JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Implementation of the SARIMA Statistical Algorithm for Sales Prediction in a Nigerian Medium Scale Enterprise

Balogun G. B., Muhammed S. M., Babatunde A. O., Hussein, S. O.
Department of Computer Science,
University of Ilorin, Nigeria

ABSTRACT

A sales prediction is a projection of future sales figures for a business over a given period of time. Forecasts are predicated on conjectures drawn from examining the trends or behavior of historical sales data. Sales forecasting has an impact on most of the factors that lead to impact optimization, profit maximization, and cost minimization. An efficient forecasting system can help a business raise profits, increase equipment utilization, reduce inventory, and achieve more changeability. But the majority of medium-sized businesses, like the Nigerian fast-food sector, have not yet adopted this technology or business model, which results in poor planning and ultimately, early closure or bankruptcy. This led to the research's objective, which is to apply the statistical method to anticipate sales in a medium-sized fast-food company. Data collection, data exploration, and statistical model creation were used to accomplish this. The machine learning model was constructed using the statistical approach known as Seasonal Autoregressive Integrated Moving Average, or SARIMA. To use its features, the model was made available as a web application. The outcome demonstrates that, if implemented, it would allow Nigeria's fast-food industry, as well as any other medium-sized industry in the nation, to make well-informed projections, maximize profits and as a result guaranteeing their survival in the competitive business environment of the developing world.

ARTICLE INFO
Article History
Received: June, 2024
Received in revised form: June, 2024
Accepted: July, 2024
Published online: September, 2024

KEYWORDS
Statistical Algorithm, SARIMA, Sales Prediction

INTRODUCTION

According to the 2010 survey report on micro, small, and medium enterprises (MSME), the Nigerian industrial sector uses an average of 56.4% of its capacity to produce a variety of goods (Nwokocha, V. C., Nwankwo, C. E., Nwosu, I. G., & Madu, I. A. 2020). Debbie (2004), referenced in Ebitu et al. (2016) and Victor Nwokocha et al. (2020), states that small and medium-sized businesses account for around 97% of all businesses in the sector and generate approximately 50% of employment and 30% of manufacturing production. This demonstrates the significance of small and medium-sized businesses to our economy.

Small businesses are defined as those having a workforce of more than ten but not more than forty-nine employees and total assets (excluding land and buildings) of more than five million naira but less than fifty million naira (Aminu, S. A. 2018). According to Aminu, S. A. (2018), medium-sized businesses are those that employ between 50 and 199 people overall and have total assets (excluding land and buildings) of more than fifty million but not more than five hundred million naira. The definitions differ according on the criteria taken into account, which differ based on the nations and areas as well. Small and medium-sized (SME) businesses also take market reach, ownership structure, and management structure into account.

In addition, medium-sized businesses are among the most influential to a nation's economy. The money these businesses make from their transactions is what allows many nations to flourish. It is also an effective way to end poverty for both individuals and communities. About 95% of Nigeria's manufacturing industries are made up of small and medium-sized businesses (Tahir & Inuwa, 2019).

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

Fast food chains in Nigeria are primarily medium-sized businesses (Ologbenla, P., 2021). They have positively impacted Nigeria's economy by creating jobs, providing for consumer demands, and hiring labor. As businesses, they must prepare for the future by allocating resources and making decisions based on business conduct. Due to poor business decisions brought on by food spoilage and quality decline, among other factors, several medium-sized businesses, including fast food chains, have filed for bankruptcy or are no longer in operation. In order to increase market share and productivity, sales forecasting is a crucial field in the food industry that has gained a lot of popularity recently (Boyapati, S. N., & Mummidi, R. 2020). Therefore, in order to anticipate future sales based on its training existing data gathered from prior transactions, technologies such as the SARIMA statistical algorithm must be used.

Numerous techniques, such as machine learning, domain expertise, and statistical techniques, are used to make predictions. A statistical technique called Seasonal Autoregressive Integrated Moving Average (SARIMA) uses patterns and trends from a time-related dataset to predict future events.

When there is a seasonal pattern of occurrence, it functions properly. It is also simple to understand and explain. It looks for Seasons (S) in datasets and uses AutoRegressive (AR) to look for the pattern of an event's recurrence. Lags are defined as the difference between the occurrence and the Moving Average (MA), or the Integration (I), which measures the average values between the occurrence and its prior value.

The SARIMA approach is utilized in industries for time series forecast, such as weather, sales, service demand, and so forth, since it provides better predicting results for seasonal time series without outlaying data (Falatouri et al., 2022).

The management methods of medium-sized businesses are the subject of this study. It affects the parties involved in the fast food industry's operations as a medium-sized business. To make important decisions about resource allocation, scheduling, planning, and human resource management, among other things, business leaders must forecast future sales. These well-informed choices are what will keep firms viable. The project, in which the Nigerian government is a key stakeholder, has the potential to generate income and assist in tax collection, so contributing to the growth of the country's economy if executed appropriately.

The difficulties faced by SMEs in Nigeria's industrial sectors include an unwelcoming business climate, inadequate capital, high credit rates, unreachable collateral requirements, a lack of management expertise, and restricted access to contemporary technology (Nwokocha et al., 2020). Over time, the government has taken steps to address some of these issues through innovations and initiatives. Several government agencies have started a few initiatives to support the sector's expansion and sustainability. Among these are the National Enterprise Development Program (NEDP), the Real Sector Support Facility (RSSF), the Small and Medium Industry Equities Investment Scheme (SMIEIS), and the Small and Medium Enterprises Credit Guarantee Scheme (SMECGS) (Nwokocha et al., 2020). These programs address things like financing, interest rates, the need for collateral, and other things. The nation's SMEs are still not developing to the full potential, though, and further research in the areas of technology and management can lead to solutions that will help the business sector expand. The application of sales forecast is one such technological advancement.

