IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY
ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 4)
Available online at: https://www.ijariit.com

Over-dues forecasting using ARIMA Technique

A. Kalyan Aravind Kumar
aravindmcs1@gmail.com
Indian School of Business, Hyderabad, Telangana

ABSTRACT

The work presented in this paper establishes an enrichment in modeling and forecasting over-dues for Beverages manufacturing company. A time-series modeling technique used to forecast over-dues for ABinBEV (Beer manufacturing company). Our work demonstrates how historical over-dues data utilized to predict future over-dues. The historical over-dues information used to develop several Autoregressive Integrated Moving Average (ARIMA) models by using Root mean squared error (RMSE) and the most suitable ARIMA model found to be ARIMA (2, 1, 0). and validation performed by comparing the accuracy of the models with three types of accuracy criteria, which are Mean square error (MSE), Root Mean Squared Error (RMSE), and Mean absolute error (MAE).

Keywords: *Forecasting, Time Series Modeling, ARIMA*

1. INTRODUCTION

1.1 About Company

Anheuser-Busch InBev SA/NV, commonly known as ABInBev, is a Belgian-Brazilian multinational drink and brewing company based in Belgium. ABInBev has a global functional management office in New York City. It has around 630 beer brands in 150 countries. The annual businesses for the company in 2019 were US$52.3 billion. The ABInBev had realized US$45.5 billion in revenue in 2016. For any FMCG organization dealing with vendors and distributors also receiving the payments for the invoices is one of the biggest concerns for ABInBEV; there are 14000 customers. On the one hand, if the invoices not received on the due date, the organization's cash flow reduces, and operating day-to-day business will be difficult. On the contrary, if the company takes punitive measures for defaulters, then it may hit the business relationship, volume, and prospects from the customer. The only sensible middle-path is to follow-up regularly for the overdue invoices. But with 14000 customers, following up becomes a daunting task for the company. So ABInBEV has to have a systematic approach to forecast every month:

a) What will be the overdue amount – so that they can plan the budget accordingly
b) Identify which customers are most likely to default - so that they can closely monitor and follow-up with them, thereby reducing overdue and increasing cash flow.

1.2 Objectives

Develop a monthly forecasting model that will have 95% accuracy. To help the business identify the Customers & reason for Overdue so that they can follow-up with the defaulted customers and improve cash flow.

1.3 Data Received

Open invoices data from Jan'14 to Dec'18, yearly data for total business and total payments and data for high-risk customers and their location.

1.4 Proposed Solution

In this paper, an attempt made to forecast over-dues for the next leading year. And the model explained for forecasting is an Autoregressive Integrated Moving Average (ARIMA) model. The primary motivation for choosing the ARIMA model in this study for the forecasting is because this model considers non-zero autocorrelation between the successive values of the time-series data.

2. LITERATURE REVIEW

Khalid Yunus et al. [1], in his paper, the writer explained a revised Auto-Regressive integrated moving average (ARIMA) modeling technique, which could apprehend time correlation and possibility distribution of measured wind-pace time-collection records offered. The method includes frequency decomposition (splitting the wind-speed data into a high frequency (HF) and low-frequency (LF) components) and checking further to differencing and energy conversion that used within the ARIMA modeling system.

Vaccaro A et al. [2] the paper suggests a hybrid design for power price forecasting. The proposed architecture consolidates the reward of the easy-to-use and relatively simple-to-tune Auto-Regressive Integrated Moving Average (ARIMA) model and the approximation endowment of local learning techniques. The architecture is robust and more reliable than the standard forecasting methodologies. This design combines a reliable built-in linear model (ARIMA) with an adaptive dynamic corrector (Lazy Learning algorithm). The corrector model sequentially updated to accommodate the whole architecture according to different market conditions. Comprehensive simulation investigations show the effectiveness of the suggested hybrid learning methods for forecasting prices for the electricity market, Ontario, Canada.

Guoqiang Liu et al. [3], in software program reliability, understanding the nature of the failure records is excellent learning. Generally, these records identified by utilizing the testing software or testing methods. But this author proposed a hybrid variant for ordinary and lengthy software program failure time forecasting on this paper. The hybrid version consists of two ways. One is the Singular Spectrum Analysis (SSA) technique, and the second one is ARIMA. In this version, the time series of software failure method initially disintegrates into numerous sub-series matching to some tendentious and oscillation (periodic or quasi-periodic) segments and noise by using the usage of SSA after which every sub-series anticipated through the best ARIMA version. Lastly, a correction system is directed for the sum of the forecast results to ensure the residual to be random natural extracts.

