PREDICTIVE ANALYSIS OF VARIOUS STATISTICAL MODELS USED IN CLOUD ENVIRONMENT

Akrati Sharma[1*], Priti Maheshwary[2]

Abstract

In the current era of the Digital Industrial Revolution, also known as Industry 4.0 or 4IR, the digital world has access to a vast amount of data, including data from mobile devices, social media platforms, businesses, the Internet of Things (IoT), cyber-security systems, health records, etc. To design and build appropriate cloud-based smart and automated applications and conduct an intelligent predictive analysis of these data to determine the possibilities of future outcomes based on historical data, data patterns, and the useful insights from data to make educated predictions about future events or trends. The key to anticipating such an analysis is having knowledge of machine learning (ML), artificial intelligence (AI), and statistical models. Predictive analysis is carried out using a number of statistical models in cloud computing circumstances. Some of these can be best fitted using state space models, SARIMA (seasonal ARIMA), exponential smoothing models, ARIMA models, and so on. In order to promote data-driven decision-making in the dynamic and changing landscape of cloud technologies throughout this digital industrial revolution, the study examines the advantages, disadvantages, and application of these models. The research provides insight into the effectiveness and applicability of various statistical methodologies, enabling decision-makers to make well-informed decisions for improving workload management, resource allocation, and operational efficiency in cloud environments.

Keywords- ARIMA, SARIMA, Statistical Models, Predictive Analysis

[1*]Research Scholar, RNTU, Bhopal, Akratisharma999@Gmail.com
[2]Proffesor, RNTU, Bhopal, pritiaheshwary@gmail.com

***Corresponding Author:** Akrati Sharma
*****Research Scholar, RNTU, Bhopal, Akratisharma999@Gmail.com

DOI: - 10.53555/ecb/2022.11.12.273

INTRODUCTION

In today's data-driven world, the ability to anticipate future outcomes and trends has become a foundation of decision-making across various domains. Predictive analysis become powerful methodology rooted in data science, offers a systematic approach to unlocking insights from historical data and patterns to forecast potential future events. This practice connects the potential of data, statistical algorithms, and advanced machine learning techniques to determine the likelihood of forthcoming outcomes. With its capacity to inform strategic planning, resource allocation, and operational efficiency, predictive analysis holds a significant promise, particularly within the dynamic realm of cloud computing environments.

Cloud computing has revolutionized the way organizations manage and deploy their IT resources. The dynamic and elastic nature of cloud technologies presents both opportunities and challenges in terms of optimizing resource utilization, effectively managing workloads, and ensuring operational robustness. Amidst this backdrop, the integration of predictive analysis techniques takes on a critical role, offering a data-driven lens to understand, anticipate, and adapt to the complex dynamics inherent in cloud environments.

The exploration of predictive analysis in the perception of cloud computing environments, examining the selection of statistical models that drive this predictive method. Among these models are the renowned ARIMA (Auto Regressive Integrated Moving Average) model, the insightful Exponential Smoothing Models, the encompassing SARIMA (Seasonal ARIMA) approach, and the versatile State Space models, among others. Through an in-depth analysis of these models, this research endeavors to unravel their individual strengths and weaknesses, illuminating the diverse contexts in which they find applicability.

As cloud technology continues to evolve, decision-makers handle with the complexities of managing resources and optimizing performance. In this search, the findings of this study promise to cast a spotlight on the performance metrics and suitability factors of various predictive analysis models. Such insights are poised to empower decision-makers and cloud administrators with the information needed to make informed choices that lead to resource efficiency, effective workload management, and ultimately, enhanced operational outcomes.

The predictive modeling examines the underlying mechanisms, historical effectiveness, and potential for adaptation within cloud computing environments. By encouraging a deeper understanding of predictive analysis, aspires to contribute to the ongoing discourse surrounding data-driven decision-making in cloud technologies.

BACKGROUND STUDY

Cloud computing has emerged as a transformative paradigm in the field of information technology, providing flexible and scalable resources for various applications. Predictive analysis of statistical models has gained significance in cloud environments, enabling effective resource allocation, workload prediction, and performance optimization. This literature survey aims to explore the contributions of various researchers in the realm of predictive analysis using statistical models within cloud computing.