It is thought that the survival rates of medium-sized enterprises in developing nations like Nigeria would be significantly increased with accurate sales projection to support decision-making.

LITERATURE REVIEW

The Sarima Algorithm

A statistical model called Seasonal Autoregressive Integrated and Moving Average (SARIMA) is used to forecast or make time-related predictions. The only information needed by the algorithm is the date as a feature and the expected result as the target. When the data has a seasonal distribution, it is employed to address time-related regression difficulties in supervised learning. There are several techniques available for resolving time series issues with certain data distribution series;

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

these algorithms have evolved as further study is done to explore potential improvements. Examples of these algorithms are:

Auto Regression: The time series forecasting algorithm known as Auto Regression (AR) is utilized. In order to generate future predictions over a certain temporal granularity, the AR method uses noise and prior values. To indicate the number of previous values, it utilizes the value (p). Other algorithms like ARMA, ARIMA, SARIMA, and others employ it. In order to accomplish this, observations are multiplied by the pertinent AR coefficients φ after being used as inputs to a regression equation with data from the p time steps prior to the current one. Furthermore, the sum is augmented by adding the series mean, represented by μ, and white noise, or random error, represented by ω.

The equation below represents the AR(p) model (Noor, T. H., Almars, A. M., Alwateer, M., Almaliki, M., Gad, I., & Atlam, E. 2022),

$y_t = \mu + \sum_{i=1}^{p} (\varphi_i y_{t-i}) + \omega_t$ Equation 2.0
(Noor et al., 2022).

Moving Average: To estimate future occurrence in a given time granularity, Moving Average (MA) takes historical error values and noise into account. Algorithms like as ARMA, ARIMA, SARIMAX, and others employ the letter (q) to indicate the number of prior occurrences. It consists of three separate parts, which are as follows: The first variable, represented by μ, is the mean of the series; the second variable, represented by θ, and the model residuals, represented by ω, are the sum of a limited number of MA coefficients; white noise is represented by ωt (Noor et al., 2022). The equation goes as follows:

$y_t = \mu + \sum_{i=1}^{q} (\theta_i \omega_{t-i}) + \omega_t$ Equation 2.1
(Noor et al., 2022).

Auto Regressive Moving Average: This algorithm, also known as Auto Regressive Moving Average (ARMA), makes time-related forecasts by combining the ideas of auto regression and moving average. It is used for stationary data that lacks trends, and it indicates prior values for AR and MA with (p,q), respectively. It is applied to stationary data forecasting. In terms of mathematics, it is expressed as:

$y_t = \mu + \sum_{i=1}^{p} (\varphi_i y_{t-i}) + \sum_{j=1}^{q} (\theta_j \omega_{t-i}) + \omega_t.$ (2.2)
(Talal H. Noor et al. 2022)

The Auto-Regressive Integrated Moving Average (ARIMA) is a statistical measure of autocorrelation. It makes use of integration as well as AR and MA techniques. Finding the difference to keep the average and variance over earlier predictions is made easier with the aid of integration. Because of this feature, it can predict non-stationary data and data with trends with accuracy. When a process Xt satisfies the equation below, it meets the ARIMA condition (p, d, q) (Noor et al., 2022).

$\nabla^d X_t = (1 - B)^d X_t$ (2.3)
(Noor et al., 2022).

The previous values are represented with (p,d,q) for AR, Integration, and MA respectively. These are calculated using Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF). With formulas:

$AIC = -2\,log(L) + 2k = -2\,log(L) + 2(p + q + P + Q)$ Equation 2.4
(Noor et al., 2022).

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

$BIC = -2\,log(L) + k\,ln(n) = -2\,log(L) + (p + q + P + Q)ln(n)$ Equation 2.5 (Noor et al., 2022).

It is used for the forecast of data that has trends.

Seasonal Auto-Regressive Integrated Moving Average (SARIMA): This algorithm takes advantage of the data's seasonality while utilizing the ARIMA technique. The notation used to represent it is (p,d,q) for non-seasonal AR, Integration, and MA, and (P, D, Q)s for seasonal AR, Integration, MA, and seasonal. The algorithm's seasonality causes a little change in how AR and MA are represented mathematically, which looks like this:

$\varphi_p(B)\Phi_p(B^s)W_t = \theta_q(B)\Theta_q(B^s)\omega_t$ Equation 2.6 (Noor et al., 2022).

According to Noor et al. (2022), P denotes the order of the seasonal AR model, D the number of seasonal variations, Q the order of the seasonal MA, and s the length of the season (periodicity). Furthermore, according to Noor et al. (2022) the ωt and B stand for the backward shift operator and the white noise value at period t, respectively. The technique is used to a wide range of problem domains and problem types.

The project's goal is to forecast food sales in Item 7 using the SARIMA statistical technique. In order to accomplish the goal, there are processes and approaches that must be followed in order to create the forecasts.

RESEARCH METHODOLOGY

The aim of the project is to use SARIMA statistical algorithm to predict food sales in Item 7. There are procedures to follow to achieve the aim, these procedures involve techniques and steps to take to make the predictions.

