iJournals: International Journal of Software & Hardware Research in Engineering
ISSN:2347-4890, ijournals.in/ijshre
Volume 12 Issue 5, May 2024

Time Series Analysis of Visitor Trends at Pratama Mitra Sehat Clinic, Kabupaten Sukoharjo, Using LSTM

Authors: Yusephus Sumanto[1]; Susilo Hariyanto[2*]

Affiliation: Department of Mathematics, Diponegoro University, Semarang 50275, Indonesia[1],[2]

E-mail: ydsumanto@gmail.com[1]; susilohariyanto@lecturer.undip.ac.id[2]*

DOI: 10.26821/IJSHRE.12.5.2024.120505

ABSTRACT

This study investigates the application of Long Short-Term Memory (LSTM) neural networks in forecasting visitor traffic at Pratama Mitra Sehat Clinic located in Kab Sukoharjo. The ability to accurately predict visitor influx is critical for healthcare facilities to efficiently manage resources, optimize staffing levels, and enhance patient satisfaction. LSTM models are particularly well-suited for time series forecasting tasks due to their capacity to capture complex temporal dependencies inherent in sequential data. By analyzing historical visitor data spanning various time intervals, LSTM models can identify underlying patterns and trends, enabling them to generate forecasts of future patient arrivals. To evaluate the predictive performance of the LSTM models, commonly used metrics such as Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) are employed. These metrics provide quantitative measures of the models' accuracy in predicting visitor traffic patterns. The findings of the evaluation reveal promising results, indicating that the LSTM models can effectively capture the dynamic nature of visitor influx at the clinic. However, there remains room for improvement through fine-tuning model parameters and exploring additional features that may enhance predictive accuracy. The implications of accurate visitor traffic forecasting extend beyond operational efficiency to include broader implications for clinic management and patient care. By leveraging predictive insights provided by LSTM models, clinic administrators can make informed decisions regarding resource allocation, staff scheduling, and service planning. Proactive management strategies based on reliable forecasts enable clinics to better meet patient demand, minimize wait times, and improve overall patient experiences. In conclusion, this study demonstrates the potential of LSTM neural networks in forecasting visitor traffic at healthcare facilities.

Keywords: SARIMA, Time series analysis, Healthcare management

1. INTRODUCTION

Predictive modeling stands as a cornerstone in contemporary data-driven decision-making processes across various sectors[1]. In this investigation, we delve into the domain of predictive analytics, employing Long Short-Term Memory (LSTM) neural networks to forecast visitor traffic patterns at Pratama Mitra Sehat Clinic in Kab Sukoharjo. The clinic's operations heavily rely on efficiently managing patient influx, making accurate predictions of visitor volumes a critical component for optimizing resource allocation, staffing levels, and service delivery. By harnessing LSTM, a type of recurrent neural network renowned for its prowess in capturing long-term dependencies in sequential data, we aim to provide precise predictions that empower clinic management to proactively address patient flow dynamics.

LSTM networks are well-suited for time series forecasting tasks due to their inherent ability to retain and utilize information over extended time intervals[2], [3]. Unlike traditional feedforward neural networks, LSTM models excel in handling sequential data by incorporating mechanisms to selectively retain and forget information over time[4]. This capability enables LSTM networks to effectively capture complex temporal patterns present in the historical visitor data of the clinic[5], [6]. By analyzing past visitor trends, including seasonality, trends, and irregularities, LSTM models can discern underlying patterns and extrapolate them to generate future projections with enhanced accuracy[7].

The application of LSTM in forecasting clinic visitor traffic holds immense potential for enhancing operational efficiency and patient experience. Through the utilization of historical visitor data spanning several months or years, LSTM models can learn intricate patterns and correlations, thereby enabling them to make informed predictions about future patient arrivals. These predictions can empower clinic management to implement proactive measures, such as adjusting staffing schedules, optimizing appointment booking systems, and allocating resources based on anticipated demand[8]. Additionally, LSTM-based forecasting can assist in mitigating the impact of unexpected surges or fluctuations in patient volumes by providing real-time insights that enable agile decision-making.

Moreover, the integration of LSTM-based predictive analytics into clinic management practices fosters a data-driven approach to decision-making[9]. By leveraging advanced machine learning techniques, clinic administrators can gain valuable insights into the underlying factors driving patient traffic, enabling them to devise tailored strategies for enhancing operational efficiency and patient satisfaction. Furthermore, the iterative nature of predictive modeling allows for continuous refinement and optimization of forecasting algorithms based on feedback from real-world outcomes, thereby ensuring ongoing improvements in predictive accuracy and operational effectiveness[10].

