Global Scientific JOURNALS

GSJ: Volume 12, Issue 7, July 2024, Online: ISSN 2320-9186

www.globalscientificjournal.com

TIME SERIES MODEL TO FORECAST CHURCH GROWTH.

By Anthony, Kain Komando, PhD Student, University of Port-Harcourt, Nigeria.

Supervisors Prof B.O.Omijeh and Dr. U.A Okengwu

Abstract This present generation has witnessed unprecedented advancement in the field of Information and Communication Technology (ICT) so much such that even the small medium enterprises are getting their businesses online. The church which is one of the fulcrums of the society is not also left out in the wake of this technological development. Many churches are more and more getting their activities online to get a wider coverage in order to fulfil their operational mandate of reaching out to the whole world.

As part of my contribution to this growing search, my study is aimed at designing a time series model to forecast church growth in terms of attendance and financial income using machine learning. The time series models that will be considered will cut across the Exponential Smoothing Models consisting of Single, Double and Triple Exponential Smoothing, the Autoregressive Integrated Moving Average (ARIMA), the Seasonal eXogenus ARIMA (SARIMAX) and the Prophet. These models are well established and widely used in predictions. While the Exponential Smoothing Models are good at capturing short term changes in trends and seasonality which are key components in the time series decomposition. The ARIMA, SARIMAX which is an improved version of the ARIMA with eXogenus variables are good at capturing patterns, trends and seasonality of the data using a combination of the past values, differencing and errors.

Keywords:"forecast", "attendance", "financial income", " time series", " Exponential smoothing", "ARIMA", "SARIMAX "and "Prophet".

1. Introduction.

There is no doubt that we are in the world that is powered and sustained by Information and Communication Technology (ICT) . ICT has been identified as one of the basic pillars on which the modern society stands (Ayeni, Akinyemi and Lawal, 2020).Today, with the help of ICT almost everything is achievable ranging from completely automated

industries to AI human beings. Aluko (2014) rightly stated that with the current revolution in the ICT world, the possibility of all areas of the world being affected by it was very high.

Churches that had neglected the revolution in the ICT world were forcefully compelled to comply when the COVID laws restricting social gathering and limiting church attendance to not more than fifty persons per sitting had to resort to ICT tools to transmit their services to members who viewed from the comfort of their homes.

Numerical growth (attendance) and financial income are core factors that every church cannot do without taking them into consideration. Their statistics are extremely valuable. Knowledge of these factors enhance planning and resource optimization. It guides in allocating resources more efficiently ensuring that no department is neglected. It enables

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2350

GSJ© 2023

www.globalscientificjournal.comimproved congregational engagement through available statistics that are used for

planning and proper execution. Also, decisions on planning for future programs and allotment of resources are data driven and not just on mere assumptions.Being able to develop a software that can forecast these factors will go a long way to facilitate the numerous decisions the church as an organization will need to take in order to forestall any eventuality. By examining all these models, we will be able to get the best that can enhance the process of decision making which is a critical aspect of effective planning.

The models will need to be fitted with previous time series data collected from the church being used as a case study according to the timestamp. The data will be processed and subjected to different feature engineering. The outcome of the various models will shed light on the most appropriate one that can be used for financial income and attendance,

2, Scope of the Research.

Redemption Ministries Port Harcourt in Nigeria will be used as a case study for forecasting growthin terms of attendance and financial income for over a period of time. Machine learning approach will be used in the predictive analysis using any of the models listed above to establish the best fitted for the system.

3 Reviewof Related Works

Being able to effectively manage church growth in terms of attendance and financial income is crucial to every modern-day church. The use of Time series forecasting together with machine learning techniques improves the extent to which forecasting can be improved in order to enable church administrators plan and improve their establishments with accuracy and efficiency. This literature review attempts to summarize some most recent studies on this subject of time series forecasting models and the use of machine learning to develop a software that can accomplish this.

