XLII Simpozijum o novim tehnologijama u poštanskom i telekomunikacionom saobraćaju – PosTel 2024, Beograd, 26 – 27. novembar 2024.

https://doi.org/10.37528/FTTE/9788673954899/POSTEL.2024.06

COMPARISON OF ARIMA AND LSTM FOR POSTAL TRAFFIC VOLUME FORECASTING

Nataša Milosavljević[1], Miloš Milenković[2], Nikola Knežević[2], Nebojša Bojović[2]
[1]University of Belgrade - Faculty of Agriculture, natasam@agrif.bg.ac.rs
[2]University of Belgrade – The Faculty of Transport and Traffic Engineering, Belgrade, Serbia, m.milenkovic@sf.bg.ac.rs, n.knezevic@sf.bg.ac.rs, nb.bojovic@sf.bg.ac.rs

Abstract: *The postal services market is characterized by fluctuating demand, particularly for letter mail and parcel services. Accurate forecasting of postal traffic volumes, especially for these service types, is essential for improved resource allocation, operational efficiency, financial planning, and cost control. In this paper the volume of domestic postal traffic was modelled with a focus on letter mail and parcels. Monthly time series data, comprising of 156 observations for each traffic type, are used to train and test various forecasting models. Both, traditional ARIMA models and Long-Short-Term Memory (LSTM) neural network models, are fitted to each time series. The best model for each series was selected based on the lowest Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) values.*

Key words: *postal traffic, letter mail, parcels, LSTM, Arima.*

1. Introduction

In modern times postal companies face significant challenges, including declining shipment volumes, increasing competition from digital services and private courier companies, and the rising expectations of consumers for fast and reliable delivery. Nowadays, postal companies are facing significant challenges including declining shipment volumes, increasing competition from digital services and courier companies as well as the rising expectations of consumers for fast and reliable delivery. In this very intense competitive landscape and under the increasing demand expectations, forecasting of postal volumes is essential, enabling postal companies to optimize resource allocation, maintain efficient operations, and meet financial objectives. Effective forecasting of postal volumes on short or medium time basis allows a Postal company to anticipate demand shifts, adapt to market changes and remain competitive by improving service quality and customer satisfaction.

The Post of Serbia (PoS) is currently undergoing significant strategic and organizational changes as it transitions to a joint-stock company. This restructuring process involves workforce reduction and adapting operations to meet evolving market trends. These changes are aimed at increasing efficiency, enhancing competitiveness, and aligning with the demands of a modern postal market. Forecasting of postal traffic volume is crucial during the PoS restructuring phase, as it provides data-driven insights for making strategic and organizational changes. Forecasting also supports the adaptation to new market trends, guiding decisions on service diversification and enabling the company to stay competitive in a shifting market landscape. By anticipating future demand, the company can implement sustainable changes that enhance operational efficiency and financial stability.

In this paper, we analyze historical data on domestic letter mail and parcel volumes handled by the PoS. Specifically, using monthly volume data from January 2010 to December 2022 (more recent data are not available), we compare two classes of time series forecasting techniques: the traditional parametric ARIMA model and the non-parametric Long Short-Term Memory (LSTM) neural network. This comparison aims to evaluate the predictive accuracy of each approach for forecasting postal traffic volumes.

The paper is organized as follows. The next section outlines the methodology for modeling and forecasting postal traffic volumes. In Section 3, the proposed models are tested and compared using two time series representing the letter mail and parcel volumes for the PoS. Finally, concluding remarks and directions for future research are presented in the last section.

2. Methodology

This research is based on evaluation of parametric ARIMA methods and non-parametric LSTM models for modelling the volume of domestic postal traffic in PoS. Diagram illustrated on Figure 1. summarizes methodology applied in this paper. Alternative ARIMA models adn LSTM configurations were tested on a training part of both time series (January 2010 – June 2020) , whereas the comparison of selected optimal configurations was condusted on a 20% of test sample (July 2020 – December 2022). Models with best forecasting performances for a given time series can be used for forecasting in forthcoming periods.

