Journal of Civil Engineering, Science and Technology

Volume 15, Issue 1, April 2024, 18 – 29

A COMPARATIVE STUDY OF CATBOOST AND ARTIFICIAL NEURAL NETWORKS IN ENHANCING TRIP GENERATION MODELLING FOR ILORIN CITY

Oreoluwa Temidayo Biala*
Department of Civil and Environmental Engineering, Federal University of Technology Akure, P.M.B 704, Akure, Ondo State, Nigeria

Date received: 20/10/2023 Date accepted: 29/01/2024
*Corresponding author's email: oreoluwabiala@gmail.com
DOI: 10.33736/jcest.6196.2024

Abstract — Trip generation plays a crucial role in transportation planning, and the choice of an appropriate model is essential for predicting future travel patterns. This study focuses on comparing the suitability and performance of CatBoost and ANN for trip generation (production and attraction) modelling of Ilorin City. By incorporating Ilorin household and trip characteristics, population data, and maps, this study evaluates the performance of the models. The two models demonstrated high accuracy and performance. In terms of trip production, the CatBoost model displayed exceptional accuracy, attaining an R-squared value of 0.99999992016446, accompanied by an impressively low mean squared error (MSE) of 3.93870930136429e-05. In contrast, the neural network exhibited a slightly lower accuracy of 0.999873850524181, with an error value of 0.0581313408911228. Similarly, for trip attraction, the CatBoost model showcased remarkable accuracy and precision, achieving an accuracy score of 0.9999999999999994 and an extremely low error value of 2.26762031965784e-13. The neural network model demonstrated an accuracy of 0.99999999990335 and a negligible error value of 0.000000041994. These findings underscore the strong predictive capabilities of both models for trip production and attraction, with the CatBoost model notably excelling in achieving nearly flawless accuracy and minimal error values across both aspects in Ilorin. Further research can explore the application of other advanced machine-learning techniques and combine their strengths to enhance the accuracy and robustness of trip-generation models.

Keywords: model, CatBoost, neural network, trip generation, machine learning

1.0 INTRODUCTION

In recent years, machine learning has gained significant attention as a powerful tool for predictive modelling in various domains, including transportation planning [1]. Predicting future travel demand is a crucial aspect of long-term transportation planning, aiding in devising strategies to cater to upcoming transport requirements. These strategies encompass land use regulations, pricing initiatives, and augmenting transportation infrastructure such as highways and transit services [2, 3]. With countless travellers making diverse choices regarding their travel times, destinations, and modes, the travel demand is substantial. These choices are influenced by various factors like family circumstances, traveller characteristics, and transportation options [4]. Trip generation, a fundamental aspect of transportation planning, refers to the estimation of the number of trips originating or ending in a particular zone [5]. With the surge in transportation demands attributed to factors like population growth, urbanization, and income per capita, the imperative lies in the development of transportation infrastructures [6]. At the core of effective transportation planning lies the pivotal concept of travel demand modeling. This process involves forecasting future travel patterns through the collection and analysis of present travel behavior data, forming the basis for creating predictive mathematical models [7]. These models hold a critical role in guiding decision-making for infrastructure enhancement and optimization. The conventional approach to travel demand modeling adheres to a sequence of four distinct stages: trip generation, trip distribution, modal split, and trip assignment. Among these, trip generation entails estimating the count of journeys originating from or destined for specific traffic analysis zones within a defined study region, constituting a fundamental facet of the broader travel demand modeling process [5].

Machine learning methods of forecasting are a focus in the current transportation industry and have been used successfully. Machine Learning models can handle large datasets and capture complex patterns and non-linear relationships [8]. However, they require careful feature selection, preprocessing, and validation to ensure accurate predictions [9]. The performances of machine learning models strongly rely on manually chosen features, and there are currently no generally accepted rules for selecting the right features [10]. The performance of a machine learning model is influenced by several factors such as the nature of the particular transportation-related problem at hand, the nature and complexity of the input datasets and the number of influential features (independent variables). The support vector machine model [11], Bayesian networks [12], the K-nearest neighbors model [13–15], the random forest model [16], as well as extreme machine learning and ensemble-based procedures [17], consistently yield satisfactory results for transportation-related problems. Other studies [18–25] effectively employed machine learning models and neural networks for transport related tasks.

Artificial Neural Network (ANN) is a subset of Machine Learning that can learn and model large datasets based on algorithms that simulates the human brain neurons to perform complex tasks easily with high accuracy [26–28]. Neural networks usually consist of three fundamental layers, namely, the input layer (which represents the quantity or values of the input variables), the hidden layer(s) (which are the intermediary nodes containing a number of neurons that receive and split the input values space into regions, multiplying each input value with numeric values called weights and passes each weighted value and a bias value unique to each neuron layer to an activation function which decides the final output value), and the output layer (which receives an output from the activation function of each preceding node) [29].

On the other hand, CatBoost is a novel form of gradient boosting algorithms or machines (GBMs), which are a type of ensemble learning algorithm designed for machine learning tasks and offer improvements over previously created ML tools in terms of speed, accuracy, robustness and scalability [30]. CatBoost is an open-source GBM that can be implemented in Python Programming Language [30, 31]. GBMs have been shown to perform accurately in practical regression and classification tasks in many empirical studies [32, 33]. CatBoost is an improvement on previous GBMs which overcome problems of target leakage, and it can be successfully and accurately applied to both small and diverse datasets [34, 35].

Several researchers have worked on using either CatBoost or ANNs, or both in various applications and purposes. Different types of neural networks have been used in previous transportation research, including Multi-Layer Perception (MLP) neural network for traffic flow predictive modelling [10], convolutional neural network for pedestrian traffic estimation [36], radial basis function neural network (RBFNN) [37], and convolutional long short-term memory (LSTM) neural network [38, 39]. Other ANN-based transport-related predictions include those of [10, 24, 40–44]. [27] worked on the application and performance evaluation of RBFNN for trip generation modelling of Akure, Nigeria.

