Deep learning : Spatial-Temporal road traffic data congestion using Agglomerative Clustering

S. Senthilarasi, Research Scholar,
Department of Computer Science,
School of Computing Sciences
Vels Institute of Science, Technology & Advanced Studies (VISTAS)
Pallavaram, Chennai, India
senarasi80@gmail.com

Dr. S. Kamalakkannan, Associate Professor,
Department of Information technology,
School of Computing Sciences
Vels Institute of Science, Technology & Advanced Studies (VISTAS)
Pallavaram, Chennai, India
kannan.scs@velsuniv.ac.in

ABSTRACT

In the long run of road traffic data especially in urbanization due to people migration because of employment and other factors still it is a challenging issue in daily life to recover the congestion. In our paper , the causes of road traffic data and its measures have been shown. The algorithm to alternate the route for people in case of traffic has been displayed in flowchart. The different types of traffic sign detection to alert the people in common on road traffic has been discussed with colors. The estimation of road traffic congestion is computed using necessary formulas has been shown. Various measures have been calculated show the higher accuracy of agglomerative clustering compared with other models. The experimental study has been shown in Real Time traffic data of Chennai city from the year 2015 to 2021. In addition , drastic measures to control road traffic congestion has been discussed for present and future.

Keywords: *Traffic Congestion – Route Monitoring Algorithm – Traffic Sign Detection – Congestion Estimation – Agglomerative Clustering Schemes –Experimental Study – Radical measures to reduce congestion – Performance analysis*

I. INTRODUCTION

In recent times, deep learning techniques has gained rapid development, more number of experimenters in ITS started to acquire neural networks to predict extreme level traffic congestion. The more number of studies still mostly engaged with flow of traffic and its congestion[1].

At present , traffic congestion in citified road network retains a important and difficult task for researchers in transportation, particularly in improved traffic navigation systems as well as traffic management systems. Because of rapid development in big data and AI technology, it is considered that Deep Learning is one of the up-to-date congestion approaches. With the use of powerful clustering models it is applicable to have a considerable growth in range of spatial-temporal road traffic data[2].

Detecting and surveillance of traffic in real time along with extended-term judgment is preferable both for decision making and the common people. The increasing growth of population in big cities creates the ever high level requests of public transport acts as one of the important contributing elements of traffic bottleneck complications over the years. A longer travelling time makes the commuters to suffer a problem in relation to plan their journey in a smooth way. Due to dense population in developing countries, major cities are mourning from traffic congestion. Recognition and forecast of on-road traffic blockage plays an important part in an improvement of ITS(Intelligent Transportation Systems)[3].

In the state roads are strangled-vehicle population which has doubled within the ten years and in recent times, crossed the 3 Crores in number. Out of this mark, $1/5^{th}$ – 60 lakh vehicles are lying on Chennai roads that have been fighting for space. The automobiles crowd explosion has weakened traffic snarls, decreasing the average speed of vehicles from 33 to 20km/hour on city roads. In a recent survey of traffic congestion it has been found that travellers lost 5% of their income because of extensive fuel costs and delays of traffic[4].

A. Traffic Congestion:

In our country's economy, Traffic Congestion has both direct and indirect effect as well as residents health. Due to traffic congestion, nearly 1 million is spent everyday in the matter of opportunity amount and fuel consumption. It also

979-8-3503-4803-3/23/$31.00

affects an individual in society. Loss of time particularly during busiest hours, tension and the increased contamination to the climate change includes few essential aspects created because of traffic jam.

In general, five parameters such as Volume of traffic, Density of traffic, Occupancy, Index of traffic congestion and Time to travel are used to monitor and forecast the road traffic congestion[4].

B. Capacity and Control:

More precise and related real-time traffic details is able to direct a lot of developments in many areas.

