IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , Journal Volume 11, Iss 04, 2022

MACHINE LEARNING MODELS FOR REAL-TIME TRAFFIC FLOW PREDICTION IN SMART CITIES

Rachna Rajput, Harmandeep Kaur
Guru Kashi University, Talwandi Sabo

Abstract

Using the Road Traffic Prediction Dataset and the PeMS dataset, this study assesses the traffic prediction performance of machine learning (ML) and deep learning (DL) models. These datasets were used to train and evaluate a variety of machine learning and deep learning models, such as MLP NN, Gradient Boosting, Random Forest, GRU, LSTM, Linear Regression, and Stochastic Gradient. The models' accuracy was evaluated using performance measures such Explained Variance (EV) score, Root Mean Squared Error (RMSE), R-squared (R2), and Mean Absolute Error (MAPE). The outcomes demonstrate the efficacy of MLP NN and Gradient Boosting in traffic prediction, with the former performing well on both datasets. The study emphasizes how crucial model selection and dataset selection are to enhancing traffic management system forecast accuracy.

Keywords: traffic congestion, neural networks, machine learning, deep learning, traffic flow prediction, adaptive traffic control, smart traffic light controllers.

1. INTRODUCTION

In huge urban areas, traffic guideline is truly troublesome. There are a few nations in the globe that have embraced Intelligent Transportation Systems (ITS) to lessen the costs connected with gridlock [1]. While creating ITSs, traffic stream expectation models are useful. To accomplish transportation strategy objectives and goals, for example, request the board or need measures for public transportation, a control and data framework utilizes coordinated correspondences and information handling innovation to work on the development of individuals and merchandise [2]. It does this by further developing wellbeing, diminishing street blockage, and really dealing with the event of clog [3].

Traffic stream expectation has a few purposes in region the board and city transportation. To gauge the stream volume at a future time utilizing information gathered from at least one perception stations during past periods, the traffic stream expectation issue is a period series issue [4]. The objective of this undertaking is to utilize a traffic stream expectation calculation to help the framework to gauge traffic. Contingent upon what the client looks for, the framework can propose things to them. Many elements associate progressively to produce gridlock [5]. These elements remember changes for traffic volume over the long run, the plan of the streets, the environment, mishaps, and upkeep exercises on the streets, among others. The system's ability to display traffic flow and weather conditions on the roads will benefit the general public by reducing the likelihood of accidents and enhancing road safety [6] [7]. The purpose of this study is to forecast urban traffic flow using machine learning (ML) techniques. As a type of artificial intelligence, machine learning entails creating computer algorithms that get more accurate as they process and absorb large quantities of data [8]. Many applications can benefit from machine learning's flexible ability to learn from previous data sets [9]. The applications of machine learning ideas allow for the prediction of traffic flow. If there is a change in environmental factors (such as construction or repair work, changes in road structure, or changes in weather), current approaches are unable to produce accurate forecasts [10]. Consequently, it's critical to create a prediction system that makes use of a greater variety of the factors that contribute to traffic congestion. The analysis of appropriate traffic flow characterization in urban road scenarios is the main subject of this investigation, with a particular emphasis on long- or short-term projections. The system automatically learns how to estimate the traffic by feeding it historical and time-series data to train machine learning algorithms [11]. Precise estimation of traffic patterns is crucial in contemporary transportation infrastructure. Many applications that require accurate future traffic information benefit from it.

2. REVIEW OF LITERATURE

Goyal, Bore, Gori, Y., Mayuri, K., Rao, A. L. N., & Krishna (2023, September) [12] demonstrated the practicality of using ML and DL calculations with intelligent traffic signal regulators, as evidenced by the good execution measures of all the calculations. Machine learning is a set of computations and measurable models that computers employ to perform predicted tasks. Machine learning might be utilized for various exercises, including quantitative trade, face acknowledgment, discourse acknowledgment, clinical assurance, traffic guaging, and that's only the tip of the iceberg. GPS courses have been used widely lately to decide the negligible portion of traffic in huge urban communities, because of the help of focal traffic-observing frameworks. With the data accumulated, an idea portraying the traffic examples of the city could be made, which could then be used to expect traffic and work with an obstruct concentrate on from now on. Subsequently, the main focal point of this try is traffic stream determining.

