International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 1221-1225 www.ijaem.net ISSN: 2395-5252

A Data Mining Approach to Traffic Congestion Prediction Using the K-Nearest Neighbour Algorithm

R. Narmatha, Dr. M. Prakash, Dr. V. Vennila, M.Pavithra.

Assistant Professor, Government College of Engineering, Dharmapuri.
Associate Professor, School of Computing,SRM Institute of Science and Technology, Kattankulathur, Chengalpattu.
Associate Professor, K.S.R. College of Engineering, Tiruchengode
Assistant Professor, SSM College of Engineering, Namakkal.

Date of Submission: 25-11-2022 Date of Acceptance: 06-12-2022

ABSTRACT
The main concept of this project is to create a model to predict traffic congestion in a particular location on a particular day, which is to avoid traffic-related issues like traffic jams and traffic accidents. Due to increasing urbanization, there is a rapid growth in the number of vehicles on the road, resulting in traffic congestion. The condition of a road section where traffic demand exceeds available road capacity is known as "traffic congestion." With increasing traffic, people in cities face many hurdles that affect their day-to-day activities. In this project, we have designed a model to predict traffic congestion using data mining by implementing the K Nearest-Neighbour (KNN) algorithm, so that based on the historic traffic data, the model predicts the traffic congestion values for the required locations on a particular day at a specified time interval. In this project, a prediction model is created to predict traffic congestion on a particular road based on its spatial and temporal data. The data can be stored in a CSV file on a DBMS for use in predicting traffic congestion by using machine learning techniques. The proposed methodology first of all uses K-NN to form clusters of related items from within the traffic dataset.
KEYWORDS: K-NN, IMAP, CSV, traffic congestion.

I. INTRODUCTION

A traffic congestion prediction system allows you to get information about traffic-related issues like traffic jams in a particular location of the road network on a particular day. Due to urbanization, there has been a steady increase in the number of vehicles on the road, which results in traffic congestion. Traffic congestion is regarded as one of the greatest problems facing today's world. When congestion occurs in one part of the road network, it will affect the traffic flows on the surrounding roads. This may lead to heavy traffic jams, which will affect people's work and time. A society-wide consensus has emerged to handle traffic-related problems scientifically and reasonably. Building transportation infrastructures can only alleviate traffic congestion to a certain extent and for a limited time. Implementing traffic guidance and control, properly using road resources, and giving full play to vehicle functions are all significant ways to improve transportation efficiency, reduce traffic congestion, and improve traffic safety. In terms of the Indian situation, Indian cities are dealing with a variety of traffic issues, including mixed traffic, traffic congestion caused by noise and air pollution, and an increase in the number of vehicles on the road network. This project's goal is to create a model for predicting traffic congestion in a particular road network on a particular day using the K-NN (K-Nearest Neighbour) algorithm, which makes use of spatial and temporal correlation found in the historical data and provides information to the user about the traffic congestion on the particular road network.

II. LITERATUREREVIEW

Based upon the various traffic-related issues, researchers used the YOLO (you only look once) algorithm to overcome traffic congestion, which uses the image of the vehicles from CCTV cameras at traffic intersections as input for Calculating real-time traffic density utilizing image processing and object detection. But there is no database management in that system [1]. Focusing on the issue of DBMS, we created a model to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 1221-1225 www.ijaem.net ISSN: 2395-5252

predict traffic congestion at a particular road junction on a particular day using the K-Algorithm, which makes use of historical data and provides information to the client about the traffic on the particular road [2]. A traffic congestion prediction system allows you to get information about traffic-related issues like traffic jams in a particular location of the road network on a particular day [3]. The client gets information about traffic congestion through the mail for a particular road junction on a particular day and time [4].

III. PROPOSED SYSTEM

After analyzing the drawbacks of the existing system, we have proposed a traffic congestion system using the K-NN (K-Nearest Neighbour) algorithm to predict traffic congestion in particular areas.

• In this project, a prediction model using data mining is proposed that implements the K-Nearest Neighbour (KNN) algorithm, which makes use of the spatial as well as the temporal correlations found in the traffic dataset to predict traffic congestion.

 The Traffic congestion on a particular road is based on its spatial and temporal data, the data can be stored in CSV files on DBMS for predicting the traffic congestion by using machine learning techniques.

 The datasets are given as input for training the model. The required fields can be extracted from the given datasets.

