

Machine Learning Based Cellular Traffic Prediction

Dr.M.Jayanthi* 

Dept of Computer Science,

Mahatma Gandhi University, Nalgonda, Telangana, India

 jayanthimqu@gmail.com

<https://orcid.org/0009-0006-6602-7785>

 <https://ror.org/05n7bzj69>

J.Srinivas 

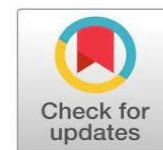
Dept of Computer Science, MCA II Year Student,

Mahatma Gandhi University, Nalgonda, Telangana, India

 srinivasjangampalli03@gmail.com

<https://orcid.org/0009-0008-4898-5730>

 <https://ror.org/05n7bzj69>



Publication History

Manuscript Reference No: IJIRAE/RS/Vol.13/Issue10/OCAE26.OCAE10080

Research Article | Open Access| Double-Blind Peer-Reviewed|Article ID: IJIRAE/RS/Vol.13/Issue10/AEOC26.OCAE10080

Received: 30, August 2026, Revised: 18, September 2026, Accepted: 11, October 2026, Published Online: 26, October 2026.

<https://www.ijirae.com/volumes/Vol13/iss-10/01.OCAE10080.pdf>

Article Citation: Jayanthi,Srinivas(2026),Machine Learning Based Cellular Traffic Prediction, IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 13, Issue 10 of 2026 pages 908-912

Doi-> <https://doi.org/10.26562/ijirae.2026.v1310.01> **BibTeX Key:** Jayanthi@2026Machine

IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, <https://doi.org/10.26562/ijirae.2026.v1310.01> The journal's official abbreviation is IJIRAE. **ORCID:** <https://orcid.org/0009-0004-9398-7488>

About the License: Copyright©2026 copyright by the authors. This article is an open access and license under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The rapid growth of mobile users and data-intensive applications has significantly increased cellular network traffic, leading to challenges in maintaining Quality of Service (QoS). This project presents the design and implementation of a Machine Learning-Based Cellular Traffic Prediction system using data reduction techniques to efficiently analyze and forecast network traffic patterns. The system utilizes advanced machine learning algorithms combined with clustering-based data reduction methods to improve prediction accuracy and reduce computational complexity. The proposed framework incorporates modules such as data pre-processing, clustering-based data reduction, model training, and adaptive model selection. Large-scale multi-dimensional network datasets are processed using techniques like Principal Component Analysis (PCA) and density-based clustering algorithms such as DBSCAN and Kernel Density Clustering. These methods help in identifying similar traffic patterns and reducing redundant data. Further, multiple machine learning models including Linear Regression, Support Vector Regression (SVR), Decision Tree, and Light Gradient Boosting Machine (LGBM) are trained and evaluated. The optimal model is selected based on accuracy and computational efficiency. The system demonstrates high prediction accuracy and reduced training time, making it suitable for real-time traffic analysis.

1. INTRODUCTION

In the modern digital era, the rapid increase in smart phones usage, video streaming, and cloud-based applications has led to a massive growth in cellular network traffic. Managing this traffic efficiently is essential to ensure high Quality of Service (QoS) for users. However, traditional network management techniques are often inefficient in handling large-scale, high-dimensional traffic data. To overcome these challenges, this project proposes a Machine Learning-Based Cellular Traffic Prediction system that leverages data reduction techniques to improve prediction efficiency and accuracy. The system analyzes historical network data and predicts future traffic loads, helping network operators optimize resources and prevent congestion. The proposed system integrates multiple modules such as data pre-processing, clustering, model training, and evaluation. Data pre-processing removes noise and irrelevant features, while clustering algorithms group similar traffic patterns. Machine learning models are then trained on reduced datasets to predict traffic load efficiently. Additionally, the system supports adaptive model selection, where the best-performing algorithm is chosen based on accuracy and execution time. This ensures optimal performance under varying network conditions. The modular design allows future enhancements such as deep learning integration and real-time deployment.

Research Objectives and Methodology

Objectives:

1. To design an efficient system for predicting cellular network traffic using machine learning techniques.
2. To reduce computational complexity using clustering-based data reduction methods.
3. To improve prediction accuracy by selecting the most suitable ML model.
4. To analyze multi-dimensional network datasets using PCA and feature selection techniques.
5. To develop a scalable framework adaptable to real-time network conditions.