- Reduced congestion
- Increased O-D matrices(traveller plans)
- Detection of traffic queue
- Increased incident management
- Enhancement of current infrastructures through a improved use of the existing road network
- Changing network traffic control
- Better information services eg., Details of traffic , dynamic route direction, road information signs etc.
- Better information quality transferred to individual drivers
- Better vehicle fleet management
- Minimized time of driving by reducing costs
- Scheme for future investments
- New outlooks in modeling transport; real-time data is able to setup changing transport models able to give forecasts in a short duration of time
- Minimizing fuel consumption(lesser CO_2 emissions) and discharges of air[11].

II. ROUTE MONITORING ALGORITHM

It is solely responsible for perceiving the route where the vehicle should be clear from traffic at duration of one-minute. If there exists any traffic congestion alerts near the vehicle (or) if any other alert on the existing vehicle route, the algorithm searches and allots a new route which consists of “No traffic congestion alerts”. In case, if the process of searching a new alert, way to clear route does not succeed or else the expiry of search time, the similar route is maintained[5]. The choice of alternate route in case of traffic congestion has been shown in Fig. 1.

III. TRAFFIC SIGN DETECTION

It is a specified application region of object detection. Traditional methods of Traffic sign detection depends on the details of shape and color[6].

A. Detection of Shape:

The count of shape edges are computed with an application of Douglas- Peucker algorithm. Formerly, the shapes are found with the the key bounding box to separate the Regions of Interest(ROI) from its remaining environment. Either an outer triangle (or) circle of the contours found are touched by the bounding boxes. In triangle sign, there are two triangles. The bounding box is touched by the outside triangle where the inside triangle is inner to the outside triangle and cannot handle the bounding box.

B. Detection of Color:

The most important color is noted in automatic traffic control system. The red color represents signal light denotes STOP. A red arrow denotes STOP till the green signal or arrow appears. Against a red arrow, a turn cannot be made. The yellow colors warns the red signal is around to appear. It also signifies the commuters has to stop a while to have a safe movement of vehicles. The Green color signifies a sign of nature and peace. The light resembles the concept opposite of danger. Also, this color calms the eyes and provides safe transit for commuters[7].

IV.ROAD TRAFFIC WITH CONGESTION ESTIMATION

In order to receive a traffic situation of driver’s nearby area, the method utilizes present GPS trailing of automobile users utilizing Android based Application. The Global Positioning System locates information transferred by the automobiles are clustered in application (local clouds). The present data transferred by the Application involves location of vehicles (latitude & longitude) of drive, fastness, duration, route etc.

The vehicles positions of GPS data are retrieved from the implementation is shown in a 2D- Problem space of size(n*n). The vehicle’s information is marked into motorist’s local map of roadways. The familiar Agglomerative Clustering is next executed to generate groups of road traffic with minimal Euclidean distance. The larger count of automobile positions within a group denoted greater density otherwise more traffic jam.

On the basis of similarity or dissimilarity measures , it is a procedure to divide the specific group of N points into x groups under the concept of N-dimensional Euclidean space clustering .