CatBoost has found previous applications in various domains, including finance-related scenarios [45, 46], medicine and biology [47], electrical utilities [48], meteorology [49], manufacturing [50], and agriculture [51]. [52] used CatBoost and ANN for financial distress prediction, noting that CatBoost outperforms ANN for such scenario. [53] proposed card fraud detection models using CatBoost and deep neural networks, and suggested that both machine learning methods yielded highly accurate results, with CatBoost performing better. [54] used a hybrid of LSTM neural network and CatBoost for electricity price forecasting. Their study results showed that the hybrid model achieved lower errors compared to single models such as MLP neural network, ensemble tree, support vector regression (SVR), LSTM and gated recurrent units (GRU) neural networks.

In the field of transportation or traffic engineering, a study by [55] successfully applied CatBoost in predicting traffic accident impact range and proved the high performance accuracy and remarkable predictive capability of CatBoost algorithm. [56] suggested that CatBoost outperforms deep neural networks in the case of driver time arrival prediction for transportation planning. Other CatBoost applications in transportation engineering include detection of anomaly in IoT-based smart homes [57], and driving style detection [58].

This study aims to evaluate and compare the predictive performance of machine learning models for trip generation forecasting of Ilorin, Nigeria. The research specifically focuses on employing two prominent machine learning techniques: CatBoost and ANN, to develop accurate and efficient models for predicting trip generation patterns in urban areas. The results of the comparative analysis provide valuable insights into the efficacy of CatBoost and ANN models in predicting trip generation patterns. The research findings contribute to the advancement of transportation planning methodologies and provide urban planners with valuable tools for making informed decisions related to trip generation estimation.

2.0 MATERIALS AND METHODS

2.1. The Study Area

In 2006, the estimated population of Ilorin was 777,667. Projections further indicated that this number surged to 1,287,267 by 2022, positioning the city as the seventh-largest in Nigeria in terms of population [59, 60]. The city is situated at geographical coordinates of 8.4679° N and 4.5656° E and encompasses a total of twenty distinct political wards. Notably, Ilorin boasts a well-established intra-city public transit system alongside a meticulously maintained major road network. The visual representation of Ilorin's city map is shown in Figure 1.

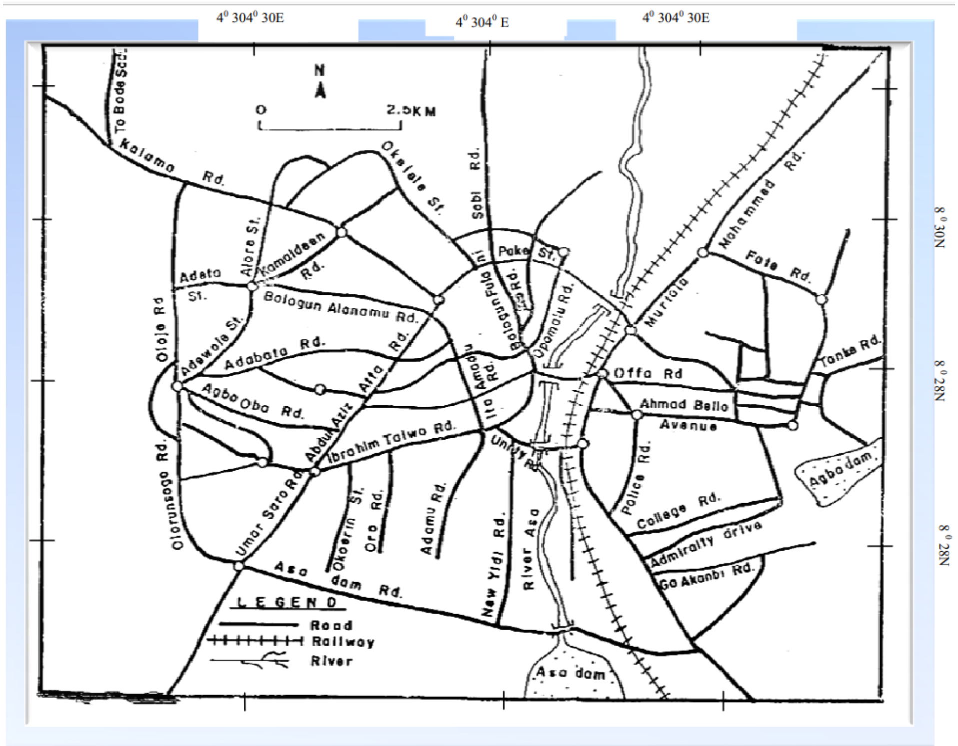


Figure 1 Map of Ilorin City

2.2. Data Collection and Analysis

This study integrates both secondary and primary data sources to form its foundation. The secondary data pool encompasses statistics related to population, land use attributes, and geographical maps. These resources were procured from entities such as the Kwara State Chapter of the National Populations Commission, other governmental agencies, and Geographic Information Systems (GIS). In contrast, the primary data pertains to intricate details encompassing household demographics, socio-economic characteristics, traffic volume and composition, as well as essential characteristics of travelers' trips including weekly trip counts, purposes of trips, and preferences of routes.

The collection of primary data was carried out through household surveys, with a specific focus on origin-destination surveys. This was achieved through the administration of questionnaires to households. The administered questionnaires were tailored to the already established zones. The process of conducting the survey embraced the systematic random sampling technique [61]. This approach involves selecting every "n" household (e.g., every tenth household) on a given street for the purpose of questionnaire administration. To ensure proper management of the overall sample count per zone, Yamane's adjusted formula was utilized [62].

The data analysis process employed in this study predominantly centered around the utilization of Microsoft Office Excel (latest version: 365) as the principal analytical tool, complemented by the selective use of Microsoft Word (365) and Microsoft Powerpoint as necessary. Additionally, the Geographic Information System (GIS) software also played a crucial role in facilitating a wide range of tasks, including map generation, analysis, and production.

In order to enable a systematic and efficient spatial observation, coupled with quantitative analysis of the intricate interplay between land use and economic factors influencing travel characteristics in Ilorin, the city was discretely partitioned into eight distinct zones using the ArcGIS software. These delineated segments, referred to as traffic analysis zones (TAZs), were meticulously crafted to ensure consistency in administrative boundaries, closely aligned with Ilorin's political wards. By adopting this strategic approach, the analytical process gains the capacity to integrate and explore the multifaceted dimensions of land usage patterns, population distributions, and administrative delineations. This integration significantly enhances the study's capacity to comprehensively examine the travel characteristics within the city.

