International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advanced Time Series Modeling in Digital Payments: Harnessing Seasonal Patterns for Enhanced Forecasting

Chinni Krishna Abburi[1*]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: In today's digital payments landscape, accurate forecasting of future trends has become essential. By obtaining reliable forecasts, organizations can gain numerous benefits: they can detect and prevent fraud, enhance operational efficiency, improve customer retention, optimize marketing campaigns, and deliver a more personalized customer experience. Given these advantages, companies are increasingly focusing on predictive analytics, making it crucial to select the right model for analyzing data effectively. In this study, we trained and tested five different models to evaluate their performance and efficiency on a digital payment's dataset, which exhibits strong seasonal trends. Through this approach, we aim to determine the most suitable model to support strategic decision-making and drive business success in the competitive digital payments sector.

Keywords:*Time Series, Prophet, Exponential smoothing, SARIMA, XGBoost, LSTM, Model Evaluation.*

1. Introduction

The rapid growth of digital payments is transforming the global financial landscape. With approximately two-thirds of the world's adult population now using digital payment methods, the market has reached an estimated $9.47 trillion in 2023. This impressive growth is expected to continue, with projections indicating an 11.79% annual growth rate that could see the market reach $14.79 trillion in the coming years.

As digital payments become increasingly prevalent, organizations in this domain are turning to analytics to enhance their services and gain a competitive edge. Among various analytical tools, forecasting has emerged as a critical focus area. Effective forecasting allows companies to anticipate market trends, consumer behavior, and potential challenges, enabling them to make informed decisions and develop proactive strategies [4, 6].

The importance of forecasting in the digital payments industry cannot be overstated. It helps organizations optimize resource allocation, improve fraud detection, enhance customer experiences, and refine marketing strategies. However, the rapid increase of forecasting models has created a new challenge: selecting the most appropriate model for specific datasets and business needs.

One of the key factors that significantly impacts forecasting in the digital payments sector is seasonality. Payment patterns often fluctuate based on various seasonal factors such as holidays, pay cycles, and annual events. This seasonality can greatly affect the accuracy and reliability of forecasting models, making it crucial for organizations to choose models that can effectively capture and account for these patterns.

To address these challenges various models have been developed and adopted by digital payments industry.

1.1. Prophet Model:

The Prophet model is designed for time series forecasting, particularly suited for data with strong seasonality and multiple seasons of historical observations. It incorporates both weekly and yearly seasonality and adjusts for holiday effects, making it effective for seasonally driven forecasts. Initially developed to handle periodic patterns, Prophet has since been extended to manage more complex seasonal trends, providing robust forecasting capabilities in time series analysis.

1.2. Exponential smoothing:

Exponential smoothing is a forecasting method that predicts future values by applying weighted averages to past observations, prioritizing more recent data [1]. This method is particularly useful for short to medium-term forecasts in dynamic environments like digital payments, where recent trends significantly affect outcomes.

1.3. SARIMA:

Autoregressive Integrated Moving Average (ARIMA) is a widely used time series forecasting method that captures trends and patterns in non-seasonal data through autoregression, differencing, and moving averages. Seasonal ARIMA (SARIMA) extends this model to handle seasonal patterns, making it suitable for time series data with both trends and regular seasonal variations, such as monthly or quarterly cycles.

[1] *Business Intelligence Engineer, Visa Inc., TX, USA*
ORCID ID:0009-0007-9495-0558
**** Corresponding Author Email: abburichinnikrishna@gmail.com***

1.4. Extreme Gradient Boosting (XGBoost):

XGBoost is a high-performance machine learning algorithm that uses gradient boosting techniques [2]. It's widely favored in data science and machine learning competitions for its efficiency and flexibility, especially when working with structured data. While XGBoost excels in classification and regression tasks, it's versatile enough to handle various supervised learning problems [5].

1.5. Long Short-Term Memory (LSTM):

LSTM is a type of Recurrent Neural Network (RNN) architecture designed to address the vanishing gradient problem that traditional RNNs face when dealing with long-term dependencies in sequential data.

2. Analysis

This study examines five forecasting models—Prophet, Exponential Smoothing, SARIMA, XGBoost, and LSTM—applied to a dataset from digital payment transactions consisting of over 1,000 records with a notable seasonal effect on sales volume. Accurate forecasting in such datasets necessitates a robust preprocessing approach to enhance model performance. In this analysis, data preprocessing was carried out to address several critical aspects. Null values were identified and handled, outliers were detected and treated, and unnecessary columns were removed, retained only columns relevant to the forecasting task. Duplicates were eliminated to ensure data integrity, and data formatting was adjusted according to the requirements of each model. Through comprehensive data cleaning and preparation, this study aimed to optimize the performance of each model, providing a foundation for precise evaluation metrics and an accurate assessment of forecasting effectiveness in digital payment transactions.

