Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021)
IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

Implementation of Spatial-Temporal Road Traffic Data using Agglomerative Clustering

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

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

2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) | 978-1-6654-1272-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICCS51141.2021.9432370

***Abstract*—At present, urbanization is extensively increased in its infrastructure and employment. Transportation of people from one place to another with the use of vehicles on roads parallel at the same time of travel leads to traffic congestion due to several parameters of traffic data. Control of traffic density is the challenging issue among the common people and traffic police simultaneously. The paper signifies the on road traffic and its safety points, types of traffic congestion and methodologies, implementation of agglomerative clustering (bottom-up approach and a part of hierarchical clustering) in different linkages and validity measures of clusters. In addition, data visualizations on traffic data are shown in series.**

Keywords: Road traffic, Traffic congestion, Traffic count methodologies, Agglomerative clustering, Spatial- Temporal data visualizations

I.INTRODUCTION

In search of employment, life style and comfort of people from remote places that migrate to urban areas, which causes a road traffic during the movement of people from one place to another place by using vehicles. Such urbanization in developing countries led to high population survival in metro cities like Chennai, Bangalore, Mumbai and Delhi. Because of the massive usage of vehicles for transportation on roads, a severe problem called "traffic congestion" has increased in urban areas [1].

A. Road Traffic and its Safety Points:

It is a fact that, authorities and common people have been informed about the necessity for an improved road safety and traffic consciousness. Yet, quite members of traffic safety points are present in our country.

- Commoner's Carelessness
- Road's inferior condition
- Insecure vehicle design
- Road safety standards in progress
- Mediocre government
- Insufficient proper law administration
- Shortage of emergency services

Commoner's Carelessness – It is absolutely true that, common people doesn't care about the road and safety points though they know that, breaking the rules leads to a risk factor.

Road's inferior condition – It is because of low infrastructure seen on road designs and improper maintenance of roads.

Insecure vehicle design – It is common to accept that motor vehicle's manufacturers not showing good attention to the safety measures of their vehicles to run on roads as they compared to sale.

Road safety standards in progress – The absence of road signs (or) markings which are necessary should be rectified in advance.

Mediocre Government – Commoners and motorists are often suffering from poor standards lying on road safety points due to insufficient proper plan of road projects authorized by the government. It is often seen by people, road projects in progress normally washed off in monsoon.

Insufficient proper law administration – The proper law has to come in active force to implement traffic and road safety points.

Shortage of emergency services – At the time of emergencies, a quickness of response team not readily available to hospitalize the sufferer.

With many challenges arise in cities due to transportation framework utilized by people residing in urban areas on everyday routine. The association of increased population and increased private vehicles for transportation known as Catalytic agent which leads to traffic congestions and high emissions of CO_2 spread and effect in environment. In spite of spending time and money in traffic jams, it could be used in productive work[2].

Countries in earning low and middle-income are affected in worst due to crashes nearly 3% of their Gross Domestic Product. Few sources are assigned to the people in usage of the roads, their manners, whereas the remaining sources are related with infrastructure of roadways, motor vehicles and implementation of laws. The crash patterns and its cost varies from state-to-state, region-to-region. The essential features in the development of road safety plans are thorough

study of crash models and its core causation elements. The execution of safety laws associated with wearing of seatbelt, helmet, speed of vehicles and its safeness might be properly matched to the righteous critical important sectors (or) countries-states. In India, road crashes are high in progress and stands the second largest system in the world[3].

B. Traffic Congestion:

Due to the extensive growth of population in urban areas and their requirement for drives and transportation,

- Category 1: Recurrent Congestion
- Category 2: Non-recurrent Congestion[5]

