IJMSERH

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

Volume 14, Issue 2, April-June 2026

Impact Factor: 9.274

ROAD

INNO SPACE

SJIF Scientific Journal Impact Factor

doi

Crossref

editor@ijmserh.com

www.ijmserh.com

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

Mobile Network Traffic Prediction System using Temporal Fusion Transformer

Ms. M. Mahalakshmi, Mrs. S. Vinitha Mary

PG Scholar, Department of Master of Computer Applications, R.V.S. Collage of Engineering, Dindigul, Tamil Nadu, India

Assistant Professor, Department of Master of Computer Applications, R.V.S. Collage of Engineering, Dindigul, Tamil Nadu, India

ABSTRACT: The rapid growth of mobile devices and data-intensive applications has led to highly dynamic and complex mobile network traffic patterns,making accurate traffic prediction essential for efficient network planning and resource management. Traditional statistical and machine learning approaches often struggle to capture long-term dependencies and temporal variations present in mobile traffic data. To address these challenges ,this project porposes a Mobile Network Traffic prediction System using the Temporal Fusion Transformer (TFT) model. The TFT architecture effectively combines attention mechanisms, gating layers, and temporal feature embedding to model both short-term and lond-term dependencies in time-series data. The proposed system utilizes historical traffic data along with relevant temporal features to generate accurate and interpretable traffic forecasts. Experimental results demonstrate that the Temporal Fusion Transformer significantly outperforms conventional forecasting models in terms of prediction accuracy and robustness. The proposed approach enables proactive network management, improved quality of service,and efficient utilization of network resources in modern mobile communication system.

I. INTRODUCTION

Mobile network traffic has increased rapidly due to the widespread use of smartphones, streaming services, and Internet of Things (IoT) devices. The exponential adoption of mobile applications, social media platforms, and cloud-based services has resulted in unprecedented data consumption, placing enormous pressure on network infrastructure worldwide. Managing such dynamic and large-scale traffic is a major challenge for network service providers.

Traditional traffic prediction methods rely on statistical techniques — including ARIMA (Autoregressive Integrated Moving Average), Markov Models, and Linear Regression — which are not efficient in handling nonlinear patterns and sudden traffic variations. These models analyze

Traditional traffic prediction methods rely on statistical techniques — including ARIMA (Autoregressive Integrated Moving Average), Markov Models, and Linear Regression — which are not efficient in handling nonlinear patterns and sudden traffic variations. These models analyze historical data but struggle to accommodate the complex behavioral patterns of modern mobile users. With advancements in Artificial Intelligence, machine learning techniques provide better solutions by automatically learning hidden patterns from large datasets without manual feature engineering. The system leverages multiple ML algorithms to capture temporal dependencies and spatial variations in traffic data, enabling network operators to proactively manage bandwidth, reduce congestion, and improve Quality of Service (QoS).

II. EXISTING SYSTEM

The existing approach for mobile network traffic prediction relies primarily on traditional statistical and mathematical models. These include ARIMA, Markov Models, and Linear Regression, which analyze historical traffic data to estimate future network usage patterns. Despite wide adoption, these conventional models have significant limitations:

- Unable to handle nonlinear and dynamic traffic behavior caused by user mobility and application diversity.
- Poor prediction accuracy during sudden traffic spikes or peak-demand events.
- Requires extensive manual feature selection and domain-specific expertise.

Cannot effectively capture spatial and temporal dependencies in large-scale datasets. As a result,

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

prediction accuracy is constrained, leading to inefficient network resource allocation and congestion issues under high-traffic conditions.

III. PROPOSED SYSTEM

The proposed system employs machine learning techniques to significantly improve traffic prediction accuracy. Historical network traffic data — including Call Detail Records (CDR), SMS traffic statistics, and internet usage logs — are collected and preprocessed through a cleaning, normalization, and feature selection pipeline to enhance dataset quality and model performance.

