International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 5, September- October 2017 ISSN 2278-6856

MODIFIED MODEL OF PREDICTING TRAFFIC USING KNN AND EUCLIDEAN DISTANCE

Navreet Kaur[1], Meenakshi Sharma[2]
[1]Research scholar (CSE)
Rayat Group Of Institutions, Railmajra, Punjab, India

[2]Assistant Professor (CSE)
Rayat Group Of Institutions, Railmajra, Punjab, India

Abstract-*Adverse situations creep in as traffic enhances on road. This leads to significant problems for users. These problems include delay and accidents. Traffic problem is difficult to address but users can be given prior information about on road traffic so that user can take appropriate action in terms of choosing path. This research paper deals with traffic prediction to predict on road traffic using KNN and Euclidean distance mechanism. The mechanism is implied on dataset derived from online source(UCI). For demonstration three lanes are considered for prediction. Implementation is done within MATLAB. The obtained accuracy of prediction is high and mean square error is low through the proposed literature.*

Keywords-Accuracy, Euclidean Distance, KNN, Mean Square Error, Prediction, Traffic

I. INTRODUCTION

Traffic prediction system is used facilitate better prediction and reducing delay by utilizing channels effectively. This is critical in current situations as traffic is enhancing by leaps and bounds. On road traffic in rural and urban areas have continues deviations. [1]discussed deviations cause variation or disturbance in prediction process. In order to resolve the problem traffic prediction system must be presented with stationery time series values. In order to accomplish this, dataset values are fed into the models like ARIMA. [2]ARIMA Model is based on auto regression phase that leads to prediction values having relatively high mean square error.

Another technique which is followed involves Manhattan distance with [3]KNN approach. Manhattan distance is calculated by subtracting the pivot point distance from rest of the values present within the dataset. Values so obtained are compared against the threshold value represented with K. In case value of K is less than the calculated distance then obtained result discarded, indicating traffic is not similar at that point of time. Prediction accuracy using this mechanism is better as compared to simple ARIMA model. Proposed literature works on traffic prediction using [4], [5]KNN with Euclidean distance. Implementation is done within MATLAB and accuracy obtained through the mechanism is better as compared to existing approach without Euclidean distance. The steps involved in the proposed work is listed as under

- Obtain the dataset from UCI website related to traffic.
- Take value of K from user as input
- Subtract the current value from rest of the values in dataset
- Compare obtained values against K.
- In case K is greater than obtained difference then group them in common cluster
- Predict the traffic according to common groups so obtained.

Rest of the paper is organised as follows: section 2 present the literature survey, section 3 gives the proposed system, section 4 gives the performance analysis and result, section 5 gives conclusion and future scope and last section gives the references.

II. LITERATURE SURVEY

Technology is enhancing and is widely used in monitoring traffic. Traffic related issues occur more frequently due to late discovery hence causing delay in critical activities. Technology however helps in predicting traffic at the early stages and hence preventing delays caused by traffic.[6] Proposes Internet of Things to create android application for predicting traffic in urban and rural areas. Internet of things(IoT) is widely used for collecting information regarding different parameters from the road and then techniques are applied to predict the traffic. Techniques used for prediction which was considered in this paper include KNN, [7]Random Forest, [8]Euclidean distance and [9]ARIMA. The prediction generated varies depending upon the accuracy of data presented. Accuracy of data presented depends upon the sensors. Sensors if malfunction may produce inaccurate data. Fault tolerance hence is critical in such situations. This literature[10], [11] explore the applications of IoT for collection of data, techniques used to process data presented by sensors and then fault tolerance capabilities possessed by different techniques used to analyse data presented by sensors. Information generated from sensors is stored within dataset. As more and more information in terms of attributes are collected, size of dataset increases. The collected information is then analysed for accuracy and future prediction.