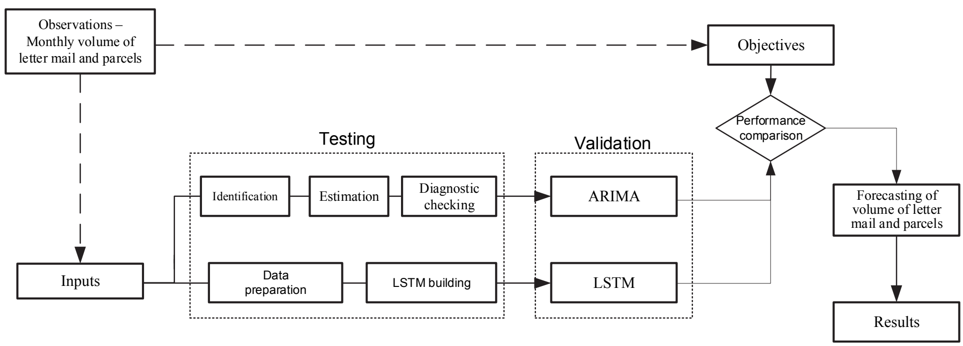

Figure 1. Comparative analysis of ARIMA and LSTM methods for modelling of letter mail and parcel services

2.1 Autoregressive Integrated Moving Average (ARIMA)

ARIMA method represents one of the most important parametric univariate techniques for time series modelling (Box et al., 2008). ARIMA methods are composed of autoregressive model (AR – Auto Regressive), model of moving averages (MA – Moving Average) and ARMA model as a combination of AR and MA (Milenkovic et al., 2019). AR model includes time shifted members of noise or residuals. Neccessary condition for application of ARIMA models is a stationary time series. Letter „I" (Integrated) in the abbreviation ARIMA means that differentiation of 1. order has been conducted in order to make the time series stationary.

Equation which represents $ARIMA(p,d,q)$ model for time sequence Y_t can be represented as:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t \tag{1}$$

where p represents the order of AR process, d is the differentiation order and q is the order of MA process. ε_t represents a white noise sequence which is assumed as a normally distributed variable with zero mean and variance σ^2. B represents the backshift operator, whose effect on a time series Y_t can be summarized as $B^d Y_t = Y_{t-d}$. If time series has trend, seasonal pattern and short time corerelacitons a Seasonal ARIMA (SARIMA) model can be used. Besides the three main components already described above, there is a need for seasonal differencing in order to make a seasonal time series stationary (Milenković et al., 2016). The generalized form of SARIMA(p, d, q)× $(P, D, Q)_s$ model for a series Y_t can be written as (Suhartono, 2011):

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{2}$$

ARIMA modelling for a given time series includes model identification, parameter estimation and model validation (Box et al., 2008). Box-Jenkins methodology is illustrated on Figure 2. Analysis of autocorrelation (ACF) and partial autocorrelation functions (PACF) results in identification of an appropriate ARIMA model. For the sake of elimination of subjectivity and improving the process of determining the final order of ARMA model, Akaike Information criteria (AIC), Akaike Information Criteria with correction (AICc), Bayesian Information Criteria (BIC) as well as normalized BIC can be applied. The phase of estimation includes fitting of model and time series as well as estimation of parameters by maximum likelihood method. The last step includes validation of selected model by diagnostic checking of stationarity, invertibility and presence of redundance in model parameters (Milenkovic et al., 2016).

2.2 Long Short Term Memory (LSTM) models

Long Short-Term Memory (LSTM) neural networks are a type of recurrent neural network (RNN) designed to handle the vanishing gradient problem, which is common in traditional RNNs. LSTMs are particularly effective for time series prediction tasks due to their ability to learn long-term dependencies (Sepp and Schmidhuber 1997; Manowska et al. 2021). Memory blocks that are responsible for memorizing, and manipulations between blocks are done by special multiplicative units called gates (Figure 3.).

