The Eurasia Proceedings of Science, Technology,
Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2025

Volume 37, Pages 63-76

ICEAT 2025: International Conference on Engineering and Advanced Technology

SmartSARIMAX: An Advanced Model for Bandwidth Prediction in Data Networks

Ahmed M. Kareem
University of Diyala

Muntadher Khamees
University of Diyala

Alaa Taima Albu-Slaih
University of Al-Qadisiyah

Abstract: In the data network field, particularly in the domain of fast evolving data networks, it is necessary to have a good bandwidth estimating for resource planning and user guarantee because the system of weft or dynamics between data flows is increasingly and progressively complex in the structures topology. In this paper, we present a novel forecasting method known as SmartSARIMAX (S_SARIMAX) based on the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) approach to estimate bandwidth consumption. S_SARIMAX incorporates additional variables, including user behaviour patterns and historical bandwidth trends, to accurately simulate complex seasonal and network traffic trends. Our model is rigorously tested with real-world datasets, dramatically improving prediction accuracy over standard methods. The results show that the S_SARIMAX model provides reliable predictions to support strategies to stimulate network management processes with an MAE and RMSE as forecasting metrics and the proposed model outperforms the comparable model by more than 90%. This study presents essential contributions to bandwidth prediction, offering a strong asset for network operators to predict the demand, plan capacity and develop the users' Quality of Experience (QoE).

Keywords: Bandwidth prediction, Time series analysis, ARIMA, SARIMAX, Network performance.

Introduction

In physical layer communications, the term "bandwidth" is suited to the spectral width of the electromagnetic signals and the propagation behaviour associated with communication systems. In data networks, bandwidth refers to the maximum data transfer rate a network link or path can support. The data network bandwidth estimation model will be discussed in this article. The other major topic of central relevance to this standard is bandwidth; this measures the amount of data that can be carried from one point to another in a given timeframe, especially in the context of packet networks (Prasad et al., 2003).

Recent advancements in mobile networks and streaming technologies enable users to access live content through mobile devices (Bentaleb et al., 2020). About 4.9 billion internet users worldwide consume billions of hours of online video daily (Loh et al., 2022). Consequently, streaming has become the primary type of traffic in communication networks. Measurements are necessary for diagnosing network errors, optimizing network performance in best-effort networks, and adaptive mechanisms in applications like streaming video (Johnsson et al., 2023). The available bandwidth (avail-bw) is an essential metric in many scenarios , including capacity provisioning, routing, traffic engineering, quality of service management, streaming applications, server selection, etc. (Dovrolis & Jain, 2002). A link's available bandwidth (ABW) is the unused part of its capacity (Strauss et al., 2003).

This involves measuring and analyzing all variables that impact bandwidth; they are then statistically summarized to obtain the ultimate prediction. The success of predictive models is highly dependent on the selection of temporal and user behaviour-related features. In this context That is the Baseline Model (Mathormad, n.d.), a plain ARIMA (Autoregressive Integrated Moving Average) model with no independent variables that shows how fundamentals might affect bandwidth. While this provides insight to basic tendencies, it neglects user interactions and other environmental variables that influence the bandwidth usage pattern.

The SmartSARIMAX model based on the Seasonal Autoregressive Integrated Moving Average with Exogenous eXogenous variables (SARIMAX) method is proposed to improve the accuracy of predictions. Considering external factors like user activity, the SmartSARIMAX outperformed the Baseline Model in terms of prediction and capturing the seasonality. S_SARIMAX is a robust solution providing more advanced modeling features relying on the SARIMAX method that can support the addition of exogenous variables to handle seasonality and externalities influencing bandwidth usage. Second, although the MAE and RMSE are simple for explanation, the S_SARIMAX can be more user-friendly model as an interpreterble model to provide meaningful and reliable forecasts since it guides interpretable prediction results that matter to decision makers. In addition, giving high-quality visualizations makes results more interpretable and helps convey results across different types of audiences properly.

The paper is organized into five main sections: Section 1 describes related works; Section 2 frames the dataset and summarizes feature selection techniques and data preprocessing methods; Section 3 discusses the architecture of S_SARIMAX; Section 4 discusses the model evaluation based on scale-dependent metrics; and Section 5 provides a summary of this paper.

Related Works

In the past couple of decades, the field of bandwidth prediction has received considerable attention due to the rising demand for effective means of network management and improving user experiences. Much work has been done examining different methodologies and frameworks to support the complexity of bandwidth estimation, especially with regard to multimedia services and real applications. We provide an overview of the following significant research contributions that employ a range of methodologies, such as traditional statistical models, machine learning, and hybrid approaches.

There is an urgent need to predict bandwidth while adapting to changing network conditions. The data consists of live video conferences recorded to provide significant samples(Gottipati et al., 2024). The proposed model (Ivy) adopts the appropriate algorithm selection in line with network changes based on offline meta-learning, achieving an 11.4% improvement in (QoE) compared to similar works based on meta-heuristics. İn similary, (Khairy et al., 2024) improving the quality of experience (QoE) adds importance to building a bandwidth estimation approach for real-time communication (RTC). Voice/video calls represent data used across Microsoft Teams. The model was created to bridge the gap between simulation and actual user experience, depending on offline reinforcement learning, leveraging the data's realism and emphasising user-relevant metrics.

Kougioumtzidis et al. (2022) aims to improve the prediction and management of Quality of Experience (QoE) for multimedia services through predictive models of perceptual experiences, focusing on video streaming and gaming. QoE has been model led and predicted by machine learning approaches using common factors influencing QoE. In comparison Labonne et al. (2020), suggests a solution for predicting the bandwidth usage of the network links using a machine learning technique and examining both ARIMA, Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) models for building these techniques. The LSTM model achieves a prediction error of less than 3%, significantly outperforming the other models and ensuring the ability to detect and prevent congestion.

Using long short-term memory (LSTM) networks for a realtime mobile bandwidth prediction, this study is a continuation of work similar to Mei et al. (2020) by incorporating multi-scale entropy analysis and model switching. The base RLS model is compared with the LSTM models trained offline to capture temporal patterns and predict future bandwidth, and the LSTM model performs better accuracy in various mobility scenarios The basic RLS model is compared to an off-line trained LSTM models, which has temporal pattern memory capabilities and has learned to forecast future bandwidth. In (Bentaleb et al., 2020) instead, it aimed at increasing bandwidth measurement and prediction accuracy in low-latency streamingsystems. The bandwidth prediction techniques in the ACTE architecture are implemented as an estimate of the current available bandwidth through Sliding Window Moving Average and Recursive Least Squares (RLS), which improves the adaptation decisions made in real-time video sessions to improve overall QoE.

Motivated by Al-Issa et al. (2019), where the paper presents BCIDASH+, a smarts streaming framework using time-series prediction to improve HTTP adaptive streaming over wireless networks. Implementing two forecasting approaches, the model offers optimal bitrate levels from clients according to network conditions, improving video delivery performance that can address quality fluctuations and reduce re-buffering events using an extensive evaluation in a wireless testbed environment. Similarly (Vasilev et al., 2018), this study utilizes machine learning algorithms, namely the Bayesian Networks and the NeuraI Networks, that estimate the QoE factors from network quality of service metrics. This study proposes a combination of hidden variable extraction and information about the context to achieve more accurate predictions concerning the re-buffering ratios and variations in video quality.

