ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/karthika-gopalakrishnan/72

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science
https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1

*Original Article*

# Account Payables Prediction using Time Series Modelling Techniques: ARIMA and SARIMA in Financial Industry

**Karthika Gopalakrishnan***

Data Scientist, USA

**Citation:** Karthika Gopalakrishnan. Account Payables Prediction using Time Series Modelling Techniques: ARIMA and SARIMA in Financial Industry. *J Artif Intell Mach Learn & Data Sci 2022*, 1(1), 222-224. DOI: doi.org/10.51219/JAIMLD/karthika-gopalakrishnan/72

**Received:** 02 February, 2022; **Accepted:** 08 February, 2022; **Published:** 20 February, 2022

***Corresponding author:** Karthika Gopalakrishnan, Data Scientist, USA

**Copyright:** © 2022 Gopalakrishnan K., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Forecasting time series data using historical values has been gaining momentum in recent days. Forecasting is used in various industries like Finance, Healthcare, Energy and Consumer Analytics. Many financial institutions have leveraged forecasting mechanisms to predict their cash flows in future for arriving at effective management decisions. For cash flow analysis, this paper focusses on Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA Models in Time Series Forecasting. The models are trained on the historical payables data and used to predict the payables in the test data.

**Keywords:** Statistical methods, Time series forecasting, Payables prediction, ARIMA, SARIMA, Financial services

## 1. Introduction

Corporate sectors are harnessing the power of Machine Learning algorithms to enhance decision-making in management. One of the highly sought-after applications in the financial sector is Account Payables Prediction, enabling businesses to anticipate and strategize impending financial responsibilities. Precise forecasting of accounts payable provides organizations with crucial insights into their data, enabling informed financial decisions and preventing unexpected expenditures. This accuracy in Payables Prediction empowers management to monitor expenses effectively and plan for the upcoming fiscal year. The modeling of payable transactions involves employing stochastic time series modeling techniques

The time series forecasting method proves particularly valuable when limited information exists about the impact of explanatory variables on the output. In this

Modeling approach, the dependent variable or output relies solely on its historical values. Once a model is established, it becomes a tool for predicting future values. Time series methods have undergone extensive study in the field of forecasting and continue to see ongoing enhancements. Among the well-explored models, ARIMA (Autoregressive Integrated Moving Average) stands out, widely acknowledged for its simplicity in implementation and utilization of the box-Jenkins methodology.

In this paper, ARIMA models, both the seasonal and non - seasonal variations, have been studied to predict the daily account payables using historical data. The dataset is preprocessed, and the historical data is trained using ARIMA models to predict the payables in future. Finally, the performance of the model is judged by necessary error analysis.

## 2. Literature Review

ARIMA is among the simplest yet popular statistical methods that have been applied to time series model. ARIMA popularity can be credited to the effort of Box and Jenkins. ARIMA model is developed by integrating two different forms of linear regressions, the Autoregressive (AR) and the Moving Average (MA). AR (p) and MA (q) model is expressed as.

Here, a1,…,ap, and m1,…, mq are the parameters of autoregressive and the moving average portions respectively; the constant in the expression is denoted by c. p and q represents the order of respective AR and MA portions. White noise is denoted as et. In equation (2) ut, ut − 1,…,ut − q represents white noise (error) terms. Expectation of y(t) is represented by μ. Integrating the two models represented by (1) and (2) using same data of the training set, the ARIMA (p, q) becomes.

Here p represents the autoregressive and q represents the moving average terms. For this model the prime requirement is to make the time series data statistically stationary in terms of mean, variance, and autocorrelation. In case the data displays nonstationary, the data must be differenced. This helps to convert the data into a stationary time series data resulting in ARIMA (p,d,q) where d refers to the degree of differencing.

The account payables show the obvious periodic transactions resulting from seasonal changes. These seasonal changes can be dealt with using the SARIMA model. This model is formed by incorporating a seasonal term in the ARIMA model. The SARIMA model in the general form is expressed in the form SARIMA (p,d,q) (P,D,Q)S. In this expression p signifies the order of autoregressive while P signifies the order of seasonal autoregressive. Order by difference is represented by d and D represents the seasonal difference. The term q and Q represent the order of MA and SMA respectively. S represents the seasonality. The seasonal portion of the model will be depicted by the seasonal lags of the correlation function plots.

## 3. Methodology

### 3.1. Data preprocessing

The real time Account Payable transactions provided by the City of Oxnard Finance Department for the fiscal year 2021 and preceding four fiscal years are considered for this paper. The data is available at https://data.oxnard.org/Financial-Data-and-Reports/Accounts-Payable-Transactions/ay6a-gmt6/about_data. The dataset contains daily payables transactions, check number, vendor name, description, and invoice number. **Table 1** shows the overview of the data present in the dataset. The account payables are shown as a time series using python in **Figure1.**