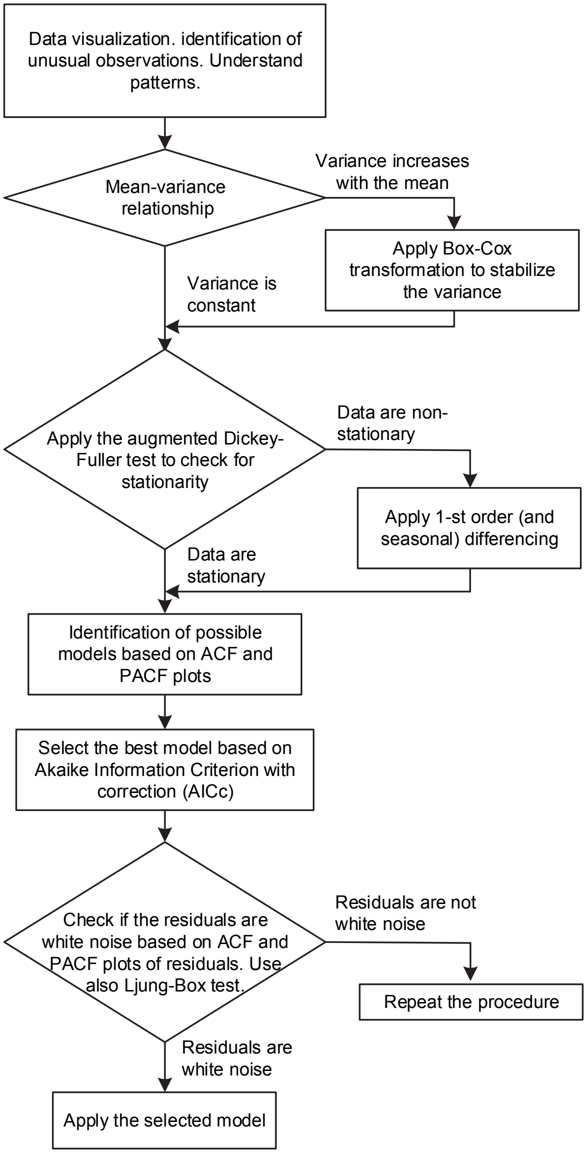

Figure 2. Box-Jenkins methodology for ARIMA model selection

The gates control the flow of information (Ma et al. 2015; Hrnjica and Mehr 2020). The forget gate decides which information to discard from the cell state. It is defined as:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

Where f_t is forget gate activation at time step (t), σ is sigmoid activation function, W_f is a weight matrix for the forget gate, h_{t-1} is hidden state from the previous time step, x_t Input at the current time step and b_f is bias term for the forget gate.

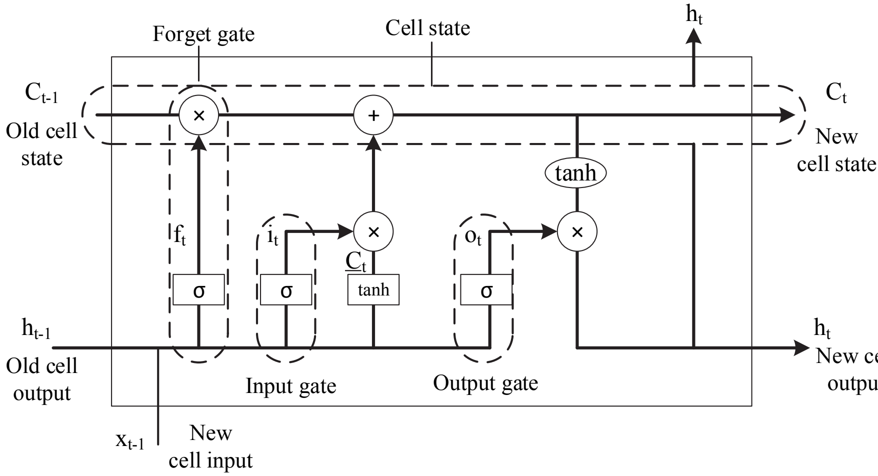

Figure 3. Memory cell and gate units of LSTM memory block (Milenkovic et al., 2023)

The input gate determines which values from the input should be updated in the cell state. It consists of two parts: the input gate layer and the candidate values.

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \quad (2)$$

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \quad (3)$$

Where i_t is the input gate activation at time step (t), $\tilde{C}_t$ is candidate cell state, W_i is weight matrix for the input gate, W_C is weight matrix for the candidate cell state, b_i is bias term for the input gate and b_C is bias term for the candidate cell state. The cell state is updated using the forget gate and the input gate:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

C_t is the cell state at time step (t), whereas C_{t-1} represents cell state from the previous time step. The output gate determines the output of the LSTM cell. It is defined as:

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \quad (5)$$

$$h_t = o_t \cdot \tanh\left(C_t\right) \quad (6)$$

Where o_t is output gate activation at time step (t), h_t is hidden state at time step (t), W_o is a weight matrix for the output gate and b_o represents the bias term for the output gate.