Like ways Yue et al. (2017), The LinkForecast framework uses machine learning, specifically random forest algorithms, and combines upper- and lower-layer information to predict cellular link bandwidth in LTE networks. It has enhanced accuracy in predicting diagrams with prediction errors for average prediction diagrams ranging from 3.9% to 17.0% in different scenarios. This work provides a machine learning-based method to predict QoE in SDN, based on techniques such as Decision Trees and Neural Networks (Abar et al., 2017). It is an attempt to provide realtime behind-the-screen user experience analysis by comparing the performance of varying computer through benchmarking their reading depth parameter computes including (but not limited at) Pearson's Correlation Coefficient and Root-mean-square deviation.

In MLASH, a machine learning optimized approach to enhance video rate adaptation when using HTTP streaming is introduced (Chien et al.,2015). MLASH: MLASSH enhances user's video quality and brings about significant network-resource-saving through the admission of utilized information on video quality, and dynamic encoding rate scaling based on impact analysis in response to network state change employing a world-wide representative test data set. This document Charonyktakis et al. (2015) describes a primarily modular mechanism for predicting QoE in VoIP services through different ML-based experiments. The MLQoE mechanism aims to achieve telecommunication service improvement concerning predictors depicting the effect of network performance on user experience and by selecting the best-performing algorithms for prediction.

Similarly (Aroussi & Mellouk, 2014), the study analyzes the correlation between QoE and quality of service. To this end, different machine learning algorithms are used for modelling. This adds new perspectives to correlation and models to improve prediction algorithms used to understand the user experience on wireless devices based on network performance data. By the same token (Alreshoodi & Woods, 2013), this paper surveys techniques significantly improving the accuracy of prediction of the quality of service and QoE relationships in multimedia services.

The study also analyses several classification algorithms (including Decision Trees and Support Vector Machine) and proposes how different performance metrics contribute to better integrating QoS into the evaluations of QoE. The actual paper (Mushtaq et al., 2012) looks into the relationship between quality of service and QoE in streamed video, with a stress on the ways that network parameters drive user satisfaction. The paper presents a comparative analysis of various machine learning algorithms for QoE prediction and shows that the usage of Decision Trees and Random Forest classifiers is efficient for this task. System for Requisite Bandwidth Estimation Using SVM We proposed system of estimating the Requisite Bandwidth using Machine learning algorithms is to be given in Chen et al. (2007). It contrasts two ways of probing models and introduces a normalization technique to improve estimation, even in the absence of similar samples seen during training. We demonstrate that this proposed method can be used to accurately estimate bandwidth compared to pathChirp and Spruce tools using NS-3 emulation data. Table 1: Summary of the studies on these issues is shown.

Methodology

In this section, we explore our systematic approach to developing and validating S_SARIMAX for bandwidth forecasts. This includes detailed descriptions of the dataset and feature selection methods, as well as how to preprocess the data before using. Congratulations! Good point that the data is only one piece of a complex puzzle, and this method focused on maturity in both terms of accumulated external variables and simply better models.

Table 1. Compressive of related work

Reference	Methodology	Key Findings	Strengths	Limitations
Gottipati et al., 2024	Employing offlline meta-learning for bandwidth prediction	Enhancing QoE by more than 11.4%.	Depending on dynamic network for selection bandwidth prediction algorithms.	Ambiguity in clarifying the relationship between QoS and QoE.
Khairy et al., 2024	Offline RL prediction model.	Using data's realism for estimation available bandwidth	bridge the gap between simulation and actual user experience	Focusing on the user side may not reflect a comprehensive view of the network
(Kougioumtzidis et al., 2022)	Predictive modeling using ML	Developed models to quantify QoE in multimedia	Comprehensive analysis of user experience	Limited to specific multimedia services
(Labonne et al., 2020)	Machine learning (ARIMA, MLP, LSTM)	LSTM model significantly outperformed others	High prediction accuracy	Focused on specific network links
(Mei et al., 2020)	Long Short-Term Memory (LSTM)	Realtime bandwidth prediction	Captured temporal patterns effectively	Requires extensive training data
(Bentaleb et al., 2020)	Sliding Window Moving Average	Improved adaptation decisions in streaming	Effective in low-latency environments	May not generalize to other applications
(Al-Issa et al., 2019)	Time series forecasting	Optimized bitrate levels for adaptive streaming	Effective in addressing quality fluctuations	Evaluation limited to wireless testbed
(Vasilev et al., 2018)	Bayesian Networks and Neural Networks	Predict QoE factors from quality of service metrics	Integration of hidden variables	May complicate model interpretation
(Yue et al., 2017)	Random forest algorithms	Improved cellular link bandwidth prediction	Significant accuracy improvements	Limited to LTE networks
(Abar et al., 2017)	Decision Trees and Neural Networks	Realtime QoE prediction in SDN	Evaluates performance through multiple metrics	Lacks empirical validation
(Chien et al., 2015)	Machine learning framework	Enhanced video rate adaptation	Utilizes diverse datasets	Complexity in implementation
(Charonyktakis et al., 2015)	Various machine learning algorithms	QoE prediction in VoIP services	Modular approach	Specific to VoIP
(Aroussi & Mellouk, 2014)	Various machine learning algorithms	Analyzed QoE and quality of service relationship	Comprehensive survey	Lacks new experimental data
(Alreshoodi & Woods, 2013)	Review methodologies	Improved prediction accuracy for QoE/quality of service	Highlights effective performance metrics	Limited to comparative analysis
(Mushtaq et al., 2012)	Machine learning approaches	Investigated QoE/quality of service correlation	The empirical study provides insights	Focused on specific multimedia contexts
(Chen et al., 2007)	Support Vector Machines (SVM)	Enhanced bandwidth estimation	Outperformed existing tools	Relies on simulation data

Dataset and Features Selection

Data collected through predictive models for bandwidth estimation requires effective feature selection to improve model accuracy/performance. The datasets size for training and testing were (6234565, 6) and (371619, 5) respectively. Fig. 1 Time series of bandwidth _total from January 2017 to March 2019. More clarification on how to handle data, the input data into the S_SARIMAX model includes six columns: BANDWIDTH_TOTAL, MAX_USER, HOUR_ID, SERVER_NAME, UPDATE_TIME, and ZONE_CODE (which would be removed during preprocessing). After the preprocessing, this set is systematically reduced to 5 main variables in the feature selection, which has a high impact on the bandwidth-consuming behaviour, such as BANDWIDTH_TOTAL, MAX_USER, HOUR_ID, SERVER_NAME, and UPDATE_TIME, see Table 2. These elements can be temporal trends and user behaviours, which influence bandwidth usage so that the model can use those for better accuracy in prediction analysis. On the other hand, the Baseline Model is filled with the same 6 of the original dataset, but in this case, all of them remain after preprocessing, see Table 3.