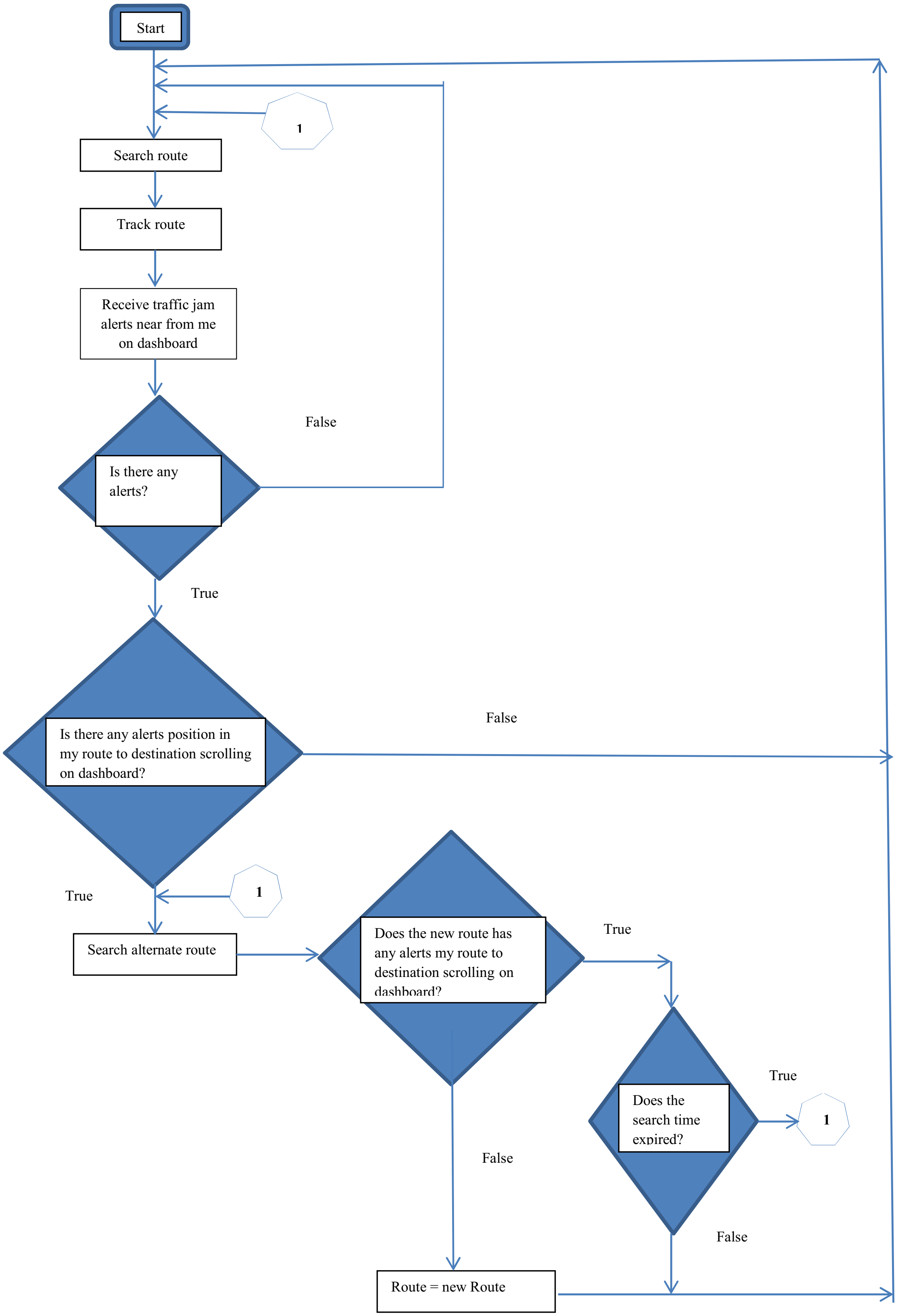


Fig. 1: Monitoring alternate route based on traffic

Let, a set S represents Y points (n1,n2,n3,….nY) and x clusters represents (a1,a2,a3,…ax) then

$$A_i \neq \phi \text{ for (each i = 1, 2, 3, … } x)$$

$$A_i \cap A_j \text{ for (each i = 1, 2, 3, … } x)$$

$$\text{for (each j = 1, 2, 3, … } x) \text{ and } i \neq j$$

$$\& \bigcup_{i,j=1}^{x} A_i, A_j = S \quad \text{………….(1)}$$

The network of roads can be defined as a neighborhood matrix, whereas every intersection point is expressed by means of hole and every margin express a road fragment. The basic weights are presumed to be one(1) if 2 vertex are connected each other directly that defines the travel time estimation[13].