Three main areas comprise the project procedures: developing the model, deploying the model as an online application, and providing cloud hosting for the application for users.

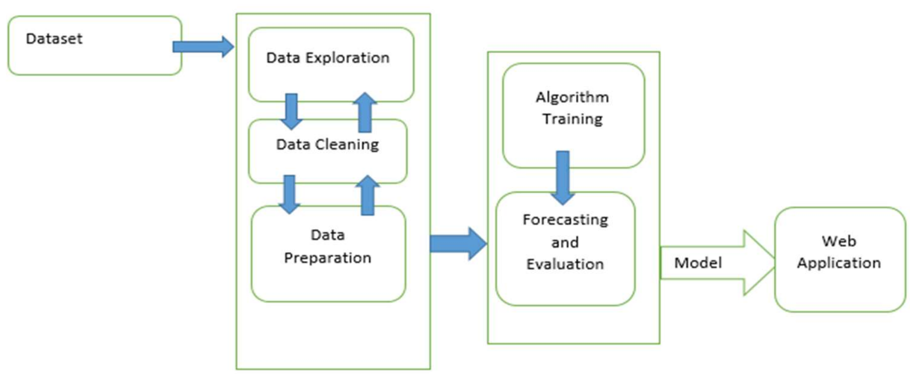


Figure 3.1: Proposed Project Workflow.

Since the project makes use of machine learning, commands will be sent to the computer via the Python programming language. Additional Python packages, such as pandas, numpy, matplotlib, seaborn, sci-kit learn, and streamlit, will also be utilized. The web application will be hosted on the cloud platform using either Heroku or Streamlit Share. The course will employ the Jupyter Notebook for model creation and Visual Studio Code for web application building as its Integrated Development Environment (IDE).

Dataset.

Dataset was obtained from a local fast food. The pandas Python package was utilized to produce an additional dataset. Using the pandas Python package and the Python

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

programming language, sales behavior was taken into account when creating the data. A Comma Separated Value (CSV) file containing 2922 rows (instances) and 5 columns (features and a target) held the generated data. The Sales column in the dataset is the goal column, and the Date, Day, Month, and Year columns are the features. Except for the Date column, which is represented as date-time data, the majority of the columns in the collection are of the integer data type. The data features and related information are shown in the table below.

Table 3.1: Columns Description

Feature Name	Column Description	Values Description	Data Type
Date	It contains the date represented as YY-MM-DD	2015-01-01 - 2022-12-31.	Date-time
Day	It contains the numerical representation of the days in a week	0 – 6, with 0 representing Sunday and 6 representing Saturday.	Integer
Month	It contains the numerical representation of the months in a year.	1 – 12, 1 represents January, and 12 represents December.	Integer
Year	It contains the numerical representation of the year the sales were made	2015 – 2019.	Integer
Sales	It contains the total amount of sales made in a day	50000 as the minimum sale value and 999760 as the maximum value.	Integer

Tools and Resources Used

The resources and tools employed for the project differ according to the implementation category under consideration. The Python programming language and its libraries were utilized to create data, construct a SARIMA model, and launch the model as an online application. GitHub and Onrender were employed for the deployment process.

The model was built using the web-based Integrated Development Environment (IDE) Jupyter Notebook, and it was deployed as an online application using Visual Studio Code.

Data Generation

The University of Ilorin's academic calendar affects the restaurant's sales. She added that sales are higher on weekends than they are throughout the week. These realizations were used in creating the data with the help of the Python libraries numpy, pandas, and random.

Feature columns

The input or determinant columns, also known as feature columns, are the ones that were employed to ascertain the output or aim. The date, day, month, and year columns are the created input columns. The pandas date range method was used to generate the date column, which ranged from January 1, 2015, to December 31, 2022. The day ranged from 0 to 6, with 0 denoting Sunday and 6 representing Saturday. The month was represented by numbers, with 1 denoting January and 12 denoting December. The year column contained values ranging from 2015 to 2022. This is depicted in the following figure.

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Table 3.2: Data Features

	Date	Month	Year	Day
0	2015-01-01	1	2015	3
1	2015-01-02	1	2015	4
2	2015-01-03	1	2015	5
3	2015-01-04	1	2015	6
4	2015-01-05	1	2015	0
...	...	...	...	...
2917	2022-12-27	12	2022	1
2918	2022-12-28	12	2022	2
2919	2022-12-29	12	2022	3
2920	2022-12-30	12	2022	4
2921	2022-12-31	12	2022	5

Target Column

The target column was generated using the information provided by the Administrative manager of Item-7 okeodo branch. Based on the insight, there is a huge drop in sales value from July to October, as implemented with python random method. While the weekend days have higher values compared to other weekdays. The intuitive generation implementation code and final dataframe are shown below.

```
sales = []

for i in range(len(df)) :
    if 6 < df['Month'][i] < 11 and 0 < df['Day'][i]< 6 :
        sales.append(random.randint(50000,200000))
    elif 6 < df['Month'][i] < 11 and (df['Day'][i] == 6 or df['Day'][i] == 0) :
        sales.append(random.randint(100000,300000))
    elif (df['Month'][i] < 7 or df['Month'][i] == 11 or df['Month'][i] == 12) and 0 < df['Day'][i]< 6 :
        sales.append(random.randint(400000,800000))
    else :
        sales.append(random.randint(700000,1000000))
```

```
df['Sales'] = sales
```

Figure 3.2: Intuitive Target Generation Code

Table 3.3: Resulting Dataframe

```
df.head(20)
```

	Date	Month	Year	Day	Sales
0	2015-01-01	1	2015	3	451757
1	2015-01-02	1	2015	4	598938
2	2015-01-03	1	2015	5	444546
3	2015-01-04	1	2015	6	810450
4	2015-01-05	1	2015	0	972384
5	2015-01-06	1	2015	1	713535
6	2015-01-07	1	2015	2	450752
7	2015-01-08	1	2015	3	731828
8	2015-01-09	1	2015	4	538281
9	2015-01-10	1	2015	5	493180
10	2015-01-11	1	2015	6	888701

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Export Data generated as Comma Separated Value format.