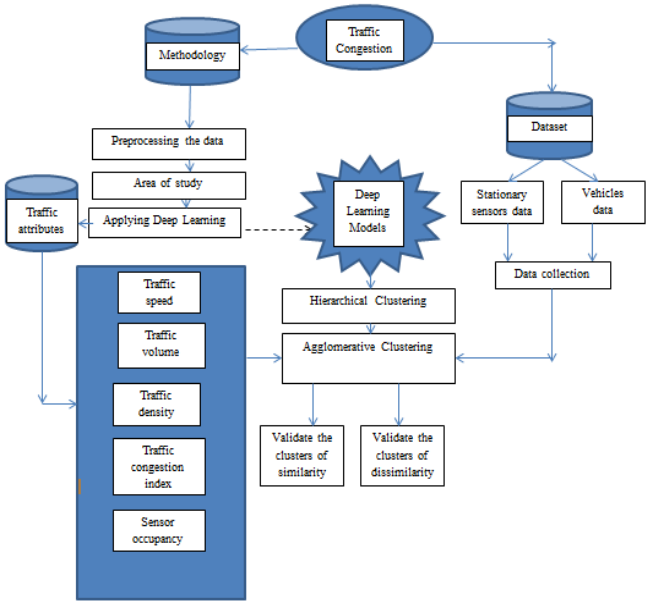


Figure 1: Proposed structure of Traffic Congestion

Intrusive Methods(IM):

Basically, a set of intrusive methods contains a data recorder and sensor that are either placed in or on the road.

Category 1: Recurrent Congestion:

The outcomes of features that creates uniform requirement flow on the transport system like driving, purchasing (or) weekend drives. The essential drives are primarily responsible for the pinnacles in traffic circulation flows, indicating merely half the traffic congestion in developed cities are reiterating at particular intervals of the day, particular portion of the transport system.

Category 2: Non-recurrent Congestion:

Non-recurrent congestion is connected to the existence and effectualness of event reaction plans such as accidents and abnormal weather conditions. The higher the vehicular traffic at particular road divisions increases the probability of accidents due to the uncertainty of traffic levels. The spatial focus of traffic creates a overload on transport infrastructures to a certain extent where traffic congestion can proceed to the complete immobilization of traffic, since both public and commuters shares the same infrastructures of roadways[6].

vehicular traffic issues such as congestion of road traffic and its safety measures are severely increasing on daily basis, since it reflects in financial growth and humans' lives. The vehicular traffic on roads are represented by evaluation, details retrieved from several traffic detectors [4].

Traffic congestion is a flaming matter in almost all developed cities because of a rapid growth of moving vehicles. The categories of traffic congestion are primarily consists of two:

C. Traffic Count Methodologies:

In general, traffic count technologies can be subdivided into two groups

- Intrusive Methods
- Non-intrusive Methods

rate density of the substance, which makes the appearance of potential difference between the electrodes.

Advantage:

It is used to count speed and weight of the vehicles crossing the lanes of roadways.

IM- Magnetic Loops:

It is the most standard advancement used in collection of traffic data. The loops are implanted in pavement to produce a magnetic field in a formation of square.

Advantage:

Details are transformed to a device that counts has been kept on the roadside

Disadvantages:

Because of limited life expectancy, the devices get affected by large vehicles, whereas not affected due to bad climate conditions.

The operation and servicing the device costs are too high

Non-intrusive Methods (NM):

These methods are generally based on distant observations (readings)

NM - Manual counts:

It is one of the most customary methods to accumulate traffic data by trained observers which cannot be effectively acquired via automated counts. For example – vehicle occupation rate, pedestrians and classification of vehicles. The commonly used equipment's in most cases are balance sheet, numbering boards in mechanical and electronic systems.

NM- Passive and active infra-red:

The existence, rapidity and kind of vehicles are found from the place of observation where the infrared power is emitted. The major difficulties are the outcomes of poor

accomplishments in case of bad weather and lane coverage constraints [7].

This paper illustrates the different metrics in evaluation of the optimal number of clusters within specified time is derived in agglomerative clustering along with the concept of DBSCAN. Also, the visualization of traffic data in different areas of Chennai region has been shown in plots.

II RELATED WORK

Recently, a quite number of papers have been published by various authors in Spatial-Temporal combination on road traffic data by using different Clustering techniques. The following tabular form depicts the study of different papers related to road traffic [8],[9],[3],[2],[10-14],[15-19] .

