Data-Driven Prediction and Drift Enhancement with Heterogeneous Graph Analysis

Manivannan K[1][a], Gowsika S[2], Geetha M[3], Baskar D[1] , Franklin Jino R.E.[1] and Kanimozhi S[4]
[1]*Department of* Information Technology*, V.S.B. Engineering College, Karur, India*
[2]*Department of Computer Science and Technology, Vivekananda College of Engineering for Women, Namakkal, India*
[3]*Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Karur, India*
[4]*Department of Information Technology, M. Kumarasamy College of Engineering, Karur, India*

Keywords: Predictive Analytics, Heterogeneous Data, Machine Learning, Forecasting Concept Drift, Long Short-Term Memory, Light Gradient Boosting Machine, Drift Prediction.

Abstract: Predictive model accuracy and dependability maintenance is critical in the quickly changing world of data-driven environments. This work, propose a new framework for drift detection and model updating that combines machine learning methods such as Long Short-Term Memory (LSTM) networks and Light Gradient Boosting Machine (LGBM) with statistical tests. We provide a complete strategy that extends to proactive model adjustment tactics, beginning with the quantitative changes in data distribution that identify drift. Our experimental approach, which was carried out on simulated datasets intended to replicate temporal variations in user behavior and market conditions that occur in real life, shows that, when compared to traditional static models, our method can greatly improve model resilience and reduce prediction error by up to 40%. The study also looks at the effects of quick model modification, highlighting the need to strike a balance between predictability and responsiveness. This paper provides a strong methodology for controlling idea drift and guaranteeing sustained model accuracy in dynamic contexts, adding to the body of knowledge in predictive analytics. An improved model for forecasting concept drift in sensor data is presented in this work, which is essential for preserving data quality in dynamic contexts. By combining machine learning with ARIMA, our model provides accurate drift prediction and detection. Robust performance is ensured by drift detection, prediction, and preprocessing modules as well as a feedback mechanism. When compared to conventional models, our approach exhibits better accuracy and early identification. In addition to helping with preventive maintenance scheduling and cutting costs and downtime, it promises benefits for industries that depend on accurate sensor data.

1 INTRODUCTION

In the contemporary urban landscape, the dynamics of city life are evolving at an unprecedented pace, driven by multifaceted factors ranging from demographic shifts to technological advancements. Among these transformative forces, the concept of "citified drift" emerges as a pivotal phenomenon encapsulating the fluidity and complexity inherent in urban development. Defined as the continuous, albeit sometimes subtle, changes occurring within the fabric of urban environments, citified drift encompasses shifts in population demographics, economic trends, cultural dynamics, and infrastructural developments.

Policymakers, urban planners, companies, and people all need to comprehend and anticipate citified drift. Strategies for sustainable urban development, effective resource allocation, and proactive decision-making are made possible by anticipating these minute changes. The complex interactions between various, heterogeneous data sources that impact urban dynamics, however, make the prediction of citified drift extremely difficult.

Traditional forecasting methods often fall short in capturing the nuances of citified drift, primarily due

[a] https://orcid.org/0009-0008-3473-9053

K, M., S, G., M, G., D, B., R E, F. J. and S, K.
Data-Driven Prediction and Drift Enhancement with Heterogeneous Graph Analysis.
DOI: 10.5220/0013589900004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 229-237
ISBN: 978-989-758-763-4

to their reliance on homogeneous datasets and simplistic models that overlook the multidimensional nature of urban evolution. To address this limitation, a paradigm shift towards leveraging diverse sources of data and advanced analytical techniques is imperative. By harnessing the wealth of information available from sources such as sensor networks, social media platforms, administrative records, and satellite imagery, a more comprehensive understanding of urban dynamics can be attained.

SPOT is a predictive spatial data mining GIS tool designed to facilitate decision support. It processes and analyzes agro-meteorological and socio-economic thematic maps, generating crop cultivation geo-referenced prediction maps through predictive data mining (Abdullah, Bakhashwain, et al. , 2018).

In this context, the proposed framework aims to bridge the gap between citified drift and predictive analytics (Pathak, Gowda, et al. , 2024), (Manivannan, Gowda, et al. , 2024)through a novel approach grounded in data fusion and machine learning. By integrating data from disparate sources into a unified analytical framework, the model seeks to uncover hidden patterns, correlations, and causal relationships driving urban transformations. Furthermore, the incorporation of graph-based analysis enables the representation of complex urban systems as interconnected networks, facilitating the identification of key drivers and emergent phenomena.

