

A RESEARCH PAPER ON IMPROVING TRAFFIC CONGESTION PROBLEM BY IOT FOR SMART CITIES

SURJEET SINGH¹, DEEPAK KUMAR²

M.Tech Scholar¹, Assistant Professor²

1 CSE Dept, Sri Sai College of Engineering & Technology, Badhani Pathankot, India

2 CSE Dept, Sri Sai College of Engineering & Technology, Badhanii Pathankot, India

Abstract –

Traffic Management causes drivers' disappointment and costs billions of dollars yearly in lost time and fuel utilization. So as to beat such issues, this paper exhibits a component for Intelligent Transport Systems, which expects to identify and oversee activity clog. Restriction of movement on street systems is only lesser speeds, flourished outing time and flourished lining of the vehicles. At times when the quantity of vehicles increases than the capacity of the street, movement blockage occurs. In the urban areas of India, traffic clogging is an important issue of concern. Activity blockage happens when the request exceeds the approachable street limit. This way, the concentration is to diminish an opportunity to prepare, reroute and inform vehicles. Traffic flow forecast is the key purpose of Intelligent transportation frameworks investigate and in addition the vital condition for movement administration, control and direction. Presently conventional figure strategies and models incorporate nonparametric relapse demonstrate, exponential smoothing, time arrangement examination, counterfeit neural system, Kalman separating, movement re-enactment, Euclidian distance, dynamic activity task et cetera. The success of any framework depends on the forecast of traffic it has. The improvement in traffic can corrupt the execution of the framework. So it is important that precise estimation can be given.. So as to improve the expectation rate we propose a half breed method that uses the ARIMA model, KNN and Euclidean distance. Here and now activity stream is one of the center advances to perceive movement stream designs. In this exchange, in connection to the traits that the development of movement changes over and over, a fleeting activity stream forecast technique in light of a three-layered K-closest neighbor non-parametric relapse calculation is proposed. Especially, two screening layers in view of the similitude of shape were presented in K-closest neighbor non-parametric relapse technique, and the consequences of expectation were yield with the assistance of the weighted averaging on the complementary estimations of the likeness of shape separations and the strategy for most-comparable point remove modification.

Key Words: Traffic Congestion Prediction (TCP), IoT, K- Means, FCM, KNN, Euclidean distance etc.

1. INTRODUCTION

Traffic monitoring is one of the greatest challenges and need of the hour as well. In order to accomplish traffic monitoring various techniques has been devised. Most of the techniques focus on traffic monitoring based on clustering. Clustering is accomplished with the help of K- Means, FCM, KNN, Euclidean distance etc. The sensors are required in order to capture the information related to the flow of traffic. The actions performed by the people and the vehicles are captured using camera on the roads in case of traffic observation. This information is stored within the dataset.

Proposed literature focuses on KNN and Euclidean distance mechanism for predicting traffic. This section provides in depth into various clustering mechanisms available to be used for traffic analysis.

Road traffic blockage discovery and administration has been a test and a few nations have thought of new ideas and thoughts to distinguish clog. Today, number of vehicles is expanding at an immense rate and change in road infrastructure, which has its constraint, has not possessed the capacity to facilitate the issue. Throughout the decade, numerous creative ideas and innovations have been produced in connection to gathering continuous traffic information and utilize it for picking up learning with respect to the different parts of traffic stream. Both quantitative and subjective data can be gathered utilizing such systems. The quantitative information include: a) normal path speed b) line length c) trip time d) holding up time, and so forth.

The qualitative data incorporate, for example, distinguishing and describing blockages at different site of incidents. All the above information can be utilized as an appropriate parameter for deciding the idea of traffic stream. For instance, trip time is the normal time taken by a vehicle to traverse extend of path or starting with one intersection then onto the next and interface travel time is the aggregate time to go from an offered source to a given goal.

There is quick augmentation in the vehicles perhaps activity blockage out and about is furthermore augments. By and by, it is an extreme issue in light of the way that on street, more number of vehicles is accessible and as a result of that the stop up in rush hour gridlock extended close by the extension in sullying and extra time is misuse on street movement every day. Along these lines, activity state estimation is a champion among the most recent issue in ITS which accept a critical part to lessened the movement time, improve the activity profitability, atmosphere condition, et cetera. Diverse sensors are open in the market which are useful for the ITS. Here we affirm our work using simply couple of sensors like: IR sensor, temperature sensor, gas sensor, sensor display, et cetera.

2. SUPERVISED LEARNING

Supervised learning, with respect to electronic thinking (ET) and machine learning, is a kind of structure in which both data and needed yield data are given. Data and yield data are named for game plan to give a learning reason to future data taking care of.