Table 1: Study of Spatial-Temporal Traffic data

S.NO.	AUTHOR	TITLE OF THE PAPER	CLUSTERING TECHNIQUE	CONTRIBUTIONS	MERITS	EXPERIMENT STUDY
1	Sachin kumar and Durga Toshniwal	Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC)	Hierarchical clustering	Examines road accident data for every hour with the use of CPCC. Group the sequences of traffic data in support of clustering	Ability in finding clusters of good quality	Road accidents, count of 26 districts of Gujarat state from January 2010 to December 2014
2	E.Vijaysekar, J. anuradha, Anshita Arya, Balamurugan Balusamy, Victor Chang	A framework for smart traffic management using hybrid clustering techniques	DBSCAN clustering	Forecasts the closely crowded roads on the basis of the existent and earlier traffic congestions Proposes the different ways for the specified origin to end. Increasingly process the limitation of live flow on online traffic data.	The most convenient route with factors such as travel, traffic and distance are suggested for real –time responses	Real time live stream of data
3	Avijit Maji, Nagendra R. Velaga and Yohan Urie	Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies	Hierarchical Agglomerative Clustering	Demonstrates the suggested substructure of the case history on Indian states and Union Territories. Modification of new approaches and plans implemented in censorious parts.	The study on states of India and Union Territories are used to explain the presented structure	Case Study of India - 29 and 7 Union Territories
4	Paulo Figueiras, Gulherne Guerreiro et.al.	Real-time monitoring of Road Traffic using Data stream mining	CRISP – DM(Cross Industry Standard Process for Data Mining)	Monitors real-time road network on continuous (24X7) basis Detects traffic events in real-time and data collections from different fixed sensors	Traffic operators values are extended by providing existing information to visualize the road network which is not accessible in	349 inductive loop counters, located in Slovenian road network

					real-time	
5	Mouna Elloumi, Riadh Dhaou, Benoit Escrig, Hanene Idoudi	Monitoring Road Traffic with a UAV – based system	Uniform Method	Generates adaptive UAVs trajectories – a method on the basis of covering the running points in view. Design vehicles mobility using mobility models. Observes and collects the real-time traffic to a center to regulate traffic process.	Related to coverage and detecting events in comparison to techniques , the uniform method provides better implementation build on fixed trajectories	Taxicabs in Rome, Italy and taxicabs in sanfrancisco, USA
6	Fahdah Alalyan, Nuha Zamzami and Nizar Bougila	Model-Based Hierarchical clustering for categorical data	Hierarchical clustering	Groups categorical data built on Multinomial, Bernoulli mixture models. Comparative study among two broadly used density-based distances specifically Bhattacharya and Kullback-Leibler. Large tests on congregating text and images with the sack of visual models.	This approach is applicable for different application including hierarchy structures, count(or) binary data	First data set- 7 sectors, 4581 HTML articles divided into 7 classes Second data set – Webkb4, 4199 web pages divided into 4 clusters Third data set – NIPS dataset, 391 files in 9 different groups.
7	Dongwei Xu, Chenchen Wei et.al	GE – GAN: A novel Deep learning framework for road traffic state estimation	Deep walk – Graph embedding technique, GAN	Application of Deepwalk to acquire the spatial depiction in the road network. Provides better performance with the use of GE-GAN.	GE-GAN model executes better than KNN, Deeptrend 2.0, BGCP and LSTM models	Traffic volume data – Caltrans peMS - higher than 15000 detectors. Speed data – seattle area, 323 detectors per year
8	Sergii Babichev, Bohdan Durnyak, Iryna Pikh and Vsevoiod Senkivskyg.	An evaluation of the objective clustering inductive technology effectiveness implemented using Density-based and Agglomerative Hierarchical clustering algorithms	DBSCAN clustering OPTICS clustering and Agglomerative hierarchical clustering	Empirical evaluation of OPTICS density-based clustering algorithm within the system design. Comparative analysis among OPTICS, DBSCAN and Agglomerative hierarchical clustering algorithms.	Density-based clustering algorithms prove the high effectiveness compared to agglomerative clustering because of object group in more detail.	School of computing- data aggregation, compound, multi shapes and jain , Eastern Finland University
9	Muhammad Tauhidur Rahman, Arshad Jamal and Hassan M.Al-Alamadi	Examining hotspots of traffic collisions and their spatial relationships with land use: A GIS-Based geographically weighted Regression	Geographically Weighted Regression (GWR)	Untangles the relationship in the midst of traffic crashes with density of population and non-rural land by utilizing the GWR method. Supplies necessary directions for road safeness agencies, policy makers.	It supports perceptions used for understanding better the complex interrelationship among traffic-it's safety and use of land	Vehicular crash dataset – between the years 2009 and 2016 collected from Dammam traffic department

