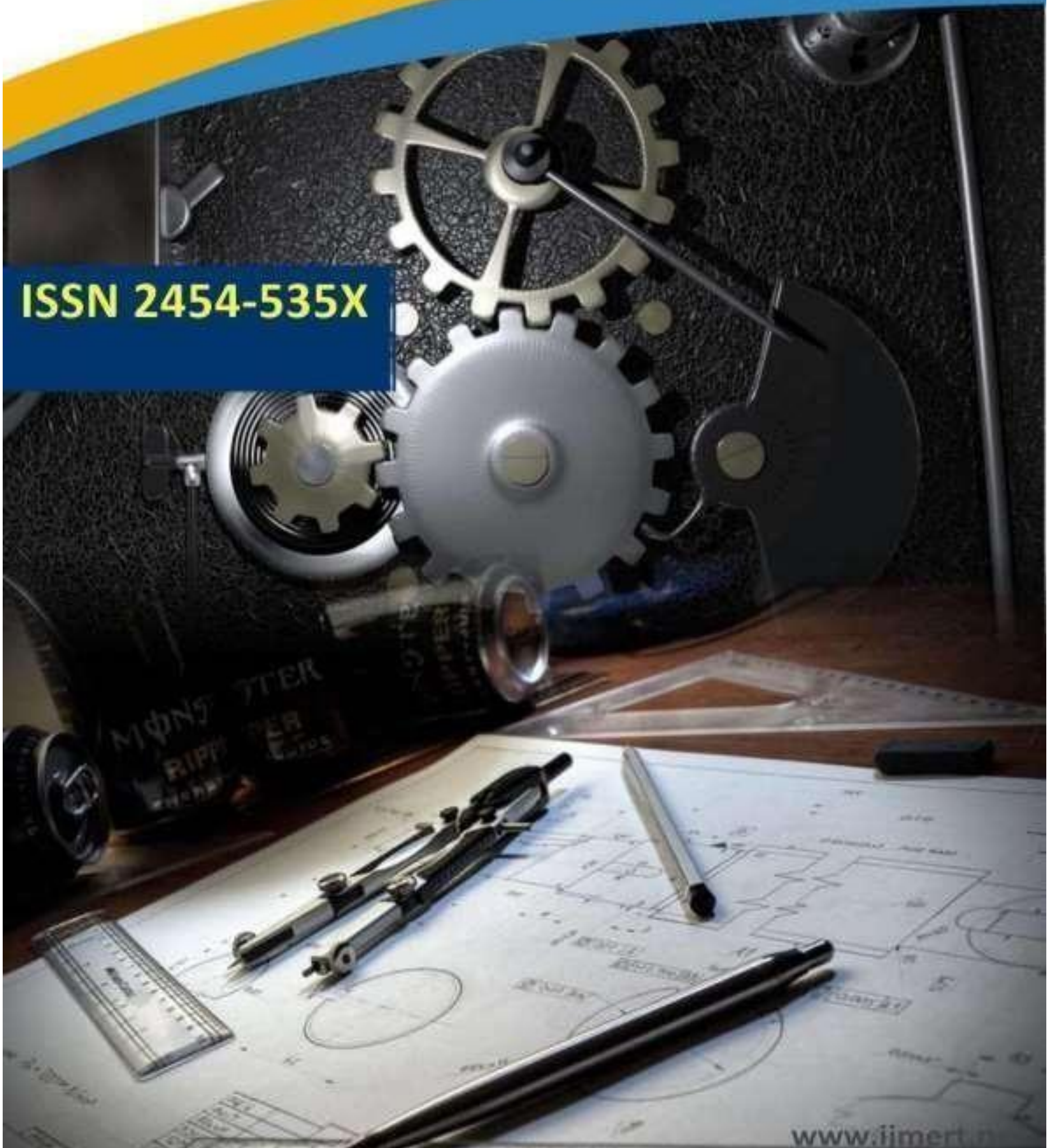
International Journal of
Mechanical Engineering Research and Technology

ISSN 2454-535X



Email ID: info.ijmert@gmail.com or editor@ijmert.net

Int. J. Mech. Eng. Res. & Tech 2017

ISSN 2454 – 5X

http://www.ijmert.com
Vol.16 Issue2, 2024

DATA FITS-A HETROGENOUS DATA FUSION FRAMEWORK FOR TRAFFIC AND INCIDENT PREDICTION

[1]MR. N. CHANDIRAPRAKASH,[2]M RAVINDRANATH,[3]DWARAPUDI LAKSHMI DEEPIKA,[4]SURA PHANINDRA,[5]NEERAJ KUMAR BOMMA,[6]MOHD AQIBUDDIN

[1]Assistant Professor,department of information technology, malla reddy institute of engineering and technology(autonomous),dhulapally,secundrabad,chandrunit26@gmail.com