3. Results and discussion

This study focuses on the observation and prediction of letter mail and parcel volumes of PoS. The historical data for both time series are presented in Figure 4. The time series data have been obtained from the Statistical Office of the Republic of Serbia (https://www.stat.gov.rs). The sample data are monthly observations of letter mail volumes and parcel volumes covering the period from January 2010. to December 2022. The first 126 monthly observations were used as a training data set, whereas the remaining 30 observations served for verification of selected models. Both types of services (letter mail and parcels) were independently investigated and the appropriate models were estimated. ARIMA and LSTM are implemented by the use of R and Python software packages, respectively.

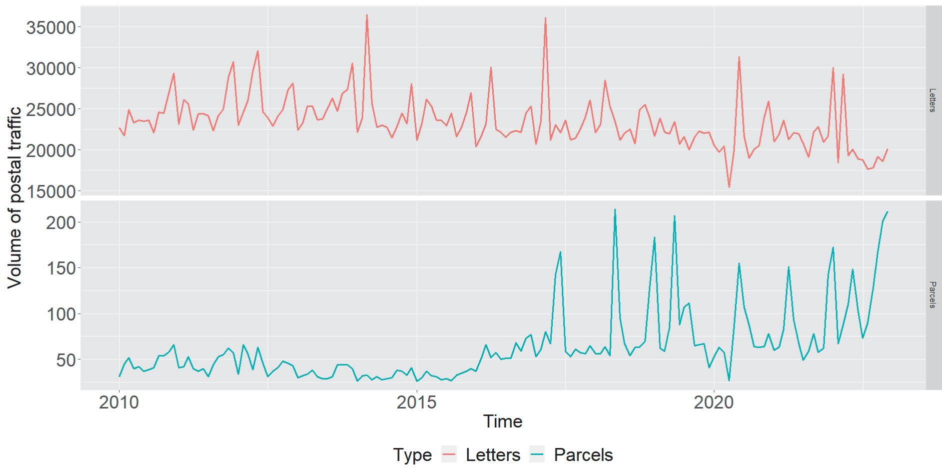

Figure 4. Historical observations of postal traffic volume – letter mails and parcels (January 2010 – December 2022).

3.1 ARIMA results

Table 1. summarizes the process of the ARIMA model selection. In the preliminary step, the time series is visually examined and the procedure for detecting outliers is applied Letter mail time series contains 3 outliers (March 2014, March 2017, June 2020). Parcels time series contains 8 outliers (May 2017, June 2017, May 2018, December 2018, January 2019, May 2019, April 2020, June 2020). All these outliers in both time series are replaced with interpolated values based on the surrounding non-outlier data, preserving the overall trend and seasonality. According to ADF test the letter mail time series is stationary (it doesn't have unit root, p-value is less than 0.05), however, by visual inspecting (there is a subtle trend in the data) and by employing KPSS test (Kwiatkowski-Phillips-Schmidt-Shin test) it can be concluded that the time series is non-stationary with a strong indication of a trend (p-value is 0.01, which means the null hypothesis of stationarity at the 1% significance level can be rejected). KPSS proves that parcel volume related time series is non-stationary too (p-value is 0.01). Based on the

results of Ljung-Box test it appears there is a significant seasonality in the letter mail time series (p-value = $1.99 \cdot 10^{-06} < 0.05$). This is proven by KPSS testing of stationarity of seasonally differenced time series (p-value = 0.0992 which means the time series is stationarized). Parcel volume time series also has a seasonal character (KPSS test of seasonally differenced time series gives p-value of 0.07).