Takaomi HIRATA et al. [4], the Time series records measure and prediction may be crucial to perceive nonlinear phenomena. Linear Models, Auto Regression(ARIMA), and nonlinear models consisting of Multilayer Perception (MLP) are well-known. a Deep Belief Net (DBN) using a specific version of Restricted Boltzmann machines (RBMs) changed conventional methods dimensionality reduction methods. In this paper, the author explained how DBNs were first used to time series forecasting systems by their original studies.

Ling Wang et al. [5], in this paper, the author described the new method called "the accumulative wear prediction method of metro wheels based on the ARIMA(p,d,q)" based on the study of the measured wear data and the wear characteristics. This project executed on the dataset that consists of "wheels of Guangzhou Metro Line 1". As per the definition of the time series modeling method of the ARIMA(p,d,q) model, the first author described the stationarity analysis. Then he continued on the transformation of the metro wheel wear data. After he also explained the application of the AIC criterion and the Maximum Likelihood Estimation method. Finally, the forecasting for the flange thickness and the diameter of the metro wheels conducted.

Theresa Hoang Diem Ngo et al. [6], the time-series data is a set of values of a distinct variable that occurs over time in a specific pattern. The usual traditional patterns are increasing or decreasing trend, cycle, seasonality, and unpredictable fluctuations. To model a time-series situation as a function of its past values, analysts generally recognize the pattern assuming that the trend will continue in the future. This paper highlighted how to identify an appropriate time series model by matching patterns by plotting the sample graph of Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF).

Eliete Nascimento Pereira et al. [7], in the forecast of time series, it has been common to assume that a particular predictive method, such as an Auto-Regressive Integrated Moving Average (ARIMA) model which generates residuals like white noise. To correct these noises, the model should produce forecasts with more accuracy. This paper puts ahead a Wavelet Hybrid Forecaster (WHF) that combines the numerical model techniques such as wavelet decomposition, ARIMA models; Artificial Neural Networks (ANNs); and the linear combination of forecasts. The WHF forecasts take into account the data from the frequency present in the given time series using the Wavelet Components (WCs) acquired by the wavelet decomposition approach.

W. Jacobs et al. [8], This research aims to forecast the values of the time series of UHT milk demand in a dairy industry by consolidating forecasting of ARIMA and MLP/RNA models and examine the effects to the individual models, explaining the combined forecast for the stock planning. Eight predictions merging techniques used and the outcomes attained by fitting the ARIMA and MLP/RNA templates compared with the results collected in the proposed combinations. The results revealed that the combination of SARIMA models and DMLP, the inverse mean square method, gave an excellent prediction.

3. ARIMA FORECASTING MODEL

Time series forecasting is a multidisciplinary systematic tool used to solve prediction problems. Its implementation is easy and adjustable because it requires only historical observations of the necessary variables. The common equation of successive differences at the d_{th} difference of X_t is as follows:

$$\Delta^d X_t = (1 - B)^d X_t$$

Where d is the difference of order and usually is 1 or 2, and B is the backshift driver.

The successive difference at one-time lag equals to,

$$\Delta^1 \mathrm{X_t} = (1 - \mathrm{B})\mathrm{X_t} = \mathrm{X_t} - \mathrm{X_{t-1}}$$

The common ARIMA (p, d, q) is concisely expressed as follows [9]:

$$\emptyset_p(B)W_t = \theta_q(B)e_t$$

Where $\emptyset_p(B)$ is an auto-regressive operator of order p, $\theta_q(B)$ is a moving average operator of order q and $W_t = \Delta d X_t$.

ARIMA model developed using Matlab R2012a (7.14) software; this software also used to prepare the data and calculate the ACF and PACF; Data construction incorporated the removal of outliers, treatment of zero readings, and interpolation of missing data. Model performance evaluated using Root Mean Square Errors (RMSE) and the coefficient of determination (R2).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(X_t - X_0)^2}{n}}$$

Where X_t is the forecasted observation, and X_0 is the actual observation.

4. DATA PREPARATION AND EXPLORATION

The monthly invoice data were processed to find customer-wise, general-ledger wise commercial loss (> 90 days) and overdue (< 90 days) amount as on the month-end. For High-risk customers, the full amount of an invoice, irrespective of it is due or not, have been considered to be overdue.