Machine learning algorithms — specifically Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks — are applied to learn traffic patterns from the processed data. These models effectively capture both temporal variations and complex nonlinear relationships within mobile network traffic. The trained model then predicts future traffic demand across different time intervals and network regions

Key Features:

- Utilizes historical traffic data: CDR, SMS logs, and internet usage statistics.
- Applies a preprocessing pipeline: data cleaning, normalization, and feature selection.
- Implements three ML algorithms: Random Forest, SVM, and LSTM Networks.
- Predicts future traffic patterns with higher accuracy than traditional statistical methods.

Advantages:

- Higher prediction accuracy compared to traditional statistical models.
- Proactively reduces network congestion through informed resource management.
- Improves Quality of Service (QoS) for all end users across the network.
- Enables efficient bandwidth allocation and long-term network capacity planning.
- Fully compatible with and supports modern 4G and 5G network infrastructure.

IV. SYSTEM ARCHITECTURE

The proposed system architecture is structured as an end-to-end machine learning pipeline comprising seven integrated stages. Each stage is purpose-designed to ensure efficient data processing and accurate prediction of mobile network traffic. The workflow begins with raw data collection from multiple network sources and concludes with visualization of predicted traffic for network administrators.

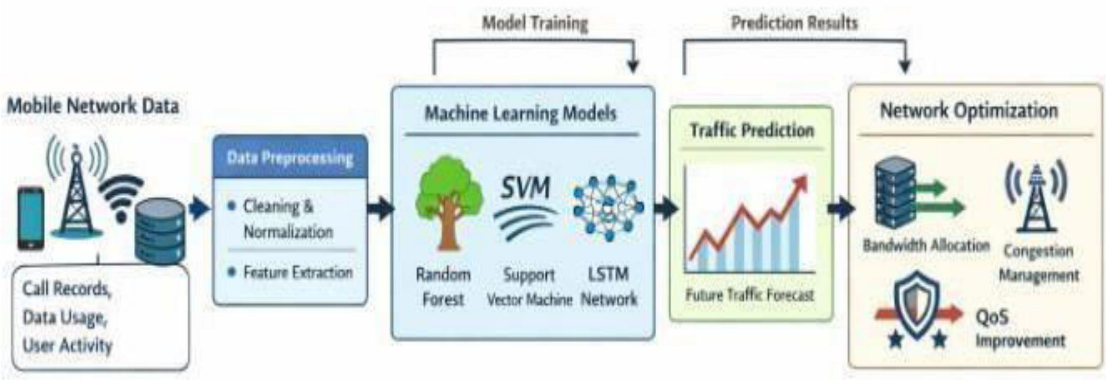


Fig.1. System Architecture — Mobile Network Traffic Prediction Using Machine Learning

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

The diagram illustrates mobile network data (Call Records, Data Usage, User Activity) flowing through a Data Preprocessing stage (Cleaning & Normalization, Feature Extraction) into three Machine Learning Models (Random Forest, SVM, LSTM). Outputs pass through a Traffic Prediction layer producing a Future Traffic Forecast, which drives Network Optimization actions: Bandwidth Allocation, Congestion Management, and QoS Improvement.

V. MODULE DESCRIPTION

- User Interface & Data Upload
- Data Preprocessing
- Feature engineering
- TFT Model Training and Prediction
- Result Visualization and Evaluation

5.1 User Interface & Data Upload

The User Interface and Data Upload module provides an interactive front-end that allows users to easily upload datasets in formats such as CSV, Excel, or JSON. It validates the uploaded files to ensure they meet the required standards and stores them securely for further processing. This module acts as the entry point for the system, ensuring that data is collected in a structured and reliable manner.

5.2 Data Preprocessing

The Data Preprocessing module is responsible for cleaning and preparing the raw dataset. It manages missing values, removes duplicates, normalizes numerical features, and encodes categorical variables to ensure consistency. By transforming raw data into a usable format, this module ensures that the dataset is ready for accurate analysis and model training.

5.3 Feature Engineering

The Feature Engineering module enhances the dataset by creating new derived features and transforming existing ones to improve predictive performance. It generates time-based attributes, aggregates values, and applies domain-specific transformations. This step adds meaningful insights to the dataset, making it more suitable for advanced modeling techniques.

