules, so it can be used in the decision support system, especially in data-poor systems such as this study. On the contrary, Naive Bayes worked in the same way as J48, though it did not have the same visual representation. In the case of RF, the small sample size had a negative result in the ability to differentiate categories, as it is mentioned in various studies that its performance increases with the increased volume of the data.

7.3 Variation Between Clusters and Level of Intervention

High-density areas (Cluster 0) seem to warrant earlier consideration based on the indicators observed, and areas that could be improved may relate to infrastructure and traffic control.

- Installation of smart signal systems.
- Improving safety and traffic control.

Cluster 2 can be the object of proactive planning, whereas Cluster 1 can simply be observed on a routine basis.

7.4 The Relationship Between Exploratory Analysis and Predictive Modeling

The K-means results assisted in deriving the categorical target variable used in the classification models, using pertinent real-world data. This combined approach helps clarify the relationship between multiple variables and traffic outcomes, thereby enabling a more nuanced interpretation of the classification results. Predictability of the traffic patterns in the urban environment can be boosted by incorporating the use of supervisory and unsupervised approaches to help in ameliorating the performance of the models.

7.5 Study Limitations

The sample was small, consisting of just sixteen urban locations, which is limited by the fact that there were limited official traffic records available in Nasiriyah. This restricts the scope of research into the plethora of factors that determine traffic movement. However, the research counters this weakness by using machine-learning procedures that still stand strong when dealing with small sample sizes, which are further enhanced by ten-fold cross-validation to strengthen the validity of the empirical results. As such, the findings must be viewed as tentative observations which form a first analysis framework that will be investigated in the future.

Also, a lack of salient behavioral covariates, namely, vehicle classification, trip intent, and travel itineraries limits the ability of the study to capture the full dynamics of vehicular movement. These variables combined in future datasets would significantly increase predictive accuracy and make the model more faithful to the real-life conditions. This study, therefore, demonstrates the need to have more comprehensive and methodically organized data collection to enable deeper and more practical traffic studies in the urban setting.

7.6 Added Value for Urban Planning

The results of this study can be used as an effective instrument to enhance innovativeness in such cities as Nasiriyah in an effort to make urban transport management more emphasized. These results can be integrated with GIS platforms to enhance spatial interpretation and strengthen evidence-based transportation planning. The findings of the study present the initial factors that contribute to the innovation processes in that they enable them to focus on electricity and urban transport management. The results reveal the aspect of data collection systems that should be improved. Moreover, the successful solutions to the digital economy challenges can be reached within a short period with the help of historical studies and the integration of GIS systems and break-even points.

In the end, these processes enable localizing the means of needs interpretation, which results in informed action and supports improved mobility and enhances overall urban efficiency.

8 Conclusion

This paper identifies the use of AI methods to analyse and categorize traffic density in an urban setting, as a case study of Nasiriyah city. The K-means clustering algorithm was applied without supervision and the aim of finding the patterns in the traffic data was met; three major clusters of traffic were identified, which illustrated the different degrees of congestion and road accidents. A key contribution of this analysis is the construction of the classification variable (traffic level) was constructed using empirically derived criteria based on field measurements.

The obtained model results proved the effectiveness of the J48 decision tree algorithm in comparison with other algorithms like Naive Bayes and RF the balance and the accuracy (81.25%), which substantiates the relevance of the algorithm in the urban context when the interpretation of the obtained results is required and the implementation is to be fast. This combined unsupervised–supervised approach demonstrates its effectiveness in supporting an intelligent decision-making system, especially in the urban setting where only limited data is available, and supports the prioritization of infrastructure and traffic management interventions.

The findings of the integration of supervisory models and exploratory trajectory analysis with hybrid methods are interesting, with K-means finding that there are noticeable differences in the studied areas. At the same time, the classification algorithms, including J48, have recognized the level of congestion, and the J48 model was the best in this situation due to the lack of data. Further studies will broaden the model to bigger geographic data as well as combine it with GIS and shared digital applications to assist in the further analysis of the commercially current urban regions. Future traffic studies should have this as a possible foundation of the analysis.