Table 1. Arima model selection

Time series	Stationarity (ADF test: p-value)	Correlogram			
		Non-seasonal		Seasonal (frequency = 12)	
		ACF	PACF	ACF	PACF
Letters	0.0156	1st lag (0.416) 2th lag (0.222) 12th lag (0.408)	1st lag (0.182) 4th lag (0.182) 12th lag (0.321) 13th lag (-0.202)	1st lag (0.345) 12th lag (-0.426)	1st lag (0.345) 12th lag (-0.339)
Parcels	0.254	1st lag (0.777) 2th lag (0.645) 3rd lag (0.648) … 12th lag (0.526)	1st lag (0.777) 3rd lag (0.302) 4th lag (-0.228) 15th lag (-0.258)	1st lag (0.576) 2th lag (0.376) … 6th lag (0.260)	1st lag (0.576) 3rd lag (0.408)

In the letter mail volume related time series, based on autocorrelation values of seasonally differenced series and significant negative autocorrelation at Lag 12 (-0.426) as well as significant partial autocorrelation at Lag 12 (-0.339) the seasonal AR and MA orders are equal to 1. For non-seasonal AR and MA components, the initial order was 1. The best model for letter mail related time series, according to AIC is SARIMA(0,1,1)(0,1,1)$_{12}$ which also has the lowest MAPE and the highest Adj R-squared. The same procedure was followed for parcels volume time series, where the best model found was SARIMA(1,1,2)(1,1,2)$_{12}$. Table 2. summarizes the performances of alternative configurations of evaluated ARIMA models for both time series.

Since the p-value of residuals for selected SARIMA model of letter mail time series is relatively high (0.3818), it can be concluded that there is no significant autocorrelation in the residuals of selected model. Ljung-Box test for the parcels volume time series also indicates that residuals show no significant autocorrelation, they can be considered white noise (p-value = $0.8539 \gg 0.05$).

The BDS (Brock-Dechert-Scheinkman) test evaluates whether a time series (or residuals in this case) exhibits nonlinearity or is purely random. In case of letter volumes, at smaller epsilon values (e.g., 863.8745), some low p-values (e.g., 0.0320 at dimension 2 and 0.0018 at dimension 4) indicate possible nonlinearity. However, for most higher epsilon values, the p-values are not significant, suggesting no strong evidence of nonlinearity or dependency at broader scales. In case of parcel volumes, given the very small p-values across most embedding dimensions and epsilon values, the null hypothesis **(**no evidence of nonlinearity**)** can be rejected**.** This suggests that the residuals are not i.i.d. (independent and identically distributed) and that there is embedded nonlinearity or structure in the residuals that has not been fully captured by the model.

Table 2. MAPE, AIC and adjusted R-squared of ARIMA models

Time series	Models	MAPE (%)	Akaike Information Criteria (AIC)	Adjusted R-squared
Letters	$SARIMA(1,1,1)(1,1,1)_{12}$	5.16	2043.38	0.49
	$SARIMA(0,1,1)(0,1,1)_{12}$	**5.10**	**2041.38**	**0.51**
	$SARIMA(1,1,0)(1,1,0)_{12}$	6.46	2073.14	0.28
Parcels	$SARIMA(0,1,2)(1,1,2)_{12}$	12.46	863.02	0.68
	$SARIMA(1,1,2)(1,1,2)_{12}$	**12.06**	**861.81**	**0.69**
	$SARIMA(1,1,1)(1,1,1)_{12}$	12.89	869.24	0.67

Table 3. Non-linearity testing for ARIMA residuals

Time series	Parameter ε/σ	Dimension (m=2)		Dimension (m=3)	
		Statistic	Probability	Statistic	Probability
Letters	0.5	2.1450	**0.0320**	1.8118	0.0700
	1.0	0.6414	0.5213	0.4976	0.6188
	1.5	1.2139	0.2248	0.4157	0.6777
Parcels	0.5	2.2258	**0.0260**	3.8923	**0.0001**
	1.0	2.6669	**0.0077**	3.8187	**0.0001**
	1.5	3.3050	**9e-04**	3.7963	**1e-04**

3.2. LSTM results

The LSTM network was built based on the Keras framework of the Python 3.10 platform. Before modelling, each training data that belonged to a time series for each border crossing was normalized or rescaled from the original range so that all values were within the range of 0 and 1 (Milenkovic et al., 2023). Then the methods and parameters of the LSTM model needed to be configured. Depending on the time series, the hidden layer was built from 50 to 100 LSTM cells, the number of iterations varied from 20 to 400 and the batch size spanned from 4 to 10. The activation function was set to the rectified linear activation function (ReLu), the loss function was MSE, and the optimizer was the stochastic gradient descent (SGD). The best LSTM model configurations for each time series with associated MAPE values for training and testing samples are given in Table 4.

