IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Web Traffic Time Series Forecasting using ARIMA and LSTM

Hemant Salunke[1], Amol Kalhapure[2], Avishkar More[3], Shreyash Jagtap[4], Prof. Tanmayee Kute[5]

Student, Dept. of Information Technology, P.E.S Modern College of Engineering, Pune, Maharashtra, India

Professor, Dept. of Information Technology, P.E.S Modern College of Engineering, Pune, Maharashtra, India

ABSTRACT: Web traffic forecasting is critical for effective resource allocation, load balancing, and improving user experience in web services. This study evaluates the performance of three prominent time series forecasting methods—AutoRegressive Integrated Moving Average (ARIMA), AutoRegressive Moving Average (ARMA), and Long Short-Term Memory (LSTM) networks—on web traffic data. ARIMA and ARMA are traditional statistical models known for their robustness in handling linear patterns and seasonality. In contrast, LSTM, a type of recurrent neural network, excels at capturing long-term dependencies and nonlinear patterns in sequential data.We conduct a comprehensive empirical analysis using a real-world web traffic dataset. The models are assessed based on their predictive accuracy, computational efficiency, and ability to handle the inherent volatility and irregularities in web traffic data. The results indicate that while ARIMA and ARMA provide competitive performance for short-term forecasts with relatively lower computational costs, LSTM demonstrates superior accuracy in capturing complex, long-term dependencies, albeit at the expense of higher computational resources.This paper contributes to the field by offering a comparative analysis of traditional statistical models and advanced deep learning techniques for web traffic forecasting, providing insights into their applicability and limitations. The findings suggest that a hybrid approach, leveraging the strengths of both methodologies, could potentially yield enhanced forecasting performance.

I. INTRODUCTION:

Accurate web traffic forecasting is essential for optimizing server resources, managing bandwidth, and enhancing user experience in web-based services. Predicting future web traffic patterns helps organizations preemptively address potential issues related to load balancing, downtime, and performance bottlenecks. Traditional statistical methods like AutoRegressive Integrated Moving Average (ARIMA) and AutoRegressive Moving Average (ARMA) have been widely used for time series forecasting due to their effectiveness in modeling linear patterns and seasonality. However, these models often struggle with the nonlinear and highly volatile nature of web traffic data. Recently, Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have gained popularity in the realm of time series forecasting. LSTM networks are particularly well-suited for capturing long-term dependencies and nonlinear patterns, making them a promising tool for web traffic forecasting. This study aims to compare the performance of ARIMA, ARMA, and LSTM models in forecasting web traffic. By analyzing their predictive accuracy, computational efficiency, and ability to handle the complexities of web traffic data, we seek to provide a comprehensive understanding of their respective strengths and limitations. This research not only highlights the efficacy of traditional and modern forecasting techniques but also explores the potential benefits of integrating these approaches to achieve superior forecasting outcomes.In the following sections, we detail the methodology, experimental setup, results, and conclusions drawn from this comparative analysis, contributing valuable insights to the field of web traffic forecasting.

II. LITERATURE SURVAY:

The system successfully rebuilt the old model and introduced new features throughout the prediction model construction, resulting in higher model efficiency. New features were combined in various ways. 1)Use the median of the provided window length in each time series as an independent feature to capture weekly, monthly, quarterly, and yearly page popularity. 2)Median of medians of various time frame windows based on the golden ratio. The study analysed the acquired results and compared the ac curacies in various scenarios to establish the value of each attribute. We'll then try to figure out how to improve an existing model by tweaking parameters. The goal of the study was to develop the best time-series forecasting model that would allow us to estimate future traffic statistics when a large enough.

dataset was available. With this goal in mind, researchers began looking for predictive models that would allow them to estimate the value of data:

A survey was performed to identify the key factors that lead to project success in construction and software development projects . For software development projects, the result shows that the ranking of priority are as following: Project planning, well defined objectives and requirements, customer involvement, throughout the process, project manager efficiency, top management's involvement, communication, efficiency, involvement of the team in achieving the objectives, project monitoring, way of solving conflict, frequent control checkpoints, cost control, project strategy.