 The Spatial K-Neighbor finds various locations in the dataset, and for each of these locations, the K Neighbors are found. K value is configurable within the system. The spatial and temporal correlation module builds a model based on the spatial and temporal properties.

 The spatial correlations are considered along with the temporal correlations with the help of the KNN algorithm to predict traffic congestion.

 The proposed solution is more accurate than the existing solutions.

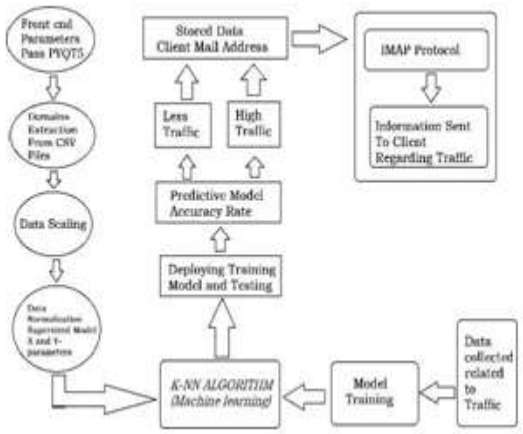

Fig: Architecture diagram for proposed system The above-shown image is the Architecture diagram for our proposed system.

The proposed system consists of these different modules:

A. Input interface module
B. Data extraction module
C. Prediction module
D. Database module
E. Mail alert module

A. Input interface module:

In this module, the input traffic data can be interfaced with software for analysis. The input data can be interfaced by using the pyqt5 front end.

B.Data extraction module:

In this extraction module, the traffic congestion on a particular road is based on its spatial and temporal data, the data can be stored in a CSV file on DBMS for predicting the traffic congestion by using a machine learning technique.

C.Prediction module:

In this module, the machine learning KNN k-nearest neighbor algorithm will predict traffic congestion. The K-Nearest Neighbor (KNN) algorithm, makes use of the spatial, as well as the temporal correlations found in the traffic dataset, in order to predict traffic congestions. This algorithm will be used the traffic-related dataset as a reference.

D.Database module:

In this module, the KNN nearest neighbor algorithm predicted traffic congestion in real-time. The traffic congestion data can be stored in the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 1221-1225 www.ijaem.net ISSN: 2395-5252

CSV file database.

E.Mail alert module:
If the machine learning algorithm predicts traffic congestion. The traffic congestion data can be stored in a CSV file database and that traffic congestion information will be sent as a mail alert to clients by using (the IMAP)
Internet Message Access protocol.

IV. WORKING WITH IMAP:

(IMAP) INTERNET MESSAGE ACCESS PROTOCOL was an Application Layer Protocol (ALP) that acts as a contractor to receive emails from the server(mail server). These protocols help the client/user to identify the type of messages or e-mails being sent or
received from any (source or destination host).IMAP is the most common for retrieving emails. IMAP follows the client/server model. It can be also named/termed as Internet Mail Access Protocol or Interactive Mail Access Protocol.

Features of IMAP:
i. Manages multiple emails and organizes those emails into different categories.
ii. Client or User can set flags to their e-mail to track the message or e-mail that the user/client has already seen.
iii. The ability to choose whether or not to retrieve e-mail from a server (mail server) before downloading the message's content.
iv. Easy to search for the mail that the user/client already received.

Working of IMAP:
Internet Message Access Protocol follows a Client/Server Architecture. It is the most commonly used e-mail protocol. Internet Message Access Protocol (IMAP) uses the (TCP) Transmission Control Protocol for communication which is to ensure that the data delivered and received are in order. IMAP is a client-server protocol that runs on other computers connected by a network. When an IMAP4 resides on the server where the user/client Mailboxes are located, the IMAP4 Protocol works.