Table 1: All possible of ARIMA models with its RMSE

Data	Numbers
Total number of accounts	14212
Accounts with at least 1 overdue	7784
Accounts with no overdue	6428
customers have been continuously overdue	4
customers have intermittent data points	7780

5. TIME SERIES ANALYSIS AND BUILDING ARIMA

Developed the forecasting model using the provided dataset. Figure 1 below represents the line plot of over-dues for AbinBEV from 2014 to 2018.

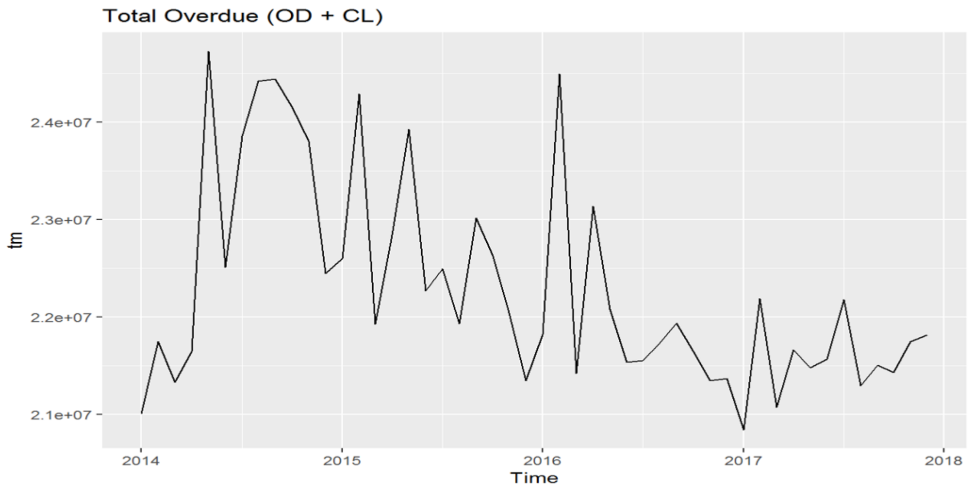

Fig. 1: Total Amount of Over-dues and Commercial Loss from 2014 to 2018.

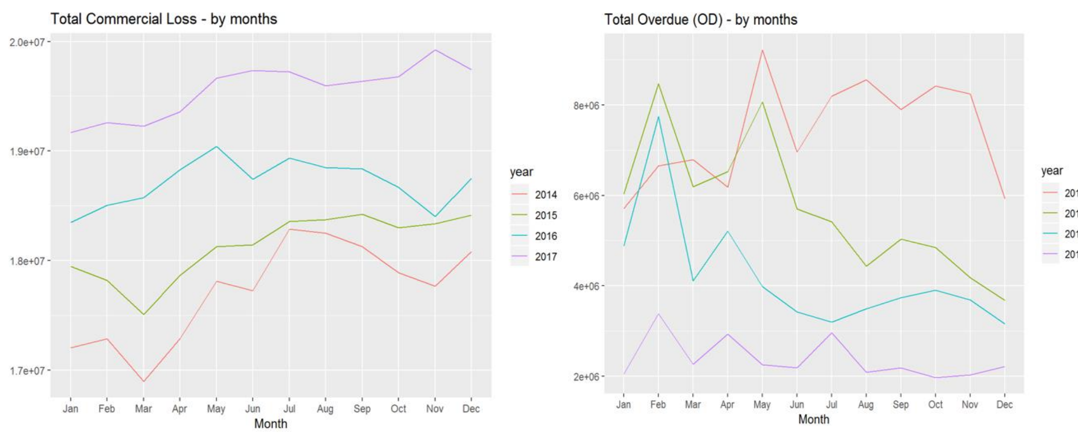

Fig. 2: Total Amount Over-dues and Commercial Loss by month.

To obtain the suited ARIMA model, RMSE is chosen as criteria since many ARIMA models can be established for one column of the dataset based on using different values of p,d, and q. Hence, the proper ARIMA model which has less RMSE. The following table will show all the feasible fitted ARIMA models with its RMSE.