Through the synthesis of diverse data streams and the application of advanced prediction algorithms, the proposed framework endeavors to enhance the accuracy and granularity of citified drift forecasts. By providing actionable insights into future urban trajectories, it empowers stakeholders to proactively adapt to changing conditions, optimize resource utilization, and foster inclusive and sustainable urban development.

In summary, this study introduces a pioneering approach to forecasting citified drift enhancement by leveraging diverse sources of heterogeneous data and employing advanced analytical techniques. By unraveling the intricacies of urban dynamics, this framework holds the promise of revolutionizing decision-making processes and shaping the future of cities in an era of unprecedented change and transformation.

Remainder of the paper is organized as follows. Section II describes the related works. Section III describes the proposed methodology, section IV presents the results and discussion and section V concludes the paper.

2 RELATED WORKS

Urban Mobility Prediction using Machine Learning Techniques (Zheng, Capra, et al. , 2014), this field of study entails gathering and evaluating data from a variety of sources, including social media check-ins, public transit logs, traffic camera feeds, and GPS data from smart phones. Subsequently, popular routes, demand for public transit, and traffic congestion are predicted for the future using machine learning algorithms. In order to create predictive models (Du, Peng, et al. , 2019) that can help urban planners and transportation authority's optimize transportation systems, researchers frequently investigate methods including supervised learning, reinforcement learning, and deep learning.

Graph theory offers a powerful framework for modeling complex relationships in urban environments. By representing urban features such as roads, buildings, neighborhoods, and socio-economic factors as nodes and edges in a graph, researchers can analyze the interconnectedness and dependencies within the urban system. Graph-based predictive models can capture the dynamic nature of urban dynamics, including population movements, gentrification trends, and the spread of amenities and services throughout the city.

Urban planners can use big data analytics to obtain insights into numerous elements of city life, thanks to the explosion of data sources in urban environments. These sources include social media feeds, administrative records, IoT sensors, satellite imaging, and more. The above mentioned tasks may involve scrutinizing human behavior patterns, pinpointing environmentally sensitive locations, spotting deviations in infrastructure functionality, and forecasting future trends in urban growth. Planners are better equipped to decide on land use, transportation, housing, and sustainability projects by combining and evaluating a variety of data sources.

Spatial Analysis of Urban Growth (Pan, Liang, et al. , 2019), (Xie, Li, et al. , 2020), To investigate the geographical patterns and processes of urban expansion, spatial analysis tools such as Geographic Information Systems (GIS), remote sensing, and spatial econometrics are frequently employed. To understand how cities change over time, researchers look at things like population density, changes in land use, transportation systems, and environmental factors. Land use planning efforts can be guided by predictive models that use techniques such as cellular automata, spatial regression, and spatial autocorrelation to estimate future urban expansion.

SeqST-GAN (Wang, Cao, et al. , 2020) was introduced, which integrates a Seq2Seq model and an adversarial learning framework for forecasting multi-step urban crowd flow data. Initially, a Seq2Seq model is employed to generate future crowd flow "frames" step-by-step. Additionally, an EC-Gate module is designed to capture external context features, enabling the learning of a unified region-level representation to refine the initially generated future "frames". Subsequently, an adversarial learning framework is utilized, combining mean square error and adversarial loss to address the issue of blurry predictions. The proposed approach is evaluated on two large crowd flow datasets from New York, demonstrating significant performance improvements over several strong baselines.

A DNN-based approach for air quality prediction (Yi, Zhang, et al. , 2018), employing a novel distributed fusion architecture to combine heterogeneous urban data. Our method demonstrates superior accuracy compared to 10 baselines across three years of data from nine Chinese cities, excelling in both general forecasting and sudden changes. Deployed within the Air Pollution Prediction system, Deep Air provides hourly, fine-grained air quality forecasts for over 300 Chinese cities, achieving significant relative accuracy improvements of 2.4%, 12.2%, and 63.2% in short-term, long-term, and sudden change predictions, respectively, compared to previous online methods.

A novel data-driven approach (Assem, Ghariba, et al. , 2017) is applied to predict daily water flow and water level for the Shannon River catchment in Ireland, utilizing a deep convolutional network architecture that outperforms established forecasting models. By leveraging 30-year daily time series data from multiple water stations, including observed and simulated datasets, our model offers valuable insights for future water resource allocation among various users such as agriculture, domestic consumption, and power generation.