This study primarily evaluates the performance of various forecasting models applied to transactional data, focusing on four key aspects: accuracy, computation efficiency, adaptability, and scalability. To assess the model's accuracy, we employ a comprehensive set of performance metrics:

2.1. Mean Squared Error:

Mean Squared Error (MSE) is one of the most important performance indicators of the models we have used in this study. It measures the averaged squared difference between the actual value and the predicted value.

2.2. Root Mean Squared Error:

Root Mean squared Error (RMSE) is the square root of MSE. It is used to calculate the average absolute error between the actual value and the predicted value. Both MSE and RMSE give insights into how well our model is performing, but RMSE would be preferred because it is easy to interpret RMSE compared to the actual values, as both are in the same units.

2.3. Mean Absolute Error:

Mean Absolute Error (MAE) measures the average difference between the model's predicted value and the actual value.

2.4. Mean Absolute Percentage Error:

Mean Absolute Percentage Error (MAPE) measures the absolute percentage difference between actual and predicted values.

Beyond accuracy, we analyzed each model's:

2.5. Computation Efficiency:

Measuring the speed and resources required for training

and making predictions.

2.6. Adaptability:

Evaluating how the model can adjust to the changes in the data patterns and any new information.

2.7. Scalability:

Determining if the model could maintain the same performance with the increasing amount of data. We aim to comprehensively understand how different forecasting models perform when applied to transactional data, considering both their predictive accuracy and practical implementation considerations. The above-multifaceted approach allows for a more nuanced comparison of the models, helping to identify the most suitable techniques for forecasting scenarios in the context of digital payments transaction analysis.

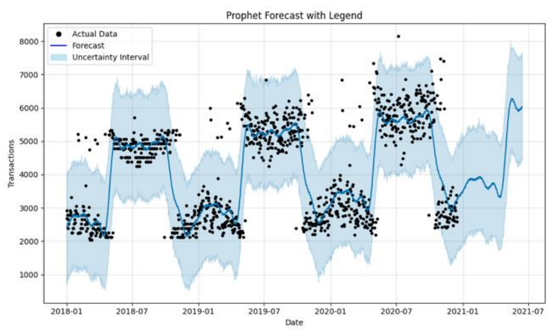

Fig. 1. Prophet Model Forecast

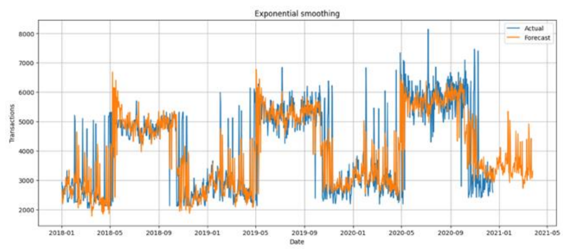

Fig. 2. Exponential Smoothing Model Forecast

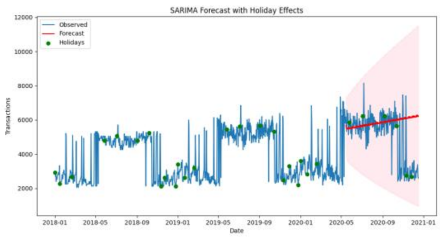

Fig. 3. SARIMA Model Forecast

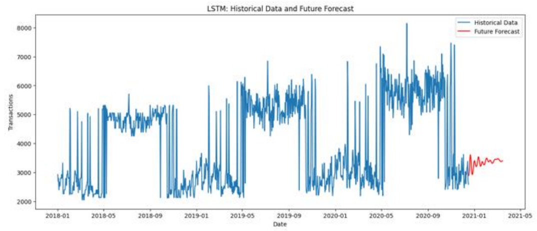

Fig. 4. LSTM Model Forecast

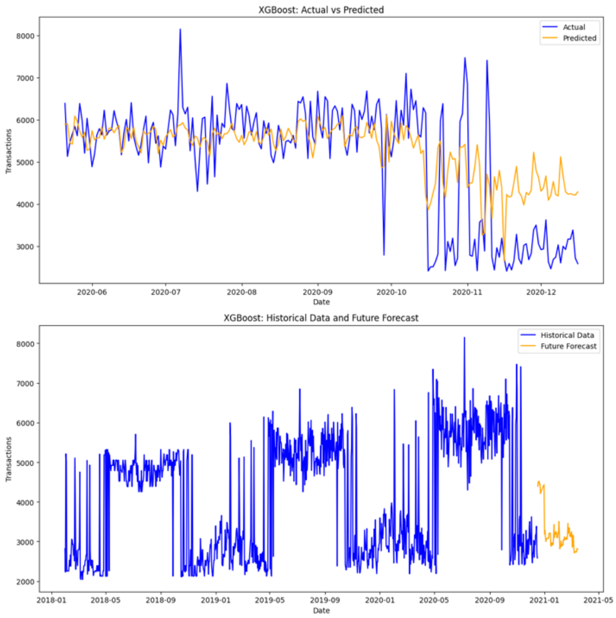

Fig. 5.XGBoost Model Forecast

Table. I Evaluation Metrics for Various Forecast Models.