(Norris 2012) [1] stated that Venerable Bob Jackson was the first researcher to study the application of statistical data in the areas of church growth as far back as 2002. Bob (2002) [2] stated in his book titled “Hope for the Church”, looked at the church attendance figures he was able through analysis interpret the future of the church in terms of attendance and also made analysis for the causes of decline with strategic responses to combat it Bob (2002) ChristophorusBenedittoAdityaSatrio et al (2020) [3], researched on COVID-19 using ARIMA and Prophet. Records of all the deaths, confirmed cases and recovered cases were taken. Comparative analysis were ran on the models and made to forecast for the next thirty days. The Prophet model performed excellently with minimal variation from the actual data. Ayeni. J.A. et al (2021) [4] Ayeni. J.A et al (2021) did a research work on Church Growth Prediction using Predictive Analytic (Linear Regression Model) Technique. The complexity of religion has not given room for more research on the quantitative aspect of growth but has seen more work done on the qualitative aspect though the religious leaders have through such work gained more knowledge, they have not been able to measure growth by any of these processes in order to ascertain if growth is taking place or not. Norris (2012) stated that Venerable Bob Jackson was the first to study the application of statistical data in the areas of church growth. By the research, Norris (2012) quoted Venerable Bob Jackson that team ministries tend to have low growth in attendance with significant decline during vacations. According to Mathwork (2020), Machine Learning is a data analytic technique that teaches computers to do what comes naturally to animals and humans. This could be interpreted to mean learning from experience. (Mathworks, 2020, Weng et al 2017) further ascertained that Machine Learning algorithms are computational methods to ‘learn’ information directly from

© GSJ

www.globalscientificjournal.comdata without depending on a predetermined equation as a model. This learning approach is

further used to identify patterns in the unfolding data sets to make critical decisions and predictions. Ayeni et al 2021 used Linear Regression Model Technique with the AI as the engine room for the development. Their research consisted of harvesting historical attendance data from Light of Christ Church from 2014 to 2019. The data was collated for mining and analysis. The Predictive model was applied to predict growth till 2027. The model developed was able to make predictions. However, limitations from the fact the it cannot be integrated for use by the third part.

The metrics of evaluation will be the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE).

Evaluation Metrics.

The metrics of evaluation used was the Mean Absolute Percentage Error and the Mean Absolute Error as shown in equations (1) and (2)

MAE = (1/n) * Σ|yi – xi| [5]

(1)

where:

• **Σ:** A Greek symbol that means “sum”

yi: The observed value for the ith observation

• **xi:** The predicted value for the ith observation

• **n:** The total number of observations

for Mean Absolute Percentage Error

[6]

(2)

Where:

• *n* is the number of fitted points,

• At is the actual value,

• Ft is the forecast value.

• Σ is sum

Kindly note:

i)

By employing the MAE and MAPE which are performance metrics, it enables us to see how well the models are performing.

ii)

They are very useful in monitoring accuracy

iii)

The combination of these two gives credibility to the process.

iv)

They enable us to compare and evaluate the prediction accuracy.

3. Methodology

3.1. Data Collection

To build accurate predictive models, data collection is a crucial first step. Historical church attendance and financial income data will be gathered, including relevant time-stamped information.

The dataset was collected from Redemption Ministries in Port Harcourt, Nigeria from 2019 to 2022 The data was collected directly from the church's recordsfrom the administrative staff

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2352

GSJ© 2023

www.globalscientificjournal.comand finance department according to their time stamp. These records were stored in excel

format and later exported to the Jupyter notebook in python. The data were carefully checked for accuracy and completeness to ensure that they captured all relevant information. Any missing or incomplete data was obtained from other sources, such as the church's financial statements or attendance logs.

The data collected spanned a period of four years, from 2019 to 2022, and consisted of over 678 records. The raw data is presented in table 3. 1, which shows the attendance and account records obtained from Sunday services, Wednesday services, and Friday services, as well as the special services held during certain weeks.

Table 3.1: Extract of Raw dataset obtained from 2019 to 2022

3.2 Pre-processing.

Before building a time series model to predict the attendance and financial income of the church, it is important to pre-process the data to ensure that it is in a suitable format for analysis. This involves a number of steps, including data cleaning, normalization, and feature engineering. In this section, we describe the data pre-processing steps that were carried out on

the raw data obtained from the church between 2019 and 2022.

i)

Data Cleaning and Feature Engineering

The following were the steps taken to perform data cleaning and feature engineering:

a). Loading the data

The data was loaded using the Pandas Python library into a Jupyter Notebook.