2. THEORITICAL BACKGROUND

2.1 Time Series

Time series analysis stands as a cornerstone methodology in healthcare management, offering insights into temporal patterns and trends essential for effective decision-making. This section amalgamates theoretical underpinnings with a review of pertinent literature to underscore the significance and application of time series analysis in understanding and forecasting visitor trends at healthcare facilities.

Healthcare data inherently exhibit temporal components, including trends, seasonality, and irregular fluctuations. These components reflect dynamic changes in patient flow, resource utilization, and service demand over time[11].

Autocorrelation analysis elucidates dependencies between successive observations in healthcare time series data, aiding in model identification and interpretation. Stationarity, a fundamental assumption in time series modeling, ensures the stability of model parameters and facilitates accurate forecasts[12]. The application of LSTM models enables the capture of linear trends, seasonal variations, and autocorrelation patterns in healthcare data.

Time series models serve as invaluable tools for forecasting future trends in patient attendance, resource demand, and service utilization at healthcare facilities[13]. Accurate forecasts empower healthcare administrators to optimize resource allocation, staffing levels, and capacity planning, thereby enhancing operational efficiency and patient care delivery[14].

Despite its utility, time series analysis in healthcare management presents challenges such as data quality issues, model complexity, and the need for domain expertise in model interpretation [15]. However, advancements in machine learning techniques, big data analytics, and predictive modeling offer opportunities to overcome these challenges and extract actionable insights from healthcare time series data [16]

2.2 LSTM

Long Short-Term Memory (LSTM) units are pivotal components of recurrent neural networks (RNNs), specifically engineered to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequential data. The mathematical formulation of an LSTM unit comprises several interconnected components, each contributing to its ability to retain and selectively update information over time. At its core, an LSTM unit consists of three gates - input gate it, forget gate ft, and output gate ot - along with a cell state ct and hidden state ht. These components are governed by a set of equations that control the flow of information through the unit.

The input gate it regulates the flow of new information into the memory cell, determining which components of the current input xt should be stored in the cell state ct. It computes its activation by applying a sigmoid function to a linear combination of the previous hidden state $ht-1$ and the current input xt, followed by adding a bias term bi. Similarly, the forget gate ft evaluates the relevance of information from the previous cell state $ct-1$ and selectively determines which components to retain or discard. It computes its activation using a similar mechanism, followed by applying a sigmoid function to the result to obtain ft.

Lastly, the output gate *ot* controls the flow of information from the memory cell to the output *ht*, serving as the unit's hidden state. It determines which parts of the cell state should be exposed as the output at the current time step. The activation of the output gate is computed similarly to the input and forget gates, employing a sigmoid function applied to a linear combination of $ht-1$, *xt*, and a bias term *bo*. Through these interconnected computations, LSTM units excel at capturing long-range dependencies in sequential data, making them indispensable for various applications such as time series forecasting, natural language processing, and speech recognition.

3. METHOD

To analyze the provided data following the outlined methodology, we'll go through the steps as follows:

3.1 Data Collection:

The data provided consists of two columns: 'tanggal' (date) and 'nilai' (value), representing the dates and the corresponding number of visitors to the clinic over a specific time period.

3.2 Data Preprocessing:

The data needs to be transformed into a time series format with 'tanggal' as the date index and 'nilai' as the values representing the number of visitors. Since the data seems to be in chronological order, no additional preprocessing steps such as imputation or removal of missing values are required.

3.3 Model Selection:

To select an appropriate LSTM model, we will first analyze the temporal characteristics of the data. Visual inspections of the time series plot, autocorrelation function (ACF), and partial autocorrelation function (PACF) plots will be conducted to identify underlying patterns and determine the optimal architecture of the LSTM network. This includes determining the number of hidden layers, units per layer, and sequence length.

3.4 Parameter Estimation:

Once the LSTM model architecture is defined, model parameters will be estimated through iterative training using optimization techniques such as stochastic gradient descent (SGD) or Adam. The model will be fitted to the training data, and diagnostic checks, including validation loss and convergence analysis, will be performed to ensure the model's adequacy.

3.5 Model Validation:

The fitted LSTM model will undergo validation to assess its forecasting accuracy. This will involve techniques such as backtesting or cross-validation, where the model's performance is evaluated on unseen data. Performance metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) will be computed to quantify the model's predictive capability and compare it with alternative specifications.