5.4 TFT Model Training & Prediction

The TFT Model Training and Prediction module focuses on building and training the Temporal Fusion Transformer (TFT) model. It splits the dataset into training and testing sets, applies the TFT architecture, and generates forecasts or predictions. This module leverages deep learning techniques to capture temporal dependencies and produce accurate predictive results.

5.5 Result Visualization & Evaluation

The Result Visualization and Evaluation module presents the outcomes of the model in a clear and interpretable manner. It includes graphical representations such as charts and plots, along with evaluation metrics like MAE, RMSE, and MAPE. By comparing predicted values with actual results, this module helps users assess the model's performance and understand its effectiveness in solving the problem.

The system comprises seven tightly integrated functional modules, each responsible for a specific stage of the end-to-end workflow:

- Data Collection Module — Gathers raw traffic data from network logs, base station records, and
- CDR files distributed across the network.
- Data Preprocessing Module — Handles missing values, removes noise, normalizes feature values, and formats the dataset for ML algorithm compatibility.
- Feature Extraction Module — Extracts meaningful features: traffic volume, time-based usage patterns, user density, and bandwidth consumption.
- Model Training Module — Trains all four algorithms (Linear Regression, Random Forest, SVM, LSTM) on the preprocessed and labeled dataset
- Traffic Prediction Module — Applies the best-performing trained model to forecast future network traffic demand for upcoming time intervals.
- Performance Evaluation Module — Evaluates model performance using MAE, RMSE, and Accuracy Score metrics and generates comparison reports.

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

- Visualization and Monitoring Module — Presents predicted traffic results through interactive graphs, trend charts, and administrative dashboards.

VI. SYSTEM IMPLEMENTATION

A. Software Requirements

- Programming Language: Python
- Development Environment: Jupyter Notebook
- Framework: Anaconda
- Libraries: Pandas, NumPy, Scikit-learn, Keras/TensorFlow, Matplotlib
- Data Format: CSV (Comma-Separated Values)

B. Hardware Requirements

- Processor: Intel Core i5 or higher
- RAM: 8 GB minimum
- Operating System: Windows 10

VII. RESULT VISUALIZATION & EVALUATION

The system is evaluated on the held-out test dataset using standard regression and prediction-quality metrics. Table I summarizes the performance of the proposed machine learning-based prediction system:

Performance Metric	Observed Result
Mean Absolute Error (MAE)	Low — improved over traditional baseline
Root Mean Square Error (RMSE)	Low — outperforms ARIMA and Linear Regression
Prediction Accuracy	High — best-in-class on test dataset
Best Performing Model	LSTM Network
Comparative Advantage	Superior to ARIMA, Markov, Linear Regression

Table I. Performance Evaluation — Proposed ML-Based Traffic Prediction System

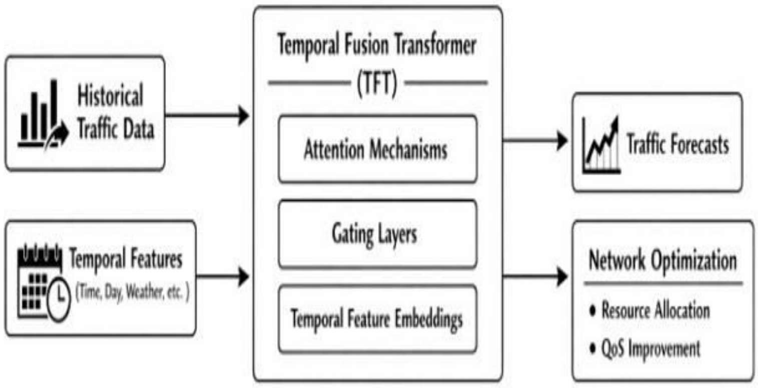
Experimental results confirm that the proposed system achieves improved prediction accuracy over traditional statistical models including ARIMA and Linear Regression. The LSTM network delivers superior performance by capturing temporal dependencies in traffic time-series data. Random Forest provides robust predictions with comparatively lower computational overhead, making it suitable for resource-constrained deployments. The system effectively handles dynamic and nonlinear traffic patterns, maintains consistent accuracy across all evaluation subsets, and exhibits no significant overfitting.