**Table 1:** Overview of account payables.

| | Date | CheckNum | VendorName | Amount | Description | AccountNumber | InvoiceNumber | AcctPerYr | AcctPerMo |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 08/25/2014 | 9999998 | MISCELLANEOUS VENDOR CX | 1373.43 | LID-560 POCKET READER 4 A | 101-2106-802.81-36 | 544 | 2015 | 2 |
| 1 | 07/09/2020 | 373810 | THE WHARF | 134.80 | SAFETY BOOTS | 101-5701-805.81-05 | 86037 | 2019 | 7 |
| 2 | 07/22/2020 | 374297 | E RECYCLING OF CALIFORNIA | -529.54 | E WASTE | 631-6301-842.82-09 | 87656 | 2020 | 8 |
| 3 | 06/25/2020 | 373524 | MOTOR VEHICLE NETWORK | 600.00 | BEVERAGE CONT RECYCLING | 631-6828-823.82-09 | 20-36202 | 2020 | 9 |
| 4 | 06/18/2020 | 373208 | HARRIS WATER CONDITIONING | 40.00 | ENGINEERING WATER | 101-3201-803.81-02 | 47638 | 2020 | 9 |

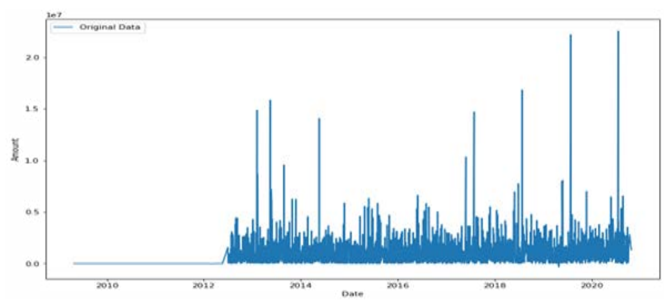


**Figure 1:** Account payables transactions as time series plot.

The payables are grouped for each date irrespective of vendor name and invoice number for time series modelling and the overview of the cumulative transactions per day are shown in **Table 2.**

**Table 2:** Overview transactions grouped by date.

| | Date | Amount |
|---|---|---|
| 0 | 2009-04-28 | -175.00 |
| 1 | 2009-06-02 | -39.93 |
| 2 | 2010-01-19 | -10.00 |
| 3 | 2010-03-16 | -44.58 |
| 4 | 2010-05-18 | -10.00 |

The time series signal is tested for stationarity using Augmented Dickey-Fuller Test (ADF) test using "adfuller" command in python kernel. The ADF statistic and p value for the signal is calculated as -7.967 and p-value as 2.846440640523167e-12. The p-value for the signal is less than 0.05 and the signal is considered stationary. There is no differencing needed to convert the signal into a stationary signal for ARIMA Modelling. The time series signal is considered as additive and the decomposition of the signal is shown in the **Figure 2**

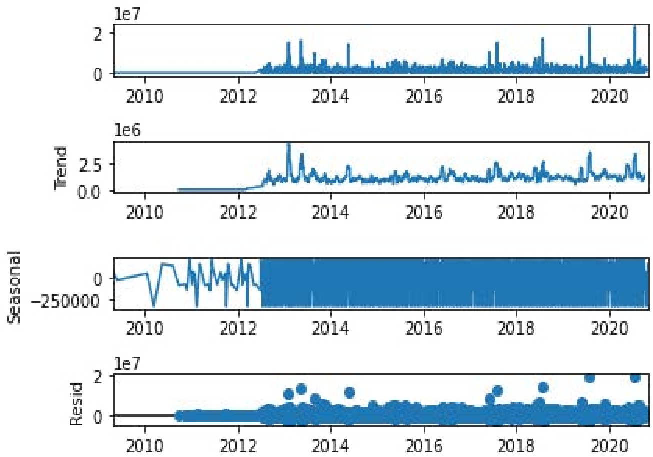


**Figure 2:** Seasonal decomposition of the data.

### 3.2. Model selection

The process of model identification involves establishing the values of the parameters p (for autoregressive components) and q (for moving average components) in both ARIMA and SARIMA models. Various methods are documented in the literature for ascertaining these parameters. In this investigation, the ACF and PACF plots have been employed to discern the optimal parameter values. The model parameters are deduced through a meticulous examination of the correlation graphs (ACF and PACF), enabling a precise estimation. The ACF and PACF plots for the data is shown in **Figure 3** and **Figure 4**. For our case, we find out, p = 5 and q=1 while using our data set. We use the value of differencing parameter d = 0.