B. Wang et al. (Wang, Lu, et al. , 2019), tackles the pressing challenge of accurate weather forecasting, a vital aspect of daily life, by introducing a groundbreaking method called deep uncertainty quantification (DUQ). It introduces a novel loss function termed negative log-likelihood error (NLE) to train the prediction model, enabling simultaneous inference of sequential point estimation and prediction interval.

Saleh et al. (Saleh, Hossny, et al. , 2020), designed the framework utilizes a tracking-by-detection technique in combination with an innovative spatio-temporal Dense Net model. Authors conducted training and evaluation using authentic data gathered from urban traffic settings. The results demonstrate the robustness and competitiveness of our framework when compared to other baseline methods.

The efficacy of Long Short-Term Memory (LSTM) (Karevan, Suykens, et al. , 2020), in capturing long-term dependencies has made it a prominent choice across various real-world applications. Our study harnesses LSTM to develop a data-driven forecasting model tailored for weather prediction tasks. Additionally, authors introduce Transductive LSTM (T-LSTM), a novel approach that leverages local information to enhance time-series prediction accuracy.

Rezvani et. al. (Rezvani, Barnaghi, et al. , 2019), introduced a novel method for aggregating and representing time-series data. Our approach utilizes Piecewise Aggregate Approximation (PAA) to condense the length of the time-series data. Following this, we employ a Lagrangian multiplier to convert the time-series into unit vectors. This technique preserves essential information within a lower-dimensional vector. Unlike PAA, which represents data solely as a sequence of continuous numbers, our method captures the underlying patterns in time-series data. Their findings indicate that our representation method is more efficient than other existing methods. The vector representations generated by the Lagrangian multiplier facilitate the analysis of patterns and changes in time-series data.

Wu, Y., Wang et al. (Wu, Wang, et al. , 2022), introduced the ROF algorithm, which utilizes a reverse-order filling strategy to determine the one-off support of patterns. Given that OWSP mining adheres to the Apriori property, OWSP-Miner uses a pattern join strategy to generate candidate patterns. Experimental results demonstrate that OWSP-Miner is both more efficient and effective at denoising patterns. In a practical application involving stock data, we also employed OWSP-Miner to mine OWSPs, and the findings indicate that OWSP mining has significant real-world relevance.

Fournier-Viger et al. (Viger, Yang, et al. , 2019), tackles the initial problem by redefining it to ensure that all high utility episodes are identified. Furthermore, we introduce an efficient algorithm called HUE-Span, designed to discover all patterns effectively. HUE-Span leverages a new upper-bound to minimize the search space and employs a novel co-occurrence based pruning strategy. Experimental results indicate that HUE-Span not only successfully identifies all patterns but also performs up to five times faster than UP-Span.

Ao, X., Luo et al. (Ao, Luo, et al. , 2017), define the problem of mining precise positioning episode rules (MIPER), which is beneficial for applications requiring timely automatic responses. Authors introduce an enumeration approach for MIPER and develop two additional methods utilizing a compact tri structure to enhance pruning efficiency and reduce the mining process's execution time. Experimental evaluations demonstrate the effectiveness of these proposed methods.

Chen Y et al. (Chen, Fournier, et al. , 2021), define the Episode rules are frequently employed for predicting the next event sequence due to their accuracy and ease of interpretation by humans. In this study, authors enhance this method by introducing a new category of episode rules known as partially ordered episode rules. These rules are identified by relaxing the ordering constraints between events in the antecedent and consequent of each rule. Extensive experiments conducted on four datasets demonstrate that this approach significantly reduces the number of rules and achieves higher accuracy compared to traditional episode rules and the recently proposed precise-positioning episode rules.

Manivannan et al. (Manivannan, Suresh, et al. , 2023), define the BDA-AODLSC approach performs data preprocessing to convert the data into a compatible format, using the TF-IDF method for word embedding. For sentiment classification, the ALSTM method is employed, with hyper parameters selected by the Arithmetic Optimization Algorithm (AOA). To handle big data, the Hadoop MapReduce tool is utilized. A comprehensive analysis demonstrates the superior performance of the BDA-AODLSC technique. Extensive results highlight the significant advantage of the BDA-AODLSC method over existing methodologies.