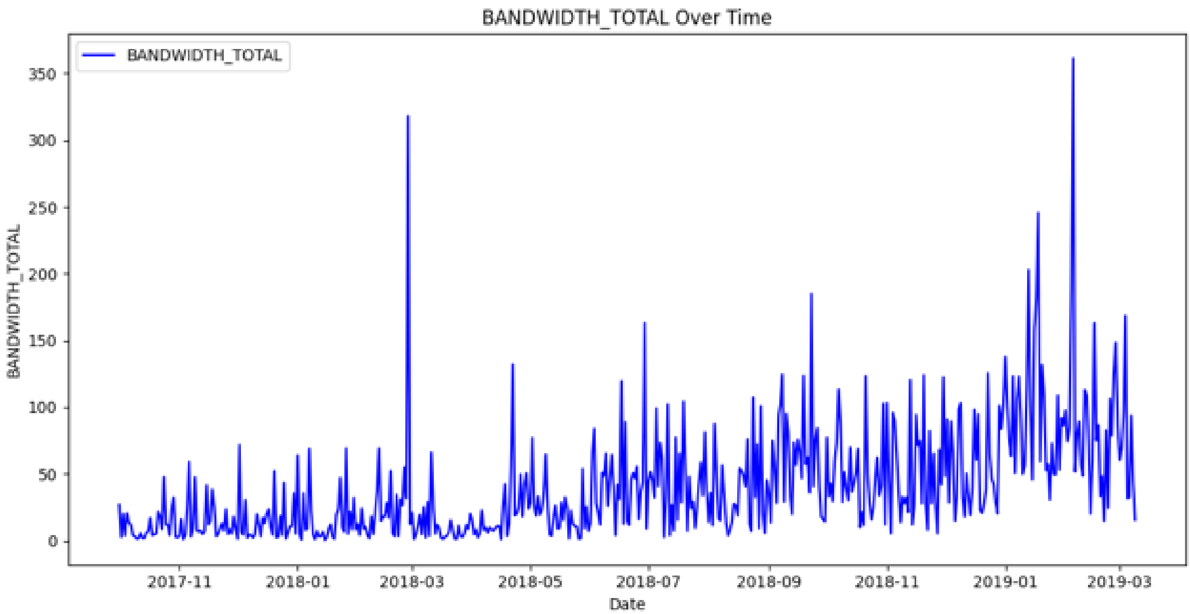

Figure1. Time series analysis of BANDWIDTH_TOTAL

Table 2. Features used in the S_SARIMAX model

Feature Description	Feature Description
BANDWIDTH_TOTAL	Total bandwidth usage over time
MAX_USER	Maximum number of users during the Period
HOUR_ID	Hour of the day
SERVER_NAME	Name of the server
UPDATE_TIME	Timestamp of the data entry

Table 3. Features used in the baseline model

Feature Description	Feature Description
BANDWIDTH_TOTAL	Total bandwidth usage over time
MAX_USER	Maximum number of users during the Period
HOUR_ID	Hour of the day
SERVER_NAME	Name of the server
UPDATE_TIME	Timestamp of the data entry
ZONE_CODE	The geographic zone of the server

Preprocessing

S_SARIMAX: Ensuring data consistency and integrity. It starts by loading the dataset and trimming whitespace from the column names, which is convenient and avoids potential issues further in the analysis. We will first convert the UPDATE_TIME column to datetime type and then set this as the index in order to analyze time series data. Notably, S_SARIMAX offers a systematic approach to dealing with duplicate records through the averaging of numeric values and keeping the first entry of non-numeric fields, which helps keep the data compact and informative.

In Fig. 2 and Fig. 3, the graphs display continuous blue lines showing fluctuations in bandwidth consumption over that Period of time and show total bandwidth usage from Jan 2017 to March 2019. They depict data processed by handling Duplicates in two ways (Handled Duplication by Keeping the First or Last Entry and Handled Duplication by Aggregating Duplicate Entries), where this Second Method was implemented. Resampling to a daily frequency is the final step in preprocessing, as it conditions the data further and fills in gaps where necessary.

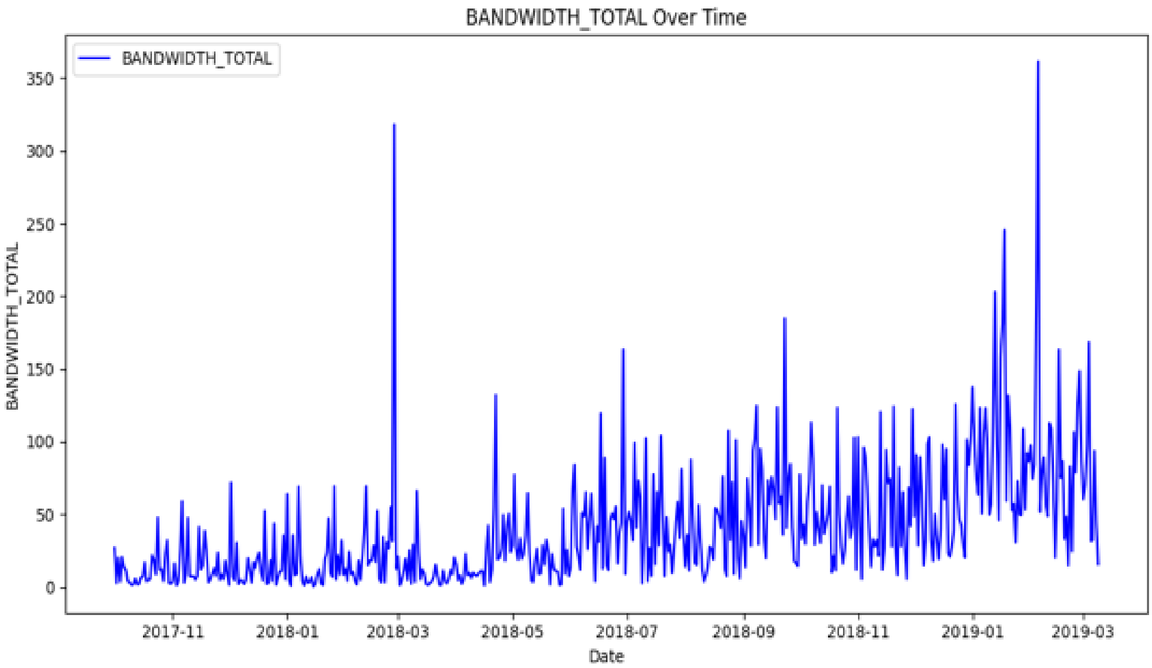

Figure 2. Bandwidth total over time (After handling duplication)