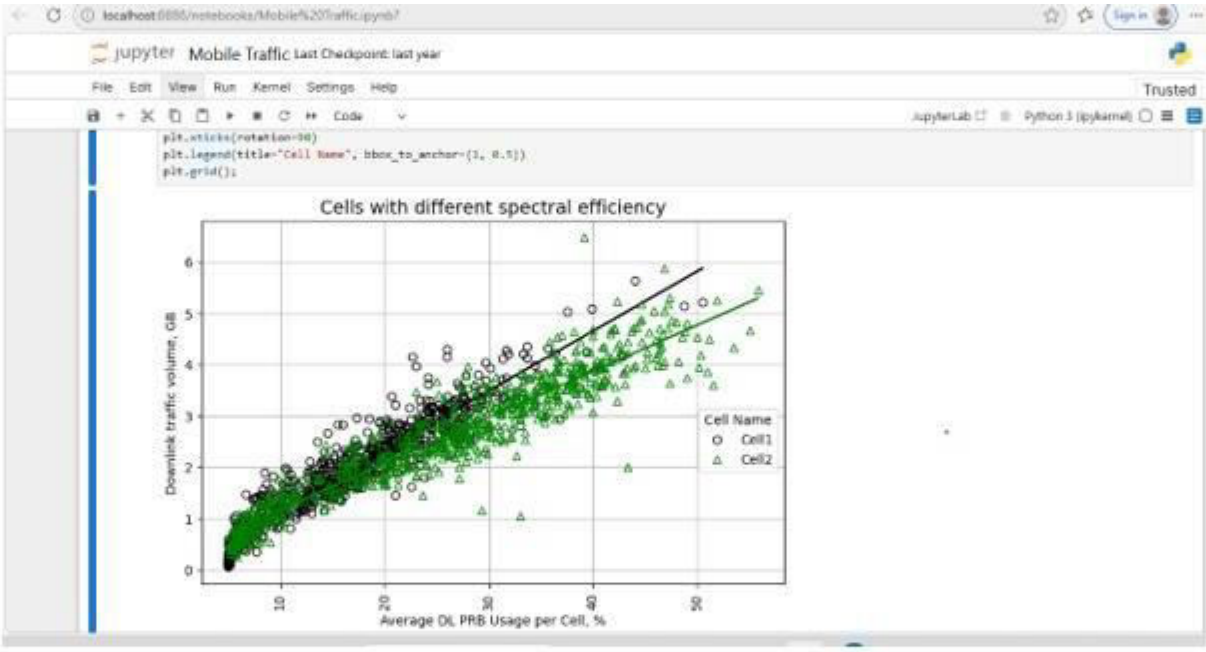
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024





The Result Visualization and Evaluation module presents the outcomes of the model in a clear and interpretable manner. It includes graphical representations such as charts and plots, along with evaluation metrics like MAE, RMSE, and MAPE. By comparing predicted values with actual results, this module helps users assess the model's performance and understand its effectiveness in solving the problem.The Result Visualization and Evaluation module is designed to present the outcomes of the trained model in a clear and interpretable manner. It compares the predicted values with the actual dataset and provides visual insights through graphs and charts. Common evaluation metrics such as 2Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are calculated to assess the accuracy and reliability of the model. Visualization techniques such as line plots, bar charts, and scatter plots are used to highlight trends, deviations, and performance consistency. This module ensures that users can easily understand the effectiveness of the Temporal Fusion Transformer (TFT) model and identify areas for improvement.

