2024 Innovations in Intelligent Systems and Applications Conference (ASYU) | 979-8-3503-7943-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ASYU62119.2024.10756993

Artificial Intelligence Driven Multivariate Time Series Analysis of Network Traffic Prediction

Gözde Filiz
Computer Sci.& Eng. PhD. Program
MEF University
Istanbul, Turkey
filizg@mef.edu.tr

Ayşenur Yıldız
Research and Development
Andromeda TV Digital Plat. Op. Inc.
Istanbul, Turkey
aysenur.yildiz@dsmart.com.tr

Erkan Kara
Research and Development
Andromeda TV Digital Plat. Op. Inc.
Istanbul, Turkey
erkan.kara@demirorenmedya.com

Suat Altıntaş
IT Sytems and Network
Andromeda TV Digital Plat. Op. Inc.
Istanbul, Turkey
suat.altintas@dsmart.com.tr

Tuna Çakar
Computer Engineering
MEF University
Istanbul, Turkey
cakart@mef.edu.tr

***Abstract*— The primary objective of this research is to employ artificial intelligence, machine learning, and neural networks in order to construct a network traffic prediction model. The analysis of network traffic data obtained from a digital media and entertainment provider operating in Turkey is conducted through the application of multivariate time-series analysis techniques in order to get insights into the temporal patterns and trends. In model development, Vector Autoregression (VAR), Vector Error Correction Model (VECM), Long-Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) algorithms have been utilized. LSTM and GRU models have performed better with low Mean Absolute Percentage Error (MAPE) and high R-squared Score (R^2). LSTM model has reached 0.98 R^2 and 8.95% MAPE. These results indicate that the models can be utilized in network management optimization as resource allocation, congestion detection, anomaly detection, and quality of service.**

Keywords—network traffic forecasting, machine learning, long-short term memory, neural networks, artificial intelligence

I. INTRODUCTION

The accurate prediction of network traffic holds significant importance in the management of network performance and allocation of resources, particularly in light of ongoing advancements in digital technology. Increasing network traffic and complexity necessitates innovative methods for network management and optimization. Artificial intelligence techniques, especially machine learning algorithms and neural network can be used to analyze network data and estimate the upcoming traffic. The objective of this study is to construct a multivariate artificial intelligence model for the purpose of forecasting network traffic. This model will inform network management optimization strategies, ultimately resulting in an enhancement of service quality. In this study, we have used decomposition, auto-correlation and partial auto-correlation multivariate time series analysis techniques to understand network traffic data from a digital content provider. We have implemented Vector Autoregression, Vector Error Correction Model, Long-Short Term Memory, and Gated Recurrent Unit algorithms to develop a network traffic prediction model.

II. LITERATURE REVIEW

The predictive analysis of network traffic is a multifaceted issue, leading to an increasing emphasis on the application of artificial intelligence (AI), machine learning (ML), and deep learning (DL) methodologies. These advanced analytical models help analyze of big data sets and recognize patterns within the network traffic to use in anomaly detection and traffic prediction.

Traditional methods for network traffic prediction includes techniques as Auto-Regressive Integrated Moving Average (ARIMA), Gaussian process, and Kalman filter [1]. Liu et al, proposed Fractional Auto-Regressive Integrated Moving Average (FARIMA) model for traffic modeling [2]. Machine learning techniques have been showed have potential in recognizing network traffic patterns by user feedback [3]. Another study focused on Random Forest (RF) and Support Vector Machine (SVM) algorithms to predict the network traffic [4]. Another study has compared ARIMA and Long-Short Term Memory (LSTM) algorithms for cellular traffic prediction and classification [5]. Hybrid models are also implemented, in their study Zhou et al. proposed a combined ARIMA and GARCH model [6].

Neural networks especially have demonstrated the potential to learn various network traffic patterns dynamically to predict network traffic [7]. Recurrent neural networks and LSTM algorithm have proposed to predict the network traffic patterns [8]. These methos are shown to be effective in operating complex and non-stationary networks. Another study has utilized reinforcement learning for network traffic prediction [9]. Use of AI, ML and DL techniques have been improved network traffic prediction providing more accurate models. Integration of these techniques with traditional network management systems can increase optimization and improve quality of service.