2.3. Trips Generation Modeling

Trip production and attraction values were calculated based on the variables acquired through the data collection process and subsequent analysis. To ensure the robustness of the selected variables, a Pearson correlation analysis and a check for multicollinearity were conducted on the survey data. Through this rigorous evaluation, the variables with the most significant impact were identified for integration into the modeling process.

The variables subjected to correlation analysis included Household Size, Average Household Income, Employment, Vehicle Ownership, Number of Females, Number of Males, Average Age, Number of Males, Ownership of Drivers' License, Number of Students, and Total Trips. These variables were systematically examined to determine their relationships and potential influences on the travel patterns observed in the study area.

Subsequently, the CatBoost Gradient Boosting Algorithm and the Multilayer Perceptron Neural Network (MLPNN) models (as illustrated in Figure 2) were employed to establish and model the intricate connection between household and travel characteristics, and the values associated with trip generation, including both trip production and attraction. These advanced modeling techniques were harnessed to capture the complex interplay between the variables (dependent and independent).

The implementation of both models was conducted using the Python Programming Language within the Google Collaboratory Notebook environment. In order to streamline and facilitate the modeling process, a comprehensive selection of programming libraries and machine learning tools were utilized. These encompassed widely-used resources such as Keras, Pandas, NumPy, TensorFlow, Sklearn, and the Google Colab library. The integration of these tools was seamlessly achieved within the open-source Google Colab Notebook environment, which was strategically chosen to optimize the efficiency and effectiveness of the modeling procedure. This environment provided a versatile platform for code execution, data manipulation, and model training, ensuring a cohesive and productive modeling experience. CatBoost module contains an already built catboost algorithm that was imported and then fitted to the data while the ANN architecture was built from various Python commands.

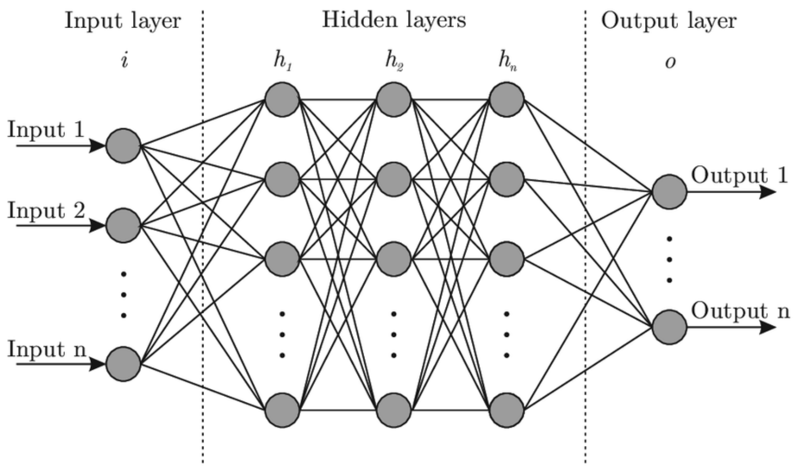


Figure 2 Typical MLPNN model architecture

2.4. Validation of the Models, Performance Evaluation and Visualisation

The research involved segregating the dataset into training and testing subsets (80-20 distribution) for model validation. This segregation followed an 80-20 distribution, with 80% of the data allocated for training and the remaining 20% designated for testing. This allocation aimed to strike a balance between model training and validation. The evaluation of the CatBoost and Neural Network models used key metrics like R-squared, Mean Squared Error (MSE), and Mean Absolute Error (MAE) to gauge predictive performance. Python libraries were utilized to create informative bar visualizations for a more intuitive presentation of performance differences between models. This comprehensive approach aimed to provide a clear understanding of the models' predictive capabilities and enable effective performance comparison.

3.0 RESULTS AND DISCUSSION

This section details the results of the data analysis, CatBoost and ANN modelling for trip production and attraction, and the performance of the models using selected evaluation metrics. The interpretation of the model performance results and visualisations are also discussed.

3.1. Delineating Ilorin into Traffic Analysis Zones (TAZs)

Ilorin city was first delineated into TAZs based on administrative boundries and practicability of household survey, data collection and traffic analysis. The administrative boundaries or political wards were used to determine the TAZS for easy spatial analysis of the city. Data collection or surveys were carried out based on the resulting traffic analysis zones. Eight TAZs were eventually taken from the 20 administrative wards as shown in Figure 3. using ArcGIS software. The figure shows the eight zones (labeled A to Z) as situated in Ilorin geographical map and their boundaries. The figure also uses different colors to distinguish betweeen the traffic analysis zones.

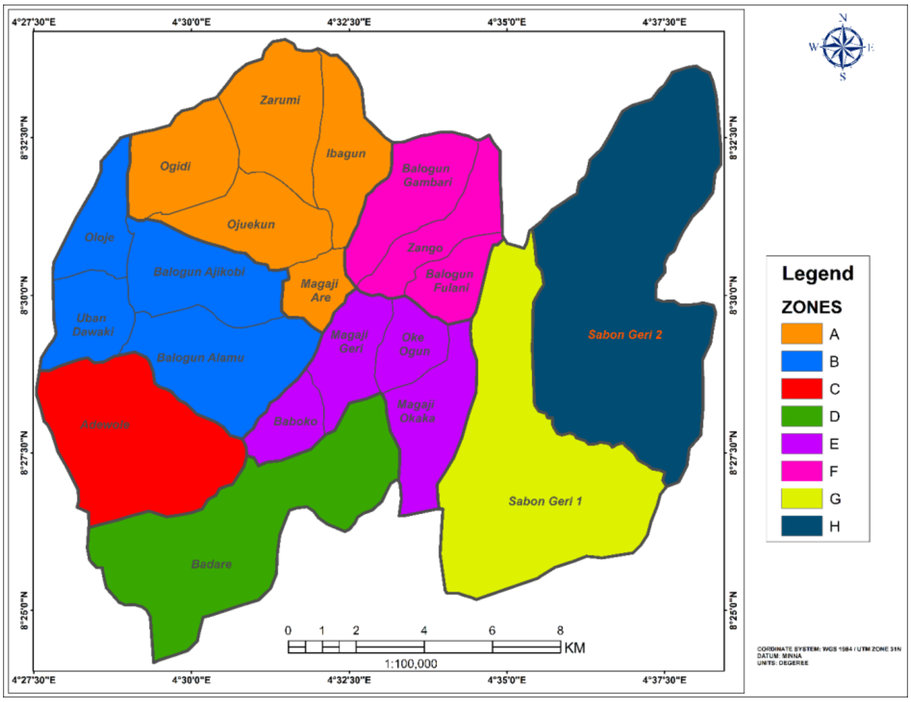


Figure 3 Ilorin TAZs, with administrative wards (Source: Ministry of Works and ArcGIS Software)