Supervised machine learning frameworks give the learning calculations known amounts to help future

judgments. Chatbots, self-driving autos, facial acknowledgment programs, master frameworks and robots are among the frameworks that may utilize either supervised or unsupervised learning. Supervised learning frameworks are for the most part connected with recovery based AI however they may likewise be equipped for utilizing a generative learning model.

Steps in Supervised Learning:

Although there are various Statistics and Machine Learning Toolbox computations for supervised learning, most use a comparative essential work process for getting a marker illustrate. (Point by point rule on the methods for get-together learning is in Framework for Ensemble Learning.) The methods for supervised learning are:

1. Data Initialization
2. Algorithm Selection
3. Model Selection
4. Validation applying strategy
5. Check curve fitting and perform above steps until condition satisfied.
6. Obtain fittest value for prediction.

3. RELATED WORK

To prove the worth of the study, literature survey is conducted describing existing work.

Chenetal describes that traffic prediction system is generally expensive in nature. The expensiveness causes this critical system to be utilized by only limited personnel. In order to tackle such issue, a MATLAB based traffic prediction system is proposed through this literature. In this mechanism data is collected from the camera, and then format is converted according to requirements. In other words normalized data is obtained. From the normalized meaning information is extracted and then predictions are generated. Real time information is not tackled and prediction due to C code may not be accurate. So some modifications in terms of accuracy are desired.

Guo et al. n.d. proposed KNN as one of the most commonly used approach used to enhance performance in terms classification. The characterization as far as neighbors was watched. Estimation of K chooses the quantity of neighbors to correspond with. Higher the estimation of K more will be the connection. This writing utilizes KNN on the criminal dataset. This paper applies KNN to enable criminological agents in distinguishing the glass to type. It additionally checks if coordinating KNN with another classifier utilizing voting can upgrade its precision in distinguishing the glass compose. The principle objective in this paper is to arrange effectively a solitary piece of glass in light of the primary segment estimated. The proposed information digging calculation for glass issue is k-closest neighbor grouping strategy which is a standout amongst the most ordinarily utilized information mining methods in design acknowledgment and characterization issues. [9]

Chen n.d. states that half and half approach for order. KNN was converged with three calculations to be

specific, Novel Byes, SVM and C4.5. K closest neighbor (KNN) calculation is a lethargic classifier that makes forecasts as per names of K closest neighbors of test cases. C4.5 calculation was an agent choice tree calculation that settles on expectations through creating a choice tree. Credulous Bayes (NB) classifier makes forecasts as per Bayesian hypothesis. Bolster vector machine (SVM) depended on factual learning hypothesis (SLT). In this paper, investigation of strategy for consolidating KNN calculation and other three classifiers is incorporated. [6]

Tan et al. 2009 concentrated diverse changes of KNN calculation and have proposed a novel calculation. This calculation was required to diminish the wastefulness of conventional K closest neighbor calculation. The proposed calculation was partitioned in two sections: Data pre-preparing: The information pre-handling part gives the weight to various traits, decides k esteem for the test tests and furthermore separates the preparation informational collection into various groups. It brings about production of model used to group future information. This part runs just once, so it would not influence the proficiency. Characterization: The genuine order of test information was done in this part. This part runs each time when it does arrangement. [20]

4. PROPOSED WORK

Proposed system uses KNN with Euclidean distance for predicting traffic. (Zhou et al. 2014)[13]Traffic prediction is based on dataset derived from online source. Traffic dataset is used for accurately determining traffic at future time intervals. The primary parameters considered from improvement include accuracy and error. The prediction process is described considering the following example where we have taken different values of K for all the three lanes. Then the corresponding error and accuracy are listed based on which graphs are generated for both error and accuracy.

a. IMPLICATION OF KNN

The K nearest neighbor is a classification mechanism that is used to make groupings of values based on an assumed value of K. In the proposed model, the difference between traffic values is evaluated with the help of Euclidean Distance mechanism and then these values are checked against the value of K. the values that are less than or equal to K are considered for further prediction.

b. IMPLICATION OF EUCLIDEAN DISTANCE

Euclidean distance mechanism is used in order to find the difference between the traffic values, considering one value at a time and then subtracting all the other values from that value, just as a loop works. In general, Euclidean distance is the minimum distance between two points in Euclidean Space. So basically, use of Euclidean distance implies that shorter distances are considered, which eventually leads to the fall in error and increase in accuracy values.

5. RESULTS AND DISCUSSIONS

Results obtained with the implication of KNN and Euclidean distance is better as compared to existing approach without Euclidean distance mechanism. Results are presented in terms of accuracy and error.