Liu et al. [1] in their study, Liu and colleagues delve into predictive analysis of cloud resource performance utilizing statistical models. The authors focus on leveraging linear regression and autoregressive integrated moving average (ARIMA) models to forecast cloud resource behavior. This work highlights the potential of statistical techniques in predicting cloud resource performance, thereby aiding in efficient resource provisioning and management.

Zhu et al. [2] investigate predictive analysis of cloud workload through the lens of machine learning models. Their study employs decision trees and random forests to anticipate cloud workload patterns. By doing so, the authors emphasize the role of machine learning in predicting workload fluctuations, enabling enhanced capacity planning and resource allocation.

Armbrust et al. [3] presents a comprehensive overview of cloud computing, laying the groundwork for understanding the challenges and opportunities in cloud environments. This landmark study underscores the necessity of predictive analysis in optimizing resource utilization, managing workloads, and addressing scalability concerns within cloud computing.

Jan et al. [4] focus on predictive analysis of virtual machine performance in cloud environments. Their research utilizes time series analysis to predict virtual machine behavior. Through an adaptive threshold-based approach, the authors propose a means of anticipating virtual machine performance, thereby enhancing overall cloud system efficiency.

Mishra et al. [5] provide a comprehensive survey of predictive analytics in cloud computing using machine learning techniques. This study explores a range of machine learning algorithms, highlighting their applications in predicting diverse cloud-related phenomena, including workload patterns, resource utilization, and performance metrics.

Gaur et al. [6] compare the predictive capabilities of statistical and machine learning techniques in cloud performance. Their research evaluates methods such as autoregressive integrated moving average (ARIMA) and support vector machines. By examining the accuracy of these models, the authors contribute to the understanding of the efficacy of different predictive techniques in cloud computing.

Wang et al. [7] delve into predictive analysis of cloud computing performance through the lens of machine learning. Their study employs neural networks and regression analysis to forecast cloud performance metrics. By harnessing the power of machine learning, this work emphasizes the potential of predictive modeling to optimize cloud resource allocation.

Liu et al. [8] explore predictive analysis of cloud computing resource utilization using machine learning. The authors propose a framework that utilizes machine learning algorithms to forecast resource utilization patterns. This study contributes to the advancement of resource management techniques in cloud environments.

Gupta et al. [9] work focuses on predictive analysis of cloud workloads using time series analysis. The authors leverage time series techniques to predict cloud workload patterns. By anticipating workload changes, this research enhances cloud resource allocation and provisioning strategies.

Kumar et al. [10] conduct a comparative study of predictive analysis models for cloud resource management. They investigate autoregressive integrated moving average (ARIMA), long short-term memory (LSTM) networks, and random forests. This comparative analysis provides insights into the strengths and limitations of different predictive models within cloud computing.

ARIMA

ARIMA [1,6] is a popular time-series forecasting model that is used to analyze and forecast data points based on their temporal patterns. It combines three components: Auto Regressive (AR), integrated (I), and Moving Average (MA). The model is capable of capturing both auto regressive and moving average relationships within a time series, as well as handling trends and seasonality through differencing.

1. Auto Regressive (AR) component: This component models the relationship between a data point and its lagged values. That is, past values of the time series were used to predict the current value. The AR(p) term represents the order of the autoregressive component, where 'p' is an integer representing the number of lagged terms included in the model.

2. Integrated (I) Component: The I component involves differencing the time-series data to make it stationary. Being Stationary is important because many time-series models, including ARIMA, assume that the data have a constant mean and variance. The differencing order, ‘d’ represents the number of times the data needs to be differenced to achieve the property of being stationary.

3. Moving Average (MA) Component: The MA component models the relationship between the current value and past forecast errors (residuals). This accounted for the influence of past errors on the present value. The MA(q) term represents the order of the moving average component, where 'q' is an integer representing the number of lagged forecast errors included in the model.

Mathematical Formulation for the ARIMA(p, d, q) model can be expressed as follows:

$$(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p)(1-L)^d Yt=c+(1+\theta_1 L+\theta_2 L2+\ldots+\theta_q L^q)\ \varepsilon t$$

Yt is the value of the time series at time 't'.

L is the lag operator.

$\phi_1, \phi_2 \ldots \phi_p$ are the autoregressive coefficients.

d is the differencing order.

c is a constant term.

$\theta_1, \theta_2 \ldots \theta_q$are the moving average coefficients.

ε_t is

the white noise error term at time 't'.