		approach for Dammam, Saudi Arabia.				
10	Andrew Moses, Parvathi R.	Vehicular Traffic analysis and prediction using machine learning algorithms	Centroid and Hierarchical methods	Generation for an evaluation metric – cross validation score for the different models. Extensive use of Grid search technique to refine the random-forest design.	It is an exact model to the dataset available in public	US department of transportation
11	Mostafa Ghazizadeh – Ahsace, Afsaneh Shamasadini-Farsangi	Developing of a New Hybrid Clustering algorithm based on Density	HDDA Algorithm	Attains reliable speed and synchronous results by the use of HDDA. Extends dividing the clusters	Avoidance of clusters in irregular disruption, ability to find clusters with varying sizes and shapes	Synthetic and artificial data set
12	Yirong Zhou, Jun Li, Hao Chen, Ye Wu, Jiangjiamg Wu, Luo Chen	A Spatio temporal hierarchical attention mechanism based model for multi-step station-level crowd flow prediction	ST-HAttn	Ability to capture the local/global interrelations of group movement in the midst of stations. Achieves spatial awareness procedures among states and regions.	Compared to the state-of-the-art techniques the performance of St-HAttn is better	New york city(MTA subway), MTA turnstile data, CitiBike trip data
13	Amin Mohammadn azar, Ramin Arvin, Asad J. Khattak	Classifying travelers driving style using basic safety messages generated by connected vehicles: Application of Unsupervised machine learning	K-Means, K-Medoids	Driving study style below a realistic moving condition in several road types for controlling BSMs caused CVs. Qualification of instant driving attitude and model by improving different temporal volatility measures.	In location-based services, the data found can be applied in for the improvement of vehicles safeness on varying types of roadways	Fuel Consumption and Emission Studies
14	Arash Khoda Bakhshi, Mohamed M.Ahmed	Principal advantage of crossed random intercepts under Bayesian hierarchical modeling to tackle unobserved heterogeneity in clustering critical versus non-critical crashes	Nested factors versus Crossed factors	Cross – Classified Random Effects Modeling (CCREM) is accomplished by the use of two-crossed random arrests to group serious crashes, including crucial (or) disabled wounds Vs non-serious crashes.	The critical crash risks consists of worthwhile variations to varying composites of environmental factors accordingly	402 miles of I-80 in Wyoming – crash report, roadway geometry –properties, weather conditions and observations of real-time traffic.
15	Mahmuda Akhtar and Sara Moridpour	A review of Traffic Congestion Prediction using Artificial Intelligence	Probabilistic Reasoning, ML algorithms	Selectively outlines the present research dealt by inflicting the several techniques of Artificial Intelligence. Collects the various models below specific divisions of Artificial Intelligence.	The methodologies are divided into three different classes for users study	PeMS – collects the state of California of traffic flow, 1-5 highway in SanDiego, California.

III IMPLEMENTATION OF DBSCAN IN HIERARCHICAL CLUSTERING

In general, hierarchical clustering procedures divides the entities into a tree of networks illustrating a datasets underneath structure of clusters, in which each network constitutes a potential cluster. Hierarchical algorithms necessitate less user-defined parameters and are indifferent arrangement of objects. The study of cluster structure makes a user to scratch the connected branches at several size of grossness and finds sub clusters within clusters.

In clustering categorical data, agglomerative (bottom-up) – a method of hierarchical clustering which necessitates to be applied to detect similarities of objects placed in clusters at several levels of emerging tree[20].

A. Linkage Criterion:

The linkage criteria establish the distance in the midst of observations set as a task of the duo distance between observations.

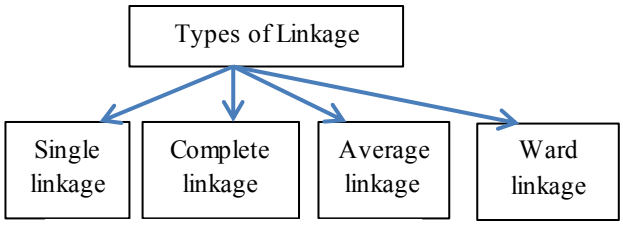


Figure 2: Linkage types

Linkage – Single:

The minimum distance between two clusters(or) the distance among the nearest objects. The linkage may create the chaining problem, by forcing clusters to each other[20].

