International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Machine Learning-Based Cellular Traffic Prediction Using Data Reduction Techniques

[1]Dr G Rama Subba Reddy, [2]Vaddi Obulesu , [3]Ajay Gujjari , [4]pattupogula Lakshmikala ,[5]Vamshi Nalapalli
[1]HOD & Professor in AIML [2,3,4] UG students
[2,3,4,5] Department of Computer Science and Engineering, Sai Rajeswari Institute of Technology

Abstract- — Estimating and analyzing traffic patterns is essential for managing Quality of Service (QoS) metrics in cellular networks. Cellular network planners often employ various approaches to predict network traffic. However, existing algorithms rely on large datasets, leading to significant time complexity and resource demands. To address this issue, we introduce a novel algorithm, AML-CTP (Adaptive Machine Learning-based Cellular Traffic Prediction), which is trained on a small, accurate dataset to enhance prediction accuracy while reducing complexity. Our methodology includes data normalization using the Min-Max Scaler, feature selection via the Select-K-Best algorithm, and dimensionality reduction through PCA. We apply density-based clustering techniques (DBSCAN and Kernel Density) to identify high-similarity clusters for training. We evaluate several machine learning algorithms, including Support Vector Machine (SVM), Linear Regression, Decision Tree, Light Gradient Boosting, and XGBoost, using a Cellular LTE dataset from an Egyptian company. The results demonstrate that the Decision Tree algorithm achieved the highest R^2 score of 96%, followed by the extension XGBoost model, which reached a remarkable R^2 score of 98%, indicating its superior performance in cellular traffic prediction.

Keywords: Cellular network traffic prediction, Quality of Service (QoS), Adaptive Machine Learning-based Cellular Traffic Prediction (AML-CTP), small dataset learning, Min-Max normalization, Select-K-Best feature selection, Principal Component Analysis (PCA), density-based clustering, DBSCAN, Kernel Density estimation, machine learning models, Support Vector Machine (SVM), Linear Regression, Decision Tree, Light Gradient Boosting, XGBoost, LTE traffic dataset, prediction accuracy, reduced computational complexity, cellular network planning.

I. INTRODUCTION

Nowadays, a rash growth of traffic data exists due to the rise of smartphone subscriptions and streaming video services. As a result, this influences the Quality of Service (QoS) of network users. Therefore, many network-level optimizations should be performed to maintain the best QoS for users. However, an optimization problem is challenging for the best QoS that adjusts the transmission power. Steering traffic from congested cells frees the physical resource blocks (PRBs) of congested cells and is considered a network that optimizes resource allocation. Hence, predicting 4G and 5G LTE-A traffic became essential to significantly enhance telecommunications QoS. Accurate monitoring and prediction of mobile traffic help improve optimizer goals; consequently, any overload or congestion in any band can be detected. Over the past decades, Machine Learning (ML) has become a crucial backbone of information technology. However, with the significant increase in data sizes, training time for models can range from hours to weeks. Hence, it poses intense pressures across computation, networking, and storage. In turn, this affects the choice of the ML model. Also, the nature of the data, which includes correlation, distribution, and homogeneity, influences the selection process of the final model. Since ML deals with the data and behaves accordingly without being programmed, data analysis is required to evaluate and estimate the benefits of optimizer goals. Practical data analysis leads to the successful choice of the most accurate prediction algorithm and saves computational time and memory usage.

Problem Statement:

- Cellular network traffic is highly dynamic and unpredictable, leading to inaccurate traffic predictions that degrade network Quality of Service (QoS) and increase the complexity of managing network resources.
- Existing algorithms for cellular traffic prediction rely on large datasets and computationally expensive models, which often result in high time complexity and resource-intensive processing, limiting their practical application.
- Network planners, operators, and end-users experience degraded network performance, including slower data

speeds, dropped connections, and inconsistent service quality, especially during peak traffic times.

- Inefficient traffic management leads to underutilization of network resources, increased operational costs for network providers, and reduced customer satisfaction, ultimately impacting profitability and user experience.
- We propose AML-CTP, an adaptive machine learning algorithm, utilizing clustering, feature selection, and dimensionality reduction to enhance cellular traffic prediction accuracy while minimizing resource and time demands.