Table 3.2 Extract from Jupyter Notebook

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2353

GSJ© 2023

www.globalscientificjournal.com**b). Rename columns and select the required columns.**

To ensure uniformity, the columns were all converted to lowercase. Since the interest is in the total attendance and total amount; the “total” and “amount” columns were selected along with the “date” column.

Table 3.3 Extract from Renaming of the columns and Selection of Required Columns

© GSJ

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2354

GSJ© 2023

www.globalscientificjournal.com**c). Create a new column to show the month/year of each record.**

Table 3.4 Extract from Jupyter Notebook showing Month/year feature Created

d). Fill null values

In the “total” feature, some null values exist. The null values were filled with numbers that had the same mean and standard deviation as the other values in the columns.

Table 3.5 Extract from Pandas showing Null Values filled in the Total Column

e) Drop the “date” column and order by the “month_year” column.

Since the new “month_year” column is available, there is no longer any need for the “date” column.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2355

GSJ© 2023

www.globalscientificjournal.comTable 3.6. Extract from Jupyter Notebook showing Unwanted Columns Dropped.

f). Visualization

The figure 3.2 below shows the visualization of the data after cleaning and pre-processing.

Figure 3.1: Visualization of the church's monthly average financial income.

Figure 3.2: Visualization of the church's monthly attendance.

3.3 Models and Procedures

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2356

GSJ© 2023

www.globalscientificjournal.comThe performance of six time series models was analysed. These include Single Exponential

smoothing, Double exponential smoothing, Triple exponential smoothing, ARIMA, SARIMAX and Prophet for forecasting future attendance and financial income.Each of the models has its peculiarities and specialization. SARIMAX in data modelling combines autoregressive, moving average and seasonal factors while ARIMA is best for combining autoregressive and moving average components. The Exponential Smoothing captures underlying trends and patterns and can account for sudden changes. The parameters of the ARIMA and SARIMAX models will be determined by plotting the ACF and PACF. With this, we will be able to check for stationarity and autocorrelation structures which are key factors. [7]. The metrics of evaluation to cross validate the authenticity of the performance of the models chosen are MAE and MAPE. These metrics have been explained in the literature. [8]

3.3.1 Plot ACF and PACF plots

Auto correlation and Partial autocorrelation plots graphically summarize the impact of observations at prior timesteps on the observations we are trying to predict. ACF plots give the q value while the PACF plots give the p value. In determining this, we look for the tail of pattern in either ACF or PACF. If the tail is crossing the blue region then it will give the potential p and q values. Where p is the number of lag observations included in the model also called the lag order and q is the size of the moving average window also called the order of moving average. The ACF measures and plots average correlation between data points and the previous values of the series measured for different lag lengths, the PACF controls any correlation between observations of a shorter lag length.

The following are python codes performed for the plotting of these functions.

3.4 Stationarity Checks

Stationarity check is critical in every time series analysis as it ascertains whether the statistical properties such as mean, variance and autocorrelation are constant. Non stationarity time series witnesses changes in these properties. The Dickey Fuller test from the statstool's library was used in testing for stationarity [9]. For the DFuller test, If test statistics < Critical Value and P-value <0.05. then the series is stationary. Where P is the number of lag observations included in the model also called the lag order. Below is an extract from the Jupyter Notebook to check for stationarity.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2357

GSJ© 2023

www.globalscientificjournal.com**3.5 Data Splitting.**

This is to enable us train the model to find a pattern in the data. while the test is used to evaluate the performance of the model on the trained dataset. The ratio used is 80:20 which is a common technique of dividing the training dataset into 80% for training of the data and 20% for the test set. This approach which is also employed in Machine Learning is critical to the performance of the model.

© GSJ

Extracts from the Python Jupyter Notebook for the splitting of the dataset.

3.6 Model Assessment

Choosing appropriate algorithms for time series prediction is crucial for the system's success. Two specialized time series machine learning algorithms will be carefully selected based on their effectiveness in handling univariate time series data. These algorithms will complement each other, covering various aspects of church attendance and financial income prediction.We shall use the Exponential smoothing algorithms, ARIMA, Prophet and SARIMAX plots to select the best fitted for attendance and financial income models. Below are extracts from the Jupyter Notebook analysing the various plots.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2358

GSJ© 2023

www.globalscientificjournal.comi)

For Financial Income Prediction

Extract of codes from Python Notebook.

ii)

For Attendance Prediction.