It’s quite a hazardous task to choose an appropriate model for performing tasks. Choosing suitable values for ‘p’,‘d’, and ‘q’ requires understanding the data and using techniques like autocorrelation and partial autocorrelation plots. These plots help identify the potential values for ‘p’ and ‘q’, while the differencing order ‘d’ is determined by the number of differencing steps required to make the data stationary. After determining the parameters, the model can be fitted to the data using techniques like maximum likelihood estimation.

SARIMA

The SARIMA model is an extended version of the ARIMA model that includes additional components to handle seasonal patterns in time-series data. It stands for the seasonal autoregressive integrated moving average. SARIMA [10,6] is particularly useful for time-series data that exhibit regular seasonal patterns such as monthly, quarterly, or yearly seasonality.

In addition to the AR, I, and MA components of the basic ARIMA model, SARIMA introduces three new components to account for the seasonal patterns: Seasonal Auto Regression (SAR), Seasonal Integration (SI), and Seasonal Moving Average (SMA).

1. Seasonal Auto Regression (SAR) Component: The SAR component models tell relationship between a data point and its lagged values at the same seasonal lag. This is similar to the AR component, but it operates on the data at seasonal intervals. The SAR (P, s) term represents the order of the seasonal autoregressive component, where 'P' is an integer representing the number of seasonal lagged terms and's' is the length of the seasonal pattern.

2. Seasonal Integrated (SI) Component: Similar to the 'I' component in ARIMA, the SI component involves differencing the data to make it seasonal-stationary. This helps remove the seasonal trend from the data. The SI(D, s) term represents the order of the seasonal differencing component, where 'D' is an integer representing the number of seasonal differencing steps and 's' is the length of the seasonal pattern.

3. Seasonal Moving Average (SMA) Component: The SMA component models the relationship between the current value and the past forecast errors at the same seasonal lag. It accounts for the influence of past errors on the present value in a seasonal context. The SMA(Q, s) term represents the order of the seasonal moving average component, where 'Q' is an integer representing the number of seasonal lagged forecast errors and 's' is the length of the seasonal pattern.

The mathematical formulation for SARIMA (p, d, q)(P, D, Q, s) model can be expressed as follows:

$$(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p)(1-\phi_1 L^s-\phi_2 L^{2s}-\ldots-\phi_P L^{Ps})(1-L)^d (1-L^s)^D Yt = c+(1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q)(1+\theta_1 L^s+\theta_2 L^{2s}+\ldots+\theta_Q L^{Qs})\,\varepsilon t$$

Where all terms are similar to the ARIMA model, with the addition of the seasonal terms and parameters ϕ_1 $\phi_2,\ldots,$ ϕ_P , θ_1 $\theta_2,\ldots,$ θ_Q for the seasonal autoregressive and seasonal moving average components.

Selecting appropriate values for the SARIMA parameters involves analyzing seasonal autocorrelation and partial autocorrelation plots in addition to the non-seasonal plots. These plots help identify potential values for 'P', 'D', 'Q', as well as the seasonal order 's'. As with ARIMA, determining the right model parameters and fitting the model to the data requires understanding the characteristics of the time series and using techniques like maximum likelihood estimation. SARIMA is a powerful tool for handling both non-seasonal and seasonal time series patterns, making it a valuable choice for a wide range of forecasting tasks.

Exponential Smoothing Models

Exponential Smoothing models are a class of time series forecasting methods that rely on weighted averages of past observations to make predictions about future data points. These models are particularly useful for data with no clear trend or seasonality and are widely used for short-term forecasting. There are three main types of Exponential Smoothing models: Simple Exponential Smoothing, Double Exponential Smoothing (Holt's method), and Triple Exponential Smoothing (Holt-Winters' method).

1. Simple Exponential Smoothing (SES): Simple Exponential Smoothing is suitable for time series data[10] without any trend or seasonality. It calculates the forecast for the next time period as a weighted average of the most recent observation and the most recent forecast. The weight given to the most recent observation and it decreases exponentially over the time.

The mathematical representation of SES is:

$F_{t+1}=\alpha\cdot Y_t+(1-\alpha)\cdot F_t$

$\mathbf{F_{t+1}}$ is the forecast for the next time period.

$\mathbf{Y_t}$ is the actual value at time t.

$\mathbf{F_t}$ is the forecast for the current time period.