The dataset generated was exported in a comma separated value format to make it easy for model building. Pandas was used to save it to C.S.V file and will be used to open it for model building. The image below shows the code snippet for that.

Model Building

The data was used to train and evaluate a SARIMA algorithm. To achieve this aim, data processing was carried out on the data to aid effective training of the algorithm. These processes include data exploration and visualization, data cleaning, and data preparation. Then it will be saved as a pickle file after training and evaluation. These steps are discussed further below.

Importation of packages and Data reading to the notebook

Packages needed for the model-building tasks were imported to the jupyter notebook IDE, these packages include the numpy, pandas, mathplotlib, seaborn, sklearn and statsmodels and so on. They all were written in python to provide various abstractions and make the model-building processes easier. Data was read into the notebook using the pandas read_csv method. It takes in the C.S.V file name and the path to the file name. The read data is then displayed as a dataframe. This is shown below.

Table 3.4: Read in DataFrame.

```
df
```

	Date	Month	Year	Day	Sales
0	2015-01-01	1	2015	3	451757
1	2015-01-02	1	2015	4	598938
2	2015-01-03	1	2015	5	444546
3	2015-01-04	1	2015	6	810450
4	2015-01-05	1	2015	0	972384
...	...	...	...	...	...
2917	2022-12-27	12	2022	1	533774
2918	2022-12-28	12	2022	2	504595
2919	2022-12-29	12	2022	3	749120
2920	2022-12-30	12	2022	4	643662
2921	2022-12-31	12	2022	5	537171

Data Exploration and Analysis.

In order to find errors and learn more about the relationships, distributions, and trends in the data, the data was examined. There are four types of analyses and explorations: graphical, non-graphical, multivariate, and univariate. Python modules such as pandas, matplotlib, seaborn, and sci-kit-learn were used to implement them. Its depiction of the daily, monthly, and annual trends in sales yields some of the outcomes. The daily trend indicates that sales peak on weekends; the month displays a decline in sales from July to October; and the year displays a virtually uniform distribution of sales over the course of the univariate analysis's years. The figures below display these.

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

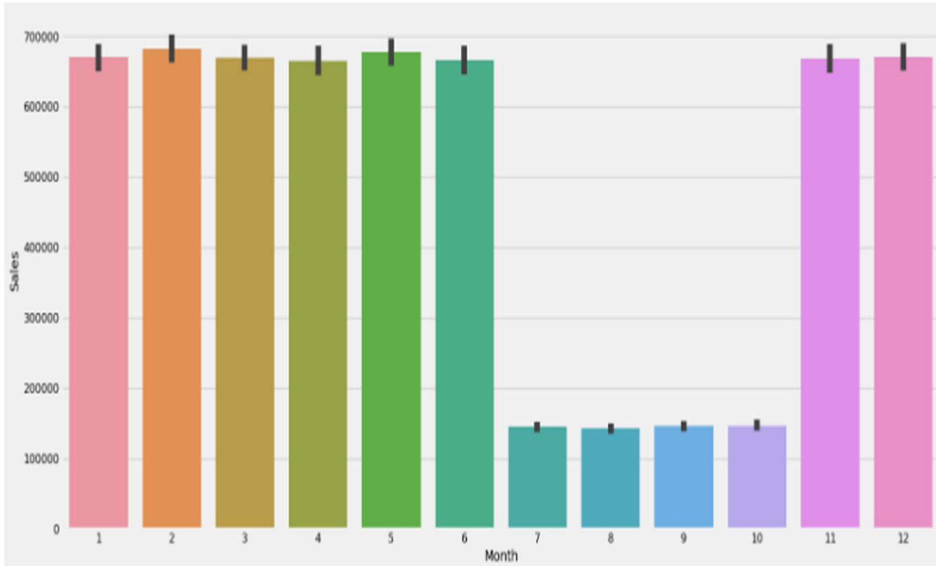


Figure 3.3 The Monthly Sales Plot.

The multivariate analysis involves the use of heatmap seaborn method to check for the correlation between numerical features and it shows no significant correlation between the features except for the month column and sales that has a negative correlation of -0.4. This is shown in the image below.