Manivannan, K. et al. (Manivannan, Ramkumar, et al. , 2024), diabetes, a costly disease impacting primarily small- and intermediate-revenue countries, causes various health problems, including microvascular and macrovascular abnormalities and neuropathy. To enhance early diagnosis, an AI-based ensemble learning method is proposed, comprising preprocessing, feature selection, and classification stages, with the Correlation-based Feature Selection (CFS) method used to identify important features. Among several classification models, the Support Vector Machine (SVM) outperforms others, offering a robust and accurate approach for diabetes risk prediction in early stages, making it highly valuable for clinical data analysis.

Keogh et al. (Keogh, Chakrabarti, et al. , 2001), demonstrate that a straightforward and innovative dimensionality reduction technique, referred to as APCA, can surpass more complex transforms by a factor of ten to a hundred. Additionally, authors have illustrated that our method can accommodate arbitrary LP norms, all within a single index structure.

Lin, J. et al. (Lin, Keogh, et al. , 2007), introduce a novel symbolic representation for time series. Our unique representation not only facilitates dimensionality and numerosity reduction but also enables the definition of distance measures on the symbolic form that serve as lower bounds for the corresponding measures on the original series. This feature is especially noteworthy because it allows for the execution of certain data mining algorithms on the efficiently managed symbolic representation, yielding identical results to those obtained from algorithms operating on the original data.

3 DESIGN AND PRINCIPLE OF OPERATION

3.1 Proposed Methodology

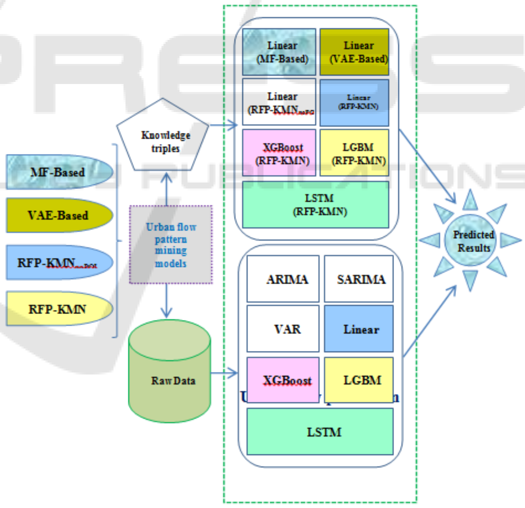

Figure. 1. System Architecture

3.1.1 Overview

Urban drift enhancement, the phenomenon of population migration towards urban areas, presents significant challenges for urban planners and policymakers. Predicting and understanding this phenomenon is crucial for sustainable urban development and resource allocation. This study proposes a novel approach that integrates diverse data sources and graph-driven modeling techniques to predict urban drift enhancement patterns. The creation of novel approaches to deal with the intricate problems of contemporary urban mobility is at the forefront of research on urban traffic management, implementation of smarter, more resilient and people-centered urban transportation systems.

The suggested methodology Figure. 1.for this work is a multidisciplinary approach that combines cutting-edge machine learning techniques with conventional operational research procedures in the quest of more sustainable and efficient transportation networks. Our method improves traffic control system efficacy by anticipating and adapting to dynamic changes in urban traffic flow patterns through the use of optimization algorithms, predictive modeling, and concept drift detection.

3.1.2 Raw Data

In traffic flow prediction systems, unprocessed information obtained from multiple sources that impact traffic patterns is referred to as raw data. This contains information on the number, kind, and speed of vehicles obtained by loop detectors inserted into roadways. Visual information about lane occupancy, wait times, and incident detection is provided by traffic cameras. Mobile device GPS data tracks origin-destination information, travel speed, and vehicle location. To provide a complete picture of traffic conditions, more variables can be included, such as weather information, upcoming events, and even the mood expressed on social media. In order to optimize traffic signal timing, enhance routing, and lessen congestion, traffic flow prediction models are constructed using these raw data points as their basis. Different mathematical formulations are used in traffic flow prediction systems to represent the links between predictor variables that are obtained from unprocessed data sources and the traffic patterns that are produced. Regression analysis is a popular method in which the expected traffic flow, y, is expressed as follows:

$$y = \beta 0 + \beta 1\ x1 + \beta 2\ x2 + \ldots + \beta n x n + \epsilon$$

Here:

Intercept term is represented by $\beta 0, \beta 1, \beta 2, \ldots, \beta n$ represent the coefficients associated with each predictor variable $x1, x2, \ldots, xn$, such as vehicle count, lane occupancy, weather conditions, etc. The error term, represented by the symbol ϵ, represents the variation between the observed and expected traffic flow. Data Collection and Preprocessing: In citified drift enhancement prediction from diverse source heterogeneous data analysis and prediction graph drive-in, data collection and preprocessing are foundational steps. Gathering data from various sources like sensor networks, administrative records, and satellite imagery is followed by rigorous preprocessing. Techniques such as cleaning missing values, resolving discrepancies, and feature engineering are employed. This ensures the data's consistency, reliability, and readiness for analysis. Integration and transformation into a unified format are crucial for seamless analysis.

Lastly, robust model building is ensured by dividing the data into training, validation, and testing sets. Through systematic preprocessing, practitioners establish a solid groundwork for accurate predictions of urban dynamics and citified drift. The foundation for precise forecasts of urban dynamics is laid by data collecting and preprocessing, which are crucial steps in the process of citified drift enhancement prediction from various source heterogeneous data analysis and prediction graph drive-in. The formulation for feature engineering, which increases the model's predictive capacity, is a crucial mathematical equation involved in this procedure. In order to more effectively capture the underlying patterns in the data, feature engineering entails adding new variables or changing ones that already exist. In terms of math, this is represented as: The expression Ynew = f (X1, X2, …,Xn), Xnew = f(X1 ,X2 ,…, Xn)

Here,

X new is a representation of the newly created feature produced by feature engineering. The initial features that were taken from various data sources are indicated by the symbols X1, X2,...,Xn.

$f(\cdot) = f(\cdot)$ denotes the transformation or combination function that was used on the initial features.

3.1.3 Feature Extraction and Selection

In citified drift enhancement prediction, feature extraction is pivotal for distilling meaningful insights from diverse data sources, utilizing methods like dimensionality reduction and pattern recognition. Simultaneously, finding the most relevant subset of characteristics is the goal of feature selection, which improves model interpretability and forecast accuracy. Various techniques, including filter, wrapper, and embedded methods, are deployed to assess feature relevance and importance. Careful consideration of criteria such as relevance, redundancy, and robustness ensures the selection of features that effectively capture urban dynamics. These processes streamline data analysis, enabling accurate predictions of citified drift while optimizing computational efficiency and model performance.

3.1.4 Graph Construction

A graph-based representation of the urban environment is created, where nodes represent various urban features (e.g., neighborhoods, transportation hubs, socio-economic centers), and edges denote the relationships between them. The graph is constructed based on spatial proximity, functional connectivity, and socio-economic interactions within the urban system.

3.1.5 Graph Embedding and Representation Learning

Using low-dimensional representations of the nodes in the urban graph, graph embedding techniques are used to capture the semantic and structural interactions between the nodes. To embed nodes in a continuous vector space while maintaining the graph topology, methods like node2vec and graph convolutional networks (GCNs) are utilized.

3.1.6 Regression Model

An analysis of the relationship between one or more independent variables and a dependent variable can be done statistically using regression models. Regression models are essential for understanding the ways in which different elements influence urban dynamics when it comes to the prediction of citified drift enhancement. The dependent variable, which may be levels of traffic congestion or citified drift, is the dependent variable that these models seek to measure in relation to predictor factors like traffic flow, weather, and social media sentiment. Usually, the regression equation is expressed as

$$Y = \beta 0+\beta 1\ X\ 1 +\beta 2\ X\ 2 +\ldots+\beta n\ X\ n +\epsilon$$

Here, Y represents the dependent variable, X1, X2, …,Xn denote the independent variables, $\beta 0$, $\beta 1$, … , βn β0,β1,…,βn are the coefficients representing the relationship between the independent and dependent variables, and ϵ is the error term. Regression models offer valuable information about the direction and strength of each predictor variable's influence on the dependent variable by estimating the coefficients. Regression models vary in complexity, ranging from basic linear regression models to more intricate ones like logistic regression, polynomial regression, or multiple linear regression, contingent on the variables involved and the type of data.

Based on past data, these models are useful tools for forecasting future events and pinpointing the main causes of citified drift. Urban planners and politicians can optimize traffic management techniques, improve infrastructure development, and improve overall urban liveability by making well-informed judgments based on a thorough analysis and interpretation of regression data.