III. MATERIALS AND METHODS

For multivariate data analysis we have implemented decomposition for trend and seasonality, autocorrelation (ACF) and partial autocorrelation (PACF), granger causality, and augmented Dickey-Fuller (ADF) test as stationarity test.

In model development phase we have used 4 multivariate time series algorithms: Vector Autoregression (VAR), Vector Error Correction Model (VECM), Long-Short Term Memory (LSTM), and Gated Recurrent Unit (GRU).

A. Data Collection and Preprocessing

For this study, we have worked with a digital media and entertainment provider operating in Turkey. The company offers digital television broadcasting and internet services. Dataset is collected using Cacti and Prtg tools via the SNMP

979-8-3503-7943-3/24/$31.00 ©2024 IEEE

protocol in every five minutes. Five-minute data is than aggregated to hourly data.

Our dataset consist of 4 features as Traffic In Volume, Traffic In Speed, Traffic Out Volume, and Traffic Out Speed for 13 different location. For data analysis and model development phases regarding this study, we have focused on the data of Acıbadem which is a location in Istanbul. Acıbadem network data has 4,392 entries for a 6-month period. Acıbadem network data have no missing values, thus no preprocessing is implemented to fill the missing values. Dataset is normalized to prepare for analysis and model development.

B. *Multivariate Data Analysis*

Decomposition for trend and seasonality of a time series is the process of disintegration of the time series to trend, seasonality, and residual parts [10]. We have used additive model to each 4 features of the data. Decomposition is used to understand the long-term trends in the series, recurring patterns and noise.

ACF method measures the correlation between different parts of the time series which shows the relation of the time series within [10]. PACF method measures the correlation for a single lag excluding all other lags, reflecting direct relation between two time intervals [10]. These analysis help to understand the dependencies of time series to select appropriate modelling techniques.

Granger causality analysis shows the predictive power or effect of one time series over another. It is used to test if the past values of one time series has significant effect of the other time series [11]. The test is implemented to understand the effects of features over Traffic In Volume feature. We have implemented the test with 5 lags and observed p-values.

ADF test is implemented to observe the stationarity of the time series which yields information about the long-term trend of the series [12]. A stationary time series have relatively constant mean and variance. It is important for the model to perform better if the features are stationary. If not, transformation techniques as logarithmic can be implemented to the feature set.

C. *Model Development*

For model development we have implemented VAR, VECM, LSTM, and GRU algorithms. The model development is started after the data is normalized. For all algorithms, the dataset is divided into two parts (80% - 20%) as training and test data. Akaike Information Criterion (AIC) is implemented to find the optimum lag, which gives information about how far the model will take into consideration the past readings. After the models are trained over the training data, it is used to make predictions for the time period of the test data. We have used performance metrics to compare the predictions of models. Performance metrics used are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared Score (R^2), and Mean Absolute Percentage Error (MAPE). MAE measures the average magnitude of the errors without considering direction of the error. MSE quantifies the mean of the squared differences between the anticipated values and the actual values. RMSE is calculated as the square root of the mean of the squared errors. It quantifies the accuracy of the model's predictions. Lower RMSE indicates a better performing model. R^2 is the coefficient of determination which represents the models capability to explain the variance of the target. An R^2 closer to 1 represents a better performing model. MAPE expresses the accuracy of the model in percentage. Lower MAPE indicates better performing model.

IV. FINDINGS

A. *Multivariate Data Analysis Results*

Observing the decomposition of Traffic In Volume, Traffic In Speed, Traffic Out Volume, and Traffic Out Speed features indicated that the trend of the series is mostly stagnant, series has daily and weekly seasonality, and has a significant amount of residuals.

ACF analysis shows that Traffic In Volume, Traffic In Speed, Traffic Out Volume, and Traffic Out Speed features have positive and high correlation for first lags indicating high autocorrelation. The features shows negatives and positive correlation after first lags showing the seasonality. PACF analysis also shows high positive correlation for first lags and decreases gradually for all 4 features. Figure 1 show the ACF and PACF analysis of Traffic In Volume feature.

