Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Urban traffic management system based on Machine Learning technology

ELBrak Imane
University of Abdelmalek Essaâdi (UAE),
Faculty of Sciences and Technologies of Tangier,
Computer Science Department,
Computer Science and Smart Systems (C3S) Laboratory
imane.elbrak@etu.uae.ac.ma

En-naimi EL mokhtar
University of Abdelmalek Essaâdi (UAE),
Faculty of Sciences and Technologies of Tangier,
Computer Science Department,
Computer Science and Smart Systems (C3S) Laboratory
en-naimi@uae.ac.ma

EL Brak Mohamed
University of Abdelmalek Essaâdi (UAE),
Faculty of Sciences and Technologies of Tangier,
Computer Science Department,
Computer Science and Smart Systems (C3S) Laboratory
melbrak@uae.ac.ma

Abstract— **The urban transportation infrastructure faces growing challenges due to the rising demand for city living and suburban areas. This trend contributes to escalating congestion, negatively impacting network service levels, and giving rise to externalities.**

The Internet of Things (IoT) plays a crucial role in enhancing smart city functions, particularly in urban traffic management, through the integration of Machine Learning clustering algorithms like K-means and KNN (K-Nearest Neighbors). In an IoT-based urban traffic management system, vehicles equipped with sensors collect data on parameters such as vehicle speed and traffic density. The collected data undergoes pre-processing to eliminate errors and duplicates, making it usable information.

In this work the K-means algorithm is applied to traffic data, grouping similar road segments based on characteristics like traffic density and average speed. These clusters represent different types of traffic and road segments, offering insights into traffic patterns across various city areas. Historical traffic data and K-means clusters are utilized to create a KNN model. This approach facilitates informed decision-making and optimal resource utilization, leading to improved traffic flow and congestion reduction. The obtained results contribute to enhancing urban traffic management systems, enabling proactive planning, and addressing congestion challenges.

Keywords: Internet of Things, IoT, ML, K-means,KNN, Smart City, Congestion.

I. INTRODUCTION

Effective traffic management is a critical aspect of urban planning, given that issues such as congestion, accidents, and delays can have significant adverse impacts on the economy, society, and the environment. However, the task of managing traffic is becoming increasingly challenging with the rapid growth of the global population [1] [2]. In such a scenario, it becomes imperative to employ modern technological tools and provide training for traffic officers to ensure efficient traffic control.

The Internet of Things (IoT) and Artificial Intelligence (AI) are dynamic and significant topics. IoT is increasingly adopted in various fields such as smart homes, smart cities, manufacturing, healthcare, agriculture, etc. IoT devices are becoming ubiquitous, forming vast interconnected networks that generate and exchange data. AI plays a crucial role in harnessing the massive data generated by IoT, utilizing machine learning techniques to extract valuable insights for informed decision-making and a better understanding of patterns and trends.

Combining IoT and AI benefits scientific research by providing real-time data access, advanced analysis, predictive modeling, and resource optimization. This enhances understanding of phenomena, solves complex problems, and opens new research opportunities across multiple domains.

In the context of connected vehicles, utilizing IoT and AI presents numerous advantages. IoT can significantly impact urban traffic management by integrating clustering algorithms like K-means and KNN. In an IoT-based urban traffic management system, sensors and connected devices collect traffic data, such as vehicle speed and density. Preprocessed data can be grouped using the K-means algorithm based on characteristics like speed and density, creating clusters of similar road segments for traffic analysis and decision-making. Complementing K-means, the KNN algorithm predicts traffic using historical traffic data and clusters obtained from K-means to form a KNN model.

When new traffic observations occur, the KNN model predicts traffic characteristics such as density, speed, or peak hours based on the nearest neighbors within the clusters. This integration of IoT and AI in traffic management enables more efficient and informed decision-making, contributing to proactive planning and addressing traffic congestion issues.

II. RELATED WORK

Due to the urbanization, the demand for good transportation system increases tremendously. Therefore, the number of vehicles inroad also has risen day by day. There are a lot of research works have been done in different applications of smart traffic management system. Identifying traffic congestion on road is one of the important aspects of it.