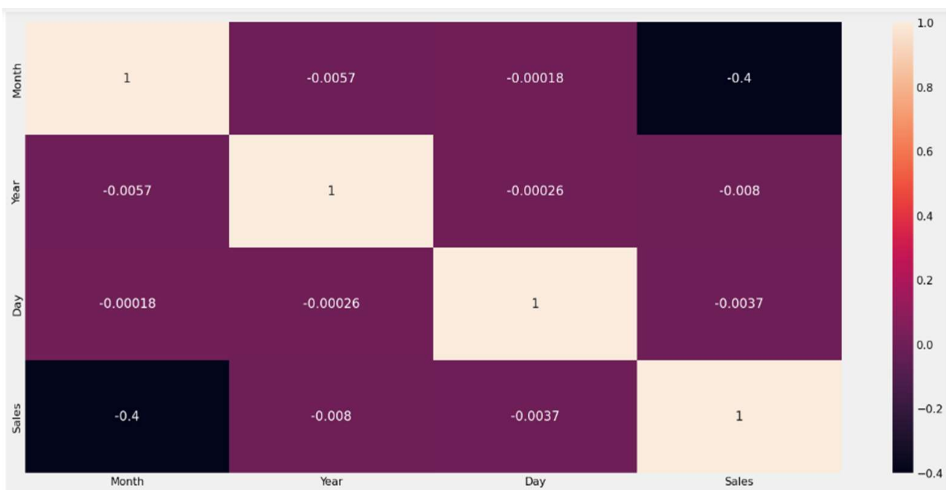


Figure 3.4 Heatmap Showing Correlations between Numerical columns.

Data Cleaning

The result from the data exploration shows that the data is clean since there is no missing values, there are no duplicate values and there are no outliers either because the features are date-related features and not really numerical features. This is shown below

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Table 3.5: Check for Missing and Duplicate Values.

```
df.isnull().sum()

Date      0
Month     0
Year      0
Day       0
Sales     0
dtype: int64

df.duplicated().sum()

0
```

Data Preparation

The pandas and statsmodels packages are two examples of Python packages that are used to prepare and alter the data before training the SARIMA algorithm on it.

After being converted to an index, the Date feature was resampled to the monthly average. As a result, there are fewer data rows and less noise and randomness overall. Only the date as an index and the target sales column were used to train the SARIMA algorithm; data columns like Day, Month, and Year are ignored. The following figure shows the resultant data.

Table 3.6: The Data Series after Transformation.

```
data.head(20)

Date
2015-01-01    647814.064516
2015-02-01    663874.750000
2015-03-01    652843.774194
2015-04-01    637566.800000
2015-05-01    689331.483871
2015-06-01    698570.266667
2015-07-01    150017.903226
2015-08-01    152681.032258
2015-09-01    139177.600000
2015-10-01    140420.709677
2015-11-01    693621.466667
2015-12-01    663249.451613
2016-01-01    679467.741935
2016-02-01    693192.758621
2016-03-01    685311.032258
2016-04-01    699455.800000
2016-05-01    683039.322581
2016-06-01    673819.766667
2016-07-01    132690.870968
2016-08-01    148255.709677
Freq: MS, Name: Sales, dtype: float64
```

RESULT

Algorithm Training and Evaluation

The prepared data was further analyzed to determine the best SARIMA parameters to fit the data, these parameters are the non-seasonal (P, D, Q) and seasonal (p,d,q,s) algorithms. The visualizations carried out include the decomposition plot, the Autocorrelation Function (ACF) plots, and the Partial Autocorrelation Function (PACF) plot, these plots are shown below