The snapshots showing error and accuracy corresponding to the lanes and different values of K for existing approach are shown as follows:

The representation of error for lane 1 in the graphical form is given as under:

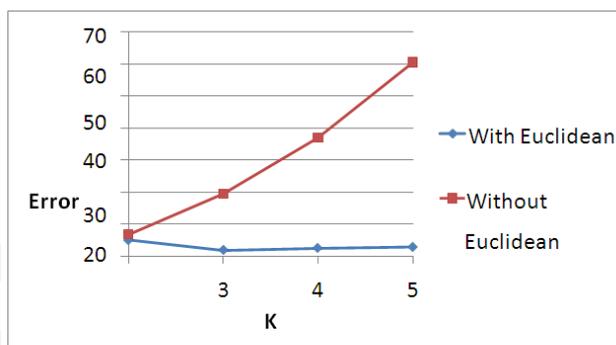


Figure 20: Error corresponding to Lane 1

The representation of error for lane 2 in the graphical form is given as under:

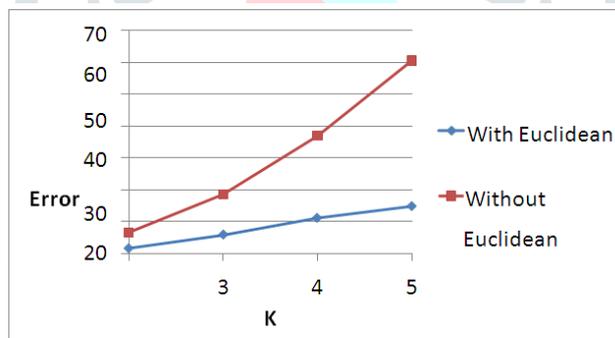


Figure 21: Error corresponding to Lane 2

The representation of error for lane 3 in the graphical form is given as under:

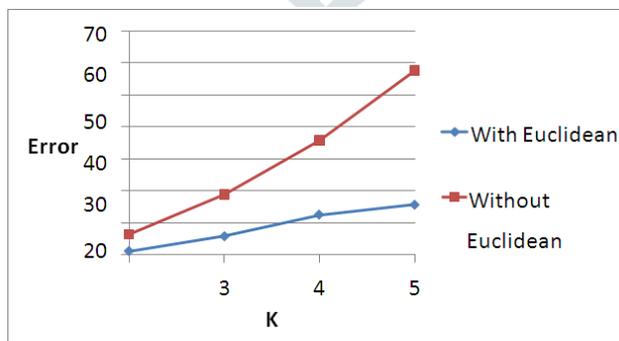


Figure 22: Error corresponding to Lane 3

Accuracy, which is the difference between the actual and the observed values, is obtained for all the lanes considered. The accuracy values for all the three lanes are listed in tabular form for different values of K as follows:

Table 4: Accuracy value corresponding to Lane1,Lane2 and Lane3 for different values of K

Lane	K	Without Euclidean Distance	With Euclidean Distance
		Accuracy	Accuracy
1	2	93.33	95
1	3	80.57	98.27
1	4	63.05	97.66
1	5	39.53	97.21
2	2	93.46	98.45
2	3	81.46	94.28
2	4	63.13	88.91
2	5	39.45	85.22
3	2	93.63	98.99
3	3	81.09	94.15
3	4	64.29	87.66
3	5	42.45	84.29

The representation of accuracy for lane 1 in the graphical form is given as under:

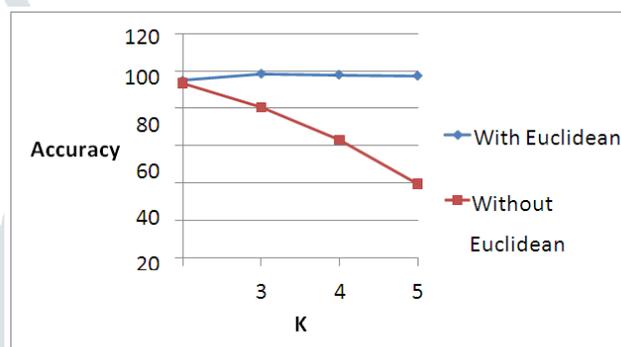


Figure 23: Accuracy for lane 1

The representation of accuracy for lane 1 in the graphical form is given as under:

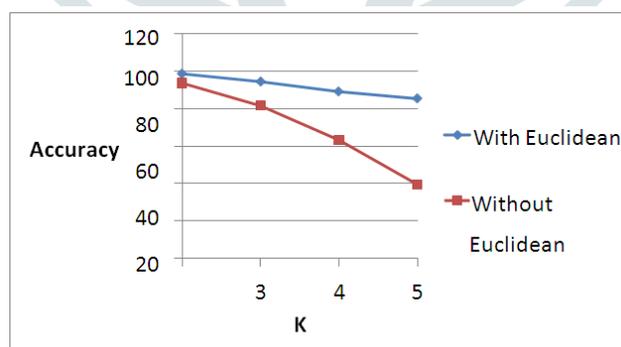


Figure 24: Accuracy for Lane 2

The representation of accuracy for lane 3 in the graphical form is given as under:

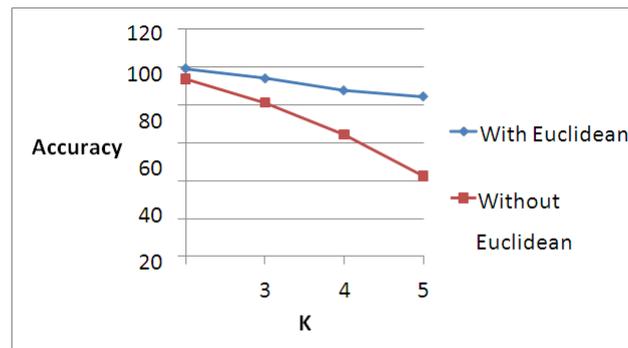


Figure 25: Accuracy for Lane 3

6. Conclusions

Traffic prediction is critical in today's environment. Traffic causes lots of mishappening including accidents, delays etc. the traffic prediction then comes to the course. The traffic prediction is a area which is researched upon by many researchers. Focus is on the reduction of delays and accident. In most of the researches ARIMA model is utilized. The time series analysis is conducted for this purpose. The problem aggravates as the size of the dataset increases. To handle the problem KNN with Euclidean distance is proposed through this literature. Traffic prediction dataset is derived from online source. The dataset is filtered using preprocessing technique. The noise and missing values are handled using the KNN and Euclidean distance mechanism proposed through this literature. The noise once handled from dataset, is fed into the KNN+Euclidean distance mechanism. Nearest neighborhood is evaluated to find difference of terms present within the dataset. Once the distance is evaluated, clusters are formed. The clusters are formed on the basis of time periods. The time periods are checked and evaluated for future prediction.

The overall process of traffic prediction is divided into set of four phases.

- First phase include determination of input and output parameters.
- Second phase consist of preprocessing including eliminating noise from within dataset
- Third phase consist of specification of values corresponding to lane number along with k for performing operation.
- Forth phase is used to evaluate results.

As existing system uses Manhattan distance with KNN, Manhattan distance is calculated by subtracting current value from rest of the values in a lane. Manhattan distance is absolute distance obtained after aggregating previous distance calculations hence error rate increases as lane traffic values increases. Euclidean distance on the other hand focuses on shorter distances and form clusters of each resulting in degradation in error values which leads to improvement in result. Results indicate that proposed system produce better results in terms of parameters MSE and accuracy. Result evaluation is according to three lanes used for evaluation. Significant improvement in terms of these parameters proves worth of this study.

7. Future scope

In future, traffic prediction by eliminating the redundancy within the dataset which is being used could be used in future work. The redundancy reduction mechanism enforce space to be conserved and subsequently traffic prediction could yield least amount of time. Critical region identification such as hospital, schools etc. can also be identified by the use of machine learning which can be merged along with the proposed system.

Traffic prediction along with the critical region identification through machine learning could be future course of action.

REFERENCES

- [1] Chen, Y. et al., 2016. 2016 International Computer Symposium A Matlab-based Traffic Prediction System. , pp.290–293.
- [2] Anusha, C. et al., 2013. Medical data mining and analysis for heart disease dataset using classification techniques. *National Conference on Challenges in Research & Technology in the Coming Decades (CRT 2013)*, pp.1.09–1.09. Available at: <http://digital-library.theiet.org/content/conferences/10.1049/cp.2013.2485>.
- [3] Azizi, N. et al., 2014. A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images. *International Conference on Multimedia Computing and Systems -Proceedings*, 0, pp.415–420.
- [4] Begum, S., Chakraborty, D. & Sarkar, R., 2015. Data Classification Using Feature Selection and kNN Machine Learning Approach. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pp.811–814. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7546208>.
- [5] Borji, A., Izadi, S. & Itti, L., 2016. iLab-20M: A Large-Scale Controlled Object Dataset to Investigate Deep Learning. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2221–2230.
- [6] Chen, D., 2017. Research on traffic flow prediction in the big data environment based on the improved RBF neural network. *IEEE Transactions on Industrial Informatics*, 3203(c), pp.1–1.
- [7] Chen, H., Dynamic Travel Time Prediction using Pattern Recognition.
- [8] Gebresilassie, M.A., 2017. Spatio-temporal Traffic Flow Prediction Spatio-temporal Traffic Flow Prediction Master ' s Degree Thesis.
- [9] Guo, F., Krishnan, R. & Polak, J.W., Short-term traffic prediction under normal and incident conditions using singular spectrum analysis and the k-nearest neighbour method. , pp.1–6.
- [10] Jan, M.A. et al., 2014. PASCCC: Priority-based application-specific congestion control clustering protocol. *Computer Networks*, 74(PB), pp.92–102