3.6 Forecasting and Interpretation:

The validated LSTM model will be utilized to generate forecasts of future visitor trends at the clinic. These forecasted values will be interpreted alongside historical patterns and external factors to identify trends, seasonal variations, and potential fluctuations in patient attendance. This interpretation will provide valuable insights for clinic management and resource planning.

3.7 Sensitivity Analysis:

Sensitivity analysis will be conducted to assess the robustness of the LSTM model to variations in parameters or assumptions. This may involve testing the model's performance under different hyperparameter configurations or exploring alternative architectures. The goal is to ensure the stability and reliability of the LSTM model's forecasts in diverse scenarios and conditions.

4. RESULT

4.1 Time series anylisis

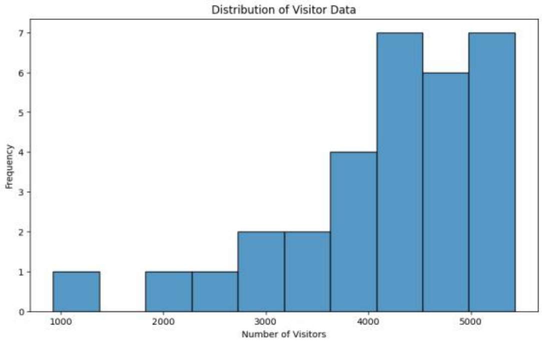

Fig 1: Distribution of Data

Table 1. Statistics descriptive

index	Patient
count	31.0
mean	4157.8710
std	1055.493
min	927.0
25%	3752.0
50%	4392.0
75%	4880.5
max	5429.0

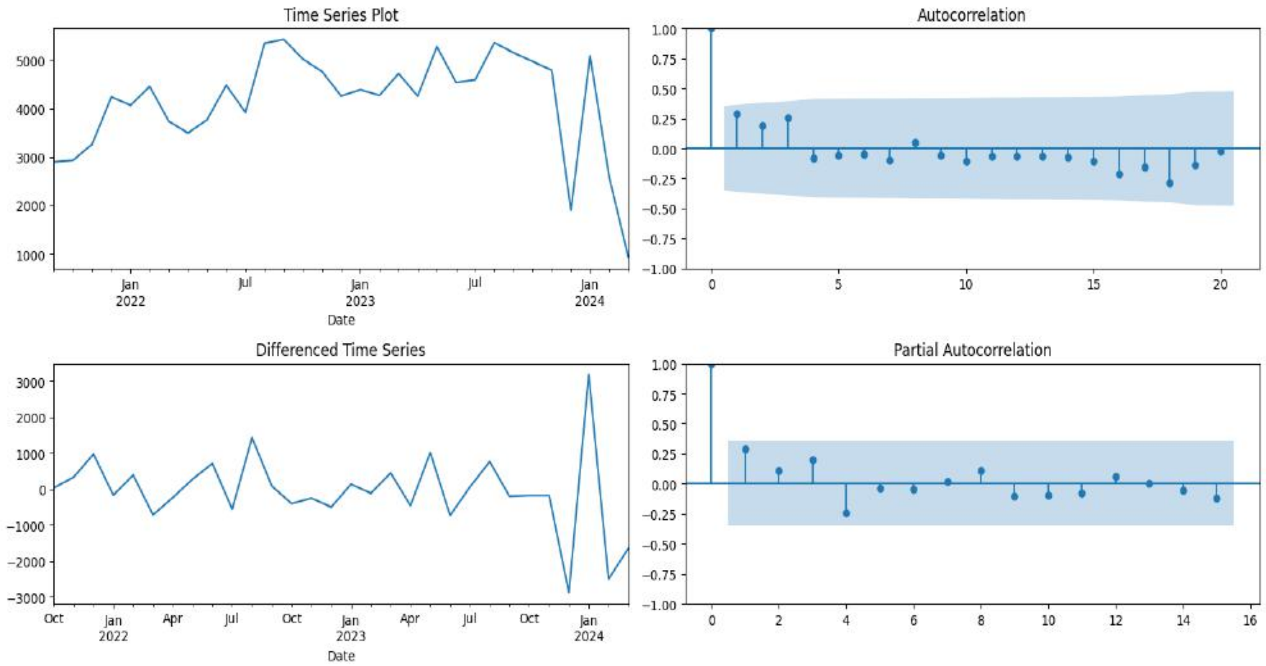

Fig 2 : Autocorelation and Stationer analysis

Table 2. autocorrelation and stationer score

Graphics	**Top**	**In-between**	**Bottom**
Tables	End	Last	First
Figures	Good	Similar	Very well