Table 4. LSTM models configuration

Time series	LSTM model configuration			MAPE (%)	
	Number of cells in hidden layer	Number of epochs	Batch size	Train	Test
Letters	75	90	8	0.20	0.34
Parcels	80	95	8	5.52	7.59

3.3. Discussion of results

In this section, the prediction accuracy of each of the two forecasting methods is analyzed. Actual observations, as well as the predictions generated by each of the proposed models for training and testing samples of domestic letter mail and parcel flows are graphically illustrated in Figure 5 (letter mail flows) and Figure 6 (parcel flows). The vertical dashed line divides the training and testing samples.

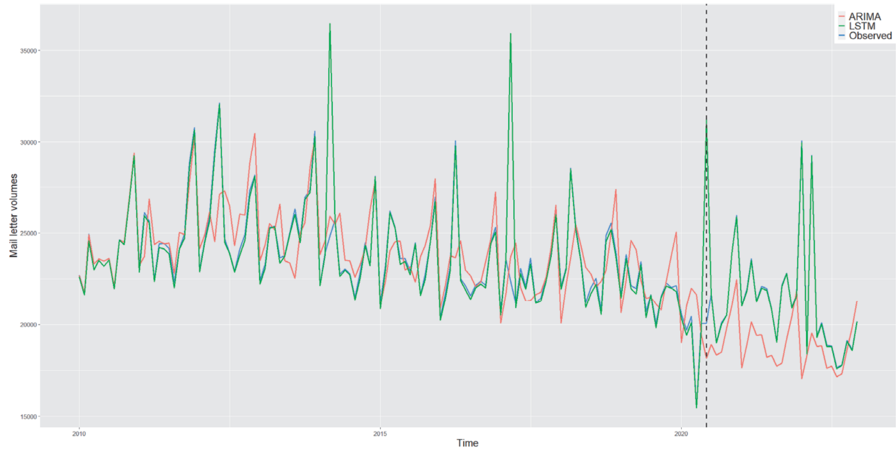

Figure 5. Letter mail flows: Fitting performances of ARIMA and LSTM

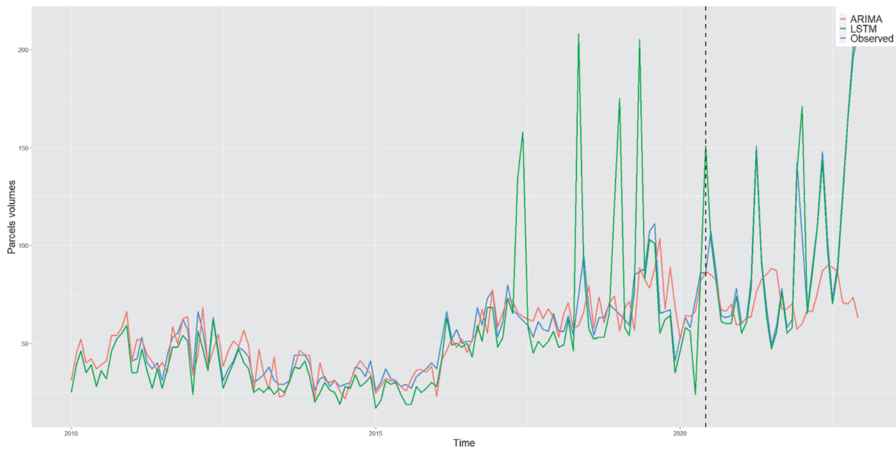

Figure 6. Parcel flows: Fitting performances of ARIMA and LSTM

Forecasting performances are tested based on Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) and the Root Mean Squared Error (RMSE) which are defined as follows (Milenković et al., 2023):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - x_i | \tag{5}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} | 1 - \frac{\overline{Y}_i}{Y_i} | \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \overline{Y}_i)}{n}} \tag{7}$$

where Y_i and $\overline{Y}_i$ represent the actual and predicted values of the time series in period i, respectively. MAE represents the mean of absolute errors. MAPE is one of the most commonly used criteria to measure forecast accuracy. It represents the sum of the individual absolute errors divided by the actual observation. RMSE represents a square root of the average squared error. Forecasting accuracy of proposed methods for letter mail volumes is given in Figure 7.