Software Requirements

- Software Platform: Anaconda
- Primary Programming Language: Python
- Frontend Framework: Flask
- Development Environment: Jupyter Notebook
- Database: SQLite3
- Front-End Technologies: HTML, CSS, JavaScript, Bootstrap

Hardware Requirements

- Operating System: Windows
- Processor: Intel i5 or above
- RAM: 8 GB or above
- Hard Disk: Minimum 25 GB free space in local drive

II. LITERATURE SURVEY

Train a central traffic prediction model using local data: A spatio-temporal network based on federated learning:

https://www.sciencedirect.com/science/article/abs/pii/S0952197623007960

Abstract: With the rapid growth of vehicle and road sensors, large amounts of traffic data can now be collected, making deep learning an effective tool for accurate traffic prediction in the Internet of Vehicles (IoV). However, most existing methods require sharing all local data to train a unified model, which raises serious concerns about data privacy, security, and communication cost.To overcome these issues, this paper proposes a Federated Spatial-Temporal Traffic Prediction Network (F-STTP-Net). Instead of sharing raw data, each local area trains its own model and only uploads model parameters to a central server, protecting data privacy.First, the road network is divided into multiple sub-areas based on their traffic characteristics. For each sub-area, a local model is designed using a Graph Attention Network (GAT) to capture spatial relationships between roads and an LSTM to model traffic changes over time. A branch-based structure is used to predict traffic volume at each intersection.Finally, the local models are combined using federated learning to build a strong global model without accessing any raw traffic data. Experiments on real-world data from the Xuchang Lotus Lake 5G autonomous driving test area show that F-STTP-Net delivers high prediction accuracy, strong generalization ability, and can be easily adapted to new areas.

Machine Learning Based Traffic Prediction System in Green Cellular Networks:

https://ieeexplore.ieee.org/abstract/document/10040347

Abstract: Accurate prediction of cellular network traffic is essential for improving network performance, quality of service (QoS), and energy efficiency. Modern mobile networks serve a large number of users and consume significant energy, making efficient resource management critical. This research focuses on developing a reliable time-series model to predict cellular traffic load in advance. By identifying user location, signal-to-interference-plus-noise ratio (SINR), and QoS requirements, the network can dynamically adjust transmit power to the minimum level needed. This approach improves service quality, reduces energy consumption, and alleviates heavy traffic load without completely shutting down base stations.

Comparison of Machine Learning Techniques Applied to Traffic Prediction of Real Wireless Network:

https://ieeexplore.ieee.org/abstract/document/9623523

Abstract: The rapid growth of connected devices has led to a significant increase in network traffic, making efficient traffic prediction essential for improving system performance. Machine Learning techniques are widely used to analyze traffic patterns and optimize network operations. This study reviews Machine Learning applications in communication networks and compares several classical models for LTE edge traffic prediction. Experimental results show that Gradient Boosting achieves the highest prediction accuracy, while Support Vector Machines offer the fastest training time. Bayesian and Huber regression models also perform well, whereas Random Forest shows weaker performance. The source code for all models is released as open access.

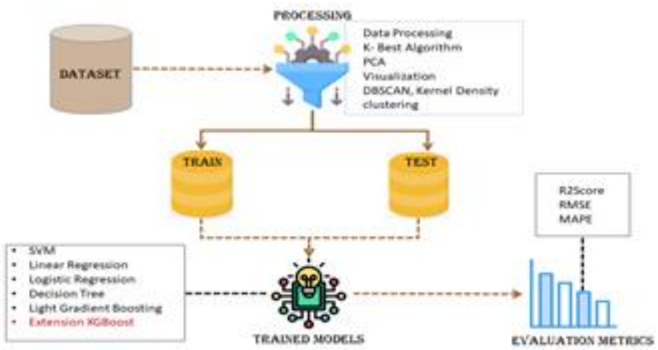
Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data:

https://ieeexplore.ieee.org/abstract/document/9623523

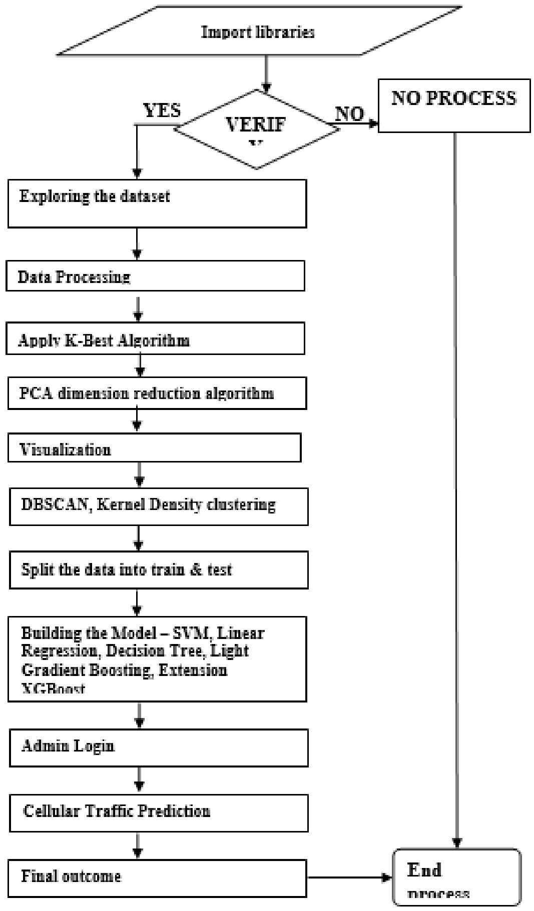
Abstract: Accurate traffic prediction is critical for intelligent cellular networks. This paper proposes STCNet, a deep learning model that uses convolutional LSTM to capture spatial–temporal traffic patterns and incorporates cross-domain data to model external influences. By clustering city areas and applying transfer learning across regions and traffic types, STCNet improves prediction accuracy. Experiments on real-world datasets show that STCNet outperforms existing methods, achieving 4%–13% performance gains through transfer learning.

III. SYSTEM DESIGN

System Architecture:



IV. DATA FLOW DIAGRAM

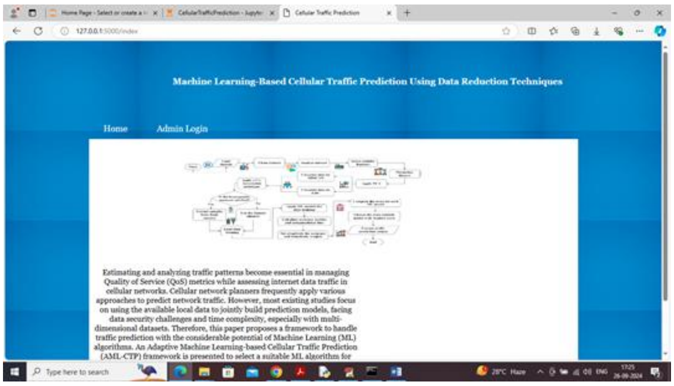


V. MODULES / METHODOLOGY

- **Data loading:** using this module we are going to import the dataset.
- **Data Processing:** The dataset is cleaned and normalized using the Min-Max Scaler algorithm, converting non-numeric values and handling missing data for a standardized dataset.
- **Apply K-Best Algorithm:** The Select-K-Best algorithm selects the top features from the processed dataset, ensuring the most relevant features for training.
- **PCA Dimension Reduction Algorithm:** PCA is applied to reduce the dataset's dimensionality, selecting uncorrelated features and minimizing redundancy.
- **Visualization:** A graph of PCA-reduced features is visualized, showing how similar data points are clustered together for analysis.

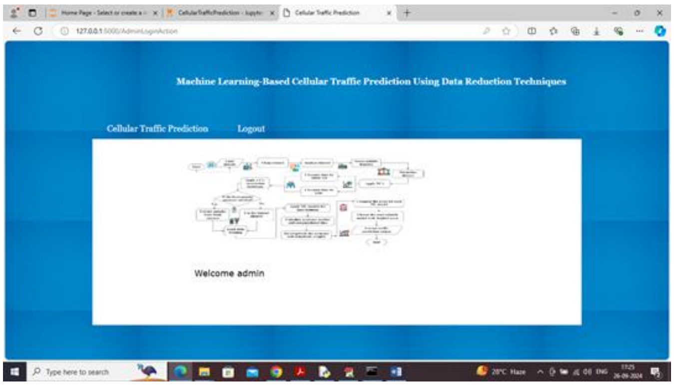
- **DBSCAN, Kernel Density Clustering:** Density-based clustering methods DBSCAN and Kernel Density are used to group similar data points and calculate similarity between clusters.
- **Split the Data into Train & Test:** The dataset is split into 80% training and 20% testing, preparing it for machine learning model training and evaluation.
- **Model generation:** Model building – SVM, Linear Regression, Decision Tree, Light Gradient Boosting, Extension XGBoost. Performance evaluation metrics for each algorithm is calculated.
- **Admin login:** In this module, admin can login into the application.
- **Cellular Traffic Prediction:** In this module user can upload the input data.
- **Logout:** User can logout after the completion of all activities.