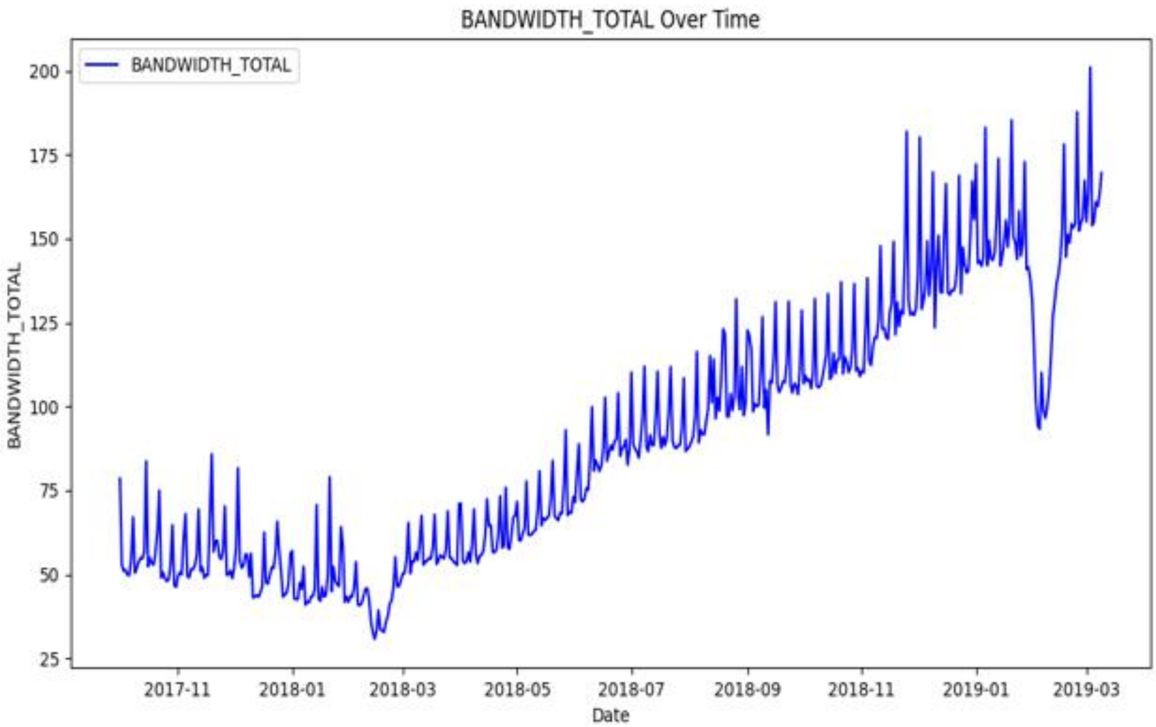

Figure 3. Bandwidth total over time (After aggregating duplicate entries)

Architecture S_SARIMAX

The S_SARIMAX and Baseline Model use different models, methods, and complexities to predict bandwidth usage. SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) is a basic model for S_SARIMAX. The SARIMAX extends the ARIMA framework by adding seasonal components to the built-in model, and it also allows for exogenous variables that affect the model predictions)(Amerise & Tarsitano, 2017; Özmen, 2021). Using historical data, it looks at the past data for target variable and its seasonal behavior with external factors such as holidays or events hence capturing complex relation amongst the data. SARIMAX greatly enhances estimation model prediction with temporal and external factors.

S_SARIMAX Overview: The SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) model that estimates BANDWIDTH_TOTAL and MAX_USER in an exogenous variable. It allows the model to take into account external parameters that have effect on bandwidth distribution, which will help in predicting by capturing seasonal and temporal trends. Both ARIMA and SARIMAX models were mixed. And in such model, I let ARIMA predict the MAX_USER values and SARIMAX predict BANDWIDTH_TOTAL according with external variables including MAX_USER, days of week, month and previous value from TOTAL_BANDWIDTH. The above-mentioned external parameters are factors that affect the object of interest and bandwidth prediction. Step 3: Model Parameter Optimization The model parameters are tuned via an outo_ARIMA function that was created, which selects the best parameters for the model using AIC, and BIC criteria by picking the best fitting model based on these with minima. Fig. 4 displays the steps of data preprocessing, model fitting, and evaluation. Every step aims to guarantee that the model is both strong and precise.

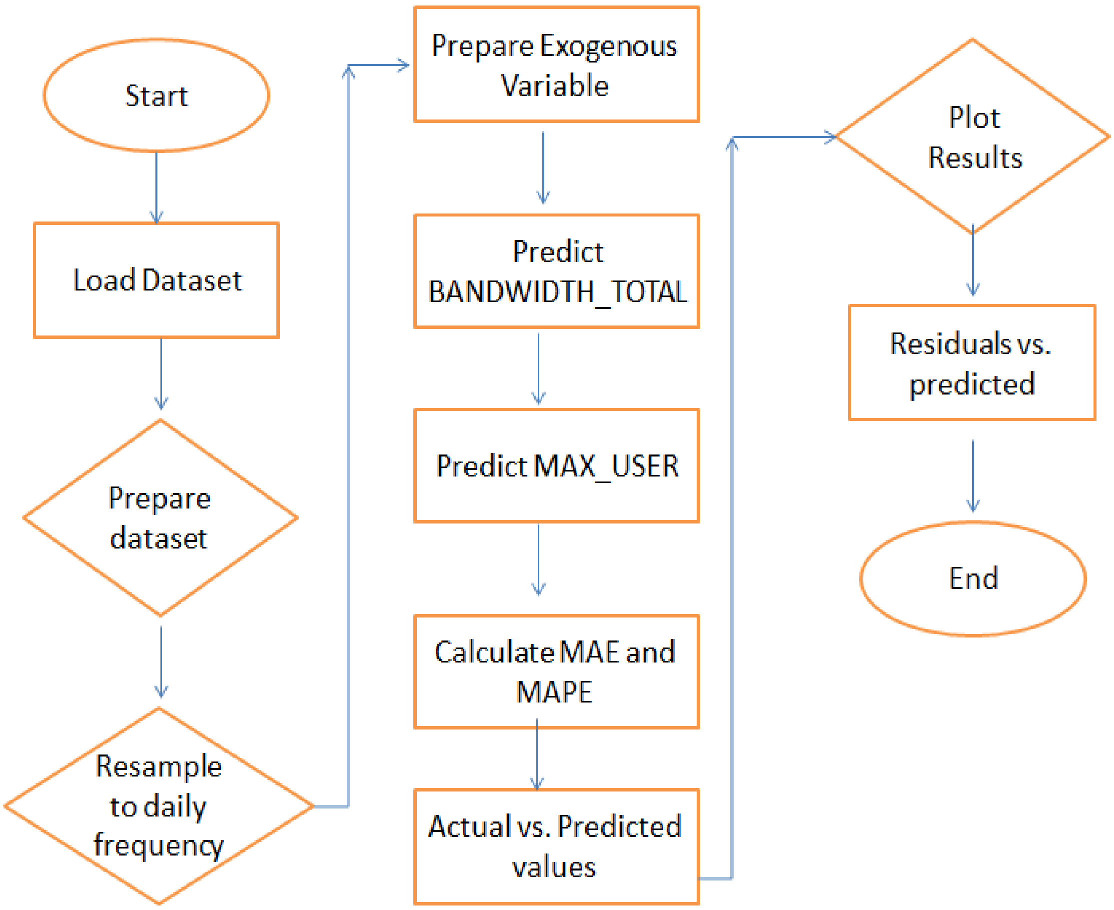

Figure 4. Flowchart of proposal model

The integrated model has an additional step: the time series, including the MAX_USER for each source, is statistically analyzed, where an ARIMA (Autoregressive Integrated Moving Average) model is fitted for MAX_USER alone based on historical data with no exogenous variable considered, where Fig. 5 shows the peak number of users over time from November 2017 to March 2019. The vertical axis shows the account on the proper unit, and the x-axis shows the date. Fig. 6 explains how the Dual-Model approach is a full framework to investigate, both quantitatively and qualit-vly, the relationship between user activity and bandwidth appropriations).