For two clusters P and Q, the single linkage departs the minimum distance among two points a and b such that a and b belongs to P and Q[21].

$$L(P,Q) = \min(D(a,b)), a \in P,\ b \in Q$$

Linkage – Complete:

The maximum distance between two clusters are useful in high-dimensional space if the objects are distantly located. The linkage is not suitable for extremely noisy datasets, as anomalies are given more weightage in the decision of clusters [20]. For two clusters P and Q, the complete linkage departs the maximum distance among two points a and b such that a and b belongs to P and Q[21].

$$L(P,Q) = \max(D(a,b)), a \in P,\ b \in Q$$

Linkage – Average:

The mean distance among group of objects is taken from two clusters. The linkage is more expensive in computation compared to other methods. It is the most popular linkage among the other linkages, which avoids the chaining problem and no special weightage to outliers [20].

For two clusters P and Q, initially for the distance among any data point a in P and any data point b in Q, arithmetic mean of these intervals are computed. Average linkage recurs the arithmetic mean value as

$$L(P,Q) = \frac{1}{n_P + n_Q} \sum_{a=1}^{n_P} \sum_{b=1}^{n_Q} D(a,b),\ a \in P,\ b \in Q$$

Where n_P = data points counts in P

$n_{Q=}$ data points counts in Q[21]

Linkage – Ward:

It is the utmost appropriate method for quantitative variables.

For two clusters P and Q, the distance is computed on the basis of the sum of squares increased and merged values.

$$\Delta(P,Q) = \sum_{a \in P \cup Q} \| \vec{X}_a - \vec{m}_{P \cup Q} \|^2 - \sum_{a \in P} \| \vec{X}_a - \vec{m}_P \|^2 - \sum_{a \in Q} \| \vec{X}_a - \vec{m}_Q \|^2 = \frac{n_{P \cap Q}}{n_P + n_Q} \| \vec{m}_P - \vec{m}_Q \|^2$$

Where $m_b^{\rightarrow}$ acts as the center of cluster b and n_b represents the numeral points in it, is referred as the combining cost of the clusters P and Q.

In case of hierarchical clustering, the growth of merging clusters starts from zero with the sum of squares calculations [21].

IV EXPERIMENTAL STUDY

The Traffic data is based on California-Traffic solution-Data from SWITRS [28]. The data includes the collisions from January 1st 2001 till the mid of October 2020. The dataset contains three main tables, consisting of collisions, parties and victims. The collision table contains 74 field names related to traffic described with 9.17m rows. The parties table holds the information about the people's age, gender and serious behavior included in collision. The victims table consists of the details related to the people got injury in the collision.

Due to high density population in Chennai- the capital of Tamil Nadu.. Every day at the moment of peak hours high traffic congestion occurs, which becomes a challenging in the concept of spatial-temporal data. In our discussion, Chennai traffic at different regions has been collected and analyzed in study are plotted in visualizations.

The novelty of this paper shows the estimation of clusters by using DBSCAN concept in the implementation of agglomerative clustering on different types of linkages within

short duration. The road traffic visualization has been shown in different plots.

A. ALGORITHM: Spatial – Temporal Road Traffic data congestion

Input: Dataset – Traffic congestion (1,2,..n)
Output: Actual number of clusters
1.Import the required libraries
2.Repeat the steps
(i) Create circles on datasets
(ii) Distributes the data to variables
(iii) Create blobs with different variances
(iv) Determine the cluster parameters
(v) Cluster parameters are set
(vi) Update parameter with dataset-specific attribute values
(vii) Normalize the data
(viii) Create cluster objects
(ix) For i ← 1 to n
(a) Cluster the similarities of data sets in specified time of different methods
(b) Cluster the dissimilarities as outliers indicated in black dots until all the clusters are formed
3. Normalize the data set to fit in variable X of DBSCAN
4. Compute the metrics of clusters in different forms
5. Validate the outcomes