3.2. Summary of Data from Household Survey

The household survey was conducted using TAZs as a basis. Employing the Adjusted Yamane's Formula for sample size calculation and the systematic random sampling method [61] for selecting households to interview, the survey included 3457 individuals from 531 households across 8 zones. Among these, 54% were students, 41% were employed, and 5% fell into both categories. The travel diaries filled out by household members documented a total of 6,219 trips, with 80.96% being home-based and 19.04% non-home-based. Furthermore, the analysis revealed an average household size of 6.51 individuals in Ilorin. Additional details are summarized in Table 1, highlighting the description of the data. Figure 4 is a pie chart representing the proportion of home-based trips versus non home-based trips in percentage and sectors. It shows that 81% of the daily trips in the city are home-based while 19% are non home-based. This means that people in Ilorin travel to or from home more than they travel between non-home locations.

Table 1 Descriptive Statistics of Variables from the Survey

Variable	Mean	S. Error	Median	SD	Range	Min	Max
Household Size	6.510	0.218	6.387	2.268	19	2	21
*Av. Income (*1000)*	239.586	19.632	238.795	228.099	2170	30	2200
No Employed	2.674	0.159	2.613	1.441	12	1	13
No of Students	3.531	0.170	3.438	1.883	13	0	13
No of Vehicles	1.290	0.118	1.284	1.149	6	0	6
No of above 12yrs	4.996	0.165	5.031	2.260	32	1	33
No of Male	3.235	0.134	3.412	1.645	12	0	12
No of Female	3.275	0.156	3.297	1.518	13	0	13
Driver's Licence	1.358	0.140	1.422	1.191	12	0	12
Total Trips	11.712	0.570	12.203	5.383	24	3	27
Home-based	9.482	0.397	9.805	4.233	23	3	26
Non-Home-based	2.230	0.184	2.370	2.941	16	0	16
car trips	6.160	0.463	6.756	4.580	19	0	19
bus trips	2.674	0.100	2.692	1.014	4	1	5
keke (tricycle)	1.360	0.053	1.359	0.615	3	1	4
motorcycle	0.693	0.052	0.702	0.652	2	0	2
walking	0.832	0.058	0.834	0.799	4	0	4

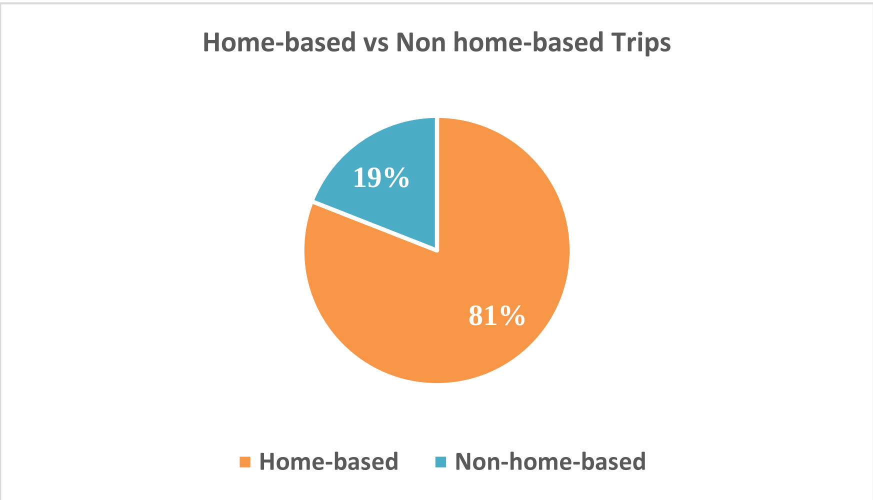


Figure 4 Pie representation of major trip purposes

3.3. Outcome of the Pearson Correlation Analysis

Table 2 shows the correlation analysis results for each pair of variables. Figure 5 shows a heat map representation of the coefficients. Once the correlation analysis was conducted, a subsequent multicollinearity check was performed to finalize the variables for the modeling process. The total number of trips (TT) served as the dependent variable for both trip production and trip attraction models. For trip production modeling, independent variables included Zonal Population and Average Household Income. The number of employed individuals was utilized as the independent variable for trip attraction modeling. These selected independent variables were identified as influential factors within their corresponding trip generation models.

Table 2 Outcome of the correlation analysis

	TT	*HS*	*AG*	*HI*	*NE*	*NS*	*NV*	*N12*	*NM*	*NF*	*DL*
Trips/Household	1.00										
Av. Household Size	0.92	1.00									
Av. Age	0.22	0.11	1.00								
Av. Income	0.60	0.61	0.64	1.00							
No Employed	0.78	0.70	0.56	0.78	1.00						
No of Students	0.40	0.58	-0.52	-0.05	-0.14	1.00					
No of Vehicles	0.72	0.56	0.74	0.64	0.88	-0.21	1.00				
No of above 12yrs	0.92	0.87	0.30	0.61	0.69	0.44	0.74	1.00			
No of Male	0.60	0.71	0.47	0.49	0.62	0.11	0.65	0.60	1.00		
No of Female	0.77	0.79	-0.24	0.44	0.44	0.71	0.23	0.70	0.13	1.00	
Driver's Licence	0.77	0.67	0.64	0.70	0.92	-0.05	0.96	0.82	0.63	0.40	1.00