The possibility with the SARIMAX model to take into account external factors that may influence bandwidth usage. It does this by accounting for the MAX_USER / BANDWIDTH_TOTAL relationship. To conclude, the differences in forecasting performance we have observed can be mostly explained by the early promotion of a more sophisticated model framework that better depicts the characteristics of the data as shown in Fig. 7.

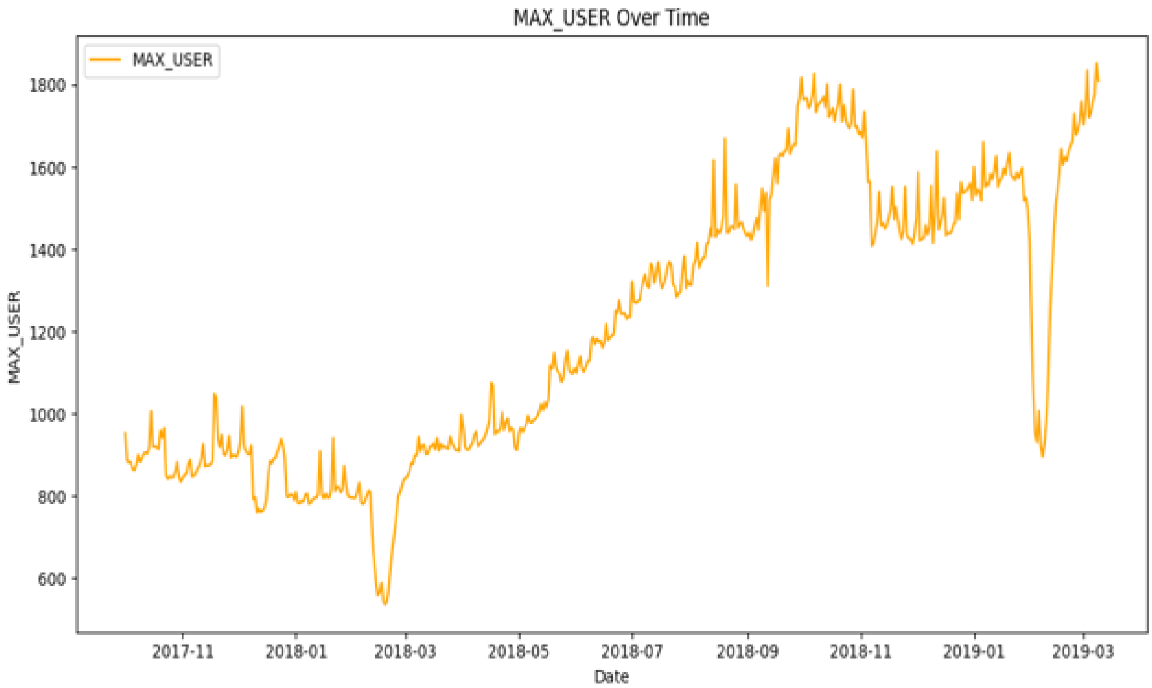

Figure 5. Maximum users over time

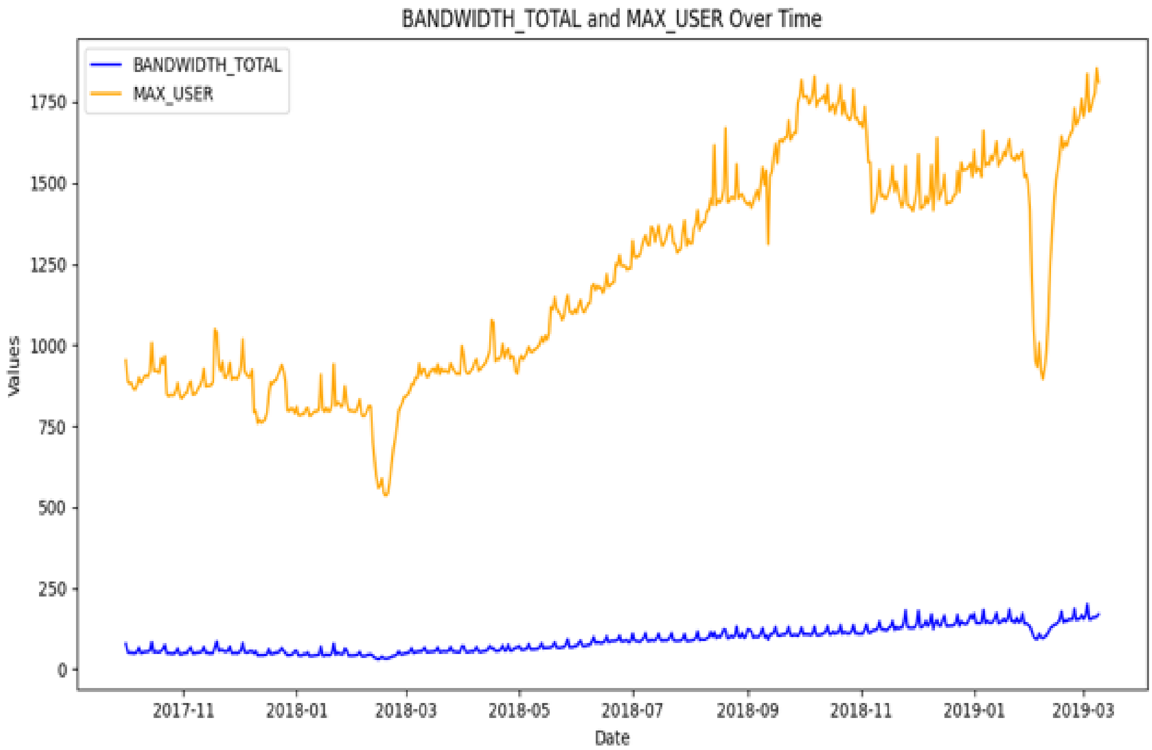

Figure 6. Bandwidth total and maximum users over time

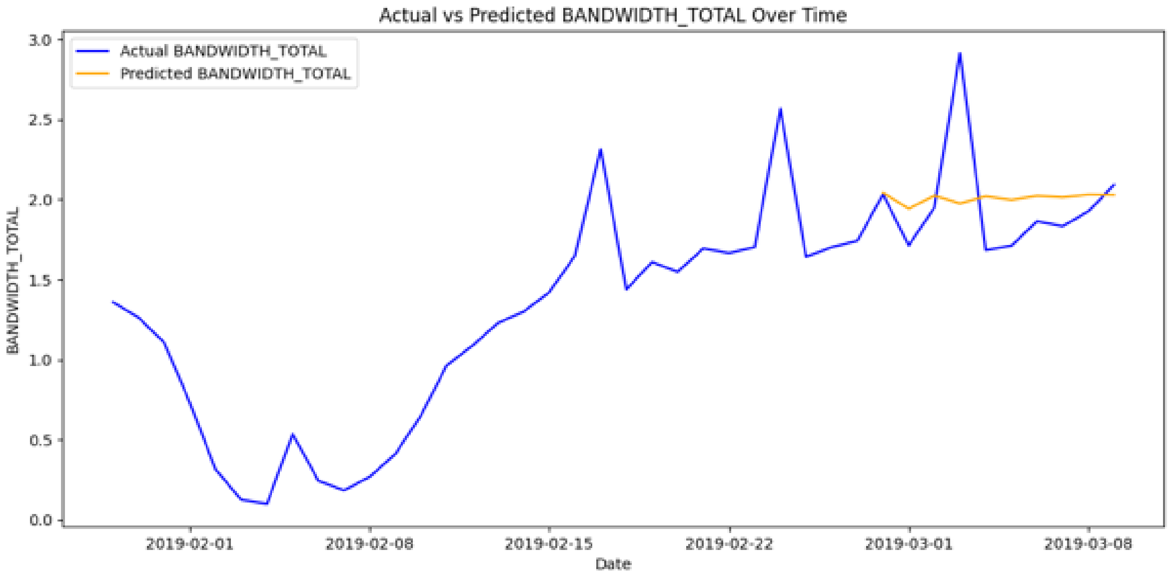

Figure 7. Actual vs predicted BANDWIDTH_TOTAL over time

Evaluation

Performance evaluation is a cross-cutting research issue. Performance metrics "error measures" are key parts of evaluation frame works across professional domains; the performance metric can be seen as a logical and mathematical theory, through which we measure how close actual outcomes are to what has been expected or predicted (Botchkarev, 2019). In general performance measures are based on scientific concepts of distance and similarity. For the machine learning regression tasks, performance metrics are calculated on comparing the trained predictions with actual (observed) data from the testing dataset. The outcomes of these comparisons can have direct impact to the decision-making on adoption machine learning algorithms for deployment. Among the most popular metrics, we can note Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Performance Metrics

Metrics derived from absolute or squared errors are called scale-dependent metrics, which share the same scale as the original data (Hyndman, 2006). They present errors in identical units (Sanders, 1997).

a) MAE is a metric that quantifies the average size of absolute errors between expected and actual values. The MAE is frequently called the mean absolute deviation (MAD) (Rakićević & Vujošević, 2015). The range of Mean Absolute Error (MAE) is (0, +∞); a lower MAE number indicates more predictive model accuracy. The benefit of MAE lies in its unit being identical to that of the original data, facilitating straightforward calculation and comprehension. The Mean Absolute Error (MAE) is frequently employed as a symmetric loss function (Flores, 1986).

The calculation process of the metric can be described as (Kim & Kim, 2020)

$$\text{MAE} = \frac{1}{n}\sum_{1}^{n}|D_{pre} - D_{act}| \tag{1}$$