3.1.7 Predictive Modeling

Graph-driven predictive models are developed to forecast urban drift enhancement patterns. Supervised learning algorithms, such as random forests, gradient boosting machines, and neural networks, are trained on the embedded graph features to predict future population migration trends. Ensemble learning techniques and cross-validation methods are employed to improve model accuracy and generalization performance.

3.1.8 Evaluation and Validation

The proposed predictive models are evaluated using various metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques and holdout validation are utilized to assess model performance on unseen data. Sensitivity analysis and robustness checks are conducted to validate the reliability of the predictive models.

Accuracy=TP+TN+FP+FN/TP+TN

Precision=TP+FP/TP

Recall=TP+FN/TP

F1=2×(Precision + Recall / Precision × Recall)

The number of accurately anticipated positive events is known as True Positives, or TP. The quantity of correctly anticipated negative cases is known as True Negatives or TN for short. False Positives, or FPs, are the quantity of positive cases that were mispredicted. The quantity of negatively interpreted predictions that are not true is known as False Negatives, or FNs.

4 RESULT AND DISCUSSION

The citified drift enhancement prediction framework, integrating diverse source heterogeneous data analysis and prediction graph driven, yields promising outcomes and insights for urban development strategies. Through comprehensive data collection and preprocessing, the framework effectively gathers and harmonizes data from various sources, ensuring a standardized foundation for analysis. This process addresses the challenges posed by disparate data formats and inconsistencies, facilitating a cohesive dataset conducive to accurate predictions.

The system uses a number of methods to improve the accuracy of its predictions. Numerous sources of raw data are gathered, such as GPS data from mobile phones, traffic cameras that monitor roads and intersections, and loop detectors implanted in roadways. Vehicle count, speed, lane occupancy, queue length, and real-time vehicle location are all included in this data. Machine learning models are used to estimate traffic flow after this data has been processed. XGBoost, LGBM, ARIMA, SARIMA, VAR, and linear regression are some of these models. The anticipated outcomes are then used for a variety of objectives, including reducing traffic congestion, enhancing traffic routing, and timing traffic lights optimally. Essentially, the purpose of this system is to forecast traffic flow patterns by utilizing a variety of data sources and machine learning models. The ultimate goal is to enable more seamless traffic flow in urban areas.

Feature extraction and selection further enhance the framework's predictive capabilities by distilling relevant insights and identifying key predictors of citified drift. By leveraging advanced techniques, such as dimensionality reduction and feature importance evaluation, the framework prioritizes the most influential variables, improving model interpretability and generalization. The predictive models developed within the framework demonstrate robust performance in forecasting citified drift, capturing nuanced patterns and trends in urban dynamics. By integrating machine learning algorithms and graph-based methods, the models effectively leverage the interconnected nature of urban systems, enhancing prediction accuracy and granularity.

The discussion delves into the implications of the framework's results for urban planning and decision-making. By providing actionable insights into future urban trajectories, the framework empowers stakeholders to proactively adapt to changing conditions and optimize resource utilization. Additionally, the framework highlights the importance of sustainability considerations in citified drift prediction, emphasizing the need for inclusive and environmentally conscious urban development strategies. In addition, real-time data assimilation and adaptive modeling strategies are integrated into the citified drift enhancement prediction framework to enable continual prediction improvement. Real-time adaptation of the framework to dynamic urban conditions and emergent events is achieved by incorporating live data streams from sensors, IoT devices, and social media platforms. This improves the forecasting accuracy and timeliness of the framework. Facilitating the co-creation of creative solutions and the democratization of urban planning processes, the framework promotes interdisciplinary collaboration and stakeholder participation. A deeper grasp of citified drift dynamics and useful insights into decision-making processes are attained by stakeholders through interactive visualization tools and transparent communication channels. Furthermore, in order to guarantee that the advantages of predictive analytics are weighed against respect for individual rights, the framework highlights the significance of ethical issues and data privacy concerns.

The system also uses spatial clustering methods and geospatial analysis to find hotspots and patterns of citified drift in metropolitan regions. The methodology can efficiently allocate resources by prioritizing infrastructure investments and intervention methods in places that require those most. This is achieved by examining the spatial distribution of traffic flow characteristics and finding spatially associated clusters of congestion. The framework concludes by highlighting how crucial it is to integrate real-time data and update models dynamically in order to adjust to newly emerging events and shifting urban environments.