Model	Model Evaluation Metrics		
	RMSE	***MAE***	***MAPE***
Prophet	702.11	506.84	14.18%
Exponential smoothing	814.18	528.24	14.60%
SARIMA	888.52	598.23	16.61%
XGBoost	900.34	602.12	16.97%
LSTM	952.81	780.34	20.93%

As shown in Table I, we assessed various time series forecasting models for digital payments data, with the Prophet model standing out as the best performer. It achieved the lowest Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), indicating strong accuracy. The higher RMSE than MAE reflects its sensitivity to significant errors, while the low MAPE suggests that the model performed well in this context.

Prophet excels in computational efficiency, allowing for rapid fitting without extensive preprocessing, which is crucial in the fast-paced digital payments industry. Its scalability enables it to handle millions of observations and support parallel processing, making it suitable for high-volume transaction data. Additionally, Prophet easily adapts to trend changes and incorporates holiday effects and custom seasonality patterns, enhancing its applicability for digital payment datasets. Overall, Prophet's combination of accuracy, efficiency, scalability, and adaptability makes it an ideal choice for forecasting in the digital payments sector, enabling better decision-making.

Figure 2 illustrates the seasonal patterns identified in the dataset, demonstrating that the results obtained using Exponential Smoothing closely align with those of the Prophet model, as shown in Figure 1. The values of RMSE and MAE are considered good for the target variable picked in the dataset. Also, like the Prophet model, RMSE is higher than MAE, indicating that RMSE penalized large errors heavily. The MAPE value for the model is also in a good range, making the model reliable for datasets that have strong seasonality effects.

Exponential smoothing has high computational efficiency, almost like that of Prophet. However, it has moderate scalability and adaptability, as we observed that the model struggled with very large datasets and with non-linear trends and irregular data.

The SARIMA model, as presented in Figure 3, displays a trend similar to those observed with the Prophet and Exponential Smoothing models. The RMSE value was higher than the MSE, and these values provided confidence that this model performed similarly to the previous models. We observed that the MAPE value increased slightly, but it remained within an acceptable range. SARIMA demonstrates moderate computational efficiency and scalability. It shows great adaptability with data that have seasonal and trend patterns but is very limited in handling non-linear patterns. We observed a similar pattern for model evaluation metrics throughout all the models we have analyzed in the study. XGBoost, as shown in Figure 5, has a higher RMSE compared to MSE, while its MAPE falls within an acceptable range. With different datasets, XGBoost showed high scalability and high computational efficiency. It is highly adaptable with both linear and non-linear datasets and can effectively handle external factors.

Of all the models, LSTM showed deviating results. Although the trend of having RMSE greater than MAE continued, the MAPE value is greater than 20%, indicating that there needs to be some improvement in the model. In this study, we tried to fine-tune hyperparameters, incorporate additional relevant features, increase data size, and consider different preprocessing techniques. Of all the models, LSTM has very limited adaptability and poor usage of computational resources. From Table I and Figure 4 We observed that the model struggled with small datasets, making it a very low-scalability model.

3. Conclusion

This study aimed to analyze various time series forecasting models, focusing on datasets with strong seasonality effects. Each model was evaluated using specific performance metrics, revealing unique characteristics suited to different types of data. The analysis demonstrated that each model offers distinct advantages and disadvantages. An appropriate model should be chosen based on several factors, such as data characteristics, including size, complexity, and seasonality or trends. It also depends on business requirements and model adaptability to change patterns and the trade-off between the model's complexity and performance.

References

[1] Ostertagova, Eva & Ostertag, Oskar. (2012). Forecasting Using Simple Exponential Smoothing Method. Acta Electrotechnica et Informatica. 12. 62–66. 10.2478/v10198-012-0034-2.

[2] V. Gupta, P. Kumar, and A. Kumar, "A Comprehensive Review on XGBoost: Application and Performance Evaluation," Journal of King Saud University - Computer and Information Sciences, 2023.

[3] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

[4] Vibhu Verma, "Exploring Key XGBoost Hyperparameters: A Study on Optimal Search Spaces and Practical Recommendations for Regression and Classification", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 12, Issue 10, October 2024.

[5] Kolambe, Milind. (2024). Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques. Journal of Electrical Systems. 20. 575-586. 10.52783/jes.1478.