[2,3,4,5,6]UG students, department of information technology,malla reddy institute of engineering and technology(autonomous),Dhulapally,Secundrabad

ABSTRACT

This paper introduces DataFITS (Data Fusion on Intelligent Transportation System), an open-source framework that collects and fuses traffic-related data from various sources, creating a comprehensive dataset. We hypothesize that a heterogeneous data fusion framework can enhance information coverage and quality for traffic models, increasing the efficiency and reliability of Intelligent Transportation System (ITS) applications. Our hypothesis was verified through two applications that utilized traffic estimation and incident classification models. DataFITS collected four data types from seven sources over nine months and fused them in a spatiotemporal domain. Traffic estimation models used descriptive statistics and polynomial regression, while incident classification employed the k-nearest neighbors (k-NN) algorithm with Dynamic Time Warping (DTW) and Wasserstein metric as distance measures. Results indicate that DataFITS significantly increased road coverage by 137% and improved information quality for up to 40% of all roads through data fusion. Traffic estimation achieved an R2 score of 0.91 using a polynomial regression model, while incident classification achieved 90% accuracy on binary tasks (incident or non-incident) and around 80% on classifying three different types of incidents (accident, congestion, and non-incident).

I.INTRODUCTION

The "DataFITS: A Heterogeneous Data Fusion Framework for Traffic and Incident Prediction" project introduces an innovative approach to enhance transportation management and incident prediction systems. In urban environments, traffic congestion and

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

incidents can lead to significant disruptions in daily routines, economic losses, and safety hazards. Traditional methods of traffic prediction often rely on individual data sources, such as traffic sensors or historical incident records, which may lack comprehensive coverage and real-time insights. In response to these challenges, DataFITS proposes a heterogeneous data fusion framework that integrates diverse data sources, including traffic flow data, weather conditions, social media feeds, and historical incident records. By leveraging advanced data fusion techniques, such as machine learning and deep learning algorithms, DataFITS aims to provide accurate and timely predictions of traffic conditions and incident occurrences. The project's overarching goal is to empower transportation authorities and urban planners with actionable insights to mitigate traffic congestion, improve incident response times, and enhance overall transportation efficiency and safety.

II.EXISTING SYSTEM

To develop ITS applications, significant data is required from real or virtual sensors [5]. Vitor et al. [4] present a platform to collect, process, and export heterogeneous data from smart city sensors, providing different statistics and visualizations. However, their platform concentrates on securing data. Similarly, [6] proposes a smart city data platform containing information from various cities. In contrast to our framework, we focus on improving the quantity and quality of the information by fusing data, and we assess the advantages of using fused data through two ITS applications. Data fusion combines data from multiple sources, enriching spatiotemporal information [7], [8], [9], [10]. Several applications benefit from data fusion, such as emergency management [11] and path planning [12]. However, fusing heterogeneous data requires additional preprocessing to combine various data types and features [13], [14]. This investigation focuses on two applications supported through data fusion: traffic estimation and incident classification, and the methods to achieve their goals, such as data acquisition, fusion, machine learning, correlation, and different data types.

Traffic estimation is a crucial smart city application for better transportation management. This review focuses on

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

data fusion, spatiotemporal correlation, and machine learning techniques to achieve accurate and reliable traffic estimation using historical data. The increasing availability of open databases (kept by governmental authorities) and Application Programming Interfaces (APIs) to commercial applications (Bing, Google Maps, etc.) results in a vast collection of trafficrelated data, making big data an opportunity for heterogeneous data fusion [15]. The challenge is to combine stationary sensor data (e.g., traffic cameras or loop detectors) and probe vehicle information (e.g., cameras, GPS, cellular data, or vehicular sensors). Anand et al. [16] used a Kalman filter to fuse traffic flow values (from cameras) and travel time (from GPS), improving a traffic estimation approach. Many recent traffic estimation models use Machine Learning (ML) [17], [18], [19], [20], [21], [22], [23], [24], [25]. Reference [17] proposes an auto-regressive model that uses data from a traffic simulator and adapts to events like accidents.

Their results showed that estimation up to 30 minutes ahead has an error of 12%. Meanwhile, [18] employs deep learning algorithms for traffic estimation, showing an improvement of accuracy and efficiency. These approaches discuss the usage of ML to create accurate models for traffic estimation, but do not consider further methods, such as data fusion, correlation, etc.

Some ML approaches use spatiotemporal correlation to improve traffic estimation quality. In [19], a neural network(NN)-based estimation using Graph Convolutional Network (GCN) and Gated Recurrent Unit (GRU) models is proposed with full public access. The GCN captures spatial dependencies from the road network, and GRU detect dynamic changes in traffic data and captures temporal dependencies. Other NN-based approaches, such as [20] and [21], show similar improvements in accuracy using data correlation. Wang et al. [22] propose an open-source deep learning framework using GCN to estimate network-wide traffic multiple steps ahead in time. Zheng et al. [23] introduce another opensource solution, the Graph Multi Attention Network (GMAN), using an encoder-decoder architecture to provide long-term traffic estimation up to one hour ahead. These approaches also include correlation to

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

improve the discussed models and offer access to their data but do not propose a solution for collecting or fusing data. Limited literature combines data fusion, spatiotemporal correlation, and ML to estimate traffic, similar to our solution. In [26], the authors fuse traffic data from stationary and dynamic sensors, considering the spatiotemporal correlation between traffic levels of road segments.