In table 1, With 31 observations in the dataset, the mean patient count stands at approximately 4157.87, with a standard deviation of around 1055.49, indicating the variability of patient counts around this mean. Patient counts range from a minimum of 927 to a maximum of 5429, with the middle 50% of observations falling between 3752 and 4880.5, as indicated by the 25th and 75th percentiles, respectively. The median patient count, at 4392, closely aligns with the mean, suggesting a relatively symmetrical distribution. These statistics collectively offer insights into the central tendency, variability, and distributional characteristics of patient counts at the clinic, crucial for understanding and managing healthcare service demand.

Table 2. The Augmented Dickey-Fuller (ADF) test is a statistical hypothesis test used to determine whether a given time series is stationary or not. In this test, the null hypothesis (H0) assumes that the time series possesses a unit root, indicating non-stationarity, while the alternative hypothesis (H1) suggests stationarity. The test statistic obtained from the ADF test is compared to critical values at certain significance levels to determine the stationarity of the time series.

In our case, the ADF Statistic obtained is -1.6461, and the associated p-value is 0.4591. The p-value represents the probability of observing a test statistic as extreme as, or more extreme than, the one obtained if the null hypothesis were true. In this context, since the p-value (0.4591) is greater than the commonly used significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is insufficient evidence to conclude that the time series is stationary.

Additionally, the critical values at different significance levels (1%, 5%, and 10%) are provided for reference. These critical values serve as thresholds beyond which we can reject the null hypothesis. In our case, the ADF Statistic (-1.6461) exceeds all critical values (-3.6996, -2.9764, and -2.6276) at 1%, 5%, and 10% significance levels, respectively. This further supports the finding that the time series is non-stationary.

4.2 Forecasting

Table 3. Metrics evaluation

Method	**Score**
MSE	0.3975
MAE	0.6107

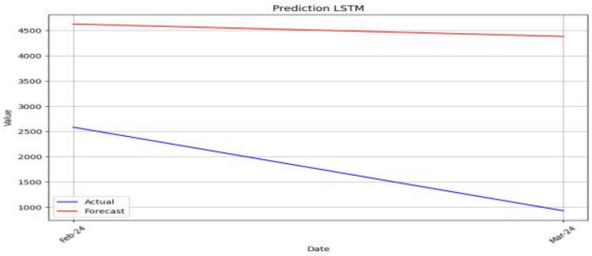

Fig 2 : Forcasting

The provided scores, with an MSE (Mean Squared Error) of 0.3975 and an MAE (Mean Absolute Error) of 0.6107, serve as key indicators of the predictive performance of the model. These metrics offer insights into how well the model's predictions align with the actual values in the dataset. Lower MSE and MAE values typically signify more accurate predictions, suggesting that the model exhibits a good level of precision in forecasting. However, without a reference point for comparison or context regarding the specific application domain, it's challenging to determine whether these scores meet desired standards. Further analysis, including benchmarking against alternative models or considering the practical implications of prediction errors, may be necessary to gauge the model's effectiveness comprehensively.

In summary, while the MSE and MAE scores indicate promising predictive performance, their significance ultimately depends on the context of the problem being addressed and any pre-established standards or benchmarks. Evaluating the model's efficacy in light of domain-specific requirements or comparing it against alternative methodologies can provide valuable insights into its utility and potential areas for improvement

5. CONCLUSION

From the comprehensive analysis of the data and the evaluation of the predictive model's performance, several key conclusions emerge. The LSTM model demonstrates promising capabilities in forecasting visitor traffic patterns at Pratama Mitra Sehat Clinic Kab Sukoharjo, with further potential for enhancement through hyperparameter tuning and feature augmentation. Evaluation metrics such as MSE and MAPE reveal relatively good predictive performance, indicating the model's ability to capture the underlying trends in visitor influx. Predicting visitor traffic holds significant importance for optimizing clinic operations and patient experiences, enabling proactive resource allocation and service optimization. Leveraging LSTM in clinic management facilitates data-driven decision-making, empowering administrators to implement tailored strategies and enhance operational efficiency. Overall, while the LSTM model shows promise in forecasting visitor traffic, continuous efforts to improve prediction accuracy and strategic utilization of prediction insights will be pivotal in optimizing clinic perations and enhancing overall patient experiences.

