International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

Web Traffic Foresting Using Time Series

Akilabanu Chikkumbi[1], Manu S. K[2], Renuka R. U[3], Sadiya Md.D[4], Ruchita R. P[5]

[1]*Assistant Professor, S G Balekundri Institute of Technology, Belagavi, Karnataka,India*
[2]*Student, S G Balekundri Institute of Technology, Belagavi, Karnataka, India*
[3]*Student, S G Balekundri Institute of Technology, Belagavi, Karnataka, India*
[4]*Student, SG Balekundri Institute of Technology, Belagavi, Karnataka, India*
[5]*Student, S G Balekundri Institute of Technology, Belagavi, Karnataka, India*

Abstract – *influence of external variables. It can effectively analyses short-, medium-, and long-term temporal patterns, making it Websites and online services often experience sudden spikes in traffic, especially during events such as product sales, result announcements, or special offers. If too many people use the system at once, it may run slowly, crash, or stop working. To address this challenge, we apply time- series–based web traffic forecasting using* ***Random Forest*** *and* ***SARIMAX*** *algorithms. These models leverage historical traffic patterns to accurately predict future load across different time intervals. The experimental results show that forecasting upcoming traffic helps service providers prepare resources in advance, optimize server capacity, and reduce the risk of traffic congestion or downtime. This predictive approach enhances system reliability and ensures a smoother user experience during peak access periods.*

By integrating Random Forest and SARIMAX for web traffic forecasting, service providers can proactively manage system load and prevent unexpected failures. This blended approach allows the system to generate better predictions and manage its resources more smoothly, which is especially helpful for online platforms that handle a lot of activity.

Keywords— Web Traffic Forecasting, Random Forest, SARIMAX, Time-Series Prediction, Load Management, Peak Traffic Analysis.

1. INTRODUCTION

Modern digital platforms increasingly rely on accurate forecasting to ensure smooth system operation and uninterrupted user experience. In environments where user activity changes rapidly—such as shopping portals during sales, result announcement websites, or high-demand online services—traffic prediction becomes essential for preventing server overload and performance degradation. Traditional reactive approaches struggle to handle such dynamic patterns because they cannot anticipate short-term fluctuations, seasonal behaviors, or sudden changes caused by external factors.
To address these challenges, advanced time-series forecasting models are widely used. **SARIMAX** is a powerful statistical model capable of capturing trends, seasonality, and the suitable for modeling web traffic or system usage that follows irregular but repetitive behaviors.

In contrast, **Recurrent Neural Networks (RNNs)** provide a machine-learning-driven approach to modeling complex nonlinear dependencies in time-series data. Because they can understand long-term patterns in data, they work well in situations where many factors influence the trends and the behavior of the data changes a lot over time.

1.1 Research Objectives

- **To analyses historical web traffic patterns** and identify trends, seasonality, and nonlinear behaviours that influence sudden spikes during peak usage periods.
- **To implement the SARIMAX model** for capturing linear relationships, seasonal variations, and the impact of external influencing factors in time-series data.
- **To develop an RNN-based forecasting model** capable of learning complex temporal dependencies and nonlinear patterns present in dynamic web traffic.
- **To compare the forecasting performance** of SARIMAX and RNN models across different time intervals, such as short-term, medium-term, and long- term horizons.

- **To design a hybrid prediction approach**, integrating SARIMAX and RNN, to achieve more accurate and stable forecasting results than individual models.

2. SYSTEM ARCHITECTURE AND METHODOLOGY

The overall architecture of our system is designed to predict future web-traffic by learning patterns from past data. The process begins by gathering past traffic data from the server, such as how many users visited the site each hour or each day. After gathering the data, it passes through the pre-processing stage, where it is cleaned, arranged in correct time order, and prepared for model training.

Once the data becomes usable, it is fed into two forecasting models: **SARIMAX** and **Random Forest**. SARIMAX works by studying time-based patterns such as trends, seasonality, and the effect of outside factors. It is helpful in understanding repeated behaviors like weekly peaks or festival-time spikes. Random Forest, on the other hand, learns by creating many decision trees and combining their outputs. This helps it capture non-linear patterns and sudden changes in traffic that simple statistical models might miss.

The method also involves splitting the dataset into training and testing portions so we can see how well the models handle data they haven't seen before. SARIMAX is trained on the ordered time-series, while Random Forest is trained using selected features such as past traffic values, moving averages, or time-based indicators. Once the models are trained, each one delivers its own set of predictions. These outcomes are then evaluated to determine which model delivers the most accurate output.

By using both models together, the system can capture long- term trends as well as sudden changes. This helps service providers plan ahead for busy periods and reduce the chances of the platform slowing down.