III. PROPOSED SYSTEM

Proposed system uses KNN with Euclidean distance for predicting traffic.[12] Traffic prediction is based on dataset derived from online source. Traffic dataset is used for

accurately determining traffic at future end intervals. The primary parameters considered from improvement include accuracy and Mean square error. [13]The prediction process is described considering the following example

Example 1: Consider sensors attached to on road side poles and collecting information about Vehicles Lane 1 (X), Lane2(Y) and Lane3(Z). Collected information forms a dataset of following structure in the form of time series

Table 1

Dataset of Traffic records with ID, Time, X, Y and Z

ID	Time	X	Y	Z
1	9:00 AM	120	150	100
1	9:15 AM	123	144	99
1	9:30 AM	122	146	101
1	9:45 AM	124	145	100
1	10:00 AM	123	146	99
2	9:00 AM	121	150	102
2	9:15 AM	125	141	101
2	9:30 AM	126	142	105
2	9:45 AM	121	149	99
2	10:00 AM	120	156	98

A. IMPLICATION OF KNN

The dataset can be analysed by the use of K-nearest neighbour technique. Let the value of K=2 and implication is assigned on X,Y and Z. The test point of x is considered to be 120. Test point y is considered to be 143.The test point of z is 99. Then KNN generates

Table 2

KNN deviation results

Id	X= $\sum\frac{X-120}{N}$	Y=$\sum(Y-143)/N$	Z=$\sum(Z-99)/N$
1	2.16	2.66	0.66
2	2.16	4.83	2

The decision boundary is created for detecting and forecasting purpose in case of KNN. The decision boundary is established by the use of rules. These rules are in the form of IF-THEN form. The threshold is assumed to be 2.5 for X, 3 for Y and 1.5 for Z.

B. IMPLICATION OF EUCLIDEAN DISTANCE

Euclidean distance is used to calculate the distance of noted dataset values from the test point. The threshold value is compared with obtained value to determine anomalies. The test point of x is considered to be 120. Test point y is considered to be 143.

Table 3

Euclidean Distance Results

Id	X= $\sqrt{\frac{(X-X_i)^2}{Total_{observation}}}$	Y= $\sqrt{\frac{(Y-Y_i)^2}{Total_{observation}}}$	Z= $\sqrt{\frac{(Z-Z_i)^2}{Total_{observation}}}$
1	2.7	3.5	0.93
2	2.7	5.53	3.12

The threshold if assumed to be 2.5 for X, 3 for Y and 1.5 for Z. The hybridization of KNN + Euclidean distance is used inthis research for optimization of parameters like accuracy and mean square error.

Proposed methodology in terms of algorithm is given as follows:

The algorithm for the proposed approach is as under

Algorithm Traffic(KNN+EUCLIDEAN)

* INPUT: Dataset with attribute values including

* Output: Prediction Accuracy

a) Perform Pre-Processing

Convert attribute values to nominal form for analysis.

b) Select classifier for traffic classification

c) Obtain values from dataset and store them for comparison

d) Specify the value of lane number for traffic prediction

e) Specify value of K suggesting neighbour distance and Euclidean distance to determine Clustering

f) Determine distance between the values of Euclidean distance which if less then K then neighbour is found.

g) Calculate accuracy and compare it with existing technique without KNN and Euclidean.

Performance analysis and results are presented in the next section.

IV. PERFORMANCE ANALYSIS AND RESULTS

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

The mean square error for Lane 1, Lane 2 and Lane 3 with existing and proposed system is shown as follows:

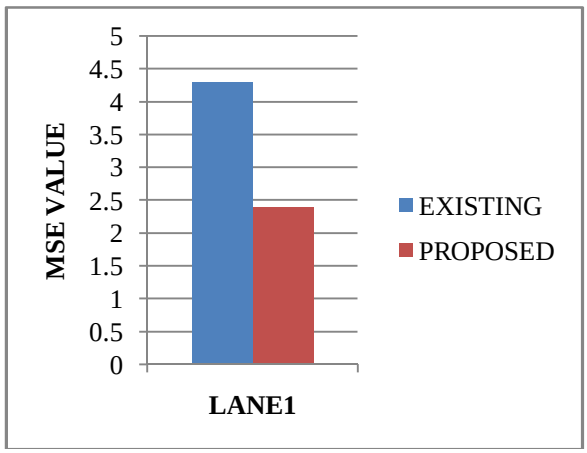

Figure 1: MSE (Mean Square Error) corresponding to Lane 1

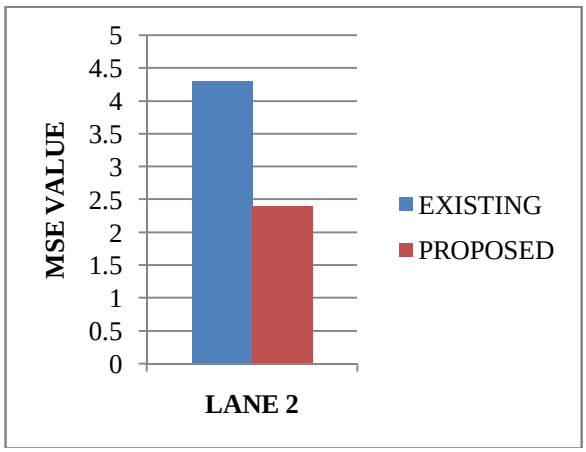

Figure 2: MSE (Mean Square Error)corresponding to Lane 2

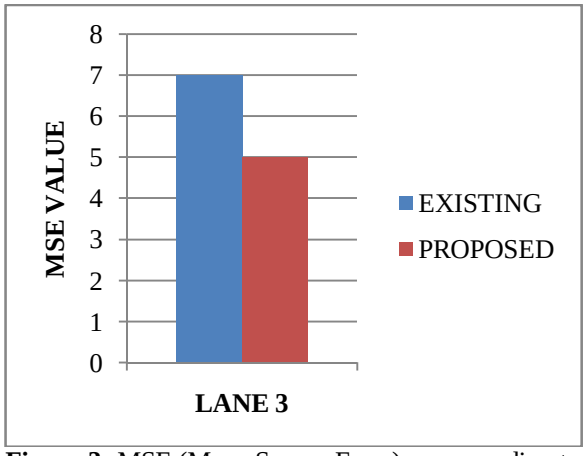

Figure 3: MSE (Mean Square Error)corresponding to Lane 3

Accuracy, which is the difference between actual and observed values, is obtained for all the lanes considered. Results in terms of accuracy for all the three lanes are given as under:

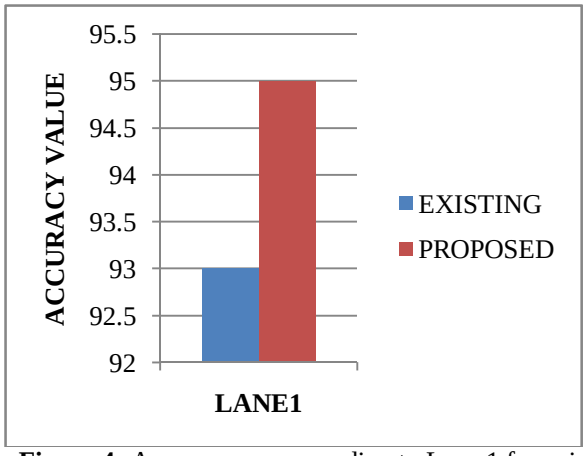

Figure 4: Accuracy corresponding to Lane 1 for existing and proposed systems

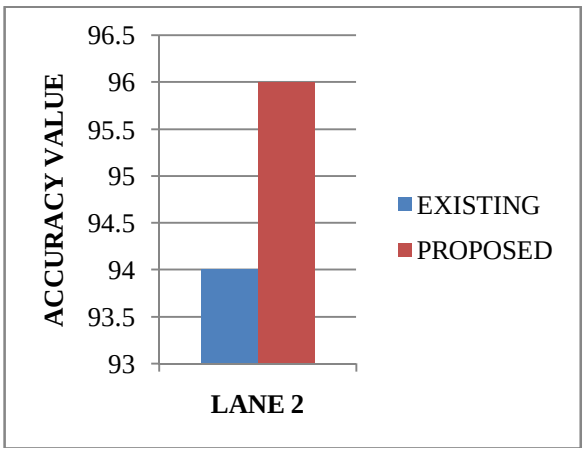

Figure 5: Accuracy corresponding to Lane 2 for existing and proposed systems

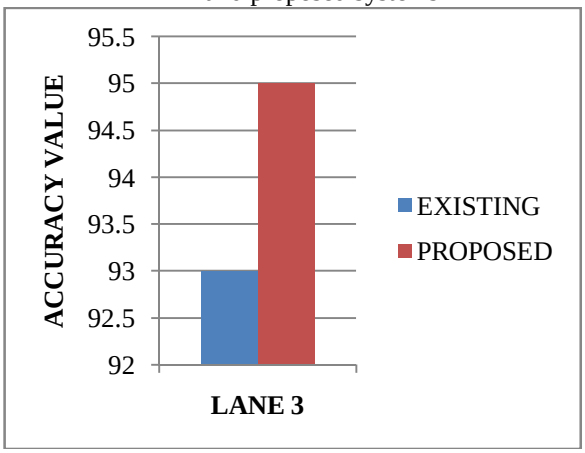

Figure 6: Accuracy corresponding to Lane 3 for existing and proposed systems

Result comparison indicates that the proposed system produces better results as compared to existing system.

V. CONCLUSIONS AND FUTURE SCOPE

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 of 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.

In future, ARIMA model with KNN and Euclidean distance can be used for enhancing accuracy.

REFERENCES

[1] G. Wang and J. Kim, "The Prediction of Traffic Congestion and Incident on Urban Road Networks Using Naive Bayes Classifier," no. November, pp. 1–14, 2016.

[2] F. Guo, R. Krishnan, and J. W. Polak, "Short-term traffic prediction under normal and incident conditions

using singular spectrum analysis and the k-nearest neighbour method," pp. 1–6.
[3] P. K. Hoong, I. K. T. Tan, O. K. Chien, and C. Ting, "ROAD TRAFFIC PREDICTION USING BAYESIAN NETWORKS."
[4] P. M. Xavier and R. Nedunchezhian, "A COMPARATIVE STUDY ON ROAD TRAFFIC MANAGEMENT SYSTEMS," pp. 108–112, 2014.
[5] U. T. Management, "U RBAN T RAFFIC M ANAGEMENT and C ONTROL S YSTEMS," no. 267321, 1999.
[6] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. Van Der Schaar, "Context-Aware Online Spatiotemporal Traffic Prediction."
[7] J. Abawajy, S. Member, M. Chowdhury, and A. Kelarev, "Hybrid Consensus Pruning of Ensemble Classifiers for Big Data Malware Detection," vol. 3, no. 2, pp. 1–11, 2015.
[8] D. Sinwar and R. Kaushik, "Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 2, no. 5, pp. 270–274, 2014.
[9] D. Banerjee, "Forecasting of Indian stock market using time-series ARIMA model," in *2014 2nd International Conference on Business and Information Management (ICBIM)*, 2014, pp. 131–135.
[10] J. Gubbi, R. Buyya, and S. Marusic, "Internet of Things (IoT): A Vision , Architectural Elements , and Future Directions," no. 1, pp. 1–19.
[11] T. Shah, N. Javaid, and T. N. Qureshi, "Energy Efficient Sleep Awake Aware (EESAA) Intelligent Sensor Network Routing Protocol," pp. 1–6.
[12] T. Zhou, L. Gao, and D. Ni, "Road Traffic Prediction by Incorporating Online Information," 2014.
[13] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. Part C*, vol. 19, no. 4, pp. 606–616, 2011.