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

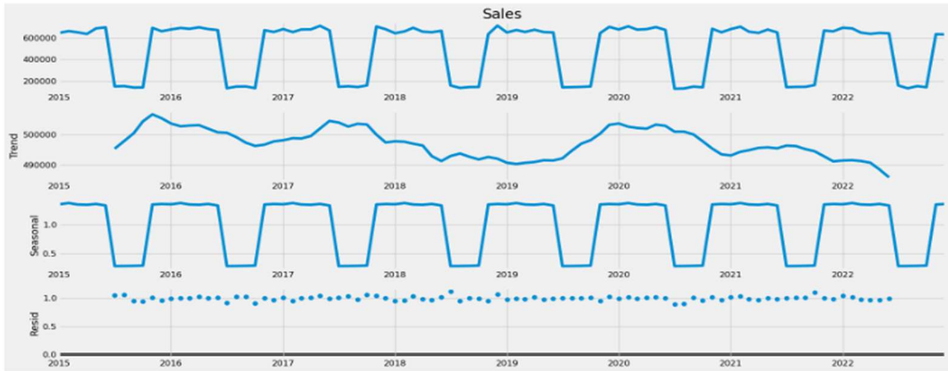


Figure 4.1: The Decomposition Plot

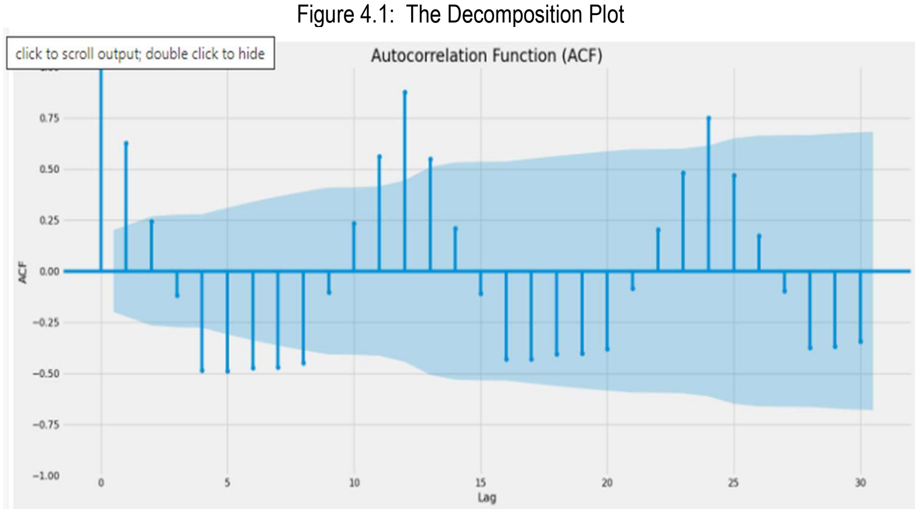


Figure 4.2: The Autocorrelation Function Plot

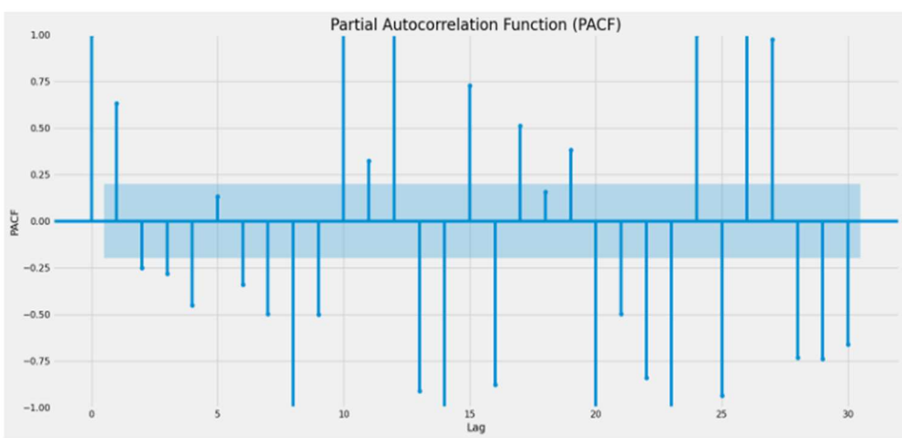


Figure 4.3: The Partial Autocorrelation Function Plot.

The grid search for time series was used to determine the other parameters after getting the seasonality as 12 from the ACF plot. These determined the seasonal auto-regression, moving average, and differencing same as their non-seasonal components. After

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

training, the parameters with the lowest Alkaike Information Criterion (AIC) score of 1593.1277 are for non-seasonal component P = 1, D = 1, Q = 1, and for seasonal component of p = 0, d = 1, q = 1 and s = 12.

These were used to train the SARIMA algorithm to become a model. This is shown in the snippet below.

```
mod = sm.tsa.statespace.SARIMAX(data,order=(1,1,1),seasonal_order=(0,1,1,12),enforce_invertibility=False)
results = mod.fit()
print(results.summary().tables[1])
```

```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.7492      0.084     -8.940      0.000      -0.913      -0.585
ma.L1          0.6184      0.084      7.350      0.000       0.454       0.783
ma.S.L12      -0.0219      0.020     -1.099      0.272      -0.061       0.017
sigma2       6.35e+08    3.7e-11   1.72e+19      0.000    6.35e+08    6.35e+08
==============================================================================
```

Figure 4.4 Algorithm Training Snippet

The trained algorithm was evaluated using both plots and evaluation metrics, diagnostic and prediction plots were used to evaluate with plots and these gave visual representations of the model behaviors. This is illustrated below.

```
results.plot_diagnostics(figsize=(16,8))
plt.show()
```

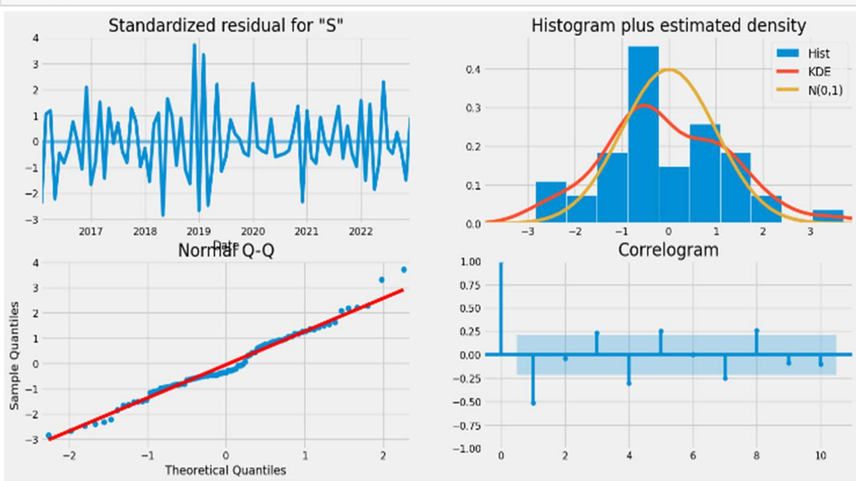


Figure 4.5 The diagnostic plot.

The diagnostic plot shows the residual data has no clear pattern, so nothing more left in it to learn.