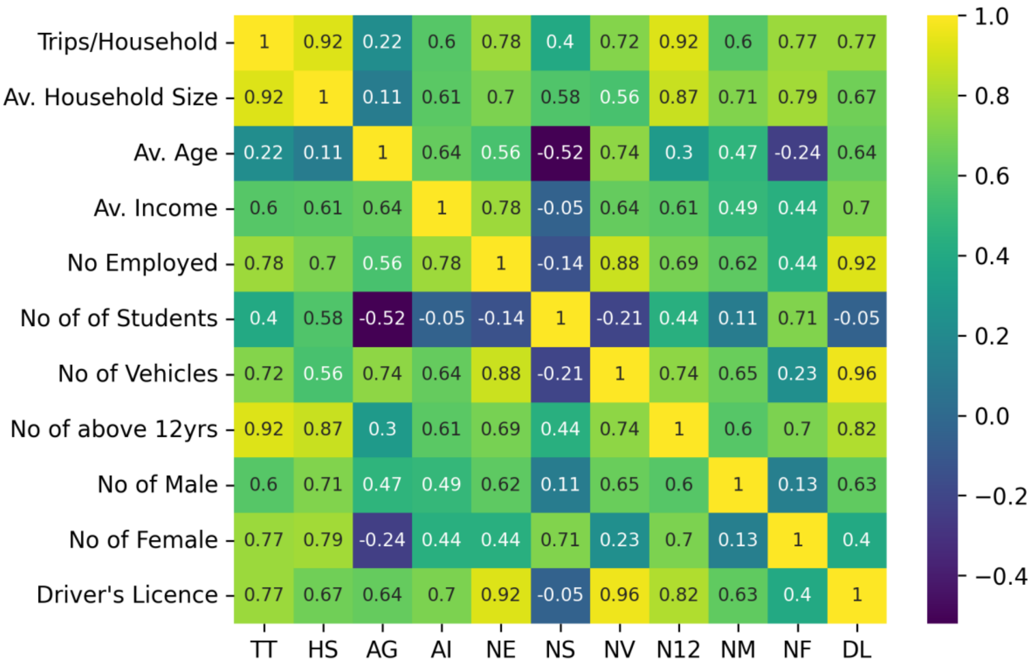


Figure 5 Heat map of the correlation analysis

3.4. Comparative Evaluation and Visualisation of the Performance of the Models

The trip production and trip attraction components of trip generation were modeled using CatBoost and MLP Neural Network techniques, leveraging the analyzed household/travel dataset. The assessment outcomes of these models are detailed in Table 3 and Table 4, featuring the evaluation metrics highlighted in the methodology section.

The outcomes indicate that both CatBoost and MLPNN models demonstrate remarkable accuracy in modeling trip generation data, encompassing both production and attraction scenarios. Additionally, these models showcase minimal error values, underlining their effectiveness in predicting trip generation values.

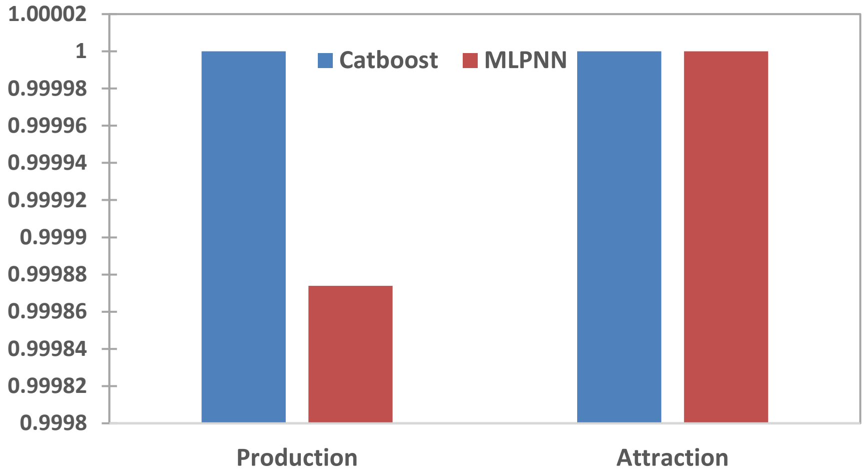
Table 3 CatBoost Models Performance

Accuracy measure	**Trip Production model**	**Trip Attraction model**
R-squared value	0.999999920164465	0.9999999999999994
Mean Absolute Error (MAE)	0.001659427440268973	3.8097102541101434e-07
Mean Square Error (MSE)	3.938709301364293e-05	2.267620319657848e-13

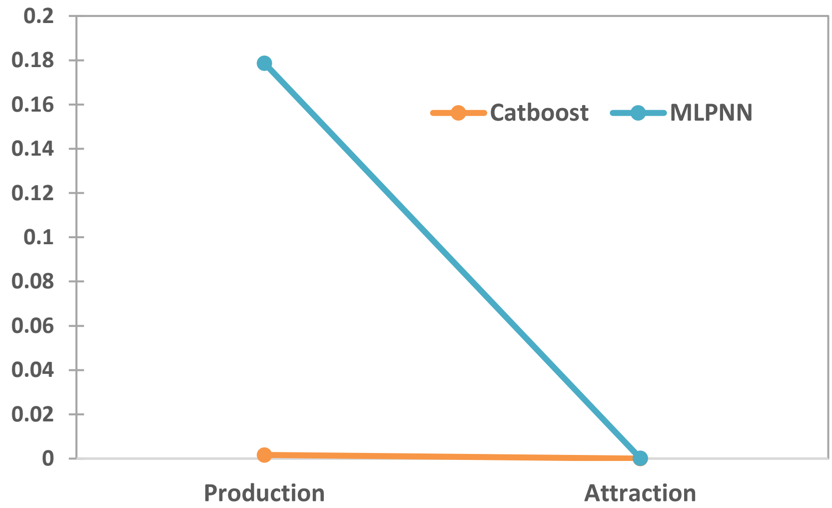
Table 4 Neural Network Models Performance

Accuracy measure	Trip Production model	Trip Attraction model
R-squared value	0.9998738505241817	0.99999999990033522
Mean Absolute Error (MAE)	0.17869409918785095	0.00015423298464156687
Mean Square Error (MSE)	0.05813134089112282	4.199481651312453e-08