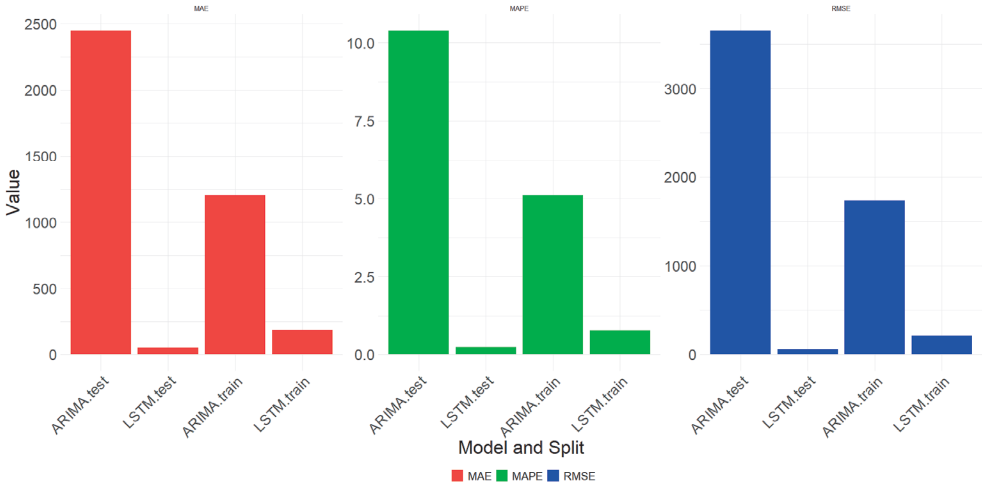

Figure 7. Letter mail volumes: Performance metrics

Forecasting accuracy of proposed methods for parcel volumes is given in Figure 8.
In both time series, in terms of performances on a training and testing sample, the LSTM shows the lower values of MAE, MAPE and RMSE. In terms of the ratio between training and testing performances in case of ARIMA the results for letter mail volumes are: RMSE (train/test): 0.19, MAE (train/test): 0.19, MAPE (train/test): 0.44 (similar results are obtained in case of parcel volumes). From these results it can be concluded that the ARIMA model much better performs on the training set, and not generalizes well to unseen data. On the other side, in case of parcel volumes, LSTM model performs

better on the test set (RMSE (train/test): 1.81, MAE (train/test): 1.81, MAPE (train/test): 3.34) (similar results are obtained in case of letter mail volumes). The LSTM's ratios being above 1 suggest that it generalizes better on the test data, indicating that it is not overfitting to the training data. In contrast, ARIMA's very low ratios (<<1) show poor generalization, as it performs much better on the training data than on the test data. This discrepancy is particularly significant for time series with nonlinear patterns, such as parcel volumes, where ARIMA fails to capture nonlinearities, leading to overfitting and poor performance on testing data.