A.Schemes in Agglomerative clustering:

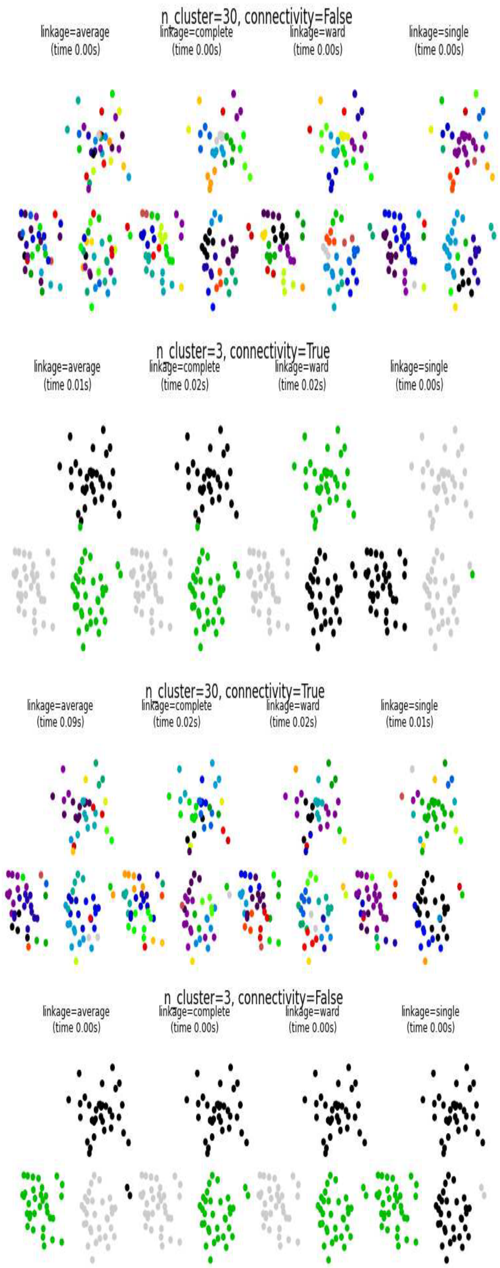
A portion of Hierarchical clustering - Agglomerative clustering is a most important and familiar method in Unsupervised machine learning. This type of scheme initializes from the datasets partition into individual nodes and step by step merging makes the existing set of commonly nearest networks into a fresh network there is left one last network, which consists of the entire data set[8]. The running time complexity of agglomerative clustering O(n3) is mathematically effective than running time complexity O(2n) of divisive clustering algorithm[9]. In finding small clusters Agglomerative clustering is best. The final result resembles different linkage of clusters formation with different clustering metrics has been shown in Fig. 2 when the algorithm completes its process[10].

Cluster analysis is the coordination of a set of patterns into clusters depends on similar features. Designs within an accurate cluster are same in maximum to each other compared to the design of different clusters.

B. Algorithm: Grouping similar data using Agglomerative Clustering

1. Import the necessary libraries
2. Read the traffic dataset
3. Create blobs with samples and its features
4. Create clusters with different linkages and affinity
5. Find True positive and False positive values
   (a) Generate the clusters of True Positives and False Positives using for loop
6. Create a graph consisting large number of homogeneous clusters
   (i) for index, linkage:
      (a) compute agglomerative clustering with parameters (type of linkage, connectivity , no. of clusters)
7. Compute the time to create clusters
8. Find the values of different metrics of clusters

The algorithm is generated with coding to cluster the similarities of road traffic data, which depends on the collision type due to crowded vehicles and its captured images on sensors. The following values represents the various clustering measures of Agglomerative Clustering.



```
Homogeneity:0.584
Completeness:0.935
v-measure:0.719
Adjusted Rand Index:0.560
Adjusted mutual information:0.712
silhouette coefficient:0.255
Calinski-harabasz score:49.161
Davies -Bouldin Score:0.871
Accuracy score TP: 94.66666666666667
Accuracy score FP: 5.333333333333333
```

Fig. 2: Generation of Clusters using Agglomerative Clustering

V. EXPERIMENTAL STUDY

The study is based on the road traffic congestion happened in various zones of Chennai metropolis. The dataset contains an impact of road traffic congestion of more than 40000 occurences in various zones of different types of victims from the year 2015 to 2021. Fig. 3.1 and Fig. 3.2 shows the total count of persons killed and injured due to on-road traffic congestion occurred in different Chennai zones from 2015 to 2021.