Tao, (2023) [13] suggested a technique for creating machine learning-based traffic flow prediction methods using several approaches and comparing the outcomes with other findings. The purpose of this study is to demonstrate how crucial it is to use machine learning technology to estimate traffic flow and how it could affect environmentally friendly transportation systems. The practical consequences of the research findings are noteworthy for urban designers, legislators, and transportation planners. The shown prediction models can aid in decision-making, allowing for preemptive steps to maximize traffic flow, lower emissions, and raise the general sustainability of transportation networks.

Profound repetitive brain networks were researched for their capability to order traffic designs in shrewd urban communities by **Ismaeel, (2023) [14].** We give an extraordinary profound intermittent brain network-based technique for characterizing traffic designs that is prepared to do effectively catching the consecutive and dynamic parts of traffic designs. The recommended model purposes a SoftMax layer to sort traffic designs and convolutional and intermittent layers to separate elements from traffic design information. The recommended model performs better compared to current methodologies concerning exactness, accuracy, review, and F1 score, as per trial information. Likewise, we offer a careful investigation of the discoveries and discuss what the proposed model might mean for savvy urban communities. The discoveries show the way that the recommended model can arrange traffic designs in shrewd urban communities with up to 95% exactness. The recommended model is evaluated utilizing a dataset of genuine traffic designs and diverged from current characterization methods.

Jafry, (2023, June) [15] Large-scale traffic data analysis is possible thanks to machine learning (ML) techniques, which may also be used to identify trends and learn from previous actions to forecast traffic flow in the future. Additionally, ML can improve traffic management by anticipating demand, controlling traffic signals, and optimizing routes. One important use of machine learning (ML) in traffic management is traffic prediction, where studies have shown that ML is more accurate in anticipating traffic congestion than conventional approaches. Machine learning is a useful tool for traffic management, especially when it comes to forecasting demand, anticipating congestion, and optimizing routes. According to studies, machine learning (ML) outperforms traditional methods in several domains, resulting in shorter travel times, smoother traffic flow, and overall better traffic management. The increasing need for sustainable and efficient transportation systems is anticipated to necessitate the inclusion of machine learning into traffic management. However, there are challenges and limitations that need to be addressed, such issues with data dependability and model interpretability. In spite of these obstacles, machine learning (ML) has promise for improving urban mobility and reducing traffic in smart, sustainable cities. To overcome these obstacles and fully utilize ML's potential in traffic control, more study is required.

3. METHODOLOGY

The Maharashtra Center Road Traffic Assumption Dataset is used in this study. Traffic sensors, like enrollment circles, transmit it. A few of public databases are crucial. Data is used for traffic planning and stop-light control restrictions. This dataset consolidates stream time series data from six urban crossing places over 56 days to anticipate vehicle passage over a day, making it ideal for ephemeral estimations. Four of the six convergences mimic a four-way junction in this survey.

3.1.Pre-treatment of Data

Due to sensor failures, missing attributes in informative collections are often zeros. Reports replace these features with the mean of the area with the missing value, but awards the missing data centers with the clear commonplace worth. The investigators considered that even if these replacement data are fake, they are better than none. However, when assumptions are made, this cycle creates spikes in genuine numbers, which increases the danger since the example occasionally shows that zero characteristics were valid. MinMaxScaler from scikit-learn scales data from 0 to 1 using the typical common spread. In this evaluation, the traffic stream for the next five minutes is calculated using the preceding hour's period series of twelve information of interest. The thirteen readings are made from overviews and used for testing and planning.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 04, 2022

Record after record is grouped and the readiness game strategy is stated randomly. Going forward, the last program segment is the outcome (Y) and the resulting segments are the information sources (X).

3.2.Neural Networks with Recurrence

3.2.1. Architecture of RNNs

GRU and LSTM are artificial discontinuous mind networks. The Keras library creates models. The data layer, which is not fixed by the number of time excursions per way. An outcome layer with neurons comparable to the sigmoid inception capacity and number of ways follows two dull layers with 64 neurons each and a 20% dropout rate.