Extract of codes from Python Notebook.

After the execution of these codes, the following results were obtained.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2359

GSJ© 2023

www.globalscientificjournal.com**4. Results and Interpretation.**

4.1 Model Prediction for Financial Income.

A. Algorithms Performance Visuals for Financial income.

i)

Single Exponential Smoothing.

Figure 4.1 Graphical Illustration of the Single Exponential Smoothing Model on the

Test Data

ii)

Double Exponential Smoothing

Figure 4.2: Graphical illustration of the Double Exponential Smoothing Model on

Test Data

iii)

Triple Exponential Smoothing

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2360

GSJ© 2023

www.globalscientificjournal.comFigure 4.3: Graphical Illustration of the Triple Exponential Smoothing Model on the

Test Data.

iv)

ARIMA

Figure 4.4: Graphical Illustration of the AutoRegressive Integrated Moving Average

on the Test Data

v)

SARIMAX

Figure 4.5: Graphical Illustration of the Seasonal AutoRegressive Integrated Moving

Average on the Test Data

vi)

Prophet

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2361

GSJ© 2023

www.globalscientificjournal.comFigure 4.6. Graphical Illustration of the Prophet on the Test Data

B. Metrics Evaluation on the Performance of the Algorithms

Table 4.1 Metrics of Evaluation of the various Algorithms for the Financial Income Model.

C: Comparing the Results of the best fitted visuals and the lowest MAE and MAPE

From Table 4.1 and Figure 4.2, we see that the Double Exponential Algorithm has the lowest values and at the same time, Figure 4.2 appears best fitted hence we adopt the Double Exponential Smoothing for the Financial income prediction.

4.2 Model Prediction for Attendance.

A: Algorithms Performance Visuals for Attendance Model.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2362

GSJ© 2023

www.globalscientificjournal.com**i)**

Single Exponential Smoothing for Attendance

Figure 4.7 Graphical Illustration of the Single Exponential Smoothing on Test data

ii)

Double Exponential Smoothing for Attendance

Figure 4.8 Graphical Illustration of the Double Exponential Data on the Test Data

iii)

Triple Exponential Smoothing for Attendance

Figure 4.9Graphical Illustration of the Double Exponential Data on the Test Data

iv)

ARIMA for Attendance.

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2363

GSJ© 2023

www.globalscientificjournal.comFigure 4.10Graphical Illustration of the ARIMA on the Test Data

v)

SARIMAX

Figure 4.11 Graphical Illustration of theSARIMAX on Test Data

vi)

Prophet

Figure 4.12Graphical Illustration of the on the Prophet on Test Data

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2364

GSJ© 2023

www.globalscientificjournal.comB. Metrics Evaluation on the Performance of the Algorithms

Table 4.2 Metrics of Evaluation of the various Algorithms for the Attendance model

C: Comparing the Results of the best fitted visuals and the lowest MAEvalue for Attendance.

From Table 4.2 and Figure 4.9 we see that the Triple Exponential Algorithm has the lowest values and at the same time, Figure 4.9 appears best fitted hence we adopt the Double Exponential Smoothing for the Financial income prediction.

4. 3. hyperparameter tuning for the selected models

Figure 4.13. Double Exponential Smoothing for Financial Income

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2365

GSJ© 2023

www.globalscientificjournal.comFigure 4.14. Triple Exponential Smoothing for Attendance.

4 6 Model Deployment

The Python Flask web framework was used to deploy the prediction models. The selected models were exported as pickle files. APIs were designed to accept POST requests. The

request body is a JSON that contains the “month”, and “year” to be predicted. The following routes were defined in the API:

i) /predict- financial income

ii) /predict-attendance

The request JSON body follows this pattern:

{

“month”: month_to_predict,

“year”: year_to_predict,

}

5. Conclusion:

Church growth forecasting is a key component to the administrators as it helps in enhanced planning and resource optimization. By having clearer understanding of the attendance and financial income patterns as portrayed by the time series models, it becomes easy for administrators to allocate resources more efficiently, such as staffing, seating arrangements, and provisions for events and programs. This optimization not only enhances the overall experience for attendees but also minimizes unnecessary costs and wastage, contributing to a more sustainable and effective operation.