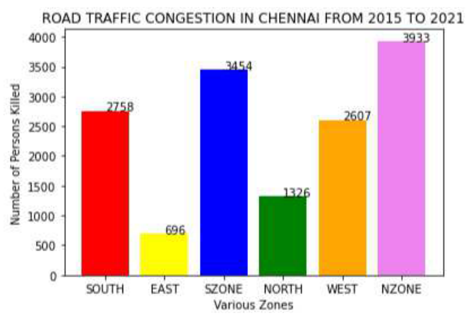


Fig. 3.1: Persons killed in road traffic congestion

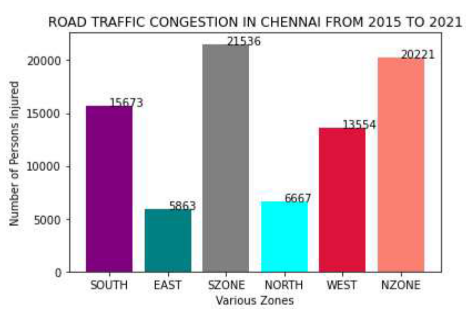


Fig. 3.2: Persons injured in road traffic congestion

On the basis of date and time format , road traffic congestion in various areas have been computed using time graph and shown in the Fig. 3.3.

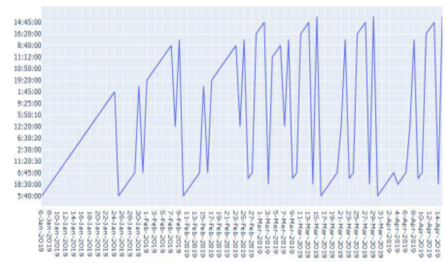


Fig. 3.3 : Traffic flow based on Temporal

VI. HARD-LINE AND RADICAL MEASURES TO RESTRAINT TRAFFIC CONGESTION

1. Development in non-motorised facilities to strengthen the usage of non- motorised facilities to strengthen an usage of motor-less transference methods
2. Development and initiation of an arrangement of mass transportation methods on recognized lanes
3. Accessibility to pile transport solution along potential of intermodal exchange terminals
4. Gradually dispose of obsolete vehicles due to inconvenience
5. Rapid and cost effectual intra-city railway system eg. Metro Rail System(MRTS)
6. Effectual, Speed and active traffic control methods to avoid traffic jams in an areas during peak hours
7. Superior road state with added overpasses and underpasses etc.
8. Rapid execution of norms such as Bharat Stage V and VI
9. Control on the large amount of automobiles in city by enforcing heavy tasks / tariff such as congestion charge, limiting entryway / usage of automobiles[11]
10. Carpooling is a better choice to minimize people subscription to the road traffic
11. Stricter execution of few of the above traffic rules can have a great effect on the quantum of road traffic[12].

VII. PERFORMANCE ANALYSIS

The performance have been achieved by an implementation of Agglomerative Clustering (GSDAC-Grouping Similar Data using Agglomerative Clustering) tested on dataset of on-road traffic congestion happened in Chennai metropolis at various zones from the year 2015 to 2021. While compared with other clustering types, most of the clustering quality measures have been proven with high results. The optimal number of clusters obtained using GSDAC has been proven with high accuracy. The types of measures has been evaluated for different clustering types including GSDAC has been shown in Fig. 4.1 to Fig. 4.9.