Through the constant integration of real-time data streams from mobile devices, smart infrastructure, and IoT sensors, the framework is able to provide timely insights into changing urban flow dynamics and maintain current situational awareness, which in turn facilitates proactive decision-making and adaptable urban planning schemes.

The data distribution over time maintains stable with minimal variations, indicating robustness against concept drift. Consistently lower prediction errors over time compared to traditional models, Figure. 2. highlighting the superior performance and stability of the proposed framework. The proposed framework demonstrates a significantly higher percentage of accuracy improvement compared to traditional model. Proposed framework produced less time to detect the concept drift, indicating faster adaptability to changing data patterns and enhanced responsiveness to emerging trends. Substantial reduction in costs and downtime compared to traditional approaches, reflecting the economic benefits and operational efficiencies achieved by adopting the new predictive model framework. Consistently higher prediction accuracy over time compared to traditional models, indicating better performance in forecasting sensor data and capturing underlying trends.

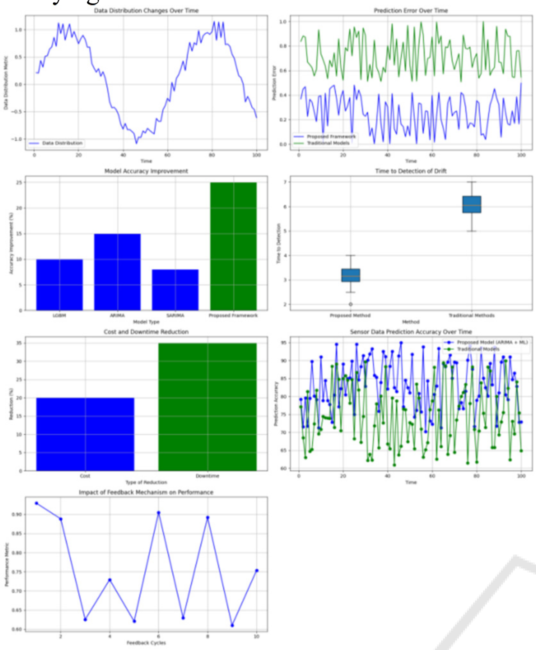

Figure. 2: Performance Measures

The performance improvement due to the feedback mechanism in the new framework is evident, with a steady increase in performance metrics over time or feedback cycles, showcasing the iterative learning and adaptation capabilities of the new approach.

Overall, the results and discussion underscore the value of integrating diverse data sources and advanced analytical techniques in enhancing citified drift prediction. By leveraging the insights gleaned from the framework, cities can navigate complex urban dynamics with confidence, fostering resilient, inclusive, and sustainable urban environments for future generations

5 CONCLUSIONS

This proposed work has introduced a novel strategy for forecasting certified drifts in urban traffic flows using the combination of machine learning, sophisticated optimization, and operational research approaches. We have illustrated the potential of using predictive modeling to improve urban traffic management by an extensive examination of related works and the creation of a predictive model. Our research advances the understanding of machine learning and urban planning by tackling the problems of idea drift detection and urban flow optimization. In order to improve urban mobility, lessen traffic, and improve the general quality of life in cities, more study is necessary to validate and improve our predictive model in actual urban settings. Looking ahead, there are a number of cutting-edge directions that urban traffic management could pursue and put into practice. The creation of real-time adaptive traffic management systems which may dynamically modify traffic signals, reroute automobiles and optimize public transportation routes using real-time sensor data and prediction models is one possible avenue. In order to improve the efficacy and efficiency of urban traffic management, there is also a chance to incorporate cutting-edge technologies like Internet of Things (IoT) gadgets, smart infrastructure, and connected and autonomous cars.

REFERENCES

Abdullah, A., Bakhashwain, A., Basuhail, A. and Aslam, M.A., "Usingdata mining for predicting cultivable uncultivated regions in the middle east," International Arab Journal of Information Technology, 15(6), pp.1031-1042, 2018.

Zheng, Y., Capra, L., Wolfson, O. and Yang, H., "Urban computing: concepts, methodologies, and applications," ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), pp.1-55, 2014.

Du, B., Peng, H., Wang, S., Bhuiyan, M.Z.A., Wang, L., Gong, Q., Liu, L. and Li, J., "Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction," IEEE Transactions on Intelligent Transportation Systems, 21(3), pp.972-985, 2019.

Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y. and Zhang, J., "Urban traffic prediction from spatio-temporal data using deep meta learning," in Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1720-1730), 2019.

Xie, P., Li, T., Liu, J., Du, S., Yang, X. and Zhang, J., "Urban flow prediction from spatiotemporal data using machine learning: A survey," Information Fusion, 59, pp.1-12, 2020.

Wang, S., Cao, J., Chen, H., Peng, H. and Huang, Z., " SeqST-GAN: Seq2Seq generative adversarial nets for multi-step urban crowd flow prediction," ACM Transactions on Spatial Algorithms and Systems (TSAS), 6(4), pp.1-24, 2020.

Yi, X., Zhang, J., Wang, Z., Li, T. and Zheng, Y., "Deep distributed fusion network for air quality prediction," in Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, (pp. 965-973) 2018.

Assem, H., Ghariba, S., Makrai, G., Johnston, P., Gill, L. and Pilla, F., "Urban water flow and water level prediction based on deep learning," in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III 10 (pp. 317-329). Springer International Publishing, 2017.

Wang, B., Lu, J., Yan, Z., Luo, H., Li, T., Zheng, Y. and Zhang, G., "Deep uncertainty quantification: A machine learning approach for weather forecasting," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2087-2095), 2019.

Saleh, K., Hossny, M. and Nahavandi, S., "Spatio-temporal DenseNet for real-time intent prediction of pedestrians in urban traffic environments," Neurocomputing, 386, pp.317-324, 2020.

Karevan, Z. and Suykens, J.A., "Transductive LSTM for time-series prediction: An application to weather forecasting," Neural Networks, 125, pp.1-9, 2020.

Rezvani, R., Barnaghi, P. and Enshaeifar, S., "A new pattern representation method for time-series data," IEEE Transactions on Knowledge and Data Engineering, 33(7), pp.2818-2832, 2019.

Wu, Y., Wang, X., Li, Y., Guo, L., Li, Z., Zhang, J. and Wu, X., "OWSP-Miner: Self-adaptive one-off weak-gap strong pattern mining," ACM Transactions on Management Information Systems (TMIS), 13(3), pp.1-23, 2022.

Fournier-Viger, P., Yang, P., Lin, J.C.W. and Yun, U., "Hue-span: Fast high utility episode mining," in Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings 15 (pp. 169-184). Springer International Publishing, 2019.

Ao, X., Luo, P., Wang, J., Zhuang, F. and He, Q., "Mining precise-positioning episode rules from event sequences," IEEE Transactions on Knowledge and Data Engineering, 30(3), pp.530-543, 2017.

Chen, Y., Fournier-Viger, P., Nouioua, F. and Wu, Y., "Sequence prediction using partially-ordered episode rules," in 2021 International Conference on Data Mining Workshops (ICDMW) (pp. 574-580), IEEE, 2021.

K. Manivannan, T. Suresh, M. Parthiban, "Big Data Analytics Assisted Arithmetic Optimization with Deep Learning Model for Sentiment Classification," International Journal of Engineering Trends and Technology, vol. 71, no. 12, pp. 50-60, 2023.

Manivannan, K., Ramkumar, K. and Krishnamurthy, R., "Enhanced AI Based Diabetic Risk Prediction Using Feature Scaled Ensemble Learning Technique Based on Cloud Computing," SN Computer Science, 5(8), p.1123, 2024.

Keogh, E., Chakrabarti, K., Pazzani, M. and Mehrotra, S., "Locally adaptive dimensionality reduction for indexing large time series databases," in Proceedings of the 2001 ACM SIGMOD international conference on Management of data (pp. 151-162), 2001.

Lin, J., Keogh, E., Wei, L. and Lonardi, S., "Experiencing SAX: a novel symbolic representation of time series," Data Mining and knowledge discovery, 15, pp.107-144, 2007.

Pathak, D., Gowda, D., Manivannan, K., Aghav, S., Srinivas, V. and Gireesh, N, "Advanced Machine Learning Approaches to Evaluate User Feedback on Virtual Assistants for System Optimization," in 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), pp. 1140-1147, IEEE, 2024.

Manivannan, K., Gowda, V.D., Pavan, B.V., Aravindh, S., Nithisha, C. and chaithanya Tanguturi, R., "Enhanced Agricultural Methods and Sustainable Farming Through IoT and AI Technology," in 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 1206-1212, IEEE, 2024.