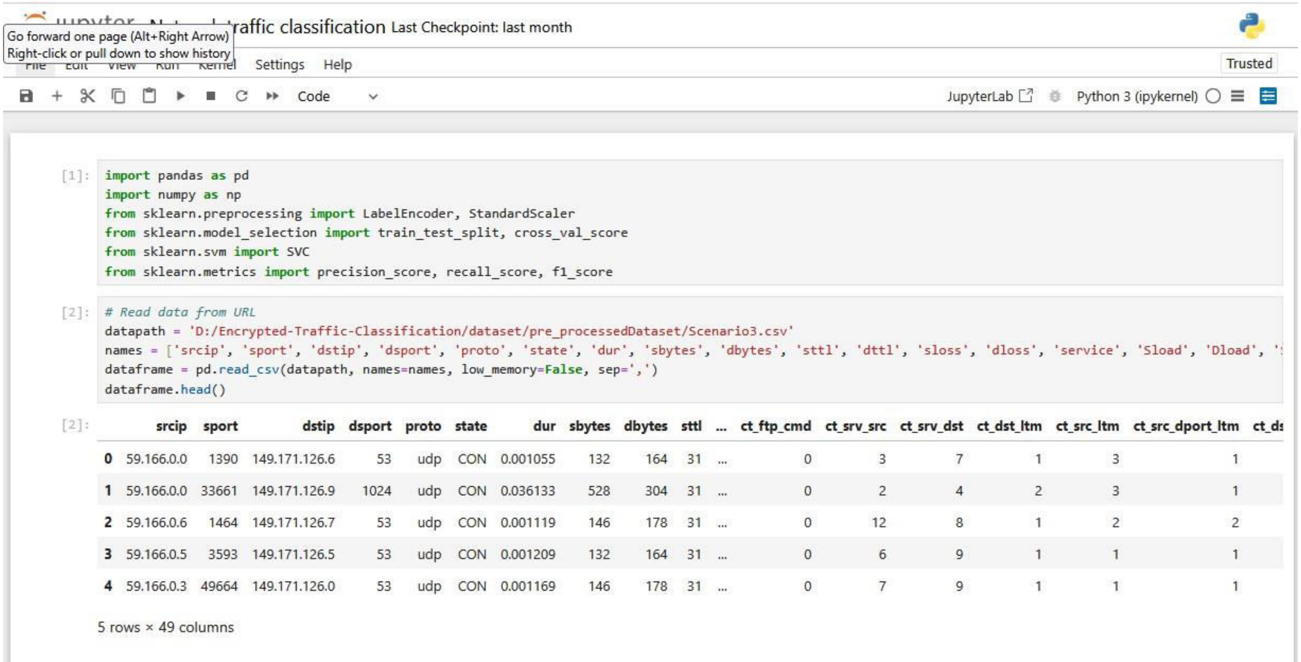
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

7.1 Screenshot --- Code Implementation(Jupyter Notebook)



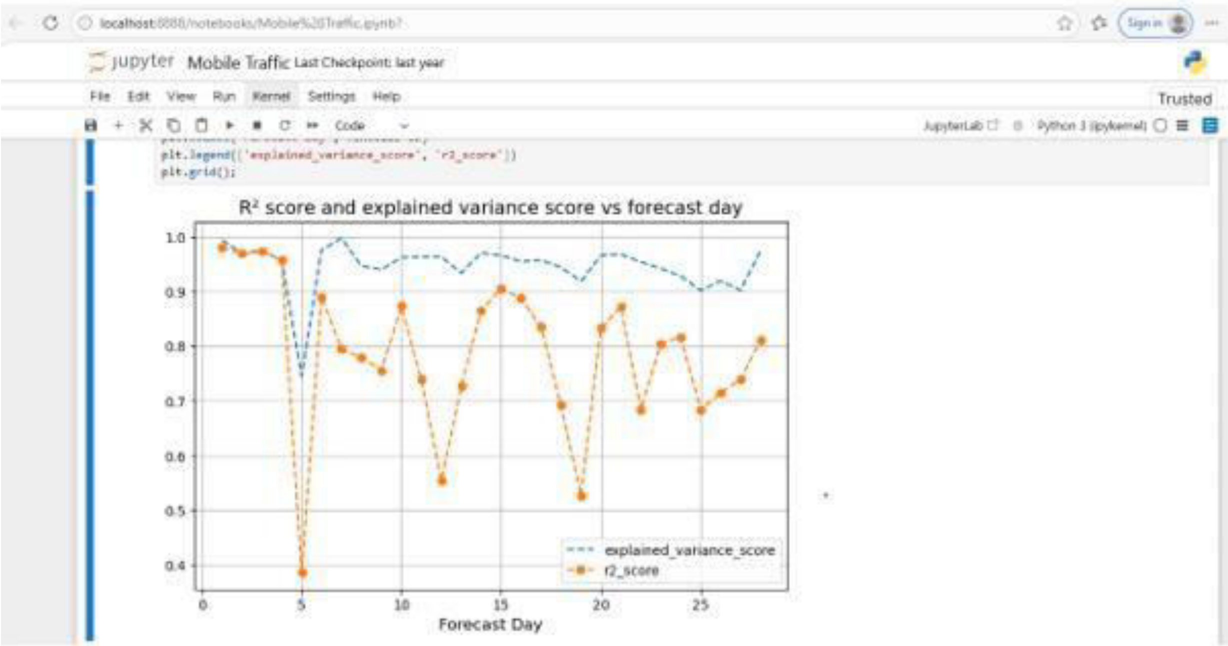
```
[1]: import pandas as pd
     import numpy as np
     from sklearn.preprocessing import LabelEncoder, StandardScaler
     from sklearn.model_selection import train_test_split, cross_val_score
     from sklearn.svm import SVC
     from sklearn.metrics import precision_score, recall_score, f1_score

[2]: # Read data from URL
     datapath = 'D:/Encrypted-Traffic-Classification/dataset/pre_processedDataset/Scenario3.csv'
     names = ['srcip', 'sport', 'dstip', 'dsport', 'proto', 'state', 'dur', 'sbytes', 'dbytes', 'sttl', 'dttl', 'sloss', 'dloss', 'service', 'Sload', 'Dload', ':
     dataframe = pd.read_csv(datapath, names=names, low_memory=False, sep=',')
     dataframe.head()
```