Table 1: Performance ratio of GSDAC with other types of clustering

Types of measures	GSDAC	Mean Shift Clustering	Spectral Clustering	BIRCH Clustering	DBSCAN Clustering	K-Means Clustering
Homogeneity	58.4	16	12	77	68.1	95.6
Completeness	93.5	16	21	41	33.8	47.2
V-measure	71.9	16	15	54	45.2	63.3
Adjusted Rand Index	56	2	3	8	65.7	38.5
Adjusted mutual information	71.2	2	3	6	39.8	60.8
Silhouette Coefficient	25.5	84.1	28	15.8	16.2	61.6
Calinski-harabasz score	49.161	2093.106	406.3	1.5	15.394	161.459
Davies-Bouldin score	87.1	22.3	12.246	15.122	2.019	84.9
Accuracy	94	34	31	13	34	19

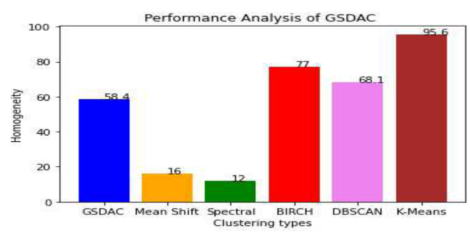


Fig. 4.1 : Homogeneity

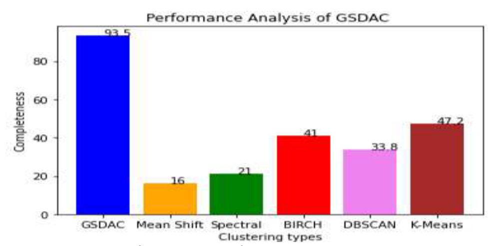


Fig 4.2 : Completeness

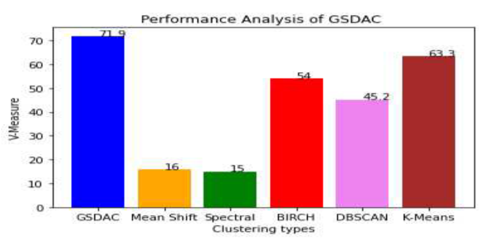


Fig. 4.3 : V-Measure

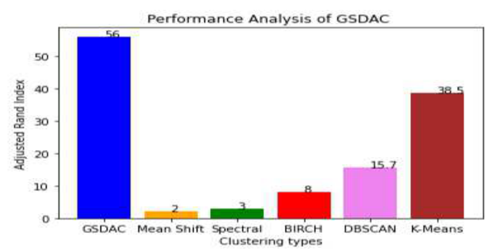


Fig. 4.4 : Adjusted Rand Index

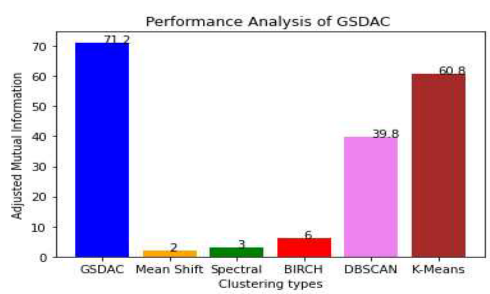


Fig. 4.5 : Adjusted Mutual Information

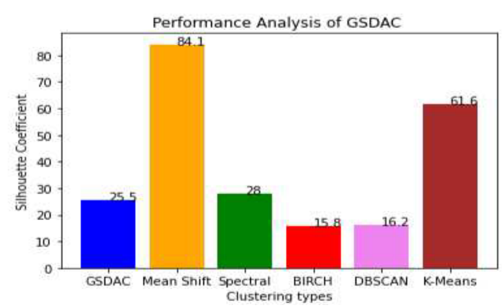


Fig. 4.6 : Silhouette Coefficient

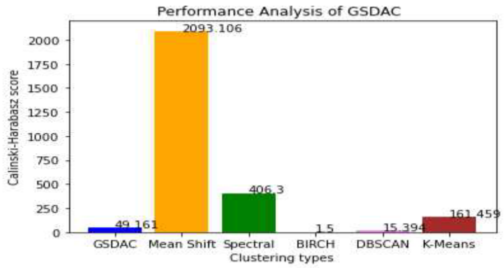


Fig.4.7 : Calinski-Harabasz score

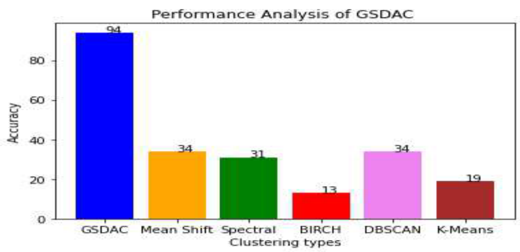


Fig. 4.9 : Accuracy

VIII. CONCLUSION

This paper describes the root causes for traffic congestion in Chennai. It represents the traffic congestion with capacity and control, route monitoring algorithm to alternate the movement of vehicles in case of traffic occurs in certain places of cosmopolis. In addition, the paper also deals with traffic sign detection on road, better representation of agglomerative clustering in spatial-temporal(vehicular traffic)data. Any other road traffic dataset can be applied in GSDAC to cluster the data and show the better performance especially in the formation of small number of clusters .

IX. FUTURE WORK

The clustering of road traffic data can be compared with other data clusters to identify the differences. The dataset can be analyzed to show the outliers while clustering. Further, this road traffic data can be implemented using RNN (Recurrent Neural Network) to yield additional outcomes.

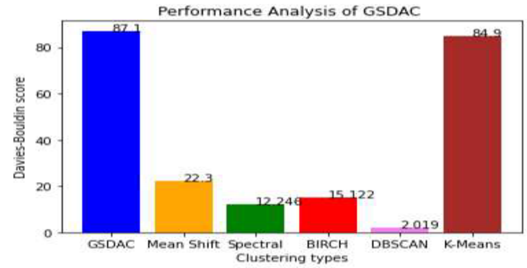