Title	Production and Author	Technical details
Web Traffic Time Series Forecasting Using LSTM Neural Networks with Distributed Asynchronous Training	Roberto Casado-Vara , Angel Martin del Rey , Daniel Pérez-Palauand Luis de-la-Fuente-Valentín and Juan M. Corchado	In this paper, we studied an architecture for web traffic forecasting based on artificial intelligence with LSTM for time series forecasting. For this purpose, new dataset to validate model and extracted the features and hidden patterns to enhance the design of the LSTM. A distributed training system was designed according to the concept of data parallelism along the lines of the Downpour asynchronous training strategy.
Web traffic time series forecasting using ARIMA model	Vrushant Tambe, Apeksha Golait , Sakshi Pardeshi, Prof.Gajanan Arsalwad.	In this research paper the students and professor develop a model that predict web traffic on certain websites . They use Wikipedia pages to predict the future forecasting. Various parameter they consider during training and testing.
An Empirical Study on Internet Traffic Prediction Using Statistical Rolling Model	Sajal Saha, Anwar Haque, Greg Sidebottom	The research paper on time series forecasting on IP network. Developer use different prediction model such as ARIMA,SARIMA, SARIMAX et. ;and analysis the output from different model output. They perform single step traffic prediction.
Network Traffic Prediction: Apply the transformer to time series forecasting	Qian Kong , Xu Zhang,Chongfu Zhang,Limengnan Zhou,Miao Yu , Yutong He, Youxian Chen, Yu Miao,and Haijun Yuan	In this paper, the transformer deep learning model is used to predict network traffic. Practical comparison proves that the training model adopted has a faster convergence speed, higher accuracy, and is easier to handle multidimensional feature data

III. METHODOLOGY:

3.1Dataset

For this project we are considering wikipedia dataset. Dataset contains historical time stamped data. Three main terms available in dataset are page name, date visited and number of visitors.

3.2Technology

ARIMA Model-

ARIMA (Auto regressive Integrated Moving Average model) is a statistical research technique that employs time series data to better comprehend or forecast future trends. An autoregressive integrated moving average model is a sort of regression analysis that assesses the strength of one dependent variable in relation to other changing variables. The model's goal is to predict future securities or financial market movements by analysing the differences between values in a series rather than the actual values. The complete model is as follows:

Φ (L) (1 – L)d yt = θ(L) εt

- Here, p, d and q ϵ Z+, and can be referred to as the order of AR, I and MA parts of the ARIMA model.
- yt refer to the input data, or, the points of observation at time t.
- εt refers to the white noise at any given time t.
- φ (L) are the lag polynomials, and L is the lag operator.
- Term‘d’ refers to the degree of ordinary differencing, it is used to make the time series stationary.

LSTM Model- LSTM networks are a type of recurrent neural network designed to handle long-term dependencies in time series data, particularly useful for nonlinear patterns in web traffic forecasting

1. Forget Gate: Decides what information to discard from the cell state. Ft = σ(Wf·[Ht−1, Xt] + bf)
2. Input Gate: Determines what new information to add.

It = σ(Wi.[Ht-1,Xt] + bi)

C’t = tanh(Wc.[Ht-1,Xt] + bc)

3. Cell State Update: Updates the cell state.

Ct = Ft.Ct-1 + It.C’t

4.Output Gate: Computes the output.

Ot = σ(Wo.[Ht-1,Xt] + bo)

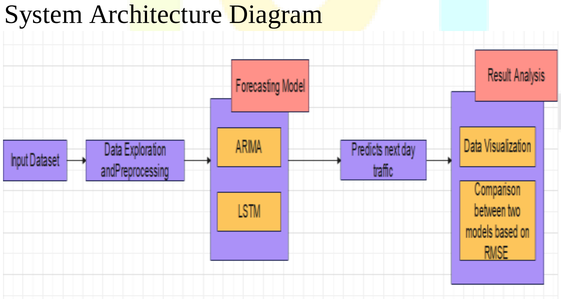
Ht = Ot.tanh(Ct)

Here, Xt is the input, Ht-1 is the previous hidden state, Ct-1 is the previous cell state, σ is the sigmoid function, and tanh is the hyperbolic tangent function. W and b are weights and biases.

For web traffic forecasting, the LSTM model is trained on historical data to predict future traffic patterns, optimizing hyperparameters like layer count, units per layer, and learning rate to enhance accuracy.