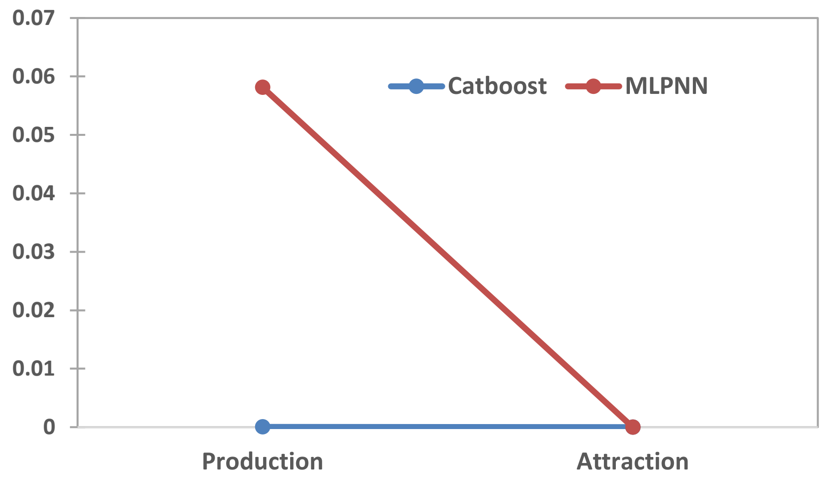
By visualizing accuracy (R-squared) and mean squared error (MSE) values, a comprehensive performance comparison of the models was conducted. WPS Office Software was used to create charts to illustrate the models' performance evaluations, as depicted in Figure 6(a)-(c). Specifically, for trip production, the visual representations unveiled closely matched and high R-square values for both CatBoost and MLPNN models, although CatBoost slightly outperforms MLPNN in accuracy, while the accuracy values were more closely matched for trip attraction (although the accuracy of CatBoost model was still slightly higher). Nonetheless, an intriguing observation emerged regarding error values: the CatBoost models demonstrated remarkably lower error values compared to MLPNN models , holding true for both trip production and attraction scenarios. In summary, both machine learning models demonstrated exceptional accuracy and low error values, yet the CatBoost Algorithm emerged as the stronger performer based on error metrics, outshining the MLPNN.



(a)



(b)



(c)

Figure 6 Models' Performance Comparison for Production and Attraction Cases by: (a) R-squared values (b) Mean Absolute Error (c) Mean Squared Error

4.0 CONCLUSION

This study explored the application and performance of CatBoost and Multilayer Perceptron Neural Network (MLPNN) in modelling trip production and attraction for Ilorin's transportation system planning. Major data used for the study was obtained from household/travel surveys. The data was explored and analysed, and correlation/multicollinearity analysis was done to choose the most influential variables in the data for the actual modelling process. The data was then used to develop the CatBoost and MLPNN models for trip production and trip attraction values. The performance of the models was then evaluated using evaluation metrics such as R-squared value, mean absolute error (MAE) and mean squared error (MSE). The performance results were visualised using Python library bar plots. The two models were then compared in terms of their accuracy and error values.

The comparative analysis of the CatBoost and neural network models for trip production and attraction revealed notable insights into their predictive capabilities. For trip production, the CatBoost model exhibited an exceptional accuracy level with an R-squared value of 0.99999992016446, accompanied by a remarkably low mean squared error (MSE) value of 3.93870930136429e-05. In contrast, the Neural Network demonstrated a slightly lower accuracy of 0.999873850524181, while its error value stood at 0.0581313408911228. Similarly, for trip attraction, the CatBoost model displayed remarkable accuracy and precision, boasting an accuracy of 0.9999999999999994 and an impressively low error value of 2.26762031965784e-13. The neural network model, on the other hand, showcased an accuracy of 0.99999999990335 and a negligible error value of 0.000000041994. These results underscore the high proficiency of both models in predicting trip production and attraction. However, the CatBoost model particularly excelled in terms of achieving near-perfect accuracy and minimal error values for both trip production and attraction of Ilorin Metropolis.

Based on the comprehensive analysis conducted, it is recommended to consider the CatBoost model as the preferred choice for trip generation modelling in Ilorin Metropolis. Its exceptional performance is vital for robust transportation planning, ensuring that future travel patterns are accurately predicted.

Further research can explore the application of other advanced machine-learning techniques and combine their strengths to enhance the accuracy and robustness of trip-generation models.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Tillema, F., Van Zuilekom, K. M., & Van Maarseveen, M. F. A. M. (2006). Comparison of neural networks and gravity models in trip distribution. Computer-Aided Civil and Infrastructure Engineering, 21(2), 104–119.

[2] Heyns, W., & Jaarsveld, S. (2017). *Transportation, Land Use and Integration: Perspectives for Developing Countries. South Africa*. WIT Press.

[3] Nwafor, E. O., Aderinlewo, O. O., & Atoo, A. A. (2018). Trip Generation Model for Makurdi Metropolis, Benue State. International Journal of Civil Engineering and Construction Science, 5(1), 17–24.

[4] Chakroborty, P., & Das, A. (2017). Principles of transportation engineering. PHI Learning Pvt. Ltd.

[5] Ahmed, B. (2012). The traditional four steps transportation modeling using a simplified transport network: A case study of Dhaka City, Bangladesh. International Journal of Advanced Scientific Engineering and Technological Research, 1(1), 19–40.

[6] Ogunbodede, E. F., & Ale, A. S. (2015). The Regression Model In The Forecast Of Travel Demand In Akure, Nigeria. Analele Universităţii din Oradea, Seria Geografie, 2, 186–194.

[7] Mannering, F. L., & Washburn, S. S. (2020). Principles of highway engineering and traffic analysis. John Wiley & Sons.

[8] Ha-Minh, C., Tang, A. M., Bui, T. Q., Vu, X. H., & Huynh, D. V. K. (2022). CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure. Springer.