Based on the findings, the recommendations of the study are as follows:

1. Embrace artificial intelligence models through smart classification models in urban traffic management by incorporating smart classification models in the municipal decision support systems.
2. Use the model on large time series such as days, hours, and seasons to increase the real-time prediction accuracy.
3. Further work on the models and their realization within the current systems to enhance the control of the urban traffic and the efficiency of its operations. The results of the classification systems must be combined with the spatial analysis (GIS) to create interactive maps to be used in the sphere of public transit, signal control, and planning in the future.
4. The next generation will examine the ability to integrate with digital twins.

These results offer a provisional analytical basis to a future empirical study of the dynamics of traffic.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflict of interest.

References

- [1] K. Hölscher and N. Frantzeskaki, “Perspectives on urban transformation research: Transformations in, of, and by cities,” *Urban Transform.*, vol. 3, p. 2, 2021. <https://doi.org/10.1186/s42854-021-00019-z>
- [2] H. Dui, S. Zhang, M. Liu, X. Dong, and G. Bai, “IoT-enabled real-time traffic monitoring and control management for intelligent transportation systems,” *IEEE Internet Things J.*, vol. 11, no. 9, pp. 15 842–15 854, 2024. <https://doi.org/10.1109/JIOT.2024.3351908>
- [3] Y. Jain and K. Pandey, “Transforming urban mobility: A systematic review of AI-based traffic optimization techniques,” *Arch. Comput. Methods Eng.*, vol. 32, pp. 5381–5417, 2025. <https://doi.org/10.1007/s11831-025-10297-6>
- [4] M. Miftah, D. I. Desrianti, N. Septiani, A. Y. Fauzi, and C. Williams, “Smart sustainable cities of the future: An extensive interdisciplinary literature review,” *Sustain. Cities Soc.*, vol. 31, pp. 183–212, 2017. <https://doi.org/10.1016/j.scs.2017.02.016>
- [5] M. Mohammadi and A. Al-Fuqaha, “Enabling cognitive smart cities using big data and machine learning: Approaches and challenges,” *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 94–101, 2018. <https://doi.org/10.1109/MCOM.2018.1700298>
- [6] A. S. Pillai, “Traffic management: Implementing AI to optimize traffic flow and reduce congestion,” *J. Emerg. Technol. Innov. Res.*, vol. 11, no. 7, pp. c272–c278, 2024. <https://doi.org/10.2139/ssrn.4916398>
- [7] B. Zhou, J. Liu, S. Cui, and Y. Zhao, “Large-scale traffic congestion prediction based on multimodal fusion and representation mapping,” in *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, Shenzhen, China, 2022, pp. 1–9. <https://doi.org/10.1109/DSAA54385.2022.10032443>

