INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Web Traffic Time Series Forecasting

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: Millions of websites are available on google and other search engines. It is very challenging task for website developers and owners to engage users on their websites. If website takes high loading time there may be chance that user will not visit that website again. So it is very necessary that developer should track daily visitors on website and train model and build algorithm to predict future day visitors to reduce the loading time of website. ARIMA model is used to forecast the website traffic to predict next day visitors on website. Advantages of ARIMA model is that it is flexible and it can easily handle a wide range of time series data, as long as they are univariate, meaning they have only one variable. The website components like page name, date of visit and number of visitors etc. will be consider for forecasting.

KEYWORDS: Web traffic, ARIMA, ARMA, Time series forecasting, Data Collection and Feature Understanding.

I. INTRODUCTION

Web traffic prediction is a major concern since it has the potential to produce difficulties in the working of websites. It is responsibility of website owner to keep track on website traffic. If web service provider doesn't care about website traffic then there may be high possibility of loosing customer because customer might have to wait for long time. It is website developers responsibility to build algorithm to handle web traffic. There will be problems like website crash, slow speed of website traffic. There are different models to handle web traffic like ARIMA model. Website components like web page name, date, number of visitors are used to build the prediction model. The techniques we are implement can be extended to diverse applications in financial markets, weather forecasts, audio and video processing. The aim of this model is to predict next day website traffic that will help to build the algorithm to handle that traffic. It will reduce the loading traffic and users need to wait for long time to access the content of website.

for education purpose. Formal project planning should be integrated into higher learning institute in managing the academic projects to further improve the institute. To further understand the current project management practices of the workers in higher learning institute, a preliminary survey is conducted on the institute workers that actively involved in the software development project.

II. LITERATURE SURVEY

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	Publication and Author	Technical Details
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:

$\Phi (L) (1 – L)d\, yt = \theta(L)\, \varepsilon t$

• 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.

IV. PROPOSED METHOD

System Architecture diagram:
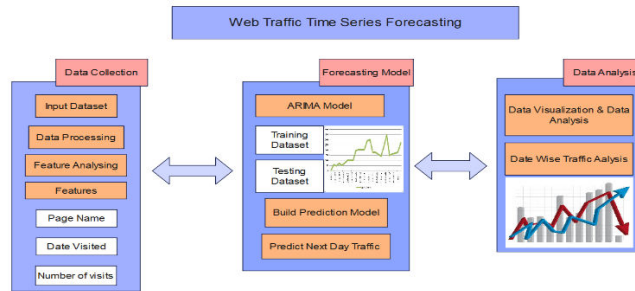
System Architecture