[9] Ahmad, T., & Aziz, M. N. (2019). Data preprocessing and feature selection for machine learning intrusion detection systems. ICIC Express Lett, 13(2), 93–101. https://doi.org/10.24507/icicel.13.02.93

[10] Zhang, Z., Li, M., Lin, X., Wang, Y., & He, F. (2019). Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. Transportation Research Part C: Emerging Technologies, 105, 297–322. https://doi.org/10.1016/j.trc.2019.05.039

[11] Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., & Han, L. D. (2009). Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. Expert Systems with Applications, 36(3), 6164–6173. https://doi.org/10.1016/j.eswa.2008.07.069

[12] Fusco, G., Colombaroni, C., & Isaenko, N. (2016). Short-term speed predictions exploiting big data on large urban road networks. Transportation Research Part C: Emerging Technologies, 73, 183–201. https://doi.org/10.1016/j.trc.2016.10.019

[13] Zhang, L., Liu, Q., Yang, W., Wei, N., & Dong, D. (2013). An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. Procedia - Social and Behavioral Sciences, 96, 653–662. https://doi.org/10.1016/j.sbspro.2013.08.076

[14] Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. Transportation Research Part C: Emerging Technologies, 66, 61–78. https://doi.org/10.1016/j.trc.2015.08.017

[15] Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. (2016). A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C: Emerging Technologies, 62, 21–34. https://doi.org/10.1016/j.trc.2015.11.002

[16] Hamner, B. (2010). Predicting Travel Times with Context-Dependent Random Forests by Modeling Local and Aggregate Traffic Flow. In 2010 IEEE International Conference on Data Mining Workshops (pp. 1357–1359). IEEE. https://doi.org/10.1109/ICDMW.2010.128

[17] Jiber, M., Mbarek, A., Yahyaouy, A., Sabri, M. A., & Boumhidi, J. (2020). Road Traffic Prediction Model Using Extreme Learning Machine: The Case Study of Tangier, Morocco. Information, 11(12), 542. https://doi.org/10.3390/info11120542

[18] Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 54, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

[19] Gu, Y., Lu, W., Xu, X., Qin, L., Shao, Z., & Zhang, H. (2020). An Improved Bayesian Combination Model for Short-Term Traffic Prediction With Deep Learning. IEEE Transactions on Intelligent Transportation Systems, 21(3), 1332–1342. https://doi.org/10.1109/TITS.2019.2939290

[20] Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. Sensors, 17(4), 818. https://doi.org/10.3390/s17040818

[21] Wang, J., Gu, Q., Wu, J., Liu, G., & Xiong, Z. (2016). Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. In 2016 IEEE 16th International Conference on Data Mining (ICDM) (pp. 499–508). IEEE. https://doi.org/10.1109/ICDM.2016.0061

[22] Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926. https://doi.org/10.48550/arXiv.1707.01926

[23] Ke, J., Zheng, H., Yang, H., & Chen, X. (Michael). (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies, 85, 591–608. https://doi.org/10.1016/j.trc.2017.10.016

[24] Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies, 79, 1–17. https://doi.org/10.1016/j.trc.2017.02.024
[25] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems, 29. https://doi.org/10.48550/arXiv.1606.09375
[26] Adeke, P. T., Inalegwu, O. J., & Jirgba, K. (2019). Prediction of Bus Travel Time on Urban Routes Without Designated Bus Stops in Makurdi Town, Benue State, Nigeria. Arid Zone Journal of Engineering, Technology and Environment, 15(2), 406–417.
[27] Eromietse, E. J., & Joseph, O. O. (2019). Comparative Assessment of Radial Basis Function Neural Network and Multiple Linear Regression Application to Trip Generation Modelling in Akure, Nigeria. International Journal for Traffic And Transport Engineering, 9(2), 163–176. https://doi.org/10.7708/ijtte.2019.9(2).03
[28] Stuart, R., & Peter, N. (2016). Artificial intelligence-a modern approach 3rd ed. Berkeley.
[29] Zheng, F. (2011). Modelling urban travel times. PhD Thesis, Faculty of Civil Engineering and Geosciences, Delft University, Netherlands.
[30] Demir, S., & Sahin, E. K. (2023). Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost. Acta Geotechnica, 18(6), 3403–3419. https://doi.org/10.1007/s11440-022-01777-1
[31] Wang, Q., Li, Z., Cai, J., Zhang, M., Liu, Z., Xu, Y., & Li, R. (2023). Spatially adaptive machine learning models for predicting water quality in Hong Kong. Journal of Hydrology, 622, 129649. https://doi.org/10.1016/j.jhydrol.2023.129649
[32] Landry, M., Erlinger, T. P., Patschke, D., & Varrichio, C. (2016). Probabilistic gradient boosting machines for GEFCom2014 wind forecasting. International Journal of Forecasting, 32(3), 1061–1066. https://doi.org/10.1016/j.ijforecast.2016.02.002
[33] Gong, M., Bai, Y., Qin, J., Wang, J., Yang, P., & Wang, S. (2020). Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin. Journal of Building Engineering, 27, 100950. https://doi.org/10.1016/j. jobe.2019.100950
[34] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363. https://doi.org/10.48550/arXiv.1810.11363
[35] Bakhareva, N., Shukhman, A., Matveev, A., Polezhaev, P., Ushakov, Y., & Legashev, L. (2019). Attack Detection in Enterprise Networks by Machine Learning Methods. In 2019 International Russian Automation Conference (RusAutoCon) (pp. 1–6). IEEE. https://doi.org/10.1109/RUSAUTOCON.2019.8867696
[36] Kataev, G., Varkentin, V., & Nikolskaia, K. (2020). Method to estimate pedestrian traffic using convolutional neural network. Transportation Research Procedia, 50, 234–241. https://doi.org/10.1016/j.trpro.2020.10.029
[37] Zhu, J. Z., Cao, J. X., & Zhu, Y. (2014). Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections. Transportation Research Part C: Emerging Technologies, 47, 139–154. https://doi.org/10.1016/j.trc.2014.06.011
[38] Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 324–328). IEEE. https://doi.org/10.1109/YAC.2016.7804912
[39] Ma, Y., Zhang, Z., & Ihler, A. (2020). Multi-Lane Short-Term Traffic Forecasting With Convolutional LSTM Network. IEEE Access, 8, 34629–34643. https://doi.org/10.1109/ACCESS.2020.2974575
[40] More, R., Mugal, A., Rajgure, S., Adhao, R. B., & Pachghare, V. K. (2016). Road traffic prediction and congestion control using Artificial Neural Networks. In 2016 International Conference on Computing, Analytics and Security Trends (CAST) (pp. 52–57). IEEE. https://doi.org/10.1109/CAST.2016.7914939
[41] Kolidakis, S., Botzoris, G., Profillidis, V., & Lemonakis, P. (2019). Road traffic forecasting — A hybrid approach combining Artificial Neural Network with Singular Spectrum Analysis. Economic Analysis and Policy, 64, 159–171. https://doi.org/10.1016/j.eap.2019.08.002
[42] Loumiotis, I., Demestichas, K., Adamopoulou, E., Kosmides, P., Asthenopoulos, V., & Sykas, E. (2018). Road traffic prediction using artificial neural networks. In 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM) (pp. 1–5). IEEE. https://doi.org/10.23919/seeda- cecnsm.2018.8544943
[43] Sun, S., Wu, H., & Xiang, L. (2020). City-Wide Traffic Flow Forecasting Using a Deep Convolutional Neural Network. Sensors, 20(2), 421. https://doi.org/10.3390/s20020421
[44] Do, L. N. N., Vu, H. L., Vo, B. Q., Liu, Z., & Phung, D. (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. Transportation Research Part C: Emerging Technologies, 108, 12–28. https://doi.org/10.1016/j.trc.2019.09.008
[45] Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting loan default in peer-to-peer lending using narrative data. Journal of Forecasting, 39(2), 260–280. https://doi.org/10.1002/for.2625
[46] Fan, J., Wang, X., Zhang, F., Ma, X., & Wu, L. (2020). Predicting daily diffuse horizontal solar radiation in various climatic regions of China using support vector machine and tree-based soft computing models with local and extrinsic climatic data. Journal of Cleaner Production, 248, 119264. https://doi.org/10.1016/j.jclepro.2019.119264
[47] Yang, H., & Bath, P. A. (2020). The Use of Data Mining Methods for the Prediction of Dementia: Evidence From the English Longitudinal Study of Aging. IEEE Journal of Biomedical and Health Informatics, 24(2), 345–353. https://doi.org/10.1109/JBHI.2019.2921418