Granger causality test showed p-values lower than 0.05 for Traffic Out Volume and Traffic Out Speed over Traffic In Volume feature indicating that these features have causality in predicting Traffic In Volume and can be used as features in the prediction models. Traffic In Speed feature on the other hand did not yield a significant result. Table 1 shows the results of Granger causality test for Traffic Out Volume feature.

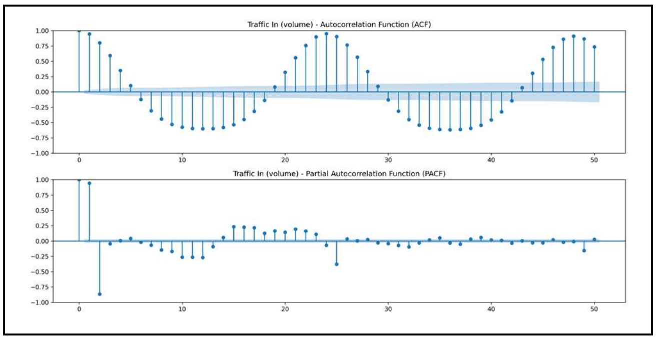

Fig. 1. ACF and PACF analysis of Traffic In Volume

For all features, ADF test resulted in p-values lower than 0.05 with absolute ADF statistic larger than critical values (CV 1%, 5%, 10%) meaning all features are stationary. Table 2 shows the ADF results of data features.

TABLE I. GRANGER CAUSALITY TEST RESULTS FOR TRAFFIC OUT VOLUME FATURE FOR LAGS 1 AND 5

Number of Lags	Test	Test Statistics	P-Value
1	SSR based F	3,564.38	0
	SSR based Chi^2	3,566.82	0
	Likelihood Ratio	2,610.88	0
	Parameter F	3,564.38	0
5	SSR based F	247.22	0
	SSR based Chi^2	1,239.23	0
	Likelihood Ratio	1,091.46	0
	Parameter F	247.22	0

TABLE II. ADF TEST RESULTS

Feature	ADF Stat.	P-Value	CV 1%	CV 5%	CV 10%
Traffic In Volume	-9.33	<0.01	-3.43	-2.86	-2.57
Traffic In Speed	-8.33	<0.01	-3.43	-2.86	-2.57
Traffic Out Volume	-9.33	<0.01	-3.43	-2.86	-2.57
Traffic Out Speed	-8.35	<0.01	-3.43	-2.86	-2.57

B. *Model Development Results*

Table 3 presents the performance characteristics of the models when evaluated on the test data. For the VAR model, maximum lag is chosen to be 30 and AIC yielded 28 as best lag order. The model resulted in 0.82 R^2 and 37.84% MAPE. For the VECM model maximum lag is also chosen to be 30 and model used AIC, Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC) along with the Johansen cointegration test to select the parameters for the best model. Model performed best with 28 lag order and 4 cointegration rank which resulted in 0.82 R^2 and 37.78% MAPE.

TABLE III. PERFORMANCE RESULTS OF MODELS ON TEST DATA

Model	MAE	MSE	RMSE	R^2	MAPE
VAR	0.080	0.008	0.092	0.824	37.84%
VECM	0.080	0.009	0.093	0.822	37.78%
LSTM	0.025	0.001	0.034	0.976	8.95%
GRU	0.027	0.001	0.038	0.969	9.83%

LSTM model is implemented as a LSTM layer followed by a dense layer. Model is implemented to use 24 hour past readings to make predictions. Using different values for parameters, we have compared 24 different model combinations. Neurons 50, 100, and 150; batch size 16 and 32; epochs 50 and 100; and optimizer Adam and RMSprop is implemented. Rectified Linear Unit (ReLU) is used as activation function. Model has reached the best results with 150 neurons, 16 batch size, 50 epochs and Adam optimizer with 0.98 R^2 and 8.95% MAPE. Figure 2 shows predictions of LSTM model along with actual values on test data for Traffic In Volume.