In [13] rhe authors aims to estimate and classify thetraffic congestion state of different road segments within a city byanalyzing the road traffic data captured by in-road stationary sensors. The Artificial Neural Network (ANN) based system is used toclassify traffic congestion states. Based on traffic congestion status,ITS will automatically update the traffic regulations like, changingthe queue length in traffic signal, suggesting alternate routes. It alsohelps the government to device policies regarding construction offlyover/alternate route for better traffic management

In [14], authors proposed a technique to classify road traffic congestion level using decision tree algorithm. They consider vehicle velocity as a parameter to identify congestion level. They collect the road traffic data using GPS device. They used sliding window technique to generate moving pattern from vehicles velocity. They have used J48 decision tree algorithm to develop decision tree model to classify the congestion leve

In [15], speed performance index was used to evaluate the road network traffic congestion states. Authors consider vehicle’s speed as an important parameter to estimate traffic states. The speed performance index is calculated from vehicle’s speed and based on index value, system classify the traffic congestion states.

[16] proposed a method which processes the data collected from automatic vehicle location system (AVL) to measure the travel time and average speed over the freeway and in turn it determines the traffic condition. In another study [17], traffic congestion estimation algorithm has been proposed where congestion features are extracted from MPEG video data and then use Gaussian Hidden Markov model to determine traffic congestion level.

III. PROPOSED SYSTEM

The urban traffic management system based on machine learning technology, with a specific application of the K-Means algorithm and the use of speed and density parameters, offers an advanced approach to solving the complex challenges of urban traffic. Collecting data from speed sensors, surveillance cameras and other relevant sources is the first step in acquiring detailed information about traffic conditions. By judiciously selecting vehicle speed and traffic density as key parameters, the pre-processing process aims to prepare this data for effective application of the K-Means algorithm. As a clustering tool, K-Means categorises city segments into distinct clusters based on speed and density patterns, revealing significant trends in traffic behaviour.

These clusters provide an in-depth understanding of the traffic patterns specific to each urban area, paving the way for more precise traffic management. In real time, the system adjusts traffic control parameters, such as traffic lights, based on current data, improving traffic flow, reducing congestion, and contributing to a more efficient driving experience in dynamic urban environments. This holistic approach demonstrates the effectiveness of machine learning, in particular the K-Means algorithm, in addressing the complex challenges of urban traffic management using speed and density as key parameters.

A. Traffic Parameters

The two main traffic parameters are used in the proposed system. These are namely **1 ,** traffic density and **2** average speed of vehicles of each road segment. These are defined as follows.

1. **Traffic Density:** Traffic Density is defined as the number of vehicles occupying a given length of a road segment. It can be expressed as:

$$d = \frac{n}{l} \tag{1}$$

where, n is the number of vehicles and l is the length of the road segment.

2. **Average Speed:** Average Speed (s_{avg}) is the sum of speed of all the vehicles divided by total number of vehicles.

where, Speed (s_i) is defined as the total distance travelled by the vehicle per unit time. Hence Average Speed can be expressed as:

$$s_{avg} = \frac{\sum_{i=1}^{n} s_i}{n}, \qquad \text{where} \quad s_i = \frac{dis_i}{t} \tag{2}$$

where, dis_i is the distance travelled by i^{th} vehicle and t is the time period

B. Dynamic Urban Traffic Optimisation: An Innovative Approach with K-Means and Machine Learning

In the context of urban traffic management based on machine learning technology, the K-Means algorithm is used to analyse and categorise traffic data into distinct clusters. Here's how the K-Means mechanism works in this particular case, using speed and density parameters.

1. Centroid initialization:

The process begins by initialising K centroids, where K represents the predefined number of clusters. These centroids are fictitious points in feature space, determined at random.

2. Assigning Points to the Nearest Cluster:

Each data point, representing a segment of the city in our case, is assigned to the cluster whose centroid is closest in terms of Euclidean distance. The characteristics here are traffic speed and density.

3. Updating centroids:

Once all the points have been assigned to clusters, the centroids are updated by calculating the average of the characteristics of the points belonging to each cluster. This readjusts the position of the centroids in feature space.

4. Reallocating Points and Updating Centroids:

The process of allocating points to clusters and updating centroids is repeated iteratively until convergence is achieved.

The process of assigning points to clusters and updating centroids is repeated iteratively until convergence is achieved. At each iteration, points may change cluster according to the new centroid locations.