[48] Coma-Puig, B., & Carmona, J. (2019). Bridging the Gap between Energy Consumption and Distribution through Non-Technical Loss Detection. Energies, 12(9), 1748. https://doi.org/10.3390/en12091748
[49] Punmiya, R., & Choe, S. (2019). Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing. IEEE Transactions on Smart Grid, 10(2), 2326–2329. https://doi.org/10.1109/TSG.2019.2892595
[50] Najm, S. M., Trzepieciński, T., & Kowalik, M. (2023). Modelling and parameter identification of coefficient of friction for deep-drawing quality steel sheets using the CatBoost machine learning algorithm and neural networks. The International Journal of Advanced Manufacturing Technology, 124(7–8), 2229–2259. https://doi.org/10.1007/s00170-022-10544-1
[51] Zhang, Y., Zhao, Z., & Zheng, J. (2020). CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. Journal of Hydrology, 588, 125087. https://doi.org/10.1016/j.jhydrol.2020.125087
[52] Jabeur, S. Ben, Gharib, C., Mefteh-Wali, S., & Arfi, W. Ben. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. Technological Forecasting and Social Change, 166, 120658. https://doi.org/10.1016/j.techfore.2021.120658
[53] Nguyen, N., Duong, T., Chau, T., Nguyen, V.-H., Trinh, T., Tran, D., & Ho, T. (2022). A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network. IEEE Access, 10, 96852–96861. https://doi.org/10.1109/ACCESS.2022.3205416
[54] Zhang, F., Fleyeh, H., & Bales, C. (2022). A hybrid model based on bidirectional long short-term memory neural network and Catboost for short-term electricity spot price forecasting. Journal of the Operational Research Society, 73(2), 301–325. https://doi.org/10.1080/01605682.2020.1843976
[55] Zhang, S., Liu, H., Yang, Y., Zhang, S., Zhang, Z., Wang, C., & Wang, M. (2023). Prediction of traffic accident impact range based on CatBoost ensemble algorithm. In A. Palanisamy Muthuramalingam & K. Subramaniam (Eds.), Second International Conference on Algorithms, Microchips, and Network Applications (AMNA 2023) (Vol. 12635, p. 54). SPIE. https://doi.org/10.1117/12.2679147
[56] Sergoyan, H. (2020). Estimating Time of Driver Arrival with Gradient Boosting Algorithms and Deep Neural Networks. Mathematical Problems of Computer Science, 53, 29–38. https://doi.org/10.51408/1963-0050
[57] Reddy, D. K. K., & Behera, H. S. (2022). CatBoosting Approach for Anomaly Detection in IoT-Based Smart Home Environment. In Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021 (pp. 753–764). Springer. https://doi.org/10.1007/978-981-16-9447-9_56
[58] Liu, W., Deng, K., Zhang, X., Cheng, Y., Zheng, Z., Jiang, F., & Peng, J. (2020). A Semi-Supervised Tri-CatBoost Method for Driving Style Recognition. Symmetry, 12(3), 336. https://doi.org/10.3390/sym12030336
[59] National Population Commission (NPC). (2006). *National Population Census of Federal Republic of Nigeria Official Gazette* (2nd ed., Vol. 96).
[60] National Bureau of Statistics (NBS) Federal Republic of Nigeria (2012). (n.d.). Annual Abstract of Statistics.
[61] Fasakin, J.O., Basorun, J. O., Bello, M. O., Enisan, O. F., Ojo, B., and Popoola, O. O. (2018). Effect of Land Pricing on Residential Density Pattern in Akure, Nigeria. Advances in Social Sciences Research Journal, 5(1). https://doi.org/10.14738/assrj.51.4003
[62] Adam, A. M. (2020). Sample Size Determination in Survey Research. Journal of Scientific Research and Reports, 26, 90–97. https://doi.org/10.9734/jsrr/2020/v26i530263