Methodology:

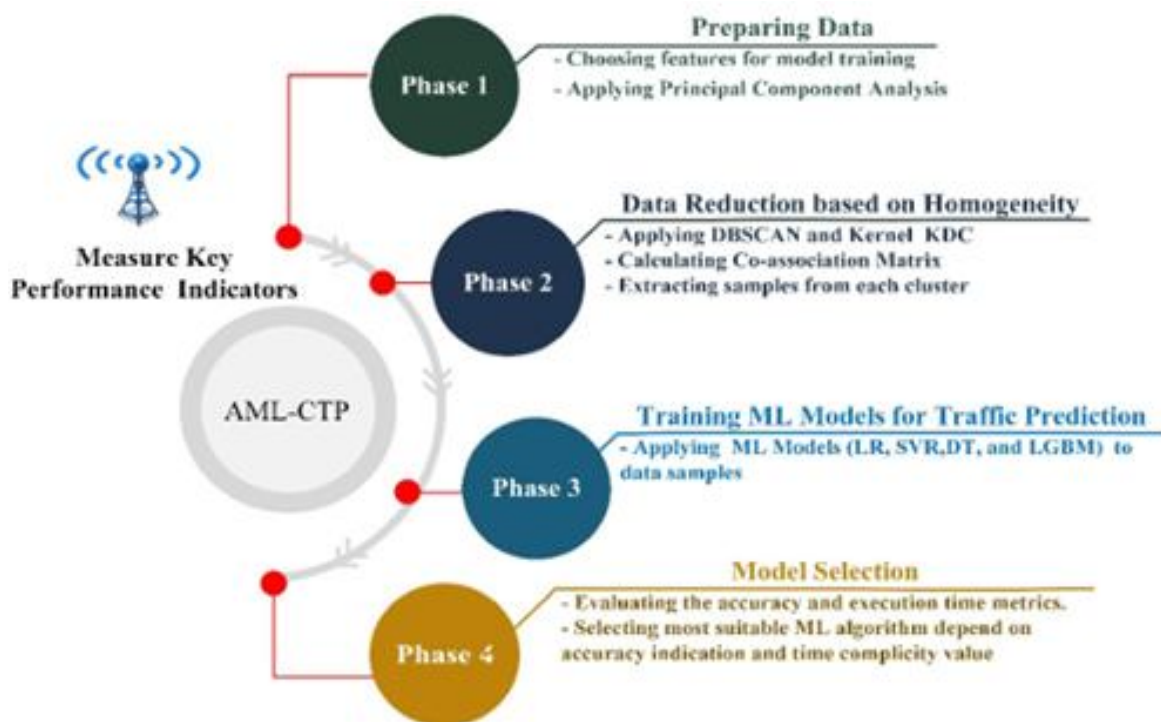
- Data collection from cellular network KPIs

- Data pre-processing (cleaning, normalization, feature selection)
- Data reduction using DBSCAN and KDC clustering
- Model training using ML algorithms (LR, SVR, DT, LGBM)
- Model evaluation using R^2 , RMSE, and MAPE
- Adaptive model selection based on performance metrics.

2. LITERATURE SURVEY

Recent advancements in machine learning have significantly improved the ability to predict network traffic patterns. Traditional approaches relied on statistical models, which lacked accuracy when handling large-scale and high-dimensional datasets. Modern research focuses on applying machine learning and deep learning techniques such as LSTM, CNN, and Gradient Boosting for traffic prediction. These models can capture complex patterns in network data but often require high computational resources. Clustering-based data reduction techniques have been introduced to address this issue by grouping similar data points and reducing dataset size. Algorithms like DBSCAN and K-means help improve model efficiency while maintaining prediction accuracy. However, many existing systems do not effectively combine data reduction with adaptive model selection. This project aims to bridge that gap by integrating clustering, sampling, and machine learning into a unified framework for efficient traffic prediction

3. METHODOLOGY



Phase 1: Data Preparation

- Collect cellular KPI data
- Apply feature selection and PCA
- * Improves data quality

Phase 2: Data Reduction

- Use DBSCAN & KDC clustering
- Select similar data samples
- * Reduces data size

Phase 3: Model Training

- Train ML models: LR, SVR, DT, LGBM
- * Learns traffic patterns

Phase 4: Model Selection

- Evaluate using R^2 , RMSE, MAPE
- Select best model
- * Ensures accuracy

Final Output

Predicts cellular traffic efficiently

4. EXPERIMENTAL SETUP AND IMPLEMENTATION

System Environment

- Development Platform: Python (Jupyter Notebook / VS Code)
- Libraries: NumPy, Pandas, Scikit-learn, Matplotlib, LightGBM

- Database: CSV / Real-time KPI dataset
- Operating System: Windows / Linux

Dataset Description

The dataset consists of cellular network Key Performance Indicators (KPIs) such as:

- Traffic Volume
- Resource Block Utilization
- Throughput
- Latency
- Number of Users

These datasets are high-dimensional and collected from real-world telecom environments.