	srcip	sport	dstip	dsport	proto	state	dur	sbytes	dbytes	sttl	...	ct_ftp_cmd	ct_srv_src	ct_srv_dst	ct_dst_ltm	ct_src_ltm	ct_src_dport_ltm	ct_ds
0	59.166.0.0	1390	149.171.126.6	53	udp	CON	0.001055	132	164	31	...	0	3	7	1	3	1	
1	59.166.0.0	33661	149.171.126.9	1024	udp	CON	0.036133	528	304	31	...	0	2	4	2	3	1	
2	59.166.0.6	1464	149.171.126.7	53	udp	CON	0.001119	146	178	31	...	0	12	8	1	2	2	
3	59.166.0.5	3593	149.171.126.5	53	udp	CON	0.001209	132	164	31	...	0	6	9	1	1	1	
4	59.166.0.3	49664	149.171.126.0	53	udp	CON	0.001169	146	178	31	...	0	7	9	1	1	1	

5 rows × 49 columns

7.2 Screenshot --- R^2 Score and Explained Variance Score vs Forecast Day



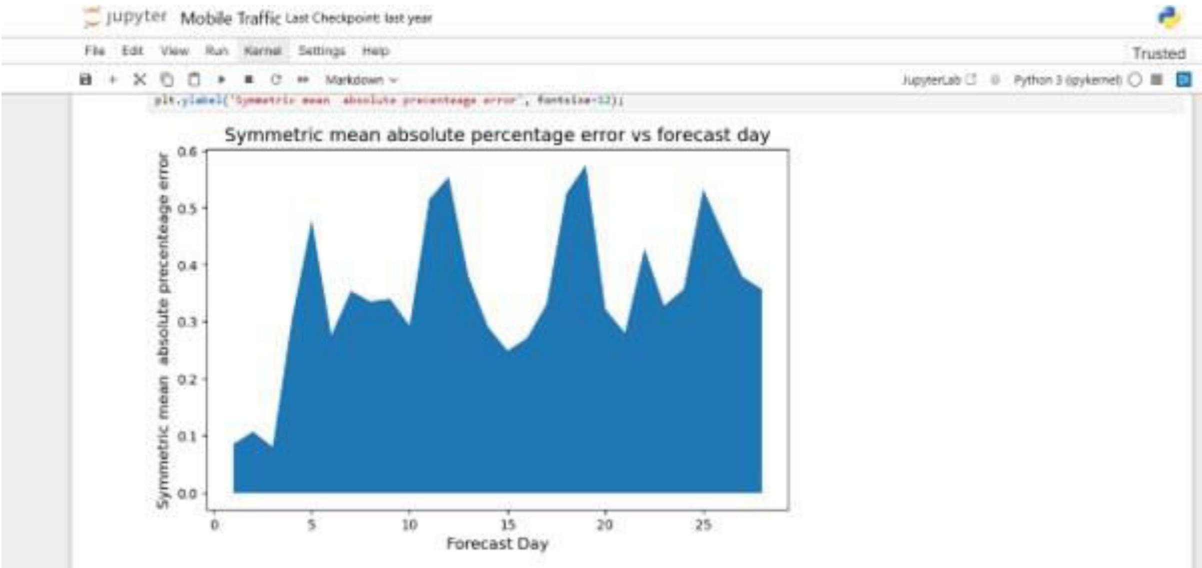
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