GRU model is implemented as a GRU layer followed by a dense layer with ReLU as activation function. Model is implemented to use 24 hour past readings to make predictions. Same parameter set is used with LSTM model, therefore 24 different model combinations are compared. Best performing model was established with 150 neurons, 32 batch size, 100 epochs, and Adam optimizer with 0.97 R^2 and 9.83% MAPE.

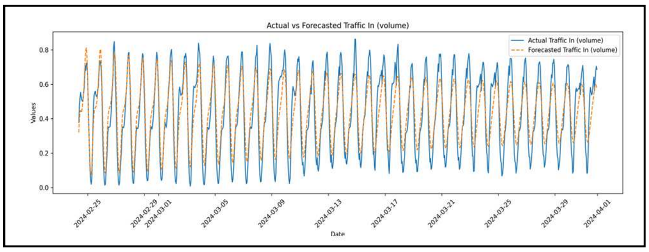

Fig. 2. Predictions of LSTM model for Traffic In Volume

In order to enhance comprehension of the precise hyperparameters employed in this investigation, Table 4 presents an overview of the principal hyperparameters relevant to the LSTM and GRU models. These hyperparameters were selected after experimenting with various combinations to achieve the best predictive performance.

TABLE IV. HYPERPARAMETERS USED FOR LSTM AND GRU MODELS

Model	LSTM		GRU	
	Searched	Best	Searched	Best
Neurons	50, 100, 150	150	50, 100, 150	150
Batch size	16, 32	16	16, 32	32
Epoch	50, 100	50	50, 100	100
Optimizer	Adam, Rmsprop	Adam	Adam, Rmsprop	Adam

V. DISCUSSION

This study is focused on predicting the network traffic to be used in network optimization for improved quality of services. The results of the study represent valuable information for potential use-case in the industry.

We have used multivariate time series analysis methods and 4 different multivariate time series modelling algorithms in the study: VAR, VECM, LSTM, and GRU. We have implemented these algorithms on Acıbadem network data and compared the results using several metrics.

Decomposition of time series indicated that the data has daily and weekly seasonality. ACF and PACF analysis demonstrated dependencies between features which are needed for the models to make meaningful predictions. Moreover, ADF test confirmed the stationarity of the features which is needed for the models perform better.

A number of measures, specifically MAE, MSE, RMSE, R^2, and MAPE, are employed to evaluate the performance of models using test data. The aforementioned metrics provide an indication of the model's predictive accuracy in relation to the observed test readings. Although all 4 algorithms performed better than chance level, especially LSTM and GRU algorithms performed best with high R2 and low MAPE values.

The LSTM and GRU models have shown better performance in capturing complex patterns in the data. The performance of the LSTM and GRU models in this study aligns with findings from the literature where these models have been recognized for their effectiveness in handling time series data with complex patterns [13-15]. These models are particularly effective in modeling temporal dependencies and log-term relationships, learning the intricate relationships between variables. The LSTM model uses memory cells to retain important information over time, which makes the model's predictive power stronger by remembering crucial patterns without losing important details. The LSTM model is widely recognized for its capacity to capture long-term dependencies. This is achieved by the utilization of memory cells that can retain information across numerous time steps. This attribute is particularly crucial for accurately predicting time series data that exhibits seasonality and trends [1]. In a similar vein, the GRU model acquires knowledge about long-term dependencies while utilizing a reduced number of parameters, hence enhancing its training efficiency. Similarly, the GRU model, while simpler in architecture compared to LSTM, has been shown to perform comparably well in various studies, particularly in cases where computational efficiency is required without significant loss in predictive power [2]. Both models have successfully captured the complex patterns and seasonal effects in the dataset, resulting in lower error rates and higher R-squared scores. This capability enhances the models' applicability in network traffic optimization and network management processes.