Actiity Diagram:
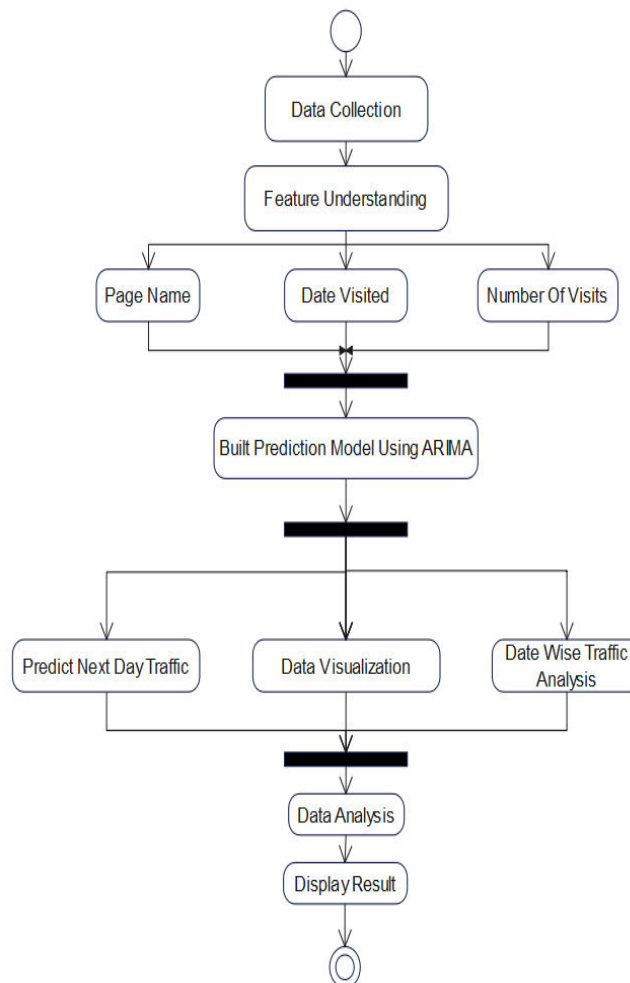
Activity diagram

V. RESULT
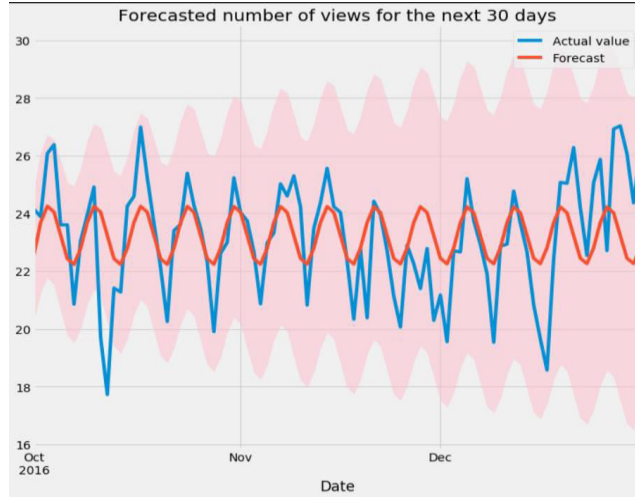
5.1Tables:
Comparison between ARIMA and LSTM model

ARIMA Model	LSTM Model
ARIMA stands for Autoregressive Integrated moving average	LSTM stands for Long short term memory
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.
ARIMA models describe the trends and seasonality in time series as a function of lagged values and Averages changing over time intervals.	LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically the default behavior

5.2 Graphs:
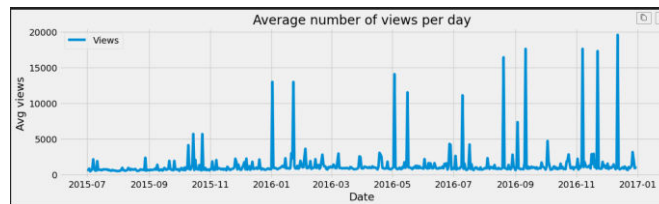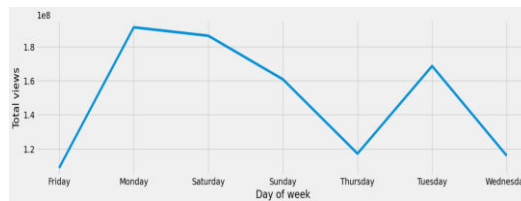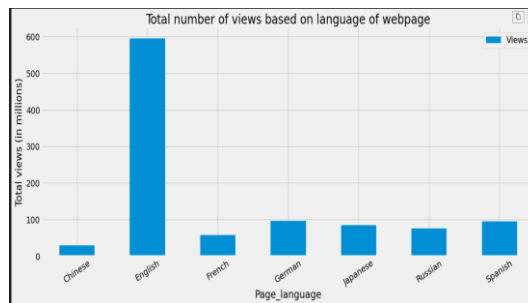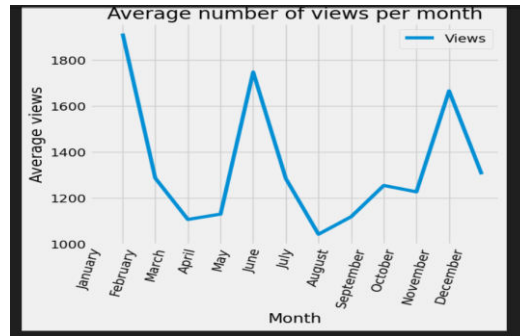
6.3 Description of result:

VI. CONCLUSION AND FUTURE SCOPE

6.1 Conclusion:
Web traffic Time series prediction can be carried out using Long Short Term Memory Recurrent Neural Network and Autoregressive integrated moving average more efficiently and accurately. Prediction of the number of users will access the website in the future is possible. The pro-posed will keep on improving as more user data is fed. Our system can be used across all websites for improving their web traffic load management and business analysis. LSTM RNN brings more efficiency to our system. Our system effectively captures seasonal patterns and long-term trends Including information about holidays, day of week, language, region might help our model to capture more correctly the highs and lows.

6.2 Future Scope:

Time Series Forecasting is one of the least explored areas and various models are evaluated to improve the accuracy of the forecast. The main focus of the proposal is to predict future web traffic to make decisions for better congestion control. Past Values are considered to predict future values. We will also seek to explore multivariate time series and offer suggestions for simplifying the decision-making process in real-time.

REFERENCES

[1] "Wikipedia Page Traffic Prediction," Navyasree Petluri and Eyhab Al-Masri, 2018 IEEE International Conference on Big Data (Big Data).
[2] "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).

[3] "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).

[4] "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."

[5] 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

[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.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Scan to save the contact details