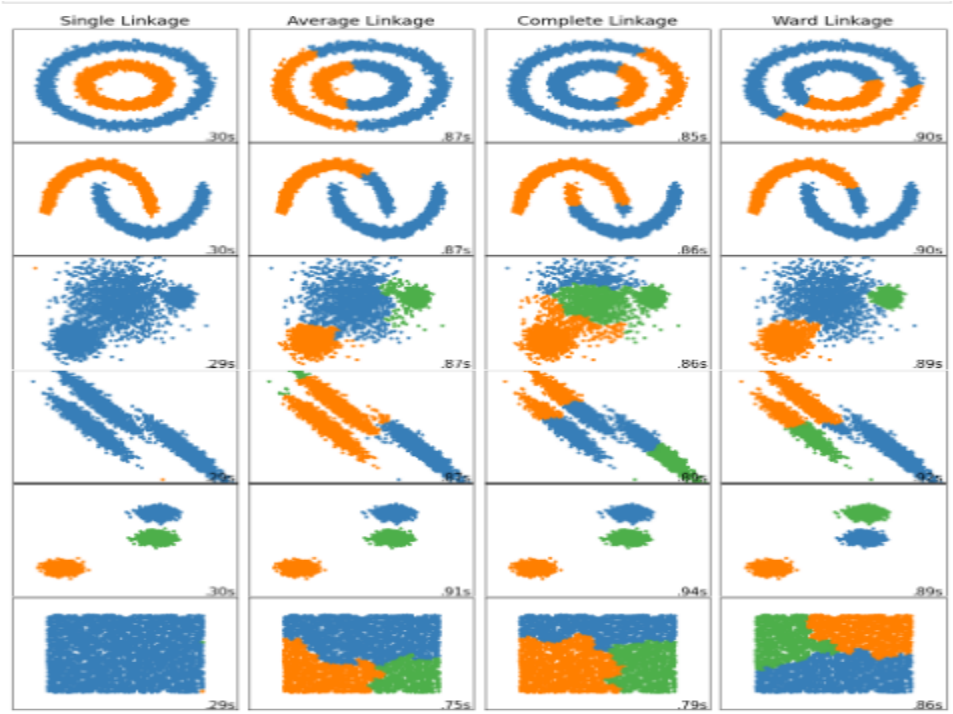


Figure 3: Different types of linkages clustering

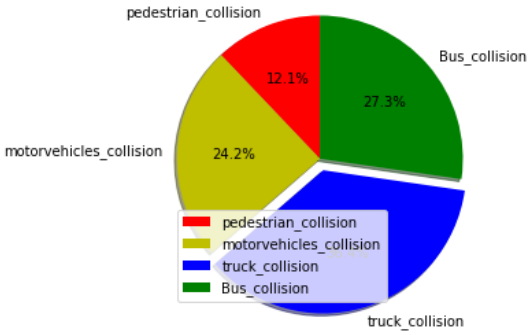
Table 2: Measures of clustering data

S.NO.	CLUSTERING MEASURES	VALUES
1	Estimated number of cluster	1
2	Estimated number of noise points	86
3	Homogeneity	0.159
4	Completeness	0.431
5	V- measure	0.232
6	Adjusted Rand Index	0.078
7	Adjusted Mutual Information	0.221
8	Silhouette Coefficient	0.177
9	Calinski – Harabasz Score	18.262
10	Davies – Bouldin Score	0.939

B. Spatial – Temporal – Traffic data visualizations:

Pie chart:

A piechart is a statistical plot in circular form which can represent only one sequence of data. Piechart is a typical method to depict the composition of groups. The complete proportion of data about the types of collision occurs on road traffic is shown in wedges of the piechart[23]. Nevertheless, it is recommended high to write down explicitly the percentage (or) numerals for every portion of the pie.



.

Figure 4 : Pie chart - Different types of collision

Joyplot:

Joyplots are piled and partly overlapping density plots. It provides a facility to lay data to compare distributions visually, particularly the changes occurs at one dimension. This type of plot is also referred as *ridgeline plot* [24]. Joyplot permits the solidity curves of several groups to intersect; it is a prominent method to envision the spread of huge groups in connection to each other. It depicts the correct information clearly to the users. In our study, road traffic occurs in Chennai regions have been shown in joy plot.