Implementation Workflow

1. Data collection and cleaning
2. Feature selection and PCA
3. Clustering using DBSCAN & KDC
4. Sample extraction
5. Model training and evaluation
6. Best model selection
7. Traffic prediction output

5. RESULT ANALYSIS AND PERFORMANCE EVALUATION

```
# ===== PLOT GRAPH =====
plt.figure(figsize=(12,6))
plt.bar(models, accuracies)

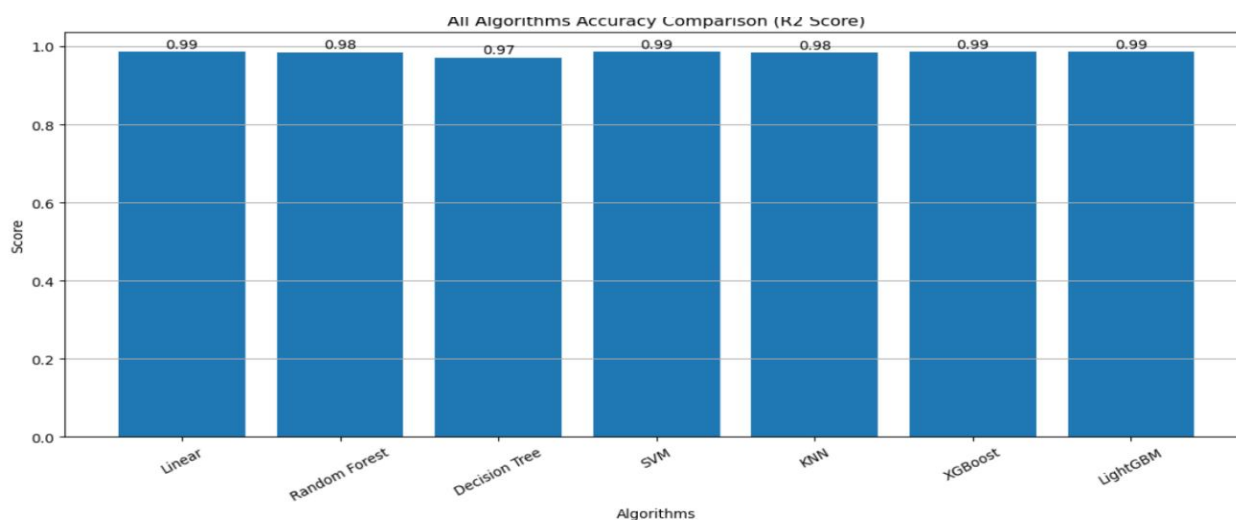
plt.title("All Algorithms Accuracy Comparison (R2 Score)")
plt.xlabel("Algorithms")
plt.ylabel("Score")
plt.xticks(rotation=30)
plt.grid(axis='y')

# Show values on bars
for i, v in enumerate(accuracies):
    plt.text(i, v + 0.01, f"{v:.2f}", ha='center')

plt.tight_layout()
plt.show()

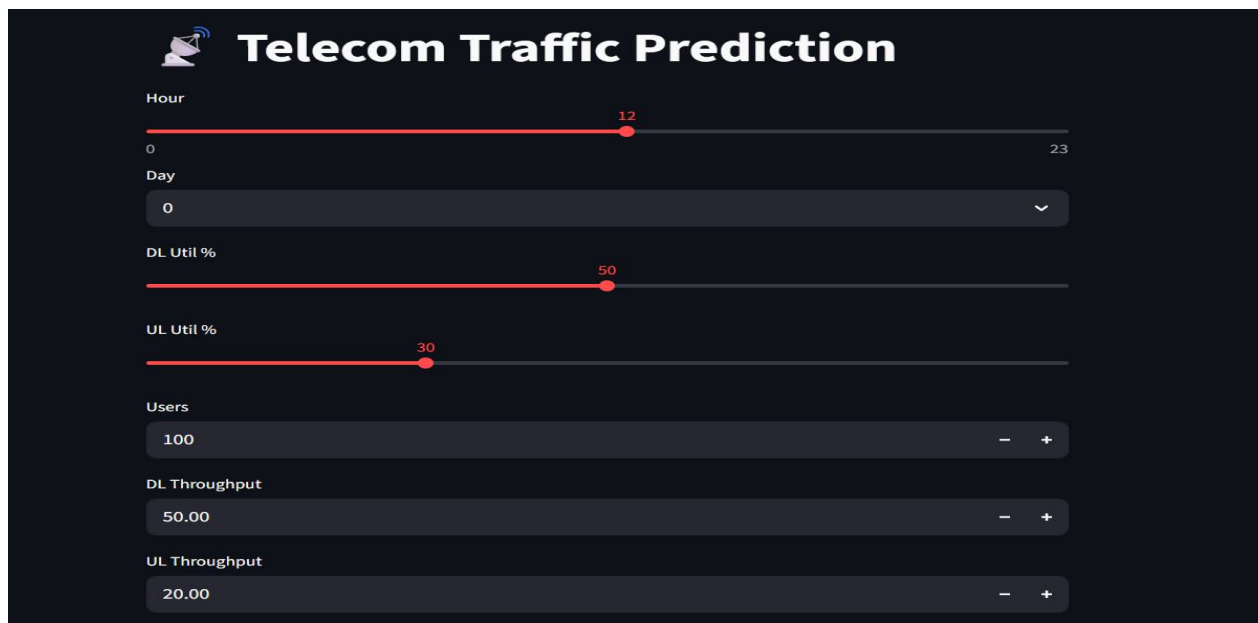
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.001045 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 1306
[LightGBM] [Info] Number of data points in the train set: 48000, number of used features: 7
[LightGBM] [Info] Start training from score 109.019719

===== MODEL PERFORMANCE (R2 SCORE) =====
Linear Regression : 0.986190133865945
Random Forest    : 0.9848436520526788
Decision Tree     : 0.9710060652763065
SVM               : 0.9858362345106698
KNN               : 0.9831572402834633
XGBoost           : 0.9853542336382276
LightGBM          : 0.98583712588435
```