This study shows how artificial intelligence based models can be used effectively on complex problems as network traffic. Results present valuable opportunities for network management, resource allocation, congestion detection, anomaly detection, and quality of service improvement. In future research, the models will be further optimized by exploring different hyperparameter tuning techniques to achieve better predictive performance. Additionally, transfer learning could be applied to adapt these models for different network environments or other related use cases, such as real-time traffic monitoring or anomaly detection in various network types. Moreover, the investigation of ensemble methods that integrate the capabilities of numerous models could potentially yield predictions that are considerably more resilient and precise. An other potential avenue is the incorporation of external variables, such as meteorological conditions or societal occurrences, into the models in order to augment their predictive capacities within particular contexts. Finally, investigating the applicability of these models in different industries, such as telecommunications or smart city infrastructures, could provide valuable insights into their broader utility and impact

ACKNOWLEDGMENT

This research project was conducted as part of the DSmart R&D Center.

REFERENCES

[1] X. Xu, C. Liu, Y. Zhao, & X. Lv, "Short - term traffic flow prediction based on whale optimization algorithm optimized bilstm_attention", Concurrency and Computation: Practice and Experience, vol. 34, no. 10, 2022. https://doi.org/10.1002/cpe.6782

[2] L. Jia-kun, Y. Shu, L. Zhang, F. Xue, & O. Yang, "Traffic modeling based on farima models", Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No.99. https://doi.org/10.1109/ccece.1999.807189

[3] B. Kuboye, A. Adedipe, S. Oloja, & O. Obolo, "Users' evaluation of traffic congestion in lte networks using machine learning techniques", Artificial Intelligence Advances, vol. 5, no. 1, p. 8-24, 2023. https://doi.org/10.30564/aia.v5i1.5452

[4] S. Sepasgozar and S. Pierre, "Network traffic prediction model considering road traffic parameters using artificial intelligence methods in vanet", IEEE Access, vol. 10, p. 8227-8242, 2022. https://doi.org/10.1109/access.2022.3144112

[5] A. Azari, P. Papapetrou, S. Denic, & G. Peters, "Cellular traffic prediction and classification: a comparative evaluation of lstm and arima", Discovery Science, p. 129-144, 2019. https://doi.org/10.1007/978-3-030-33778-0_11

[6] B. Zhou, D. He, & Z. Sun, "Traffic modeling and prediction using arima/garch model", Modeling and Simulation Tools for Emerging Telecommunication Networks, p. 101-121. https://doi.org/10.1007/0-387-34167-6_5

[7] D. Tao, M. Wan, K. Shi, L. Zhu, X. Wang, & X. Jiang, "Short term prediction of wireless traffic based on tensor decomposition and recurrent neural network", SN Applied Sciences, vol. 3, no. 9, 2021. https://doi.org/10.1007/s42452-021-04761-8

[8] A. Azzouni and G. Pujolle, "Neutm: a neural network-based framework for traffic matrix prediction in sdn", NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018. https://doi.org/10.1109/noms.2018.8406199

[9] N. Balamurugan, M. Adimoolam, M. Alsharif, & P. Uthansakul, "A novel method for improved network traffic prediction using enhanced deep reinforcement learning algorithm", Sensors, vol. 22, no. 13, p. 5006, 2022. https://doi.org/10.3390/s22135006

[10] R. Shumway and D. Stoffer, "Time series analysis and its applications", Springer Texts in Statistics, 2000. https://doi.org/10.1007/978-1-4757-3261-0

[11] L. Siew, W. Lam, S. Jaaman, & P. Lee, "Bibliometric analysis of granger causality studies", Entropy, vol. 25, no. 4, p. 632, 2023. https://doi.org/10.3390/e25040632

[12] S. Saïd and D. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order", Biometrika, vol. 71, no. 3, p. 599, 1984. https://doi.org/10.2307/2336570

[13] Weerakody, Philip, et al. "Cyclic gate recurrent neural networks for time series data with missing values". Neural Processing Letters, vol. 55, no. 2, 2022, p. 1527-1554. https://doi.org/10.1007/s11063-022-10950-2

[14] Noh, Seol-hyun, et al. "Analysis of gradient vanishing of rnns and performance comparison". Information, vol. 12, no. 11, 2021, p. 442. https://doi.org/10.3390/info12110442

[15] Hong, Binjie, et al. "Long memory gated recurrent unit for time series classification". Journal of Physics Conference Series, vol. 2278, no. 1, 2022, p. 012017. https://doi.org/10.1088/1742-6596/2278/1/012017