Internet Message Access Protocol Architecture:

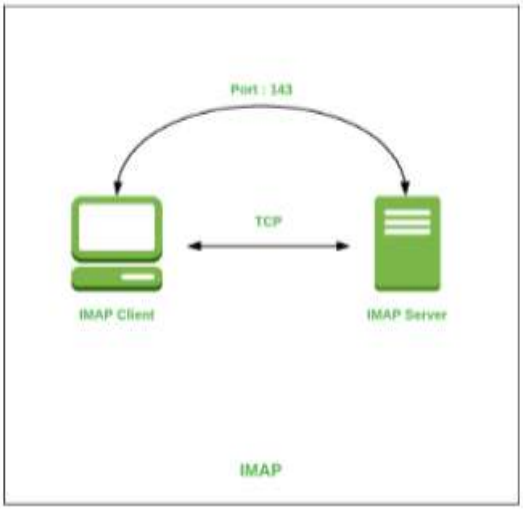

Fig no 4.1 IMAP protocol

Advantages:
 syncs every session that is maintained by the client/user.
 Secures the POP3 protocol because the email is only stored on the IMAP server.
 Clients/users can access their mail from any computer.
 The E-mails are safe even when the computer is lost or destroyed because the E-mails are stored in the IMAP server

Disadvantages:
 Maintaining IMAP is very complex.
 E-mails are available when there is an internet connection.
 It uses users' hard drive space.
 Too slow to download E-mails/messages.
 Several E-mails don't support Internet Message Access Protocol(IMAP) which makes it difficult to manage.

V. RESULT:

In this result section, we have shown the performed model for predicting traffic congestion using the K-NN algorithm.
As a result the K-Nearest Neighbor (KNN) algorithm, which makes use of the spatial as well as the temporal correlations found in the traffic dataset, predicts the traffic congestion on a particular road on a particular date.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 1221-1225 www.ijaem.net ISSN: 2395-5252

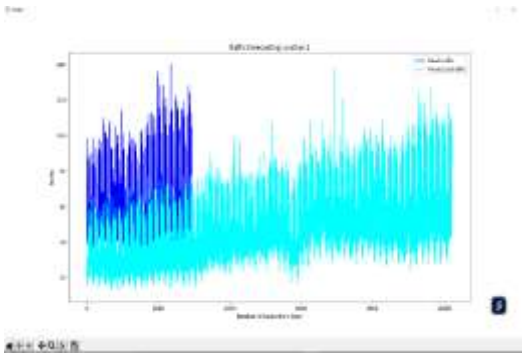

Fig: 5.1 Traffic forecasting graph for junction 1

In this fig 5.1, we have shown the screenshot of the traffic forecasting graph for junction 1.

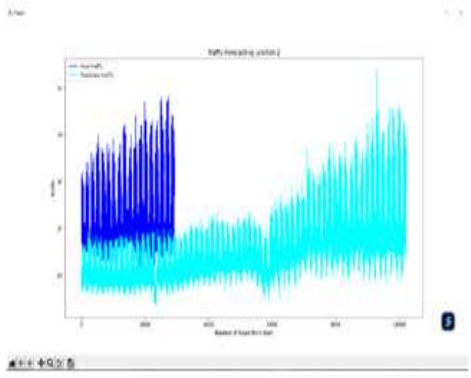

Fig: 5.2 Traffic forecasting graph for junction2

In this fig 5.2, we have shown the screenshot of the traffic forecasting graph for junction2

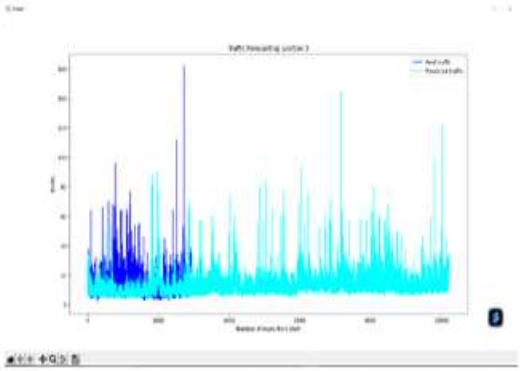

Fig: 5.3 Traffic forecasting graph for junction 3

In this fig 5.3, we have shown the screenshot of the traffic forecasting graph for junction 3.

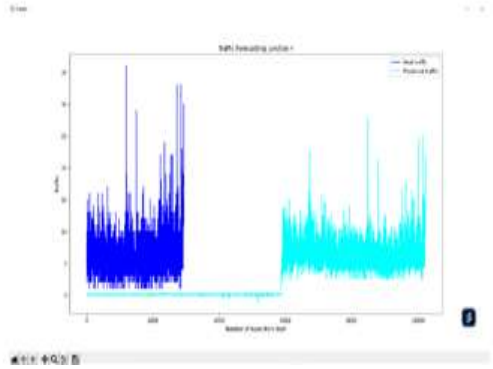

Fig: 5.4 Traffic forecasting graph for junction 4

In this fig 5.4, we have shown the screenshot of the traffic forecasting graph for junction 4.