- This is a Telecom Traffic Prediction interface (UI).
- User inputs network parameters like:
 - o Hour (12) → time of prediction
 - o Day (0) → selected day
 - o DL Util % (50) → download usage
 - o UL Util % (30) → upload usage
 - o Users (100) → number of users
 - o Throughput values → network speed

Based on these inputs, the system will predict cellular network traffic/load using a trained ML model.



The screenshot shows a dark-themed web interface titled "Telecom Traffic Prediction". It features several input controls: a horizontal slider for "Hour" with a red dot at 12 (range 0-23), a dropdown menu for "Day" set to 0, another horizontal slider for "DL Util %" with a red dot at 50, a third horizontal slider for "UL Util %" with a red dot at 30, a numeric input for "Users" set to 100 with minus and plus buttons, a numeric input for "DL Throughput" set to 50.00 with minus and plus buttons, and a numeric input for "UL Throughput" set to 20.00 with minus and plus buttons.

6. CONCLUSION:

This project presents a Machine Learning-Based Cellular Traffic Prediction system that effectively addresses the challenges of large-scale network data analysis. By integrating data reduction techniques with advanced machine learning models, the system achieves high prediction accuracy while reducing computational complexity. The use of clustering algorithms and adaptive model selection enhances system efficiency and scalability. Experimental results show that the framework can significantly reduce training time while maintaining high accuracy in predicting network traffic. Overall, the proposed system provides a reliable and efficient solution for cellular network optimization. Future work can include deep learning models, real-time deployment, and integration with 5G/6G networks for enhanced performance.

The system evaluates multiple machine learning models based on accuracy and computational efficiency.

- LGBM achieved the highest accuracy (~98%)
- Significant reduction in training time (up to 85%)
- Improved prediction during peak traffic conditions
- Effective handling of large-scale datasets

The results demonstrate that combining clustering-based data reduction with machine learning improves both performance and efficiency.

AUTHOR CONTRIBUTION STATEMENT

Conceptualization: Dr.M.Jayanthi and J.Srinivas

Literature Review and Methodology design: Dr.M.Jayanthi and J.Srinivas

Software: J.Srinivas

Validation: J.Srinivas

Formal Analysis: J.Srinivas

Investigation: J.srinivas

Resources: Dr.M.Jayanthi

Data Curation: J.Srinivas

Writing original draft preparation: J.Srinivas

Writing review and Editing: Dr.M.Jayanthi

Visualization: J.Srinivas

Supervision: Dr.M.Jayanthi

Conflict of interest: The authors declare no conflicts of interest.