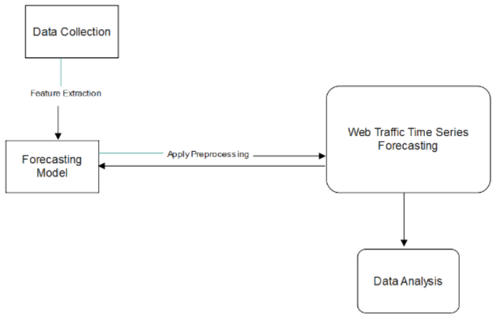
IV. PROPOSED METHOD:

System Architecture Diagram

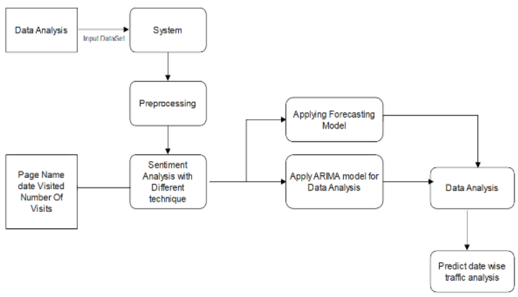


Data Flow Diagram

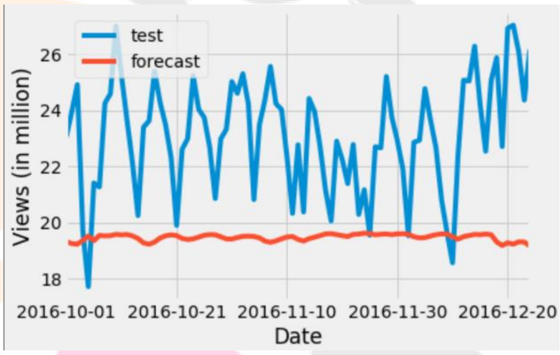
DFD1:

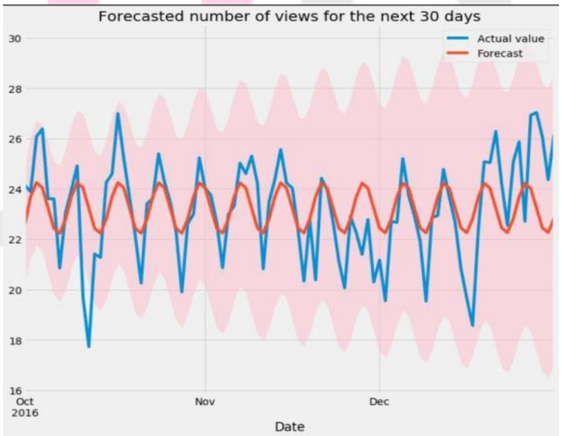


DFD2:



V. Result:



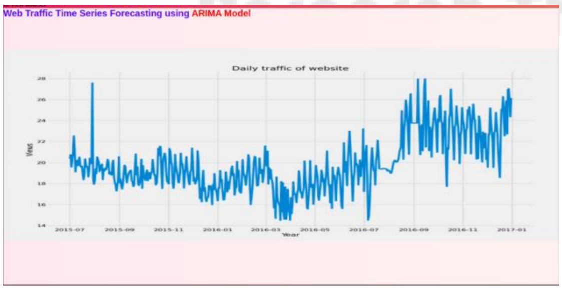


5.1 Comparison

Aspects	**ARIM**	**LSTM Model**

Full form	ARIMA stands for Autoregressive Integrated moving average	LSTM stands for Long short term memory
Basics	It is a class of statistical models for analyzing and forecasting time series data	It is a type of recurrent neural network capable of learning order dependence in sequence prediction problems.
Model Type	Statistical Model	Deep Learning Model
Data Requirement	Stationary Time Series	Sequential Time Series
Seasonality Handling	Requires manual intervention	Can learn and handle seasonality automatically
Computational Efficiency	Generally fast	Computationally intensive, especially for large datasets
Forecasting Performance	Effective for short-term forecasts	Excels in long-term forecasting, especially for complex patterns
Implementation Ease	Relatively straightforward implementation	Requires expertise in deep learning and neural network architecture

5.2 Graphs:



VI. DESCRIPTION OF RESULT

6.1 Conclusion: In this study, we compared the performance of ARIMA, ARMA, and LSTM models for web traffic time series forecasting. Despite the popularity of LSTM networks in capturing complex patterns, our findings suggest

that ARIMA and ARMA models exhibit promising efficiency and accuracy for short to medium-term forecasts. ARIMA's ability to handle linear trends and seasonality with minimal computational resources makes it a practical choice for web traffic forecasting tasks. While LSTM networks offer flexibility in capturing long-term dependencies, their computational demands may outweigh the marginal improvement in accuracy, especially for shorter forecasting horizons. Thus, ARIMA emerges as a promising and efficient choice for web traffic forecasting applications, particularly when computational resources are limited.