Finally, the User/client receives an E-mail regarding the traffic congestion in a particular junction on a particular
day which can be useful for the driver to take less Traffic Road.

Fig: 5.5 Output for E-mail page

In this fig 5.5 we have shown the screenshot of Output for E-mail page.

Fig: 5.6 e-mail containing Traffic details attachment

In this fig 5.6, we have shown the screenshot of the Output for E-mail page containing an E-mail with

traffic details attachment.

VI. CONCLUSION & FUTURE SCOPE:

A K-Nearest Neighbor (K-NN) model is provided in this project. For starters, traffic flow has unpredictability, nonlinearity, and complexity. The proposed prediction approach can be used to anticipate road traffic flow in the short term, according to performance metrics. The prediction strategy, the historical database, the search mechanism, and algorithm parameters The practicality of the K-NN non-parametric regression-based traffic flow prediction approach is over 90%, and it also re-reads that the methods are utilized in short-term traffic flow prediction. In the Future, we can improve the system by adding features like detecting the speed of vehicles using real-time data collected that can be useful to avoid Accidents in congested areas.

REFERENCES:

[1]. Huiping Cao, Nikos Mamoulis, and David W Cheung. Mining frequent spatio-temporal sequential patterns. In Fifth IEEE International Conference on Data Mining (ICDM), pages 8–pp. IEEE, 2018.

[2]. Enrique Castillo, Jos´eMar´ıaMen´endez, and Santos S´anchez- Cambronero. Traffic estimation and optimal counting location without path enumeration using Bayesian networks. Computer- Aided Civil and Infrastructure Engineering, 23(3):189–207, 2019.

[3]. Eric Hsueh-Chan Lu, Wang-Cheng Lee, and Vincent S Tseng. Mining fastest path from trajectories with multiple destinations in road networks. Knowledge and information systems, 29(1):25–53, 2020.

[4]. Xiaofeng Wang, Gang Li, Guang Jiang, and Zhongzhi Shi. Semantic trajectory-based event detection and event pattern mining. Knowledge and information systems, pages 1–25, 2019. International Journal of Advances in Engineering and Management (IJAEM) Volume 2, Issue 1, pp: 01-05 www.ijaem.net ISSN: 2395-5252

[5]. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In 20th International Conference on Very Large Data Bases, pages 478–499. Morgan Kaufmann, Los Altos, CA,2019.

[6]. Jiancheng Long, Ziyou Gao, Xiaomei Zhao, Aiping Lian, and Penina Orenstein. Urban traffic jam simulation based on the cell transmission model. Networks and Spatial Economics, 11(1):43–64, 2020.

[7]. B. L .Smith, B. M. Williams, O. R .Keith (2017). Comparison of parametric and nonparametric models for traffic flow forecasting.

[8]. Traffic flow forecasting using approximate nearest neighbor nonparametric regression. The National ITS Implementation Research Center U.S. DOT University Transportation Center. Park,L.&Rilett,R.. (2019).

[9]. Forecasting multiple-period freeway link travel times using modular neural networks. Transportation Research Record, 2019, 1617:163-170.

[10]. Nihan, L. &Holmesland, O. (2020). Use of the box and Jenkins time series technique in traffic forecasting. Transportation Research Record (2):125-143 B.V.

[11]. Dasarathy (2016). Nearest Neighbor Forms: NN Patten Classification Techniques. Los Agapitos IEEE Computer Society Press, 2015.

[12]. Yiannis, K. &Poulicos, P. (2018). Forecasting traffic flow conditions in an urban Network-comparison of multivariate and univariate approaches. Transportation Research Record 1857, 2003:74-84.

[13]. Smith, B. L. &Demetsky, J. (2019). Short-term traffic flow prediction: neural network approach. Transportation Research Record 1453, 1994:98-104

[14]. Armstrong, J.S. (2019).1 Findings from evidence- based forecasting: methods for reducing forecast error. International Journal of Forecasting 2016, 22:583-598.

[15]. Anthony, S. & Matthew, K.G. (2019). A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research part C, 2018,11(2):121-135

[16]. Guo Jianhua (2017). Adaptive estimation and prediction of univariate vehicular Traffic condition series [D]. North Carolina State University.