The models chosen are widely accepted for discrete time series and very good to forecast the immediate future.

6. Limitations.

Due to poor record keeping, we could only get financial income and attendance records from 2019 to 2022 which might have affected the efficiency of the models. The higher the dataset, the greater the impact on the projections. The models were selected based on the results obtained,

GSJ: Volume 11, Issue 11, November 2023

ISSN 2320-9186

2366

GSJ© 2023

www.globalscientificjournal.comForecast for both models lag behind with increase or decrease in trend over time. They also

do not account for dynamic changes that occur in practice and hence require constant updating.

Missing data resulting from inaccurate records were filled using statistical means and

variance. This could impact negatively on the overall

The workstation that was used is a laptop computer that does not have high computational powers and does not support Computer Unified Device Architecture (CUDA) that can fast track the computations. [10]

The results from all the models were obtained within a particular range of parameter consideration. However, we cannot ascertain if better results could be obtained from a wider range. This was due to the limitation hardware resources which also constrained the parameter range for selection thus causing a great limitation in models' ability to exploit its capability.

7. Future work.

The time series models used in this thesis are relatively few compared to the number of existing ones. further research can be conducted to have a large range of models for comparison. This will greatly enhance the prediction accuracy as better models might not have been included in this work.

The research focussed on univariate data. It has been established through research that the more the sources of data, the better the outcome of the predictions.

The possibility of deploying a hybrid model as proven in recent time series predictions is also recommended as such models take into consideration statistical models such as ARIMA and machine learning model such as Artificial Neural Network (ANN) for more efficient results. [11]

Limitations from the hardware could also be considered and improved versions used to see the outcome of the research.

7. Acknowledgment.

I sincerely acknowledge with deep gratitude the contributions of my supervisors Prof. B.O. Omijeh, Dr Mrs U.A. Okengwu, Dr Uchendu and Dr Mathew

References.

1. Kevin, Norris {2012} Church Growth Research Programme, Note by Kevin Norris,

Senior

Strategy Officer (Analyst) (Resource) Strategy and Development Unit.

Resourcing

Mission Bulletin, 2012

2. Bob, Jackson (2012) Hope for the Church, Church House Publ. ISBN: 97807155516,

published https://www.chpublishing .co.uk/books/97807155516/hope-for-the-church.

3. C. B. Aditya Satrio et al .'Time series analysis and forecasting of coronavirus disease

in

Indonesia using ARIMA model and PROPHET.' *Procedia Computer Science* vol,

179,

pp 524-532, Jan 2021. Doi 101016ij press 2021.06.036

GSJ© 2023

www.globalscientificjournal.com4. Ayeni J. A. et al. (2021), Church Growth Prediction using LRM Technique. Source

Computer

Engineering and Intelligent System ISSN 2222-1739 (paper) ISSN 2222360 Vol. 12,

No 1, 2021.

5. How to Calculate Mean Absolute Error in Excel (Step-by-Step) (statology.org)

6. Mean Absolute Percentage Error (MAPE) - Statistics How To

7 Adhikari, R; Agrawal, R.K. An introductory study on Time Series Modelling and

Forecasting

arXiv:1302.6613.

8 https://www.en.wikipedia.org/Mean_Absolute_error

9 Tutorials Point. Time Series Tutorials 2019. https://www.tutorialspoint.com/time.htm

(accessed on 21 October 2023)

10. 'CUDA Toolkit', July 2013. https://developer.nvidia,com/cuda-toolkit page 52

11. F.M.Tseng, H.C Yu and G.G Tzeng. ' Combining Neural Network model with

seasonal time series ARIMA' Technological Forecasting and Social Change. Vol. 69 no.1 pp.

71-87 Jan 2002 doi 10:1016/50040-1625(00)00113-X. Available online

http://www.sciencedirect.com/science/article/pi/S004016250000113X (Page 54)

© GSJ