2.1 Dataset Preparation

Getting the data ready is a crucial part of predicting web traffic accurately because the raw time-series information is typically messy and not ready to use. The initial stage is to arrange all the traffic records in a proper and regular time order, such as hourly or daily. If any values are missing as a result of server issues or logging errors, they are filled using simple methods like interpolation or forward filling so the timeline stays complete. Unusual spikes or drops—often caused by bots, server downtime, or special events—are identified and corrected so are also transformed or scaled to help the model learn the patterns more easily. Seasonal behaviors, like weekend dips or festival-time increases, are also understood so the model can separate normal patterns from unexpected changes. After cleaning and organizing everything, the data is divided into training and testing sets in the correct time order, helping the forecasting model learn precise and clear patterns.

2.2 Forecasting Models (SARIMAR and Random Forest)

In our research, we utilize two complementary approaches to tackle the problem. Estimating future outcomes of web traffic: **SARIMAX** and **Random Forest**. All models come with unique capabilities that help capture different patterns among the data points. **SARIMAX** is a statistical model designed to analyses and forecast time series data with clear trends and seasonal effects. It is especially useful for modeling regular patterns, such as daily or weekly cycles in web traffic, as well as the influence of external factors like holidays or special events. SARIMAX breaks down the time series into components like auto regression, moving averages, differencing to stabilize the data, and seasonal terms, which together help in making accurate predictions over short and medium time frames.

On the other hand, **Random Forest** is a machine learning model that synthesizes numerous decision trees during building and combines their predictions to improve accuracy and reduce over fitting. Unlike SARIMAX, unlike some models, Random Forest doesn't presume any specific statistical distribution and can learn complex, non-linear relationships in the data. This makes it especially effective for capturing sudden traffic spikes or irregular patterns that traditional time-series models might miss. By combining these two models, we aim to leverage both the interpretability and seasonal strengths of SARIMAX and the flexibility and non- linear learning ability of Random Forest. This hybrid approach is expected to provide more reliable and accurate web traffic forecasts, helping service providers manage resources effectively during peak demand.

2.3 Training and Evaluation

After selecting suitable forecasting techniques, the next phase is to train each model on the cleaned and structured dataset and then assess how accurately they perform. The dataset is split into two sections: the training set, which the models use to learn underlying patterns, and the testing set, which checks how well the models can predict new, unseen data.

For the **SARIMAX model**, training focuses on identifying the optimal parameters that represent the historical behavior of the time-series, including seasonality and any external influencing factors. The model studies previous traffic values along with additional inputs such as holidays or special events to understand patterns. Choosing the correct autoregressive and seasonal orders is important and is usually done using statistical tests and model-selection criteria.

For the **Random Forest model**, the training process begins with generating feature variables from the original time-series. These features may include lagged traffic values, moving averages, time-based indicators (such as hour, weekday, or month), and any other relevant inputs. The Random Forest algorithm creates multiple decision trees using different subsets of the training data and then combines their outputs, enabling it to learn complex, non-linear relationships in the data.

After training both models, forecasts are produced for the testing period. These predictions are compared with the actual observed traffic values using evaluation metrics such as **RMSE**, **MAE**, and **MAPE**. These error measures indicate how close the forecasts are to the real values and help identify which model delivers better performance.

3. RESULTS AND PERFORMANCE ANALYSIS

In our project, Users have several options to explore website traffic forecasts based on data from the past 60 days. The platform offers forecasts generated by both the Random Forest algorithm and the SARIMAX model enables straightforward comparison of how well each performs.

A key feature of the project is the accuracy comparison graph, which visually displays how effectively each model formulates predictions regarding web traffic. Users can hover over the graph to see detailed traffic predictions from both models at specific points in time, making it simple to understand the differences in their forecasts.

Additionally, the system offers the option to download the 30day forecast data exported as a CSV file to allow deeper examination or record-keeping. This downloadable file contains detailed traffic predictions for each day, generated by both models. Based on the accuracy comparison, the Random Forest and SARIMAX models each have their strengths. Random Forest tends to perform better in capturing sudden fluctuations and nonlinear patterns, while SARIMAX excels at modeling regular seasonal trends. Overall, synthesizing the observations from both models offers a well-rounded and reliable forecast that can help website administrators manage traffic effectively and prepare for peak demand periods.