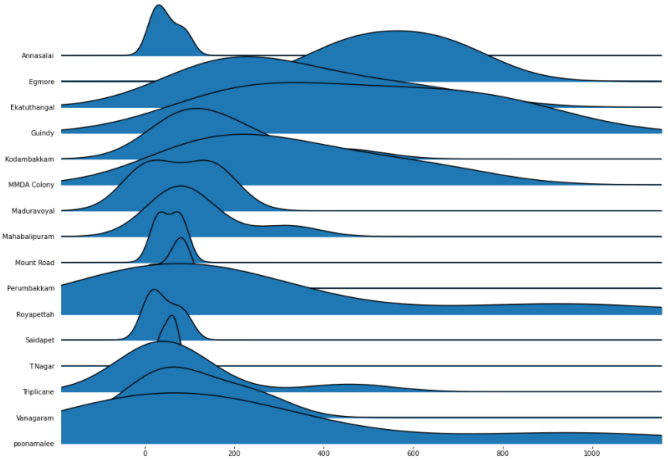


Figure 5: Joyplot- Congestion in peak hours at different places

Barchart:

A bar plot (or) a chart represents the type of data in rectangular bars based on its catetgory plus lengths and heights are comparable to the values. The bar plots can be marked either in horizontal (or) in vertical manner. It details the resemblances among the discrete categories [25] . Barchart is a typical method of envisioning objects builds on numbers (or) any one given metric. Depending on data, different colors can be used to represent items arranged in groups.

X – axis- represents the city location of Chennai

Y- axis – represents the vehicle ID

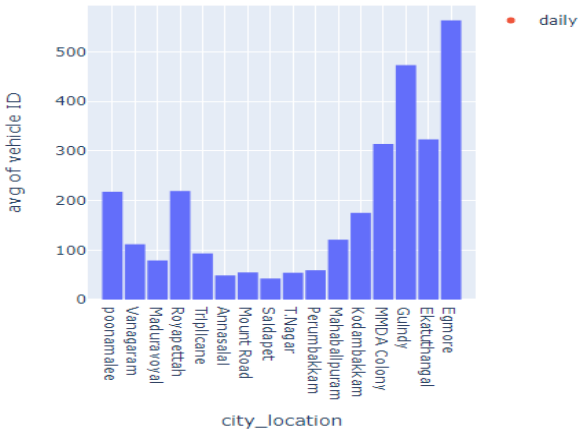


Figure 6: Barchart – Traffic congestion of vehiclesin different areas

Dendrograms:

A dendrogram is a network configuration, which can be established of a base node called root, that causes birth to various nodes associated through edges (or) branches [26]. It is an advanced feature in agglomerative clustering approach; groups' similarities altogether depends on the distance metric and arranges them in tree like nodes built on the points similarity [22].

The road traffic data in different regions of a city is shown
in terms of

- Horizontal lines – Merge of Clusters
- Vertical lines- Formation of new clusters
- Heights – The heights on horizontal lines shows the distance to be connected in forming the new clusters[27].

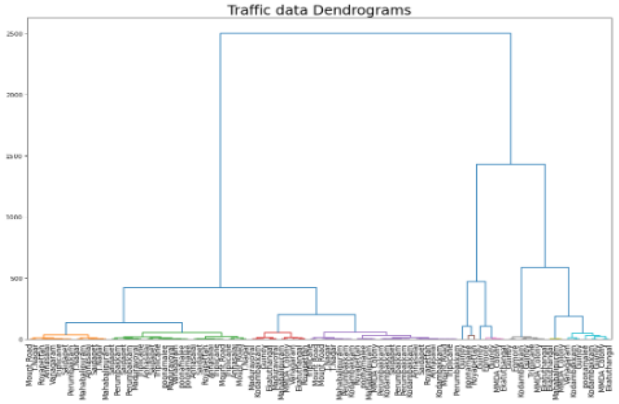


Figure 7: Dendrogram – Traffic congestion in different areas

V. DISCUSSION AND FUTURE WORKS

The road traffic data is implemented with different linkages and its cluster measures in bottom-up approach of agglomerative clustering is shown. Visualizations of traffic data is shown in different plots on the study of experiment with Chennai-road traffic. The limitations of this proposed model results in lack of analysis of data. As a future work, the agglomerative clustering has to be compared with other clustering models to prove its clustering accuracy, better representation of clusters and determining the reason of its disadvantages.

VI. CONCLUSION

This paper is focused on the on road traffic data occurring in urban areas because of population density in extreme level, comparison with remote places . The traffic congestion occurs due to parallel movement of vehicles at the same place and at the same time, the classification and its flow have been studied on California-Traffic solution-Data from SWITRS. The proposed model of Agglomerative Clustering has been applied to different linkages and evaluated the optimal number of clusters on road traffic data using validity metrics. The visualization of road traffic data has been shown in different plots. This paper benefits the users to know the evaluation of clusters using different measures applied in DBSCAN clustering implemented in Agglomerative clustering in different types of linkages. The limitations of this model results in lack of analysis of data during its process of clustering. The proposed study on this paper is just one step forward shows the better representation of clustering, where the concept has to be further enhanced by the comparative study with other clustering techniques to prove the increased level of accuracy as well as to detect the facts that arises the lack of analysis on data.