A Multiple Linear Regression (MLR) model processes the fused information to enhance traffic estimation accuracy. Unlike our solution, this approach relies solely on traffic data from sensors but does not consider different data types and sources. Zhao et al. [24] propose a general platform for spatiotemporal data fusion to enhance traffic estimation. The approach introduces a fusion method to improve accuracy by combining direct and indirect traffic-related data as input for two different ML models. The indirect traffic-related data features contain information about weather and points of interest and are used to improve the estimation quality. However, their model uses pre-existing datasets, offering no solution for data collection, and our study focuses on incident-related data, while the authors in [24] consider points of interest and weather conditions.

Disadvantages

- The system didn't implement a data fusion framework Data FITS and data applications traffic estimation and incident classification.

- The fused data from DataFITS is not cleaned, not removing all incident-related information, as it is not required by the model, and grouped into traffic areas containing one or multiple road segments.

III.PROPOSED SYSTEM

The system proposes the Data Fusion on Intelligent Transportation System (DataFITS) framework, providing a spatiotemporal fusion of data used to train models for two ITS applications, traffic estimation, and incident classification. DataFITS collects and combines real heterogeneous data (e.g., weather, traffic, incident) from various sources (e.g., open databases, map applications), preparing them by fixing errors, adapting the data structure, and finally fusing them in the exact location

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

and point in time. Our hypothesis is verified using data characterization to quantify the benefits of combining heterogeneous data sources and the proposal of two ITS applications. The performance of the two applications ratifies the benefits of larger data coverage/quality while estimating traffic and classifying incidents.

Advantages

• An open-source framework DataFITS for heterogeneous spatiotemporal data fusion, covering the acquisition, processing, and fusion of data, available in a public code repository.

• The characterization of a heterogeneous dataset combining real traffic data from two cities in Germany, collected from seven sources over nine months and provided together with the repository.

• Two traffic estimation models, one using descriptive statistics and another using polynomial regression with different parameters such as time, road type, and weather, and a comparison between single and fused datasets.

• An incident classification model trained and evaluated on heterogeneous fused data using k-nearest neighbors (k-NN), with Dynamic Time Warping (DTW) and Wasserstein as distance methods.

IV.MODULES

➢ **Service Provider**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Predicted Type, View Type Ratio, Download Predicted Data Sets, View Type Ratio Results, View All Remote Users.

➢ **View and Authorize Users**

In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

➢ **Remote User**

In this module, . User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X
http://www.ijmert.com
Vol.16 Issue2, 2024

LOGIN, after login we have to Predict Type, VIEW YOUR PROFILE.

V.CONCLUSION

In conclusion, the "DataFITS: A Heterogeneous Data Fusion Framework for Traffic and Incident Prediction" project represents a significant advancement in transportation management and incident prediction capabilities. By harnessing the power of heterogeneous data fusion techniques, DataFITS offers a comprehensive solution for addressing the complex challenges associated with traffic congestion and incident management in urban environments. The integration of diverse data sources and advanced data fusion algorithms enables DataFITS to provide accurate and timely predictions of traffic conditions and incident occurrences, thereby facilitating proactive decision-making and resource allocation by transportation authorities and urban planners. As a result, DataFITS has the potential to significantly improve the efficiency, safety, and resilience of transportation systems, ultimately benefiting both commuters and the broader community.

VI.REFERENCES

1. Lv, Y., Duan, Y., Kang, W., & Wu, C. (2015). Traffic flow prediction with big data: A deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 865-873.
2. Zhang, J., Zheng, Y., & Qi, D. (2016). Deep spatio-temporal residual networks for citywide crowd flows prediction. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI) (pp. 1655-1661).
3. Zheng, Y., Zhang, J., & Qi, D. (2017). Deep learning for spatio-temporal data mining: A survey. IEEE Transactions on Big Data, 3(4), 380-404.
4. Zhang, X., Du, S., Zhang, Q., & Wang, H. (2019). A hybrid deep learning model for traffic flow prediction based on heterogeneous data. IEEE Access, 7, 93194-93202.
5. Wang, H., Li, K., Du, X., Gao, J., & Lin, J. (2020). Multi-source heterogeneous data fusion based on deep learning for traffic flow prediction. IEEE Access, 8, 65047-65056.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol.16 Issue2, 2024

6. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI) (pp. 265-283).
7. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1725-1732).
8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.