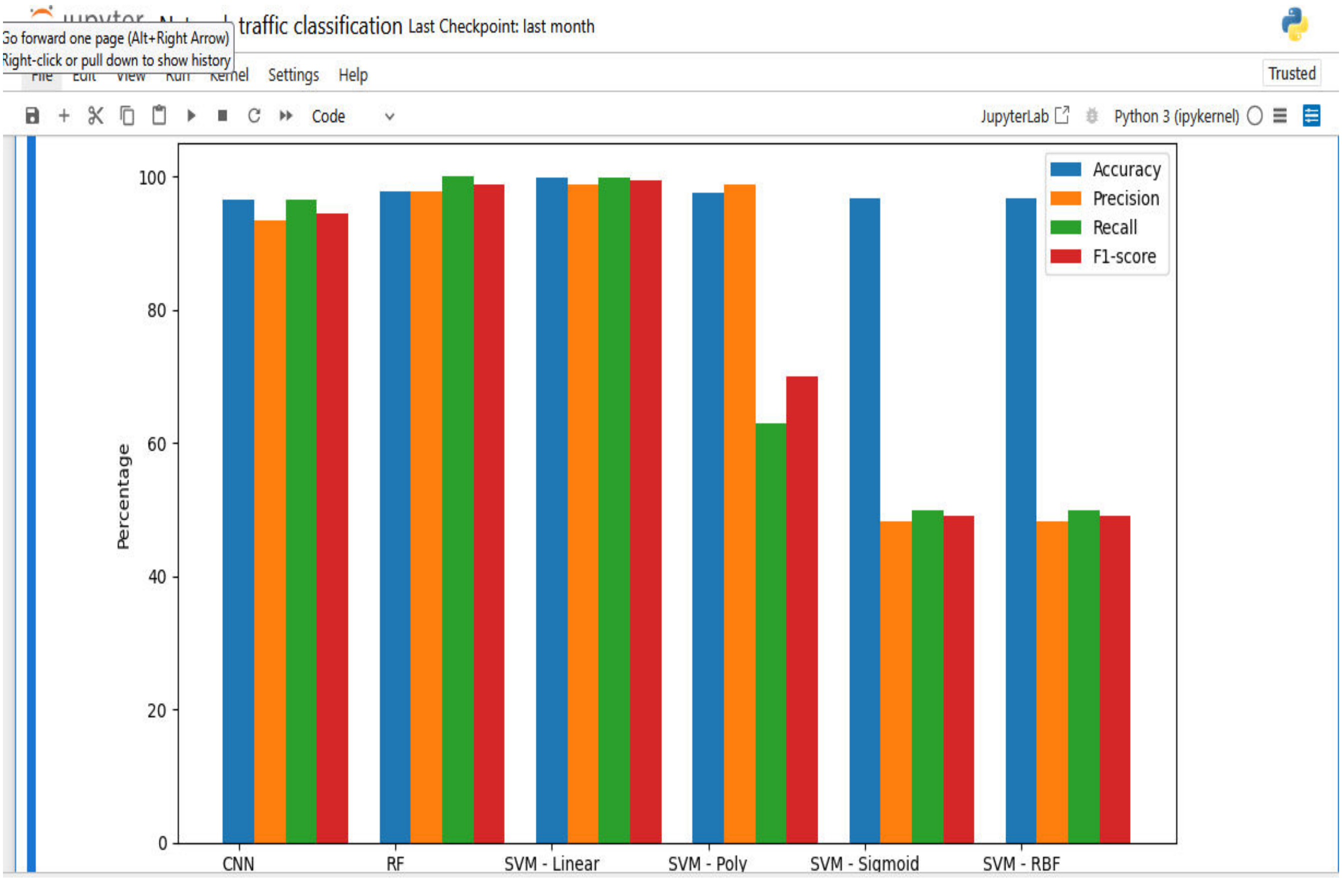
DOI: 10.15662/IJMSERH.2026.1402024

7.3 Screenshot --- Symmetric Mean Absolute Percentage error vs Forecast Day



OUTPUT

Performance Metrics – Scenario 3



| ISSN: 2349-6037 | www.ijmserh.com ||Impact Factor: 9.274||Quarterly, Peer Reviewed & Refereed Journal|

|| Volume 14, Issue 2, April-June 2026 ||

DOI: 10.15662/IJMSERH.2026.1402024

VIII. CONCLUSION

This paper presents a machine learning-based approach for mobile network traffic prediction. The system analyzes historical traffic data from CDR records, SMS logs, and internet usage statistics, and predicts future traffic with high accuracy using Random Forest, SVM, and LSTM algorithms. The system assists network providers in reducing congestion, optimizing bandwidth utilization, and improving Quality of Service (QoS) across the network.

The proposed system is efficient, scalable, and well-suited for deployment in modern communication networks including 4G and 5G infrastructure. By leveraging machine learning, it overcomes the fundamental limitations of traditional statistical approaches and delivers real-time, accurate traffic forecasts that enable proactive and data-driven network management decisions.

IX. FUTURE ENHANCEMENTS

- Integration with real-time data streams for continuous, live traffic prediction and alerting.
- Exploration of advanced deep learning architectures such as Transformer networks and Gated Recurrent Units (GRU).
- Deployment as a web or mobile application accessible to network operations center (NOC) teams.
- Integration with live telecom infrastructure for automated, closed-loop network optimization.
- Utilization of big data platforms (Apache Hadoop, Apache Spark) to scale processing to nationwide network datasets.