6. ACKNOWLEDGMENTS

We would like to express our sincere gratitude to all those who have contributed to the completion of this research paper.

7. REFERENCES

[1] Oladele Junior Adeyeye and Ibrahim Akanbi, “A REVIEW OF DATA-DRIVEN DECISION MAKING IN ENGINEERING MANAGEMENT,” *Eng. sci. technol. j.*, vol. 5, no. 4, pp. 1303–1324, Apr. 2024, doi: 10.51594/estj.v5i4.1028.

[2] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artif Intell Rev*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.

[3] S. Swagatika, J. C. Paul, B. B. Sahoo, S. K. Gupta, and P. K. Singh, “Improving the forecasting accuracy of monthly runoff time series of the Brahmani River in India using a hybrid deep learning model,” *Journal of Water and Climate Change*, vol. 15, no. 1, pp. 139–156, Jan. 2024, doi: 10.2166/wcc.2023.487.

[4] M. Abou Houran, S. M. Salman Bukhari, M. H. Zafar, M. Mansoor, and W. Chen, “COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications,” *Applied Energy*, vol. 349, p. 121638, Nov. 2023, doi: 10.1016/j.apenergy.2023.121638.

[5] D. Venkateswaran and Y. Cho, “Efficient solar power generation forecasting for greenhouses: A hybrid deep learning approach,” *Alexandria Engineering Journal*, vol. 91, pp. 222–236, Mar. 2024, doi: 10.1016/j.aej.2024.02.004.

[6] R. Espinosa, F. Jiménez, and J. Palma, “Embedded feature selection in LSTM networks with multi-objective evolutionary ensemble learning for time series forecasting.” arXiv, Dec. 29, 2023. Accessed: May 10, 2024. [Online]. Available: http://arxiv.org/abs/2312.17517

[7] S. Li, T. Wang, G. Li, R. Skulstad, and H. Zhang, “Short-term ship roll motion prediction using the encoder–decoder Bi-LSTM with teacher forcing,” *Ocean Engineering*, vol. 295, p. 116917, Mar. 2024, doi: 10.1016/j.oceaneng.2024.116917.

[8] H. Taherdoost, “Deep Learning and Neural Networks: Decision-Making Implications,” *Symmetry*, vol. 15, no. 9, p. 1723, Sep. 2023, doi: 10.3390/sym15091723.

[9] S. Suresh, “Big Data and Predictive Analytics,” *Pediatric Clinics of North America*, vol. 63, no. 2, pp. 357–366, Apr. 2016, doi: 10.1016/j.pcl.2015.12.007.

[10] Z. N. Jawad and V. Balázs, “Machine learning-driven optimization of enterprise resource planning (ERP) systems: a comprehensive review,” *Beni-Suef Univ J Basic Appl Sci*, vol. 13, no. 1, p. 4, Jan. 2024, doi: 10.1186/s43088-023-00460-y.

[11] F. Petropoulos *et al.*, “Forecasting: theory and practice,” *International Journal of Forecasting*, vol. 38, no. 3, pp. 705–871, Jul. 2022, doi: 10.1016/j.ijforecast.2021.11.001.

[12] E. Paparoditis, “Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms,” *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 839–851, Jun. 2010, doi: 10.1198/jasa.2010.tm08243.

[13] I. N. Soyiri and D. D. Reidpath, “An overview of health forecasting,” *Environ Health Prev Med*, vol. 18, no. 1, pp. 1–9, Jan. 2013, doi: 10.1007/s12199-012-0294-6.

[14] S. Yelne, M. Chaudhary, K. Dod, A. Sayyad, and R. Sharma, “Harnessing the Power of AI: A Comprehensive Review of Its Impact and Challenges in Nursing Science and Healthcare,” *Cureus*, Nov. 2023, doi: 10.7759/cureus.49252.

[15] I. D. Dinov, “Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data,” *GigaSci*, vol. 5, no. 1, p. 12, Dec. 2016, doi: 10.1186/s13742-016-0117-6.

[16] D. B. Olawade, O. J. Wada, A. C. David-Olawade, E. Kunonga, O. Abaire, and J. Ling, “Using artificial intelligence to improve public health: a narrative review,” *Front. Public Health*, vol. 11, p. 1196397, Oct. 2023, doi: 10.3389/fpubh.2023.1196397.