Table 2: All possible of ARIMA models with its RMSE

ARIMA	RMSE	ARIMA	RMSE	ARIMA	RMSE
(0 , 0 ,0)	Not fitted	(0 , 2 ,0)	2.7	(1 , 1 ,2)	1.9
(0 , 0 ,1)	2.2	(0 , 2 ,2)	Not fitted	(1 , 2 ,0)	Not fitted
(0 , 0 ,2)	2.1	(1 , 0 ,0)	2.09	(1 , 2 ,1)	3.0
(0 , 1 ,0)	Not fitted	(1 , 0 ,1)	2.6	(1 , 2 ,2)	3.1
(0 , 2 ,0)	Not fitted	(1 , 0 ,2)	Not fitted	(2 , 0 ,0)	Not fitted
(0 , 1 ,1)	1.8	(1 , 1 ,0)	1.6	(2 , 0 ,1)	2.7
(0 , 1 ,2)	1.9	(1 , 1 ,1)	1.8	(2 , 0 ,2)	2.1
(2 , 1 ,0)	**1.4**	(2 , 1 ,2)	2.1	(2 , 2 ,1)	1.5
(2 , 1 ,1)	2.0	(2 , 2 ,0)	Not fitted	(2 , 2 ,2)	1.56

Based on Table 1, below pointed were noted:

- The values of p,d, and q are between 0 and 2. because these values unlikely to be in the minus; also, these values should not be more than 2 since the estimation of the parameters will be insignificant.
- RMSE values are between 1.40 and 3.10 based on the dataset used. Hence, according to the software, the best-fitted model is ARIMA (2, 1, 0) with RMSE =1.4 for the given dataset.
- In some cases, the ARIMA model is not feasible, which means that the estimation may not be possible for the dataset, then it should be ignored.

6. CONCLUSION

This paper presented a new design that shows short-term predicting using the latest technology method. For the given dataset, an ARIMA model built over the dataset achieved to improve short-term predictions. Implemented the new technique in the case of Beverages manufacturing company data, verified its accuracy, and confirmed its performance capabilities. Around 14000 observations were collected to achieve the forecasts. And the best ARIMA model was selected based on the essential criteria, which is RMSE. Another important consideration is that the ARIMA model's forecasting accuracy diminishes gradually at this stage of the growth process, from period to period. Alternatively, Long short-term memory (LSTM) can be used to model univariate time series forecasting problems. The diminishing accuracy problems comprised of a single series of observations. The model is required to learn from the past data observations to predict the next value in the sequence.

7. REFERENCES

[1] W. Jacobs and A. M. Souza – "*Combination of Box-Jenkins and MLP/RNA Models for Forecasting Combining Forecasting of Box-Jenkins*", IEEE LATIN AMERICA TRANSACTIONS, 14(4), 2016.

[2] E. G. Kardakos, C. K. Simoglou, G. I. Chouliaras, S. I. Vagropoulos, and A. G. Bakirtzis, "*Identification of SARIMAX, SARIMA, Revised SARIMA and ANN-based Models for short-term PV formation forecasting*", IEEE International Energy Conference (ENERGYCON), Leuven 1-6, 2016.

[3] Rodrigo N. Calheiros; Department of Computers & Information systems, University of Melbourne, Melbourne, Australia. Rajkumar Buyya; Rajiv Ranjan; Enayat Masoumi, "*The Workload forecasting using the current time series ARIMA Model. And the model implications on cloud applications*", QoS IEEE Transactions on Cloud Computing, 3(4), 2015.

[4] Takashi KUREMOTO, Takaomi HIRATA, Masanao OBAYASHI, and Shingo MABU, "*New technique for Time Series Predictions using (Deep belief net) DBN and ARIMA*", International Conference on Computer Application Technologies, IEEE, 2015.

[5] D. Zhang, G. Liu, and T. Zhang, "*Software Reliability Forecasting using Singular Spectrum Analysis (SSA) and particular ARIMA Hybrid* model. The Theoretical Aspects of Software Engineering (TASE), 2015 International Symposium on *Nanjing*, 111-118, 2015.

[6] Sornpon Wichaidit, Surin Kittitornkun, "*SET50 Stock Prices prediction using the new model CARIMA*", (ICSEC) - International Computer Science and Engineering Conference, 2015.

[7] K. Bhattacharya, T. H. M. EL-Fouly, C. A. Cañizares, and *A. Vaccaro, "The new technique ARIMA Adaptive Hybrid - Local Learning Architecture for Hourly Electricity Price Forecasting*", PowerTech, 2015 IEEE Eindhoven, 1-6, 2015.

[8] Peiyuan Chen, Torbjorn Thiringer, and Kalid Yunus, "*Frequency-Decomposed Modeling based on ARIMA for Wind Speed Time Series* data", IEEE Transaction on Power Systems, 31(4), 2546-2556, 2016.

[9] Younes, M.K.; Nopiah, Z.M.; Basri, NEA; Basri, H. Medium-term municipal solid waste generation prediction by autoregressive integrated moving average. Stat. Oper. Res. Int. Conf. (SORIC 2013) 2014, 427–435. [CrossRef]