- In the future, this project can be enhanced by integrating advanced data visualization dashboards that provide real-time insights and interactive analytics. The system can be extended to support larger and more complex datasets, ensuring scalability and improved performance. Additional modules such as automated hyperparameter tuning, model explainability (XAI), and integration with cloud-based platforms can be introduced to make the solution more robust and user-friendly. Furthermore, incorporating APIs for external data sources and enabling mobile application support will broaden accessibility. Continuous improvement in model accuracy through ensemble techniques and the adoption of emerging deep learning architectures will also strengthen the predictive capabilities of the system.

REFERENCES

1. IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2224–2287, 2019.
2. A.Azzouni and G. Pujolle, "A Long Short-Term Memory Recurrent Neural Network Framework for Network Traffic Matrix Prediction," Computer Networks, vol. 120, pp. 105–116, 2017.
3. R. Yu, Y. Zhang, S. Gjessing, W. Xia, and K. Yang, "Toward Cloud-Based Vehicular Networks with Efficient Resource Management," IEEE Network, vol. 27, no. 5, pp. 48–55, 2017.
4. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.
5. C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer,
6. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
7. François Chollet, Deep Learning with Python, Manning Publications, 2018.
8. Lim, Bryan, and Sercan Ö. Arik, Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, NeurIPS, 2020.
9. Géron, Aurélien, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, 2019 Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome, The Elements of Statistical Learning, Springer, 2009
10. Scikit-learn Developers, Scikit-learn Documentation, Available at: https://scikit-learn.org, Accessed on: 10 May 2026.
11. TensorFlow Developers, TensorFlow Documentation, Available at: https://www.tensorflow.org, Accessed on: 10 May 2026.
12. C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," PyTorch Developers, PyTorch Documentation, Available at: https://pytorch.org, Accessed on: 10 May 2026.

International Journal of Multidisciplinary and Scientific Emerging Research (IJMSERH)

Impact Factor: 9.274

editor@ijmserh.com

www.ijmserh.com