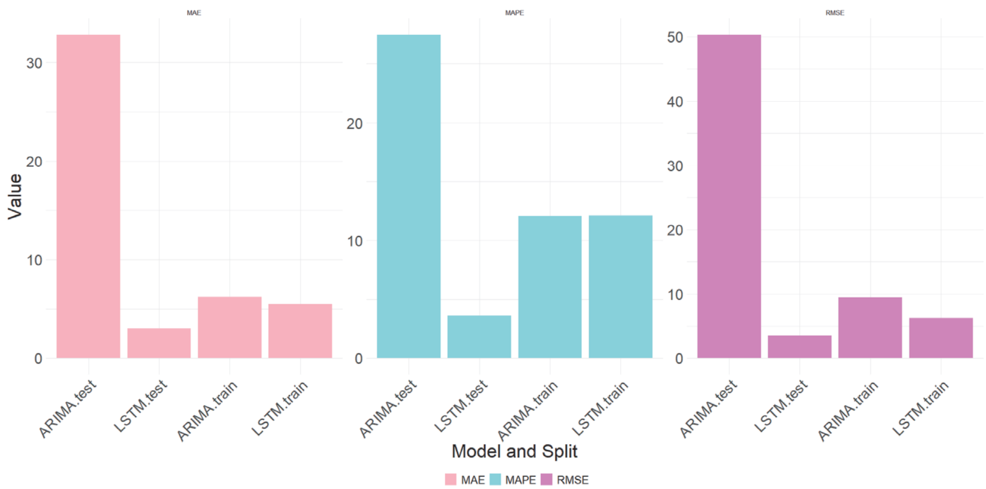

Figure 8. Parcel volumes: Performance metrics

4. Concluding remarks

In the paper, two time series modelling techniques, parametric ARIMA and non-parametric LSTM, were applied to analyze letter mail and parcel volumes in the domestic postal traffic of the Republic of Serbia. Based on the performance results, it can be concluded that LSTM outperforms ARIMA for both time series. Future research will focus on exploring alternative machine learning techniques and developing hybrid models that combine multiple prediction methods to further improve the accuracy of postal traffic forecasting.

Literature

[1] G. E. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis, Forecasting and Control, New Jersey, John Wiley and Sons, 2008.

[2] M. Milenković, N. Milosavljević, N. Bojović, and S. Val, "Container flow forecasting through neural networks based on metaheuristics", Operational Research: An International Journal, vol. 21, pp. 965-997, 2019.

[3] M. Milenković, L. Svadlenka, V. Melichar, N. Bojović, and Z. Avramović, "SARIMA modelling approach for railway passenger flow forecasting", Transport, vol. 33(5), pp. 1113-1120, 2016.
[4] Suhartono, "Time Series Forecasting by using Seasonal Autoregressive Integrated Moving Average: Subset, Multiplicative or Additive Model", Journal of Mathematics and Statistics, vol 7(1), pp. 20-27, 2011.
[5] H. Sepp and J. Schmidhuber, "Long Short-Term Memory", Neural computation, vol. 9(8), pp. 1735–1780, 1997.
[6] A. Manowska, A. Rybak, A. Dylong, and J. Pielot, "Forecasting of Natural Gas Consumption in Poland Based on ARIMA-LSTM Hybrid Model" Energies, vol. 14, pp. 1–16, 2021.
[7] Ma X., Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data" Transportation Research Part C: Emerging Technologies, Vol. 54, pp. 187–197, 2015.
[8] B. Hrnjica and A. D. Mehr, "Energy Demand Forecasting Using Deep Learning." In Smart Cities Performability Cognition Security, EAI/Springer Innovations in Communication and Computing, edited by F. Al-Turjman, pp. 71–104, 2020.
[9] M. Milenković, M. Gligorić, N. Bojović and Z. Gligorić, A comparison between ARIMA, LSTM, ARIMA-LSTM and SSA for cross-border rail freight traffic forecasting: the case of Alpine-Western Balkan Rail Freight Corridor, Transportation Planning and Technology, vol. 47(1), pp. 89-112, 2023.

Abstract: Tržište poštanskih usluga se karakteriše izrazito promenljivom tražnjom posebno u domenu pismonosnih i paketskih pošiljaka. Precizno modelovanje tokova poštanskog saobraćaja od suštinskog značaja za poboljšanje iskorišćenja resursa, operativnu efikasnost, finansijsko planiranje i kontrolu troškova. U ovom radu modelovan je obim nacionalnog poštanskog saobraćaja, odnosno pismonosnih i paketskih pošiljaka. Dve vremenske serije, od po 156 observacija, na mesečnom nivou, su iskorišćene za obuku i testiranje modela prognoziranja. Tradicionalni parametarski ARIMA modeli i neparameterski LSTM modeli neuronskih mreža su razvijeni za svaku od vremenskih serija. Odabran je najbolji model na osnovu najniže vrednosti kriterijuma korena srednje kvadratne greške, srednje apsolutne greške i srednje apsolutne procentualne greške.

Keywords: *poštanski saobraćaj, pismonosne pošiljke, paketi, LSTM, Arima*

POREĐENJE ARIMA I LSTM MODELA ZA PROGNOZIRANJE OBIMA POŠTANSKOG SAOBRAĆAJA

Nataša Milosavljević, Miloš Milenković, Nikola Knežević, Nebojša Bojović