5. Forming the Final Clusters:

Once the algorithm has converged, the resulting clusters represent areas of the city that share similar characteristics of speed and traffic density. Each cluster is now a separate entity with its own traffic patterns.

6. Cluster Interpretation:

Each cluster formed can be interpreted to understand specific urban traffic patterns. For example, one cluster might represent areas of frequent congestion with low speed and high density, while another might indicate smooth roads with moderate speed.

7. Real-time adaptation:

The K-Means model can be used in real time to continually adjust clusters and adapt to changes in traffic conditions, enabling dynamic management of urban traffic.

IV. SIMULATION PARAMETERS

As part of our project to simulate traffic congestion, here are some important parameters to take into account:

First of all, we are interested in the basic parameter, which is the average speed of vehicles on a given stretch of road. For one type of section, this is the distance between the two ends divided by the time taken. It can be closely linked to density and flow rate.

- Average vehicle speed: This parameter represents the speed at which vehicles travel along different sections of the road. It can vary according to traffic density, speed limits, weather conditions, etc.

- Traffic density: Traffic density refers to the number of vehicles present on a given section of road at a given time. It can be measured in terms of vehicles per square kilometre or vehicles per kilometre of road.

- Traffic flow: Traffic flow represents the amount of traffic that passes through a certain section of road during a given period of time. It can be expressed in terms of vehicles per hour or vehicles per minute.

- Travel time: This is the time taken to cover a certain distance on the road. It can vary according to the average speed of vehicles, traffic conditions, traffic lights, etc.

- Traffic control: Traffic control strategies, such as traffic lights, traffic signs, roundabouts, dedicated lanes, etc., can help to regulate traffic and minimise congestion.

Below is an extract from the file with all the parameters and data explored by the classification algorithm.

	datetime	predefinedlocationreference	averagevehiclespeed	traveltime	traveltimereliability	trafficstatus	densite	cat	param_extrn	vehicleprobemeas
0	2023-05-24T00:07:00+02:00	10273_D	70	10	0	unknown	46	pl	0.0	
1	2023-05-24T00:07:00+02:00	10273_G	70	10	0	unknown	44	pl	0.0	
2	2023-05-24T00:07:00+02:00	10274_D	70	9	0	unknown	35	pl	0.0	
3	2023-05-24T00:07:00+02:00	10274_G	70	9	0	unknown	38	pl	0.0	
4	2023-05-24T00:07:00+02:00	10275_D	70	15	0	unknown	28	pl	0.0	

Figure 1. Extracting data from the traffic.csv file

V. SIMULATION AND RESULTS

In this section, we test the effectiveness of our approach to solving the problem using a combination of machine learning algorithms: the K-Means method and the KNN algorithm.

First, we apply the K-Means method to divide the dataset into clusters, then use the KNN algorithm to find the source cases closest to our target case.

In the following, we describe the consistent representations:

- (a) we have a representation of speed as a function of density: as the number of vehicles increases, density increases and speed decreases

- (b) we have a representation of density as a function of speed: density can vary as a function of speed in the event that vehicles that would have been able to drive very fast in a free lane reduce their speeds.

As shown in figure 1, we can also display speed as a function of density.

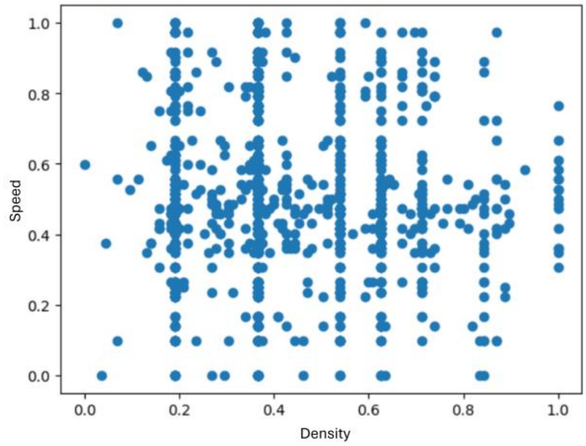

Figure 2. The ratio of speed and density data

On a portion of the data (1500 lines), i.e. on 50%, on the basis of speed and density, we have trained these data via KMeans alone, which produces eight (8) groups