3.2.2. RNN arrangement

Mean squared blunder (MSE) is used as the adversity ability in the model social affair, whereas mean outright mistake (MAE) is the estimation capacity. The Keras pack RMSprop enhancer has default values. Five percent of the planning data is used for endorsement, and 128 and 50 ages are used for preparation. The basics are done in Google Colaboratory and followed by Burdens and Inclinations.

3.3.AI Procedures

Straight Backslide, Slant Supporting Regressor, Multi-layer Perceptron Regressor, Stochastic Tendency Descendent Regressor, and Sporadic Forest Regressor are scikit-learn Python backslide models. For examine repeatability, every one of the models have default limitations and an unexpected state similar to nothing. These models require a 2D group rather than the 3D employed in RNNs, in this way 'X' is reshaped for both readiness and testing. The models receive the planning split.

4. RESULTS

To review how well the ML and DL estimations worked out, the 'y' test was first presented to a regressive scaler.

Though MAPE measures relative forecast mistakes, MAE and RMSE evaluate outright expectation blunders. For these three measures, more modest qualities signify better expectation execution. The relapse model fits better when the upsides of $R2$ and EV are more like one, which runs from zero to one. The exhibition measurements for every ML and DL model are recorded in Table 1.

Table 1: First dataset performance metrics comparison

ML/DL Model	MAE	MAPE	RMSE	R^2	EV score
MLP NN	11.263	23.26%	12.636	0.956	0.990

Gradient boosting	11.563	24.23%	12.363	0.945	0.994
Random Forest	11.845	26.23%	12.451	0.985	0.915
GRU	12.639	21.25%	12.636	0.945	0.966
LSTM	12.396	20.36%	12.452	0.978	0.945
Linear Regression	12.336	25.23%	12.285	0.933	0.932
Stochastic Gradient	12.336	22.36%	12.288	0.945	0.945