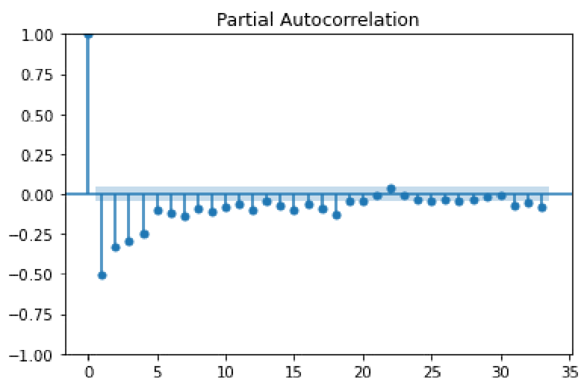


**Figure 3:** PACF plot of the time series data.

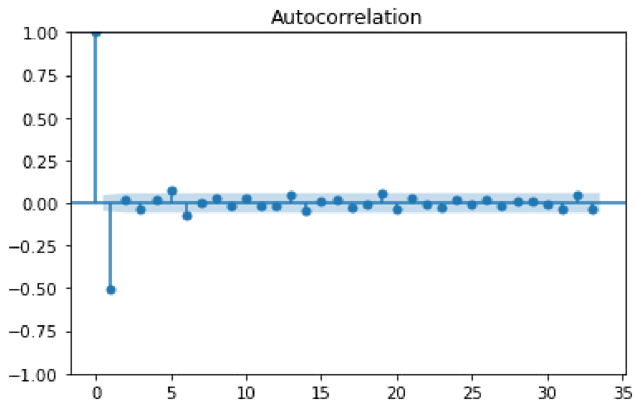


**Figure 4:** ACF plot of time series data.

### 3.3. Performance evaluation

The correctness of the model was evaluated using model accuracy measures like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

## 4. Prediction Results

Arima (5,1,0) model was used for forecasting the payables. The dataset is split into 80% train data and 20% as test data. The model was trained on historical train data and the trained ARIMA model was used to predict the values on the test data. The RMSE and MAE values calculated for the ARIMA (5,1,0) was 2241610.56 and 1253773.99 respectively. **Figure 5** shows the payables predicted by the model for the test data. The test data of payable transactions were used to forecast using the best selected SARIMA models among the ten models to find the best fit model. **Figure 6** shows the forecast results (best possible) using the model SARIMA (1 1 1) and the RMSE for the model was calculated as 2098710.87.

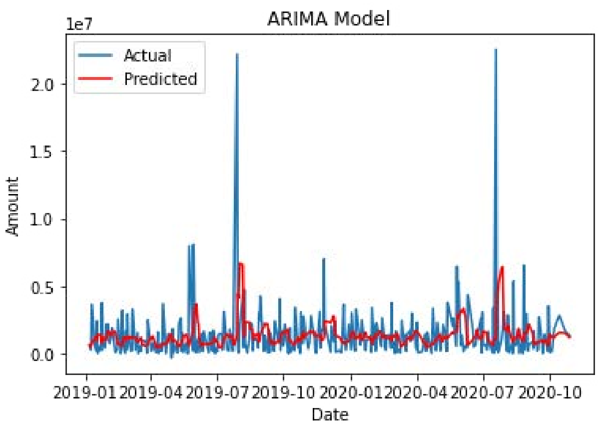


**Figure 5:** ARIMA (5,1,0) model forecast results on test data.

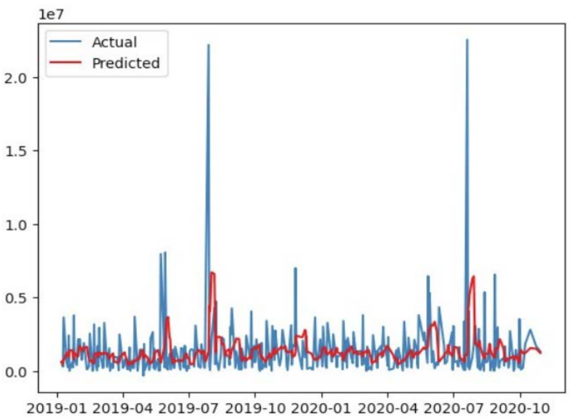


**Figure 6:** SARIMA (1,1,1) model forecast results on test data.

## 5. Conclusion

Account payables transactions available for the City of Oxnard were taken for this case study. Two types of models were applied on the data for forecasting the future values: ARIMA model and SARIMA model. Their performances were evaluated. The conclusion is based on the results obtained from the experiments conducted for this study. Based on the model evaluation metrics, better forecasting results are obtained using SARIMA model. Hence, SARIMA model would be the best fit in this case. However, the account payables transactions forecasting model can be further improved by identifying the outliers in the transactions and utilizing the other modelling techniques. The author intends to incorporate these factors in the further study of account payables forecast.

## 6. References

1. Martin T Hagan, Suzanne M Behr. The time series approach to short term load forecasting. IEEE Transactions on Power Systems, 1987; 2: 785-791.
2. D Saxena, SN Singh, KS Verma. Application of computational intelligence in emerging power systems. International Journal of Engineering, Science and Technology, 2010; 2.
3. Kakoli Goswami, Aditya Bihar Kandali. Electricity demand prediction using data driven forecasting scheme: ARIMA and SARIMA for real-time load data of assam. In: 2020 International Conference on Computational Performance Evaluation. (ComPE) North-Eastern Hill University, Shillong, Meghalaya, India, 2020.
4. Box G E P, Jenkins G M. Time series analysis: Forecasting and control. Journal of the American Statistical Association, 1970; 68: 199-201.
5. Yanming Yang, Haiyan Zheng, Ruili Zhang. Prediction and analysis of aircraft failure rate based on sarima model. 2nd IEEE International conference on Computational Intelligence and Application, 2017.