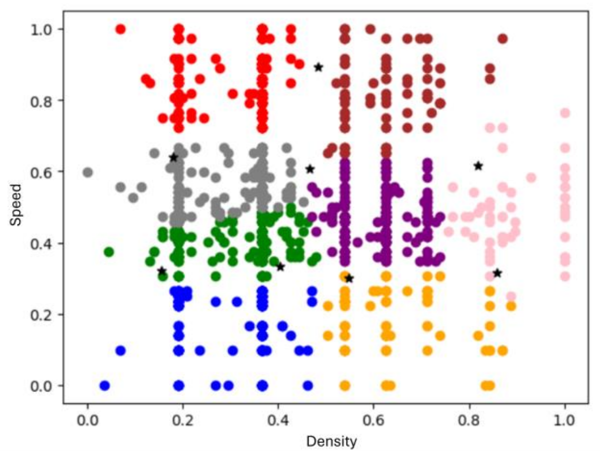

Figure 3. Application of the KMeans algorithm

There is an apparent overlap between the different groups. This is due to the large number of groups. For better clustering, a trade-off between the number of clusters and the intra-cluster distances must be made. The Elbow method [50] defines the appropriate number of clusters on which KMeans can be trained with different values of k (k is the number of iterations) and will calculate the variance (intra-cluster distance) to place the data towards the nearest cluster.

In order to obtain maximum model accuracy, we need to determine the optimal value of k, which represents the number of clusters in our data source. The Elbow method [53] is widely used to determine this value of k. After training our model with different values of k using both the training and test datasets, the optimal value of k for our model is 4, as shown in Figure 2

We therefore used the Elbow method, which combined with KMeans to produce four (4) homogeneous groups that met the classification criteria based on density and speed. Figures [2] and [3] show respectively the distance produced by Elbow as a function of the number of groups and the grouping into eight groups produced by KMeans.

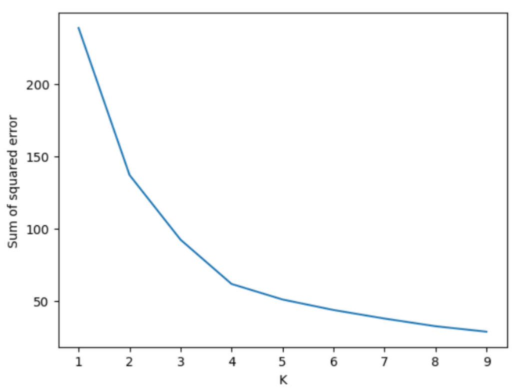

Figure 4. Application of the Elbow method

The figure shows a breakdown of the clusters and their leaders.

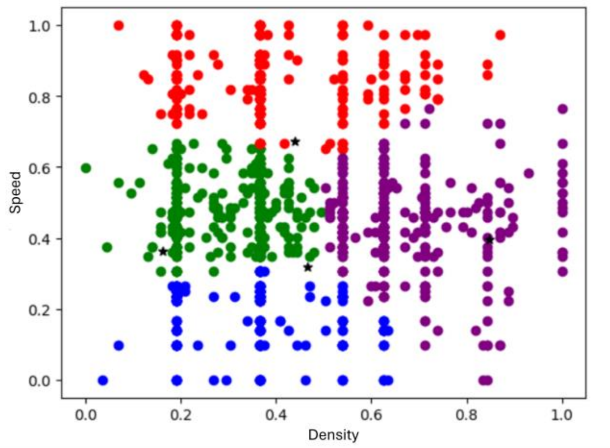

Figure 5. Number of clusters using the Elbow method

- Cluster 1 (blue): density is low and speed low, which may be linked to

The state of the road, an accident, external parameters such as fog, rain, snow, etc.

- Cluster 2 (purple): density is high and speed is average, which means that there is traffic congestion.

- Cluster 3 (green): the density is low with an average speed, which means that traffic is flowing smoothly and there is no congestion

- Cluster 4 (red): medium density at high speed, requiring careful driving and increased attention, as the presence of several vehicles can increase the risk of accidents or potentially dangerous situations. It is therefore essential to observe safety rules and drive responsibly, even when traffic density is moderate and speed is high

In order to tackle the problems linked to the traffic conditions noted in the blue and red clusters, decision-makers should analyse the parameters involved in order to make better decisions in the fight against road congestion.