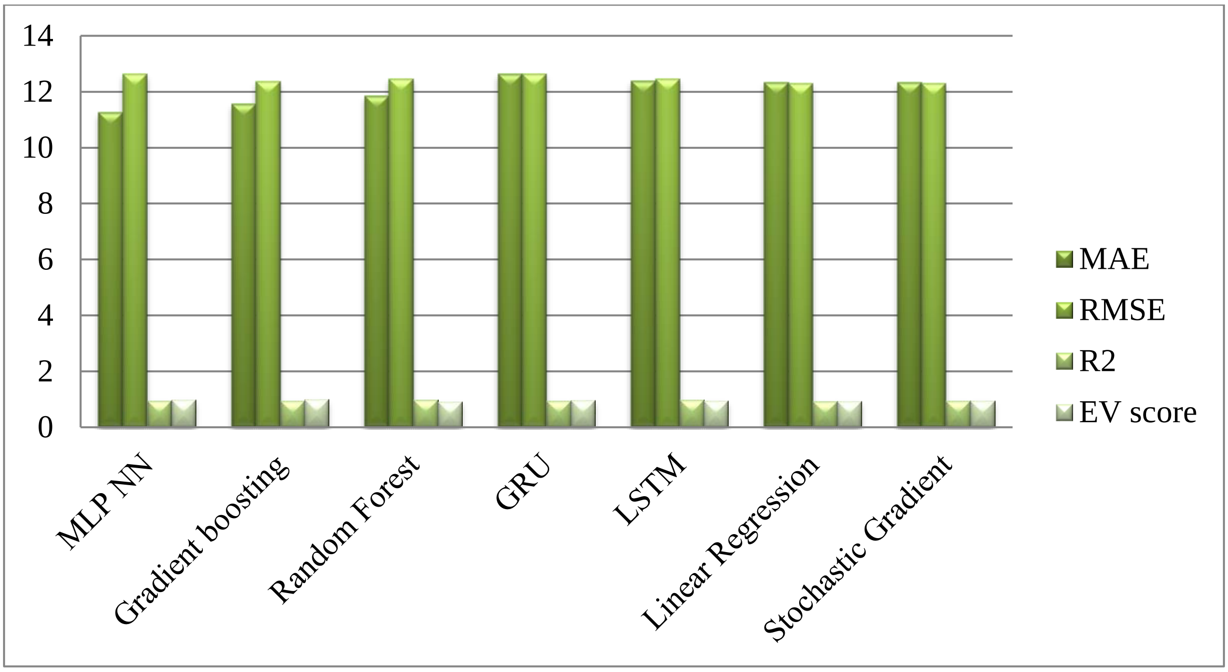

Figure 1: First dataset performance metrics comparison

The exhibition measurements of a few profound are utilized to assess these models. The discoveries exhibit the higher expectation precision of MLP NN and Slope Supporting, which acquire the most reduced MAE, MAPE, and RMSE. They further show their adequacy with great R2 and EV appraisals. While Arbitrary Timberland has solid prescient power, it performs essentially more terrible with regards to blunder measurements. Contrasting the RNN models GRU and LSTM to MLP NN and Angle Helping, they perform similarly, yet with fairly higher mistake measurements. In contrast with the best-performing models, Straight Relapse and Stochastic Slope models perform alright yet have bigger mistake measurements.

RNNs are at least a time or two arranged on distinct occasions, and the typical of each activity is obtained using ML models (scikit-learn). The erratic state allows us to compare outcomes each time.

Another dataset was employed for live testing instead of getting ready and endorsement. These new statistics come from the PeMS collection, which contains over 15,000 sensors across

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 04, 2022

California, mainly in the fourth district in Strait Locale, Alameda, Oakland. For power, *R*2 and EV score are suitable because to their dimensionlessness, adaptability, and standardization across many datasets and perspectives. Table 2 shows outside dataset ratings.

Table 2: Comparison of ML/DL Model Performance on PeMS Dataset

	MAE	MAPE	RMSE	R^2	EV score
MLP NN	8.125	18.135	9.052	0.956	0.9456
Gradient boosting	8.263	18.236	9.002	0.945	0.9561
Random Forest	8.632	17.226	9.056	0.978	0.9194
GRU	8.152	17.236	9.745	0.922	0.9569
LSTM	8.174	16.223	9.563	0.923	0.9481
Linear Regression	8.263	19.223	9.185	0.978	0.9945
Stochastic Gradient	8.336	20.331	9.563	0.966	0.9561