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

```
pred = results.get_prediction(start= pd.to_datetime("2017-01-01"),dynamic= False)
pred_ci = pred.conf_int()

ax = data["2015":].plot(label= 'observed')
pred.predicted_mean.plot(ax=ax,label='one step ahead forecast',alpha=.7,figsize=(14,7))

ax.fill_between(pred_ci.index,
                pred_ci.iloc[:,0],
                pred_ci.iloc[:,1],color='k',alpha=.2)

ax.set_xlabel("Date")
ax.set_ylabel("Mean Sales")
plt.title("Mean Sales Made")
plt.legend()
plt.show()
```

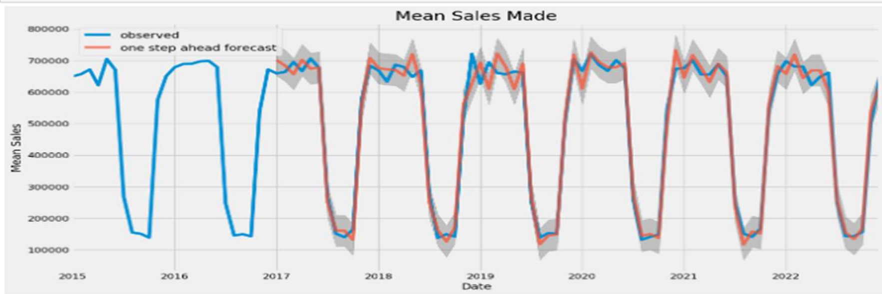


Figure 4.6 The Prediction Plot.

The plot shows how the real sales and predicted sales overlaps with an accurate representation. The metrics used in the model are the Root Mean Squared Error (RMSE) with a score of 32942.15, the Mean Absolute Error (MAE) scored 26743.59, and the R Squared score of 0.979 respectively. These are in the snippet below

```
y_forecasted = pred.predicted_mean
y_true = data['2017-01-01':]
rmse = ((y_forecasted - y_true)**2).mean()
print("The root mean squared error forecast is {} ".format(round(np.sqrt(rmse),2)))

The root mean squared error forecast is 32942.15

from sklearn.metrics import r2_score, mean_absolute_error

print(f'The mean absolute error score is {mean_absolute_error(y_true,y_forecasted)}')

The mean absolute error score is 26743.58504664447

print(f'The R squared score is {r2_score(y_true,y_forecasted)}')

The R squared score is 0.9797654365041415
```

Figure 4.7 Evaluation Metrics

Table 4.1: Evaluation Metrics and Scores

Metrics	Score
Root Mean Squared Error (RMSE)	32942.15
Mean Absolute Error (MAE)	26743.59
R2_Score	0.9797

Model Deployment as a Web Application

The developed model was made available to stakeholders as an online application to make it easier to use. Plotly, pandas, streamlit, and other Python libraries were used for this, and the Visual Studio Code IDE was used to implement them. Three components make up its implementation: input, backend, and output.

Input Components

The web application's input components are used to enter data so that forecasts can be made. They are separated into two categories: input for numerous predictions and input for a single prediction. It makes use of select_box, date_input, and number_input streamlit methods among other widgets. They are shown in the web application's sidebar. The

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

input widget for both single and multiple predictions may be seen in the pictures below.

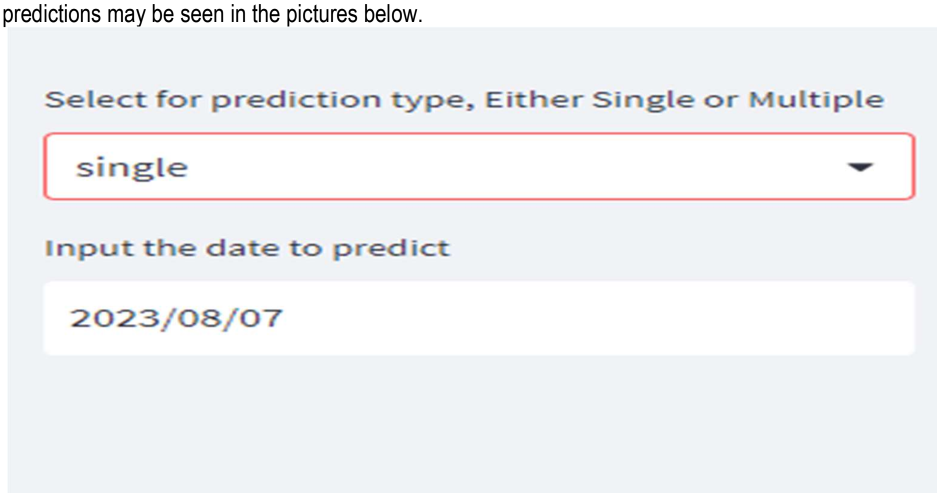


Figure 4.16 Single Prediction Input

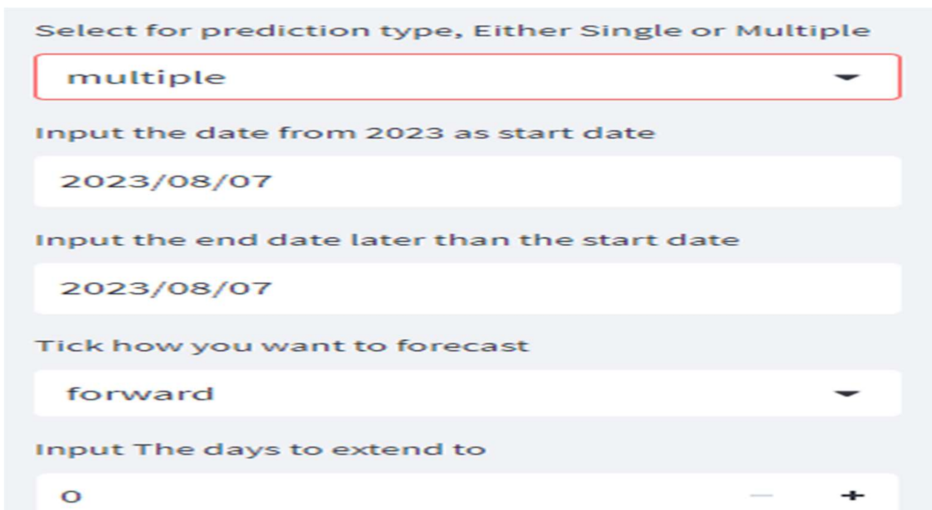


Figure 4.8 Input Widgets for Multiple Predictions.