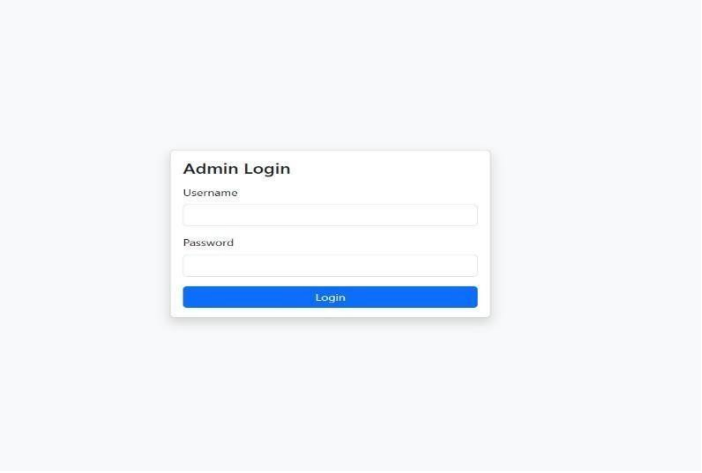

Fig 1: Login page.

Fig 2: Traffic prediction based on models.

Fig 3: Traffic prediction based on models,

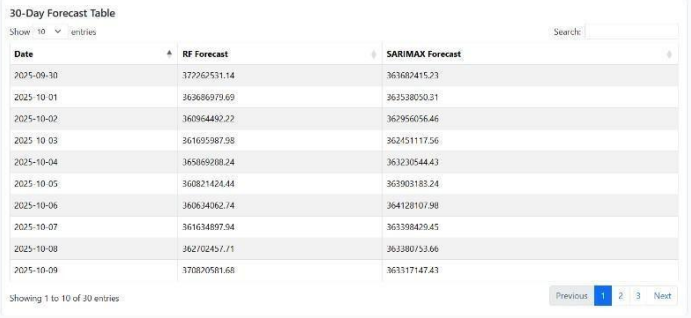

Fig 4: Traffic prediction for a specific day.

3.1 Comparative Analysis

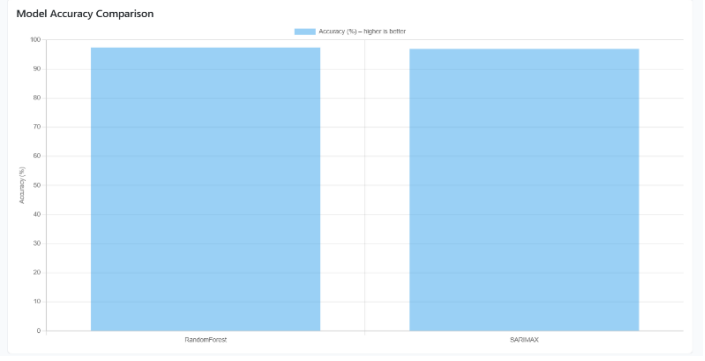

Chart 3.1: Model comparison between SARIMAX and Random Forest.

4. CONCLUSION

This project successfully uses SARIMAX and Random Forest models to forecast web traffic based on the past 30 days of data. The system helps website administrators predict future traffic, allowing better resource management during peak times. While SARIMAX captures seasonal trends well, Random Forest handles sudden changes effectively. The combination of both models, along with features like accuracy comparison and data download, makes the forecasting reliable and user-friendly.

REFERENCES

[1] Fatima, S. S. W. — "A Review of Time-Series Forecasting Algorithms for Manufacturing / Industrial Applications" (MDPI, 2024).

[2] Z. Liu, Z. Zhu, J. Gao, and C. Xu, ''Forecast methods for time series data: A survey,'' *IEEE Access*, vol. 9, pp. 91896--91912, 2021.

[3] "A Comparative Study of SARIMA, Random Forest, and ..." — International Journal / IETA (Dec 25, 2024).

[4] Zenodo / Hossain — "Comparative Analysis of ARIMA, SARIMAX and Random Forest models" (2023 / dataset: GDP example).

[5] AlSaleem, N. Y. A. — "Network Traffic Prediction Based on Time Series Modeling" (2023 PDF / conference article).

[6] Kane et al. — "Comparison of ARIMA and Random Forest time series models" (BMC / comparative study; background & methodology; classic reference updated 2020s summaries).

BIOGRAPHIES

Manu S.K is a final year Computer Science and Engineering student. She is currently pursuing her degree at S G Balekundri Institute of Technology.

Renuka R.U is a final year Computer Science and Engineering student. She is currently pursuing her degree at S G Balekundri Institute of Technology.

Ruchita R.P is a final year Computer Science and Engineering student. She is currently pursuing her degree at S G Balekundri Institute of Technology.

Sadiya Md.D is a final year Computer Science and Engineering student. She is currently pursuing her degree at S G Balekundri Institute of Technology.