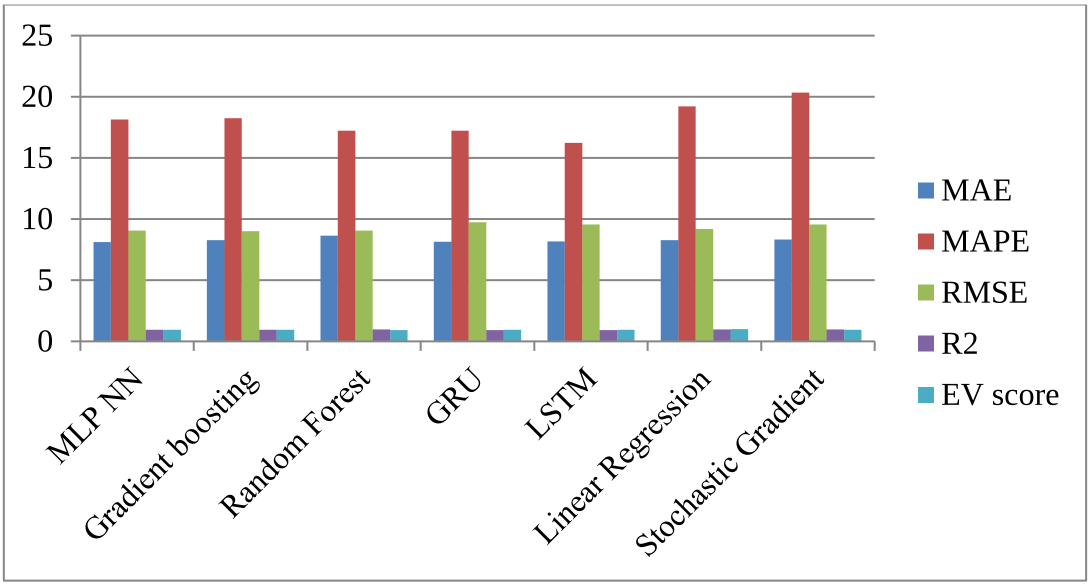

Figure 2: Comparison of ML/DL Model Performance on PeMS Dataset

The models' performance metrics on the PeMS dataset, a distinct dataset, are shown in Table 2, which evaluates the models' capacity for generalization On the new dataset, MLP NN, Gradient Boosting, and Random Forest continue to perform well, demonstrating their capacity to generalize and produce precise predictions across many datasets. With somewhat larger error metrics than the first dataset, GRU and LSTM likewise perform well on the second dataset. In comparison to the best-performing models, the Linear Regression and Stochastic Gradient models perform passably but have larger error metrics. The best-performing models overall are

MLP NN and Gradient Boosting, which show how well they can predict results from various datasets.

5. CONCLUSION

In order to anticipate traffic, this study used a public dataset and assessed the effectiveness of many deep learning (DL) and machine learning (ML) models. The outcomes demonstrated the superior performance of MLP NN and Gradient Boosting on both the original and new datasets, indicating their potent generalization and prediction powers. Although the RNN models, GRU and LSTM, fared marginally poorer but still rather well, Random Forest demonstrated competitive performance as well. Using MinMaxScaler to scale the data and handle missing values were two aspects of the pre-processing step. Input layers, two repetitive layers with 64 neurons each, a 20% dropout rate, and a result layer with neurons equivalent to the quantity of pathways contained the engineering for the RNN models. MSE was used as the error capability and RMSprop as the streamlining agent during model preparation.

REFERENCES

[1] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," IEEE Access, vol. 8, pp. 54776–54788, 2020.

[2] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method," Applied Computing and Informatics, vol. 12, no. 1, pp. 90–108, 2016.

[3] C. An and C. Wu, "Traffic big data assisted V2X communications toward smart transportation," Wireless Networks, vol. 26, no. 3, pp. 1601–1610, 2020.

[4] S. Bhattacharya, S. R. K. Somayaji, T. R. Gadekallu, M. Alazab, and P. K. R. Maddikunta, "A review on deep learning for future smart cities," Internet Technology Letters, vol. e187, 2020

[5] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of Statistical Software, vol. 33, no. 1, pp. 1, 2010

[6] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. Reddy Maddikunta, I. H. Ra, and M. Alazab, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," Electronics, vol. 9, no. 2, pp. 274, 2020.

[7] U. Gosh, "Smart city traffic pattern," [Online]. Available: https://www.kaggle.com/utathya/smart-city-traffic-patterns.

[8] C. Hinsbergen, J. W. C. Lint, and F. Sanders, "Short term traffic prediction models," in 14th world congress on intelligent transport systems, 2007.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 04, 2022

[9] J. Hoffmann, Y. Bar-Sinai, L. M. Lee, J. Andrejevic, S. Mishra, S. M. Rubinstein, and C. H. Rycroft, "Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers order in crumpled sheets," Science Advances, vol. 5, no. 4, pp. eaau6792, 2019.

[10] S. Verma and S. Badade, "Traffic prediction using machine learning," in Proceeding of national conference on machine learning, 2019.

[11] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," Journal of Statistical Software, vol. 39, no. 5, pp. 1, 2011.

[12] S. Goyal, N. M. Bore, Y. Gori, K. Mayuri, A. L. N. Rao, and A. M. Krishna, "A framework for traffic light control using machine learning and deep learning techniques," September 2023.

[13] Y. Tao, "Machine learning-based traffic flow prediction methods: A review," 2023.

[14] M. Ismaeel, "Deep recurrent neural networks for traffic pattern classification in smart cities," 2023

[15] M. T. Jafry, "Machine learning for intelligent traffic management: A review," June 2023.