b) The RMSE estimates the magnitude of average error between anticipated and actual values (Jierula et al., 2021). That is why, RMSE is nothing but the average vertical distance between the real points and the trend-line. It's just the square root of the Mean Squared Error. The RMSE information range from (0, +) a possible lower premium implies that the model prediction is more accurate. The units of RMSE correspond to the original units, enhancing its interpretability.

The metric's calculating procedure is delineated as (Kim & Kim, 2020)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(\ D_{per} - D_{act}\ )^2} \tag{2}$$

Both MAE and RMSE quantify the average prediction error size in the original data units. Compared to MAE, RMSE assigns a greater weight to significant errors due to squaring mistakes before averaging.

c) The normalized MAE is delineated as(Gustafson Jr & Yu, 2012):

$$\text{NMAE} = \frac{\frac{1}{n}\sum_{1}^{n}|D_{pre} - D_{act}|}{[mean\ Dact]} \tag{3}$$

Where the MAE is normalising by dividing the MAE by the mean of actual values.

The Results

The assessment for SmartSARIMAX is extensive; the MAE and RMSE are applied to measure the precision of predictions regarding BANDWIDTH_TOTAL and MAX_USER. That is, 0.24 MAE is were presented for BANDWIDTH_TOTAL, and 0.18 MAE were presented for MAX_USER, which resulted in a total MAE of 0.21. It fetches RMSE for BANDWIDTH_TOTAL is 0.35, and RMSE for MAX_USER is about 0.20; thus, total RMSE = 0.28. Additionally, the NMAE metric was used to present the results in relation to BANDWIDTH_TOTAL is 0.12, MAX_USER is 0.11 and the total NMAE is 0.11. With these metrics, we gain high-quality insight into the model and can have a detailed review of prediction errors, fig. 8. Further, it provides advanced debugging displays and visualizations (e.g. , histogram plots of prediction errors) to enable further analysis of performance and error distribution of the model outputs, as shown in Fig 9. The distribution and density curve assesses the reliability of the model prediction and shows possible enhancement regions.

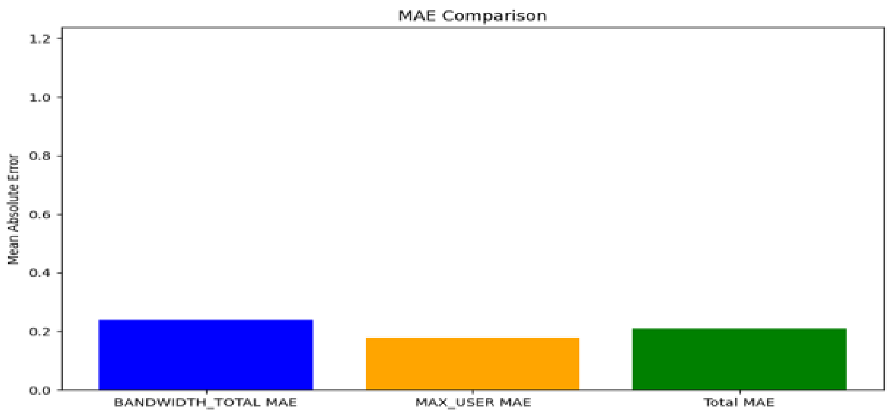

Figure 8. MAE Comparison

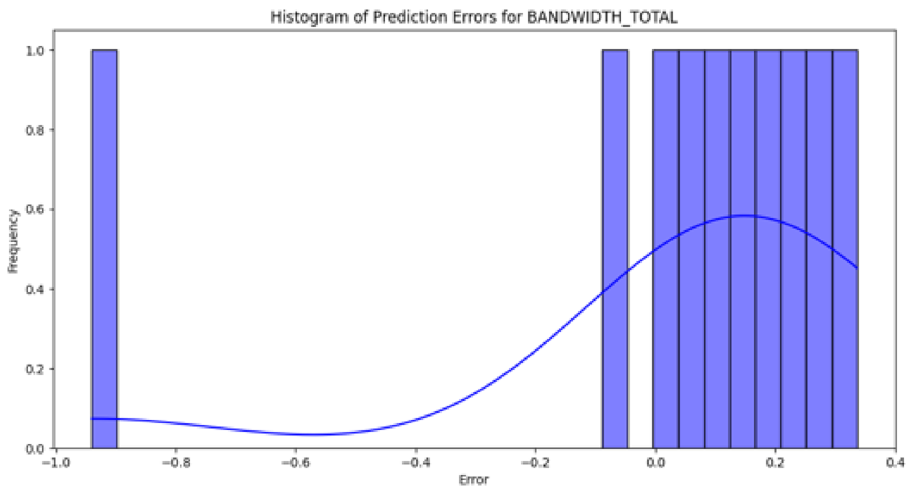

Figure 9. Prediction errors for BANDWIDTH_TOTAL

The evaluation methods utilized in SmartSARIMAX clearly show a discrepancy in the approaches to evaluating the model performance. The model shows a scatter plot for actual vs. predicted values for BANDWIDTH_TOTAL, which is a good predictor. Looking at the residuals plotted against the expected values can give us insight into how well our model is fitting.