$\boldsymbol{\alpha}$ is the smoothing parameter ($0 <= \alpha <= 1$), which determines the weight given to the most recent observation.

2. Double Exponential Smoothing (Holt's Method): Double Exponential Smoothing extends SES to account for data with a linear trend. It introduces a second smoothing parameter to estimate the trend component. This method is suitable for time series data that exhibit a consistent trend but no seasonality.

The mathematical representation of Holt's Method:

Forecast equation:

$F_{t+h}=\alpha Y_t+(1-\alpha)(F_t+T_t)$

Trend equation: $T_t=\beta(F_t-F_{t-1})+(1-\beta)T_{t-1}$

$\mathbf{T_t}$represents the trend component at time t.

$\boldsymbol{\beta}$ is the smoothing parameter for the trend ($0 <= \beta <= 1$).

3. Triple Exponential Smoothing (Holt-Winters' Method): Triple Exponential Smoothing is an extension of Holt's method that also accounts for seasonality in addition to trend. It includes three components: level (average), trend, and seasonality. This method is suitable for time series data with both trend and seasonality.

The mathematical representation for Holt-Winters' Method,

Level equation: $L_t=\alpha Y_t+(1-\alpha)(L_t-1+T_t-1)$

Trend equation: $T_t=\beta(L_t-L_{t-1})+(1-\beta)T_{t-1}$,

Seasonal equation: $S_{t+m}=\gamma(Y_t-L_{t-1}-T_{t-1})+(1-\gamma)S_t$

Forecast equation: $F_{t+m}=L_t+m\cdot T_t+S_{t+m}$.

$\mathbf{S_t}$represents the seasonal component at time t.

$\boldsymbol{\gamma}$ is the smoothing parameter for seasonality ($0 <= \gamma <= 1$).

m is the number of time periods in a season.

Selecting the appropriate smoothing parameters (α, β, γ) depends on the characteristics of the data and can be done using techniques like grid search or optimization algorithms that minimize forecast errors. Exponential Smoothing models are relatively simple and efficient for short-term forecasting, but they may not perform well on complex time series data with irregular patterns or long-term trends.

State Space models

State space models, also known as dynamic linear models, are a flexible and powerful framework for modeling and forecasting time series data. They are particularly useful for handling complex time series patterns, such as trends, seasonality, and irregularities, while providing a formal structure to incorporate external factors and perform statistical inference.

In a state space model, the underlying process generating the observed data is represented by two main components: the state equation and the observation equation. The state equation describes how the unobserved (latent) state variables evolve over time, while the observation equation links the state variables to the observed data. The state equation describes the evolution of the latent state variables. It's often modeled as a linear combination of the previous state plus some noise or innovation. This equation captures the temporal dynamics of the underlying process, including trends and seasonality. The observation equation links the state variables to the observed data. It describes how the observed data are generated from the latent state variables, typically through a linear or nonlinear relationship. This equation includes terms for the measurement error or noise, accounting for discrepancies between the observed and predicted values.

State space models often make use of the Kalman filter and smoother algorithms to estimate the latent state variables and their uncertainty given the observed data. The Kalman filter operates in a recursive manner, updating estimates of the state variables as new observations become available. The Kalman smoother, on the other hand, provides a estimate of the state variables, taking into account all observations. State space models can be adapted to a wide range of time series patterns and complexities. For instance, they can accommodate time-varying parameters, seasonality, multiple observed variables, and incorporate external predictors or exogenous variables. State space models are often estimated within a Bayesian framework, allowing for the incorporation of prior information and producing posterior distributions for the state variables and model parameters. This is particularly useful for uncertainty quantification and making probabilistic forecasts.

Comparative Study of Predictive Methods

As mentioned in table 1.1 the various parameters have been taken to compare mentioned predictive methods. The choice of the appropriate method depends on the characteristics of your data and the specific forecasting or modeling goals. Each method has its strengths and limitations, and the best choice will vary depending on the context of data analysis.