Output

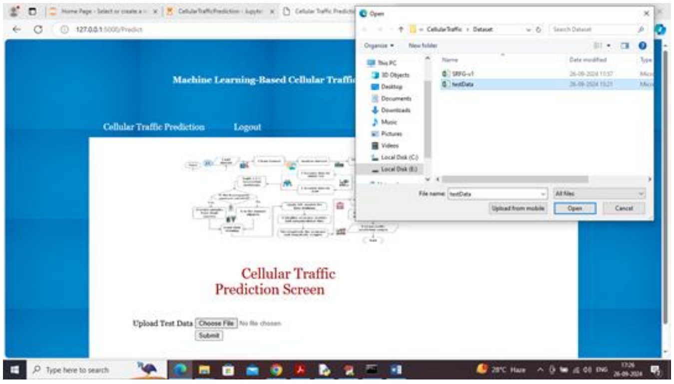


In above screen click on 'Admin Login' link to get below page

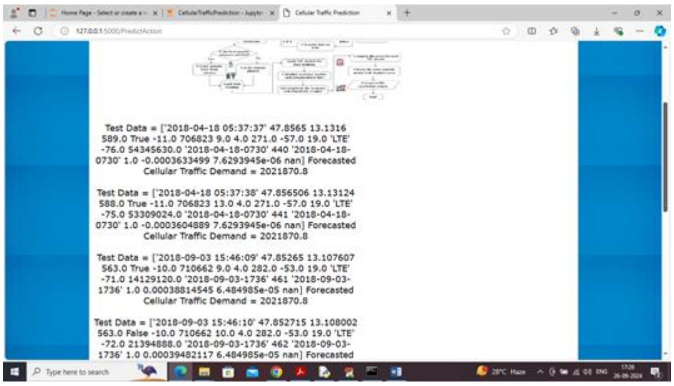
In above screen admin is login and after login will get below page



In above screen click on 'Cellular Traffic Prediction' link to get below screen



In above screen selecting and uploading test data file and then click on 'Open' button to get below page



In above screen can see test data values along with predicted cellular traffic rate

VI. CONCLUSION

In this project, we have presented an innovative approach to cellular traffic prediction using the Adaptive Machine Learning-based Cellular Traffic Prediction (AML-CTP) algorithm. By focusing on a smaller, high-quality dataset, we have significantly reduced the time complexity and resource requirements typically associated with large-scale traffic prediction. The methodology employed rigorous data preprocessing techniques, including normalization, feature selection, and dimensionality reduction, which ensured that the most relevant information was utilized for model training. After applying density-based clustering to identify high-similarity data points, we trained several machine learning algorithms to evaluate their performance in predicting cellular traffic. Among the various models tested, the Decision Tree algorithm emerged as the top performer, achieving an impressive R^2 score of 96%. Furthermore, the extension of this work with the XGBoost algorithm yielded even higher performance, with a remarkable R^2 score of 98%. These results underscore the effectiveness of our proposed system in enhancing prediction accuracy, ultimately contributing to better resource allocation and improved Quality of Service (QoS) in cellular networks.

VII. FUTURE SCOPE

In future work, we aim to further enhance the AML-CTP algorithm by exploring advanced techniques such as ensemble methods and deep learning architectures to improve prediction accuracy and robustness. Additionally, we will investigate the integration of hybrid models that combine traditional and modern machine learning approaches for better performance. Further research will also include experimenting with synthetic data generation techniques to augment the training dataset, ensuring a diverse range of traffic patterns and improving the model's generalization capabilities.

REFERENCES

1. Marios Polese, Michele Zorzi, “Machine Learning at the Edge: A Data-Driven Architecture for Cellular Network Traffic Prediction,” IEEE Communications Magazine, vol. 57, no. 4, pp. 102–108, 2019.
2. Sheng Zhang, Yong Li, and Depeng Jin, “Deep Learning Based Mobile Traffic Prediction Using Spatio-Temporal Features,” IEEE Transactions on Mobile Computing, vol. 18, no. 7, pp. 1681–1693, 2019.
3. Zhenhua Yu, Jie Xu, “A Machine Learning Approach for Cellular Traffic Prediction,” IEEE Access, vol. 8, pp. 88989–89002, 2020.
4. Ian Jolliffe, “Principal Component Analysis,” Springer Series in Statistics, Springer, 2016. (Used for data reduction techniques).
5. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2017.
6. Thomas N. Kipf and Max Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” International Conference on Learning Representations, 2017.
7. Yaguang Chen, Tianqi Chen, “Urban Mobile Traffic Prediction Using Machine Learning,” IEEE Network, vol. 32, no. 6, pp. 128–135, 2018.