Fig. 10 is immediately apparent from the model's predictive performance. Additionally, Fig. 11 provides information about the residuals plotted against the expected values; all the information is relative to the model's fit and performance Additionally, future work includes leveraging data and the possibility of modifying algorithms and preprocessing processes and addressing the limitation if the data is missing the stationary characteristic or may struggle to adapt the model to complex seasonal patterns.

On the other hand, the Baseline model may not provide this level of detail in evaluation. Although it computes some error metrics, it does not convey the level of visual checks and detailed outputs that make SmartSARIMAX attractive. The differing evaluation rigour may make it harder to pinpoint exactly where the model has weaknesses in its predictions . This also speaks to the reliability of the forecasting results, as the data in Table 4 illustrate the differences between the two models' Returns. To sum up, the very completeness of SmartSARIMAX's evaluation phase is an index of its effort to grant the correctness and interpretability of models— which deserve to be considered mandatory for effective decisions based on their predictions.

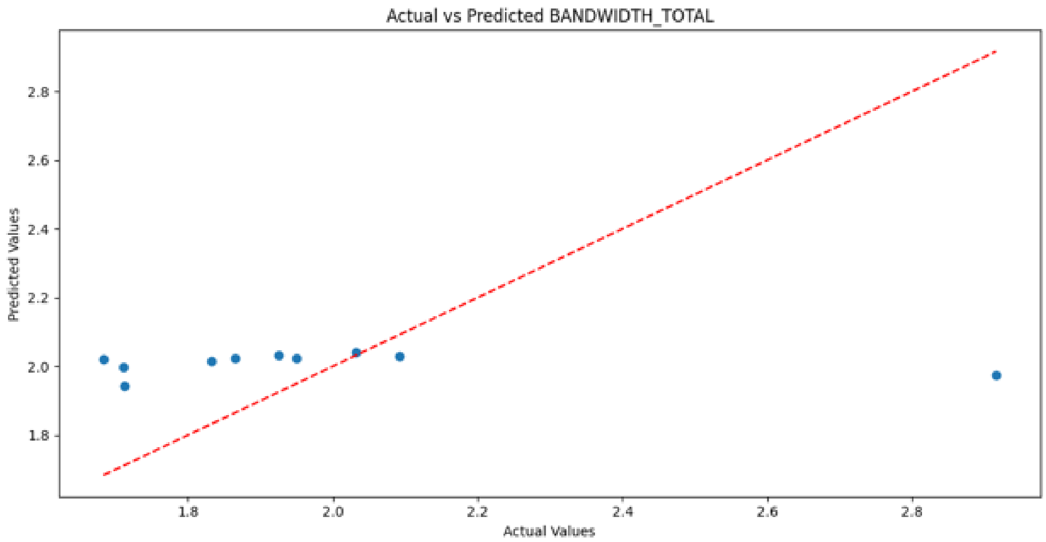

Figure 10. Actual vs predicted BANDWIDTH_TOTAL

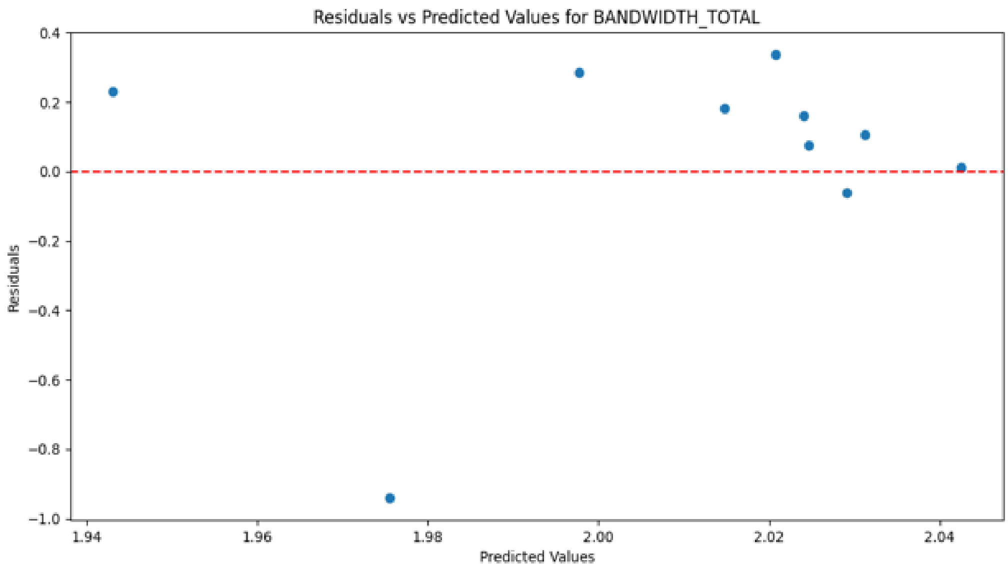

Figure 11. Residuals vs predicted values for BANDWIDTH_TOTAL

Table 4. Comparison with MAE & RMSE

Metric	SmartSARIMAX	Baseline model
MAE (BANDWIDTH_TOTAL)	0.24	5.12
MAE (MAX_USER)	0.18	6.34
Total MAE	0.21	32.92
RMSE (BANDWIDTH_TOTAL)	0.35	6.66
RMSE (MAX_USER)	0.20	8.24
Total RMSE	0.28	7.45
NMAE (BANDWIDTH_TOTAL)	0.12	0.9
NMAE (MAX_USER)	0.11	0.21
TOTAL NMAE	0.11	0.71

Conclusion

Overall, The preprocessing, modelling, evaluation, contributions analysis and compare the two optimized model with a Baseline Model is extensive. S_SARIMAX does a great job of cleaning the input data by removing extraneous whitespaces, transposing improper date-times into proper format and dropping duplicates. The daily resampling component maximises the usability and reliability of subsequent analysis. It is worth mentioning that the employed predictive models are also quite different; S_SARIMAX adopts a SARIMAX model with exogenous variables (introduced to enhance prediction through capturing bandwidth usage seasonality pattern). The Baseline model, on the other hand, opt for a simple ARIMA model that while can be good-enough-but-not-right, may overlook critical predictors and therefore predictions are not fully accurate. The second noticeable difference is the evaluation metrics being implemented, S_SARIMAX uses more comprehensive feedback methods like MAE and RMSE to provide a better understanding of the model performance. The elaborate validation and observation of trends in predictions help to better understand prediction error as well as model performance across all samples. Therefore, the proposed model offers rich potential of network traffic management and efficiency vacuum solutions indicating that rigorous studies on general protocols for band-homogeneous process dynamics are needed.

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* The authors received no financial support for the research, authorship, and publication of this article.

Acknowledgments

* This article was presented as an oral presentation at the International Conference on Engineering and Advanced Technology (ICEAT) held in Selangor, Malaysia, on July 23-24, 2025.

* The authors would like to thank the International Conference on Engineering and Advanced Technology (ICEAT) committee.

References

Abar, T., Letaifa, A. Ben, & El Asmi, S. (2017). Machine learning based QoE prediction in SDN networks. *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 1395–1400.