6.2 Future Scope: Future research could explore hybrid approaches that leverage the strengths of both ARIMA and LSTM models. Integrating ARIMA's efficiency in capturing short-term trends with LSTM's capability to model long-term dependencies could potentially yield more accurate and robust forecasts. Additionally, investigating advanced deep learning architectures, such as attention mechanisms or transformer models, could further enhance the forecasting performance for web traffic time series data. Furthermore, considering the impact of external factors like social media trends or events on web traffic patterns could lead to more comprehensive forecasting models.

VII. REFERENCES

[1] "Traffic Forecasting Using Time-Series Analysis," 6th International Conference on Inventive Computation Technologies, 2021, Mohammad Asifur Rahman Shuvo, Muhtadi Zubair, Afsara Tahsin Purnota, Sarowar Hossain, and Muhammad Iqbal Hossain (ICICT).

[2] Jianhu Zheng and Mingfang Huang, "Traffic Flow Forecasting Using Deep Learning and Time Series Analysis," IEEE Access, 2020;

Jianhu Zheng and Mingfang Huang, "Traffic Flow Forecasting Using Deep Learning and Time Series Analysis," IEEE Access, 2020; Jianhu Zheng and Ming P Montero-Manso. Fforma: Featurebased forecast model averaging

[3] "Wikipedia Page Traffic Prediction," Navyasree Petluri and Eyhab Al-Masri, 2018 IEEE International Conference on Big Data (Big Data).

[4] "Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA, and RNN," 2018 Eleventh International Conference on Contemporary Computing, Partha Sarathi Mangipudi and Rishabh Madan (IC3).

[5] "An Engineering Approach to Network Traffic Prediction Using a Time-Series Model," International Joint Conference on Artificial Intelligence, 2009. Fu-Ke Shen, Wei Zhang, and Pan Chang, "An Engineering Approach to Network Traffic Prediction Using a Time-Series Model."

[6] Montero-Manso, P.; Athanasopoulos, G.; Hyndman, R.J.; Talagala, T.S. 2020, 36, 86–92. International Journal of Forecasting, International Journal of Forecasting, International Journal of Forecasting, International Journal of Forecasting, International Journal of Forecasting,

[7] S.S. Rangapuram; M.W. Seeger; J. Gasthaus; L. Stella; Y. Wang; T. Januschowski. Deep state space models for time series forecasting. Adv. Neural Inf. Process. Syst. 31, 7785 (2018)

[8] A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tealab, A. Tea 334–340 in Future Comput. Inform. J., 334–340 in Future Comput. Inform. J., 334–340 in Future Comput.

[9] Tyralis, H., and Papacharalampous, G. Variable selection in random forest time series forecasting. Algorithms, vol. 10, no. 114, 2017. A big data and time series analysis technology-based multi-agent system for smart tourism, Chen, W.C., Chen, W.H., and Yang, S.Y.

[10] Chen, W.C., Chen, W.H., and Yang, S.Y. 2018, 8, 947, Appl. Sci.

[11] Boone, T.; Ganeshan, R.; Jain, A.; Sanders, N.R. Consumer analytics in the big data era: Forecasting sales in the supply chain. International Journal of Forecasting, 35,170–1801 (2019).

[12]. Chen, D.; Gao, M.; Liu, A.; Chen, M.; Zhang, Z.; Feng, Y. A Recurrent Neural Network Based Approach for Web Service QoS Prediction. In Proceedings of the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 25–28 May 2019; pp. 350–357.

[13]. Zhou, K.; Wang, W.; Huang, L.; Liu, B. Comparative study on the time series forecasting of web traffic based on statistical model and Generative Adversarial model. Knowl.-Based Syst. 2020, 213, 106467.

[14]. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. Int. J. Forecast. 2020, 36, 54–74.

[15]. Yang, Y.; Lu, S.; Zhao, H.; Ju, X. Predicting Monthly Pageview of Wikipedia Pages by Neighbor Pages. In Proceedings of the 2020 3rd International Conference on Big Data Technologies, Qingdao, China, 18–20 September 2020; pp. 112–115. Mathematics 2021, 9, 421 20 of 21

[16]. Bojer, C.S.; Meldgaard, J.P. Kaggle forecasting competitions: An overlooked learning opportunity. Int. J. Forecast. 2020.