- [8] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. Part B: Methodol.*, vol. 28, no. 4, pp. 269–287, 1994. [https://doi.org/10.1016/0191-2615\(94\)90002-7](https://doi.org/10.1016/0191-2615(94)90002-7)
- [9] D. Helbing, "Traffic and related self-driven many-particle systems," *Rev. Mod. Phys.*, vol. 73, no. 4, p. 1067, 2001. <https://doi.org/10.1103/RevModPhys.73.1067>
- [10] S. Zhao, Y. Xiao, Y. Ning, and D. Zhang, "An optimized K-means clustering for improving accuracy in traffic classification," *Wireless Pers. Commun.*, vol. 120, no. 1, pp. 81–93, 2021. <https://doi.org/10.1007/s11277-021-08435-x>
- [11] T. Nagatani, "The physics of traffic jams," *Rep. Prog. Phys.*, vol. 65, no. 9, p. 1331, 2002. <https://doi.org/10.1088/0034-4885/65/9/203>
- [12] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-means clustering algorithm," *J. R. Stat. Soc. Ser. C (Appl. Stat.)*, vol. 28, no. 1, pp. 100–108, 1979. <https://doi.org/10.2307/2346830>
- [13] X. Hu, X. Chen, W. Liu, and G. Dai, "Road traffic status prediction approach based on kmeans-decision tree model," *J. Eng. Proj. Prod. Manage.*, vol. 12, no. 2, pp. 108–115, 2022. <https://doi.org/10.32738/JEPPM-2022-0010>
- [14] Z. Sun, Y. Hu, W. Li, S. Feng, and L. Pei, "Prediction model for short-term traffic flow based on a K-means-gated recurrent unit combination," *IET Intell. Transport Syst.*, vol. 16, no. 5, pp. 675–690, 2022. <https://doi.org/10.1049/itr2.12165>
- [15] S. Gupta, A. Kumar, and A. Kumar, "Analysis of traffic flow congestion by integrating supervised machine learning with K-mean clustering," *SN Comput. Sci.*, vol. 6, no. 3, p. 255, 2025. <https://doi.org/10.1007/s42979-025-03761-4>
- [16] J. Wang, H. Lu, and H. Peng, "System dynamics model of urban transportation system and its application," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 8, no. 3, pp. 83–89, 2008. [https://doi.org/10.1016/S1570-6672\(08\)60027-6](https://doi.org/10.1016/S1570-6672(08)60027-6)
- [17] H. B. Faheem, A. M. E. Shorbagy, and M. E. Gabr, "Impact of traffic congestion on transportation system: Challenges and remediations—A review," *Mansoura Eng. J.*, vol. 49, no. 2, p. 18, 2024. <https://doi.org/10.58491/2735-4202.3191>
- [18] K. A. Small, E. T. Verhoef, and R. Lindsey, *The Economics of Urban Transportation*. London: Routledge, 2024. <https://doi.org/10.4324/9781315157375>
- [19] A. Richter, M. O. Löwner, R. Ebdet, and M. Scholz, "Towards an integrated urban development considering novel intelligent transportation systems: Urban development considering novel transport," *Technol. Forecast. Soc. Change*, vol. 155, p. 119970, 2020. <https://doi.org/10.1016/j.techfore.2020.119970>
- [20] C. Ma, R. He, and W. Zhang, "Path optimization of taxi carpooling," *PLoS ONE*, vol. 13, no. 8, p. e0203221, 2018. <https://doi.org/10.1371/journal.pone.0203221>
- [21] H. Pu, Y. Li, C. Ma, and H. B. Mu, "Analysis of the projective synchronization of the urban public transportation super network," *Adv. Mech. Eng.*, vol. 9, no. 6, 2017. <https://doi.org/10.1177/1687814017702808>
- [22] A. A. Musa, S. I. Malami, F. Alanazi, W. Ounaies, M. Alshammari, and S. I. Haruna, "Sustainable traffic management for smart cities using internet-of-things-oriented intelligent transportation systems (ITS): Challenges and recommendations," *Sustainability*, vol. 15, no. 13, p. 9859, 2023. <https://doi.org/10.3390/su15139859>
- [23] H. R. Sarhan, Y. A. Mousa, S. A. Hanoon, and M. S. Sachit, "Geospatial artificial intelligence-based approach for predicting slum growth, a case in Iraq," *Adv. Mech. Eng.*, vol. 7, no. 2, pp. 1146–1157, 2025. <https://doi.org/10.24271/PSR.2025.527086.2166>
- [24] M. Collotta, Y. Sun, L. Di Persio, E. S. M. Ebeid, and R. Muradore, "Smart green applications: From renewable energy management to intelligent transportation systems," *Energies*, vol. 11, no. 5, p. 1317, 2018. <https://doi.org/10.3390/en11051317>
- [25] B. Gu, J. Liu, H. Xiong, T. Li, and Y. Pan, "ECPC-ICP: A 6D vehicle pose estimation method by fusing the roadside lidar point cloud and road feature," *Sensors*, vol. 21, no. 10, p. 3489, 2021. <https://doi.org/10.3390/s21103489>