Al-Issa, A. E., Bentaleb, A., Barakabitze, A. A., Zinner, T., & Ghita, B. (2019, October). Bandwidth prediction schemes for defining bitrate levels in SDN-enabled adaptive streaming. In *2019 15th International Conference on Network and Service Management (CNSM)* (pp. 1-7). IEEE.

Alreshoodi, M., & Woods, J. (2013). Survey on QoE\QoS correlation models for multimedia services. *arXiv preprint arXiv:1306.0221*.

Aroussi, S., & Mellouk, A. (2014). Survey on machine learning-based QoE-QoS correlation models. *2014 International Conference on Computing, Management and Telecommunications (ComManTel)*, 200–204.

Bentaleb, A., Timmerer, C., Begen, A. C., & Zimmermann, R. (2020). Performance analysis of ACTE: A bandwidth prediction method for low-latency chunked streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(2s), 1–24.

Botchkarev, A. (2019). A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 45–76.

Charonyktakis, P., Plakia, M., Tsamardinos, I., & Papadopouli, M. (2015). On user-centric modular QoE prediction for VoIP based on machine-learning algorithms. *IEEE Transactions on Mobile Computing*, *15*(6), 1443–1456.

Chen, L. J., Chou, C.-F., & Wang, B.-C. (2007). A machine learning-based approach for estimating available bandwidth. *TENCON 2007-2007 IEEE Region 10 Conference*, 1–4.

Chien, Y.-L., Lin, K. C.-J., & Chen, M.-S. (2015). Machine learning based rate adaptation with elastic feature selection for HTTP-based streaming. *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Flores, B. E. (1986). A pragmatic view of accuracy measurement in forecasting. *Omega*, *14*(2), 93–98.

Gottipati, A., Khairy, S., Hosseinkashi, Y., Mittag, G., Gopal, V., Yan, F. Y., & Cutler, R. (2024). Balancing generalization and specialization: Offline Metalearning for Bandwidth Estimation. *arXiv preprint arXiv:2409.19867*.

Gottipati, A., Khairy, S., Mittag, G., Gopal, V., & Cutler, R. (2025, June). Offline to online learning for real-time bandwidth estimation. In *ICC 2025-IEEE International Conference on Communications* (pp. 1213-1218). IEEE.

Gustafson Jr, W. I., & Yu, S. (2012). Generalized approach for using unbiased symmetric metrics with negative values: Normalized mean bias factor and normalized mean absolute error factor. *Atmospheric Science Letters*, *13*(4), 262–267.

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *4*(4), 43–46.

Jain, M., & Dovrolis, C. (2002). End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. *ACM SIGCOMM Computer Communication Review*, *32*(4), 295–308.

Jierula, A., Wang, S., Oh, T.-M., & Wang, P. (2021). Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data. *Applied Sciences*, *11*(5), 2314.

Johnsson, A., Melander, B., & Björkman, M. (2023). Diettopp: A first implementation and evaluation of a simplified bandwidth measurement method. *arXiv preprint arXiv:2301.06405*.

Khairy, S., Mittag, G., Gopal, V., Yan, F. Y., Niu, Z., Ameri, E., ... & Cutler, R. (2024, April). ACM MMSys 2024 bandwidth estimation in real time communications challenge. In *Proceedings of the 15th ACM Multimedia Systems Conference* (pp. 339-345).

Kim, C. H., & Kim, Y. C. (2020). Application of artificial neural network over nickel-based catalyst for combined steam-carbon dioxide of methane reforming (CSDRM). *Journal of Nanoscience and Nanotechnology*, *20*(9), 5716–5719.

Kougioumtzidis, G., Poulkov, V., Zaharis, Z. D., & Lazaridis, P. I. (2022). A survey on multimedia services QoE assessment and machine learning-based prediction. *IEEE Access*, *10*, 19507–19538.

Labonne, M., Chatzinakis, C., & Olivereau, A. (2020). Predicting bandwidth utilization on network links using machine learning. *2020 European Conference on Networks and Communications (EuCNC)*, 242–247.

Loh, F., Wamser, F., Poignée, F., Geißler, S., & Hoßfeld, T. (2022). Youtube dataset on mobile streaming for internet traffic modeling and streaming analysis. *Scientific Data*, *9*(1), 293.

Mathormad. (2019). *AIVIVN 4: Bandwidth prediction*. Retrieved from https://www.kaggle.com/code/mathormad/aivivn-4-bandwidth-prediction

Mei, L., Hu, R., Cao, H., Liu, Y., Han, Z., Li, F., & Li, J. (2020). Realtime mobile bandwidth prediction using LSTM neural network and Bayesian fusion. *Computer Networks*, *182*, 107515.

Mushtaq, M. S., Augustin, B., & Mellouk, A. (2012). Empirical study based on machine learning approach to assess the QoS/QoE correlation. *2012 17th European Conference on Networks and Optical Communications*, 1–7.

Özmen, E. S. (2021). Time Series Performance and Limitations with SARIMAX: An Application with Retail Store Data. *Electronic Turkish Studies*, *16*(5).

Prasad, R., Dovrolis, C., Murray, M., & Claffy, K. (2003). Bandwidth estimation: metrics, measurement techniques, and tools. *IEEE Network*, *17*(6), 27–35.

Rakićević, Z., & Vujošević, M. (2015). Focus forecasting in supply chain: the case study of fast moving consumer goods company in Serbia. *Serbian Journal of Management*, *10*(1), 3–17.

Sanders, N. R. (1997). Measuring forecast accuracy: some practical suggestions. *Production and Inventory Management Journal*, *38*(1), 43.

Strauss, J., Katabi, D., & Kaashoek, F. (2003). A measurement study of available bandwidth estimation tools. *Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement*, 39–44.

Tan, Q., Lv, G., Fang, X., Zhang, J., Yang, Z., Jiang, Y., & Wu, Q. (2024). Accurate bandwidth prediction for real-time media streaming with offline reinforcement learning. *Proceedings of the 15th ACM Multimedia Systems Conference*, 381–387.

Tarsitano, A., & Amerise, I. L. (2017). Short-term load forecasting using a two-stage sarimax model. *Energy*, *133*, 108–114.

Vasilev, V., Leguay, J., Paris, S., Maggi, L., & Debbah, M. (2018). Predicting QoE factors with machine learning. *2018 IEEE International Conference on Communications (ICC)*, 1–6.

Yue, C., Jin, R., Suh, K., Qin, Y., Wang, B., & Wei, W. (2017). LinkForecast: Cellular link bandwidth prediction in LTE networks. *IEEE Transactions on Mobile Computing*, *17*(7), 1582–1594.

Author(s) Information

Ahmed M. Kareem
University of Diyala
Diyala, Iraq
Contact e-mail:*scicomphd222304@uodiyala.edu.iq*

Muntadher Khamees
University of Diyala
Diyala, Iraq

Alaa Taima Albu-Slaih
University of Al-Qadisiyah
Al-Qadisiyah, Iraq

To cite this article:

Kareem, A. M., Khamees, M., & Albu-Slaih, A. T. (2025). SmartSARIMAX: An advanced model for bandwidth prediction in data networks. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 37,* 63-76.