Fig. 4.8 : Davies – Bouldin Score

REFERENCES

[1] Jingyuan Wang, Qian Vu, Jungie Wu, Guannan Liu, Zhang Xiong, Traffic speed prediction and congestion source Exploration: A deep learning method, 2016, IEEE

[2] Guojiang shen, Chaohuan Chen, Qihong Pan, Si Shen, Shi Liu, Research on traffic speed prediction by Temporal clustering Analysis and convolutional neural network with Deformable Kernels, 2018, IEEE Access

[3] Noureen Zafar, Irfan Ul Haz, Traffic congestion prediction based on Estimated Time Arrival, Ploz one, Open Access

[4] Mahmuda Akhtar and Sara Moridpour, A review of traffic congestion prediction using Artificial Intelligence, Wiley Hindawi

[5] Nestor Cardenas- Benitez, Raul Aneuz, Aquino et.al, Traffic congestion Detection system through connected vehicles and Big Data

[6] Xie BangQuan, Werng Xino Xiong, Real-Time Embedded Traffic Sign Recognition using Efficient Convolutional Neural Network, 2019

[7] Aashrith Vennelakanti, Smriti Shreya et.al, Traffic Sign Detection and Recognition using a CNN Ensemble

[8] Daniel Mullner, Modern hierarchical, agglomerative clustering algorithms, 2011

[9] Sachin kumar and Durga Toshniwal, Analysis of hourly road accident counts using hierarchical clustering and Cophenetic Correlation Coefficient(CPCC), 2016

[10]Freecodecamp.org/news/8-clustering-algorithms-in-machine-learning-that-all-data-scientists-should-know/

[11] Joshuva Oliver A, Chennai- Alarming Vehicular traffic and the City's stand among global pioneers of sustainability

[12]Bankbazaar.com/insurance/motor-insurance-guide/simple-ways-to-reduce-traffic-congestion.html

[13] Vishwajeet Pattanaik, Mayank Singh, P.K. Gupta, S.K. Singh, Smart real-time traffic congestion estimation and clustering technique for urban vehicular roads, 2016, IEEE

[14] Zoe Bartlett and Liangxiu Han, Trung Thanh Nguyen and Princy Johnson, Prediction of Road Traffic flow based on Deep Recurrent Neural Networks, 2019, IEEE

[15] Abdelmoujib Benterki, Vincent Judalet, Manoui Choubeila and Moussa Boukhnifer, Long- Term Prediction of Vehicle trajectory using Recurrent Neural Networks, 2019, IEEE