VI. CONCLUSION

With increase in urbanization and socio-economical growth, the number of vehicles in major metropolitan cities is increasing day by day. Therefore, traffic congestion is becoming a major concern of metropolitan cities all over the world. This results in tremendous air pollution, loss of valuable time and money of citizens. Hence, traffic congestion monitoring of different road segments is very essential for analyzing the problem associated with smooth mobility. Identifying the problematic road segments within the city is one of the important job for the transport authority to assess the road condition. That will assist the government agencies or policy makers to optimize traffic rules and regulations. This work identifies traffic congestion pattern which can classify the different road segments based on traffic density and average speed of vehicles. The traffic parameters are captured by in-road stationary sensors deployed in road segments. The proposed system uses k-means clustering algorithm to categorize the different road segments.

REFERENCES

[1] Reddy, H. B. S., Reddy, R. R. S., Jonnalagadda, R., Singh, P., & Gogineni, A. (2022). Analysis of the Unexplored Security Issues Common to All Types of NoSQL Databases. Asian Journal of Research in Computer Science, 14(1), 1-12.

[2] Modi, Y., Teli, R., Mehta, A., Shah, K., & Shah, M. (2022). A comprehensive review on intelligent traffic management using machine learning algorithms. Innovative infrastructure solutions, 7(1), 128.

[3] L. Da Xu, W. He, and S. Li. Internet of things in industries : A survey. IEEE Transactions on industrial informatics, 10(4) :2233–2243, 2014.

[4] M. Han and H. Zhang, "Business intelligence architecture based on internet of things " Journal of Theoretical & Applied Information Technology, vol. 50, no. 1, pp. 90-95, 2013.

[5] P.-J. Benghozi, S. Bureau, F. Massit-Folléa, C. Waroquiers, and S. Davidson, L'internet des objets: quels enjeux pour l'Europe, Éd. de la Maison des sciences de l'homme éd., 2009, 66 p.

[6] J. D. Pessemier, "Une réflexion sur « L'internet des Objets » (IdO) ou « Internet of Things » (IoT)," 2015.

[7] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the internet of things to the web of things: Resource-oriented architecture and best practices," dans Architecting the Internet of Things, Springer, 2011, pp. 97-129.

[8] J. Hall, B. Ramsey, M. Rice, and T. Lacey. Z-wave network reconnaissance and transceiver fingerprinting using software-defined radios. In International Conference on Cyber Warfare and Security, page 163. Academic Conferences International Limited, 2016.

[9] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate. A survey on application layer protocols for the internet of things. Transaction on IoT and Cloud Computing, 3(1) :11–17, 2015.

[10] M. J. Kaur and P. Maheshwari. Building smart cities applications using iot and cloud-based architectures. In International Conference on Industrial Informatics and Computer Systems (CIICS), pages 1–5. IEEE, 2016.

[11] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the internet of things to the web of things: Resource-oriented architecture and best practices," dans Architecting the Internet of Things, Springer, 2011, pp. 97-129.

[12] R. A. Rahman and B. Shah. Security analysis of IoT protocols : A focus in CoAP. In International Conference on Big Data and Smart City (ICBDSC), pages 1–7. IEEE, 2016.

[13] Intelligent Traffic Congestion Classification System using Artificial Neural Network

[14] Abhinav Sunderrajan, Vaisagh Viswanathan, Wentong Cai, and Alois Knoll. 2016. Traffic State Estimation Using Floating Car Data. Procedia Computer Science 80 (2016), 2008 – 2018. https://doi.org/10.1016/j.procs.2016.05.521 International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA

[15]Feifei He, Xuedong Yan, Yang Liu, and Lu Ma. 2016. A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index. Procedia Engineering 137 (2016), 425 – 433. https://doi.org/10.1016/j. proeng.2016.01.277 Green Intelligent Transportation System and Safety

[16]Benjamin Coifman and SeoungBum Kim. 2009. Measuring Freeway Traffic Conditions with Transit Vehicles. Transportation Research Record 2121, 1 (2009), 90–101. https://doi.org/10.3141/2121-10 arXiv:https://doi.org/10.3141/2121-10

[17]F. Porikli and X. Li. 2004. Traffic congestion estimation using HMM models without vehicle tracking. In IEEE Intelligent Vehicles Symposium, 2004. 188–193. https://doi.org/10.1109/IVS.2004.1336379