Model	Purpose	Components	Suitable for	Parameter Estimation	Seasonality Handling
ARIMA	Time series forecasting	AR, I, MA	Data with no seasonality or trend	Estimation through auto-correlation and partial auto-correlation plots	Can be challenging
SARIMA	Time series forecasting	AR, I, MA, SAR, SI, SMA	Data with seasonality and trend	Similar to ARIMA, extended for seasonality parameters	Handles seasonality
Exponential Smoothing	Short-term forecasting	-	Data with no trend or seasonality	Manual selection or optimization	Limited to exponential trends
SES	Short-term forecasting	Level	Data with no trend or seasonality	Smoothing parameter (α)	-
Holt's Method	Short-term forecasting	Level, Trend	Data with linear trend	Smoothing parameters (α,β)	-
Holt-Winters' Method	Short-term forecasting	Level, Trend, Seasonal	Data with trend and seasonality	Smoothing parameters (α,β,γ)	Handles seasonality
State Space Models	Time series modeling/forecasting	State Equation, Observation Equation	Complex time series patterns	Bayesian approach, Kalman Filter	Flexible for various patterns

Table 1.1 Comparative studies of various predictive methods

Conclusion

In time series forecasting and modeling techniques, several powerful methodologies stand out, each helps to address specific data characteristics and analytical objectives. The comparative analysis of ARIMA, SARIMA, Exponential Smoothing models (including SES, Holt's, and Holt-Winters'), and State Space models reveals a diverse array of tools available for handling different time series scenarios.

For datasets with no visible trend or seasonality, ARIMA and Exponential Smoothing models, particularly Simple Exponential Smoothing (SES), offer straightforward options. ARIMA excels at identifying autoregressive and moving average relationships, while SES provides quick and simple forecasts for data with only a level component. Incorporating seasonality into the analysis, SARIMA and Holt-Winters' Method prove their utility. SARIMA expertly handles data with both trend and seasonality, while Holt-Winters' Method extends to incorporate level, trend, and seasonal components. These models are particularly well-suited for tasks requiring precise predictions in the presence of multiple patterns.

Exponential Smoothing models, especially Holt's Method and Holt-Winters' Method, shine in the short-term forecasting arena. These models can capture linear trends and seasonal variations effectively, making them excellent choices for situations where accurate immediate forecasts are crucial.

For more intricate time series data characterized by complex patterns and non-linear behaviors, State Space models offer a versatile framework. Leveraging the power of the Kalman filter, State Space models can elegantly accommodate changing dynamics, external influences, and various sources of uncertainty. This flexibility comes at the cost of increased complexity in parameter estimation and modeling.

In conclusion, the selection of the most suitable modeling technique hinges on a deep understanding of the underlying data characteristics, the presence of trends and seasonality, and the analytical goals. By carefully considering these factors, analysts can leverage the strengths of these methodologies to achieve accurate and insightful time series forecasts and models. Whether opting for the robustness of ARIMA, the seasonality-handling prowess of SARIMA, the simplicity of Exponential Smoothing, or the adaptability of State Space models, the right choice empowers better decision-making and more accurate predictions in the complex world of time series analysis.

References

[1] Liu, Fang et al. "Predictive Analysis of Cloud Resources Performance Using Statistical Models." Proceedings of the 2017 IEEE International Conference on Cloud Computing Technology and Science (Cloud Com), 2017.

[2] Zhu, Jing et al. "Predictive Analysis of Cloud Workload Using Machine Learning Models." 2018 IEEE International Conference on Cloud Computing Technology and Science (Cloud Com), 2018.

[3] Armbrust, Michael et al. "Above the Clouds: A Berkeley View of Cloud Computing." University of California, Berkeley, 2009.

[4] Jan, Arif et al. "Predictive Analysis of Virtual Machine Performance in Cloud Computing

Environments." International Journal of Advanced Computer Science and Applications (IJACSA), 2020.

[5] Mishra, Deepti et al. "Predictive Analytics in Cloud Computing Using Machine Learning Techniques." 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2018.

[6] Gaur, Shweta et al. "Performance Prediction of Cloud Computing Using Statistical and Machine Learning Techniques." 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019.

[7] Wang, Xinheng et al. "Predictive Analysis of Cloud Computing Performance Using Machine Learning." 2017 6th International Conference on Advanced Computing (ICo AC), 2017.

[8] Liu, Jian et al. "Predictive Analysis of Cloud Computing Resource Utilization Based on Machine Learning." 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2018.

[9] Gupta, Sourabh et al. "Predictive Analysis of Cloud Workloads Using Time Series Analysis." 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), 2016.

[10] Kumar, Harshit et al. "A Comparative Study of Predictive Analysis Models for Cloud Resource Management." 2021 International Conference on Sustainable Computing and Intelligent Systems (ICSCIS), 2021.