REFERENCES

[1] Vishwajeet Pattanaik, P.K.Gupta, S.K. Singh, Mayank Singh, Smart real-time traffic congestion estimation and clustering technique for urban vehicular roads, IEEE , 2016
[2] Paulo Figueiras, Guilherne Guerreiro, Ruben Costa, Zala Herga, Antonio Rosa, Ricardo Jardim-Gonebalves, Real-Time Monitoring of road traffic using Data Stream Mining, IEEE International Conference on Engineering, Technology and Innovation, 2018
[3] Avijit Maji, Nagendra R. Velaga and Yohan Urie, Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies, International journal of Injury control and Safety Promotion, 2017
[4] Donwei Xu, Chenchen Wei, Peng Peng, Qi Xuan, Haifeng Guo, GE-GAN: A novel deep learning framework for road traffic state estimation, Transportation Research part C, Elsevier, 2020
[5] Kapileswar Nellore and Gerhand P.Hancke, A survey of Urban traffic management system using wireless sensor networks, mdpi.com/1424-8220/16/2/157/htm,2016
[6]transportgeography.org/contents/chapter8/urban-transport-challenges/
[7] congestionmeasures/KYTC, transportation.ky.gov
[8] Sachin Kumar and Durga Toshniwal, Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coeffient(CPCC), 2016

[9] E. Vijaysekar, J.Anuradha, Anshita Arya, Balamurugan Balusamy, Victor Chang, A framework for smart traffic management using hybrid clustering techniques, 2017
[10] Mouna Elloumi, Riadh Dhaou, Benoit Escrig, Hanene Idoudi, Monitoring Road Traffic with a UAV-based system
[11] Fahdah Alalyan, Nuha Zamzami and Nizar Bougila, Model-Based hierarchical clustering for categorical data, IEEE, 2019
[12] Serii Babichev, Bohdan Durnyak, Iryna Pikh and Vsevoiod Senkivskyg, An evaluation of the objective clustering inductive technology effectiveness implemented using Density-Based and Agglomerative Hierarchical algorithms, 2020
[13] Muhannad Tauhidur Rahman, Arshad Jamal and Hassan M. Al-Alamadi, Examining hotspots traffic collisions and their spatial relationships with land use: A GIS- Based geographically weighted Regression approach for Dammam, Saudi Arabia, 2020
[14] Andrew Moses, Parvathi R, Vehicular Traffic analysis and prediction using machine learning algorithms, 2020
[15] Mostafa Ghazizadeh- Ahsace, Afsaneh Shamasadini-Farsangi, Developing a New Hybrid clustering algorithm based on density, 2020
[16] Yirong Zhou, Jun Li, Hao Chen, Ye Wu, Jiangjiang Wu, Luo Chen, A Spatio temporal hierarchical attention mechanism based model for multi-step station-level crowd flow prediction, Elsevier, 2020
[17]Amin Mohammadnazar, Ramin Arvin, Asad J.Khattak, Classifying travelers driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning, Elsevier, 2021
[18] Arash Khoda Bakhshi, Mohamed M. Ahmed, Principal advantage of crossed random intercepts under Bayesian hierarchical modeling to tackle unobserved heterogeneity in clustering critical versus non-critical crashes, Elsevier, 2021
[19] Mahmuda Akhtar and Sara Moridpour, A review of Traffic congestion prediction using Artificial Intelligence, Wiley Hindawi, 2021.
[20] Bill Andreopoulos, Clustering Categoricl data, Wiley StatsRef, 2917
[21]https://en.wikipeadia.org>wike>Hierarchical_clustering
[22]machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-masterplots-python/
[23]geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/
[24] github.com/sbebo/joypy
[25] geeksforgeeks.org/bar-plot-in-matplotlib/
[26] python-graph-gallery.com/dendrogram/
[27] joernhees.de/blog/2015/08/26/scipy-hierarchical-clustering-and-dendrogram-tutorial/
[28] kaggle.com/alexgude/California-traffic-collision-data-from-switrs