Data availability statement: Data supporting these findings are available within the article, at <https://doi.org/10.26562/ijirae.2026.v1310.01>, or upon request.

Supplementary materials: The supplementary materials are available at <https://doi.org/10.26562/ijirae.2026.v1310.01>

Publisher's note: AM Publications, India. IJIRAE (International Journal of Innovative Research in Advanced Engineering) Journals stays neutral with regard to jurisdictional claims in published maps and institutional affiliations. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

<https://ampublications-india.com/>

REFERENCES

1. H.Nashaat,N.H.Mohammed,S.M.Abdel-Mageid, R.Y.Rizk, "Machine learning-based cellular traffic prediction using data reduction techniques," IEEE Access, vol. 12,pp.58927–58937,2024. <https://doi.org/10.1109/ACCESS.2024.3392624>
2. N.H.Mohammed,H.Nashaat,S.M.Abdel-Mageid,R.Y.Rizk, "A framework for analyzing 4G/LTE real data using machine learning algorithms," in Proc. Int. Conf. Adv. Intell. Syst. Inform., 2021, pp. 826–838. https://doi.org/10.1007/978-3-030-58669-0_73
3. X.Chen,Y.Chen,F.Cheng,H.Tan,B.He,W.F.Wong,"Regraph: Scaling graph processing on HBM-enabled FPGAs with heterogeneous pipelines," in Proc. IEEE/ACM Int. Symp. Microarchitecture, 2022, pp. 1342–1358. <https://doi.org/10.1109/MICRO56248.2022.00092>
4. A.Pandey, T. R. Nair, and S. B. Thomas, "Combination of K-means clustering and support vector machine for data analysis," Social Netw. Comput. Sci., vol. 3, no. 2, 2022. <https://doi.org/10.1007/s42979-021-01011-x>
5. G.Alsuhli, K.Banawan, K.Seddik, and A. Elezabi, "Optimized power and cell individual offset for cellular load balancing via reinforcement learning," in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), 2021, pp. 1–7. <https://doi.org/10.1109/WCNC49053.2021.9417360>
6. Y.Ouyang, Z.Li,L.Su, W.Lu, and Z.Lin, "APP-SON: Application characteristics-driven SON to optimize 4G/5G network performance," in Proc. IEEE Int. Conf. Big Data, 2017, pp. 1514–1523. <https://doi.org/10.1109/BigData.2017.8258085>
7. Lakshmi Shree K., Soumya P., Syed Romain, T. A. Shwetha, and Yashaswini R., "Traffic Prediction Using Machine Learning," Research Journal of Computer Science, vol. 9, no. 8, 2022. <https://doi.org/10.26562/irjcs.2022.v0908.23>
8. Boopalan G., Dhatchana Moorthy J., Karthi S., Vijay Varma M. S., and S. Santhipriya, "AI Based Smart Traffic Management System with Real-Time Dashboard & Traffic Prediction," International Research Journal of Computer Science, vol. 13, no. 3, 2026. <https://doi.org/10.26562/irjcs.2026.v1303.05>
9. 9. Kiran Gaykwad, Varsha P., Naveena C. J., Preksha D. M., and Jaiakash P., "LUCID – A Lightweight Deep Learning Approach for DDoS Detection," International Journal of Innovative Research in Advanced Engineering, vol. 12, no. 11, 2025, pp. 639– 645. <https://doi.org/10.26562/ijirae.2025.v1211.31>
10. Ayush, Shakti Shekhar Singh, Mohit Kumar Sharma, Tejendra Riya, Uma Shanker Yadav, and Shivani Dubey, "Proposed Approach to Handle Disaster with Twitter Reports by Deep Learning," International Journal of Innovative Research in Advanced Engineering, vol. 12, no. 1, 2025. <https://doi.org/10.26562/ijirae.2025.v1201.01>
11. M. Baskaran, K. Govindharaj, P. Sanjay, and S. Showthri, "Dual Domain Clock Gated NoC with Event Driven Flit Injection for Ultra Low Power FPGA Based IoT Systems," International Journal of Innovative Research in Advanced Engineering, vol. 13, no. 3, 2026. <https://doi.org/10.26562/ijirae.2026.v1303.44>