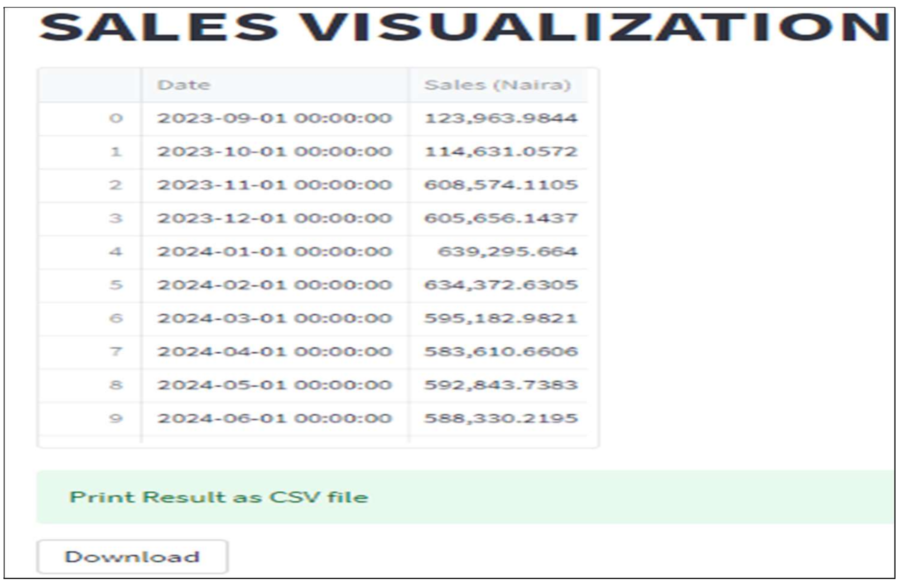
The Output Components

The algorithm's predictions are supplied for customers to use through the output components. There are single and multiple prediction output components. There are two ways it can be presented: given a single input, the result is either a date and value prediction that appears on the screen or a data frame that can be written as a CSV file. When there are several forecasts, the data frame with the date and values can either be shown as graphs, such as line plots, multiple bar plots, area plots, and contour plots, or it can be exported as a CSV file. The online application displays these in two tabs. Tab 1 has the data frame and a CSV download option, while Tab 2 displays graphs for many forecasts and a result for a single prediction. The pictures down below show these.

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Figure 4.18 Single Forecast.



	Date	Sales (Naira)
0	2023-09-01 00:00:00	123,963.9844
1	2023-10-01 00:00:00	114,631.0572
2	2023-11-01 00:00:00	608,574.1105
3	2023-12-01 00:00:00	605,656.1437
4	2024-01-01 00:00:00	639,295.664
5	2024-02-01 00:00:00	634,372.6305
6	2024-03-01 00:00:00	595,182.9821
7	2024-04-01 00:00:00	583,610.6606
8	2024-05-01 00:00:00	592,843.7383
9	2024-06-01 00:00:00	588,330.2195

Figure 4. 8Multiple Forecast with Download Option

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 12(3), SEPTEMBER, 2024
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

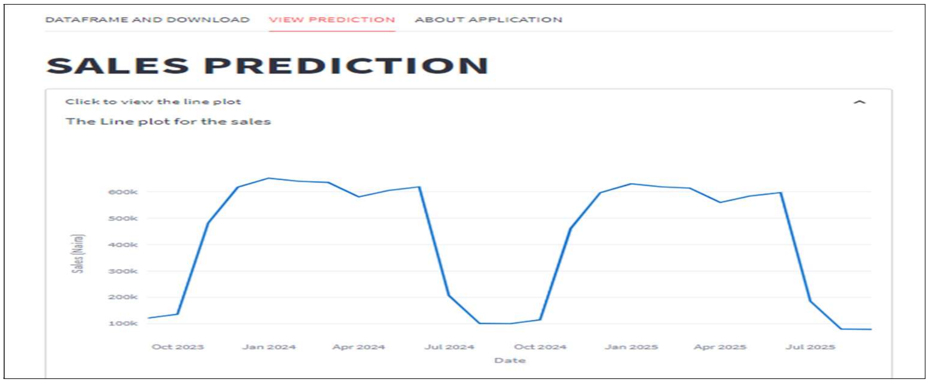


Figure 4.9 Multiple Forecast Visualization.

Hosting the Web Application on a Cloud Platform.

The web application will be hostel on render cloud platform. It is a cloud platform that has options for free host services. The link to the web application after hosting is https://fastfoods-sales.onrender.com/ .

CONCLUSION

This study addressed the urgent need for effective sales forecasting in the Nigerian fast-food industry by effectively applying the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to predict sales in a medium-sized fast-food company. According to the study, precise sales forecasts have a big impact on how well organizations run, how profitable they are, and how long they can survive in a cutthroat market. Through the creation of a web application and machine learning model, the study offers a useful tool that medium-sized businesses may use to improve their planning and decision-making procedures. The findings show that putting in place such a forecasting system might greatly increase operational effectiveness, lower the chance of shoddy planning, and eventually increase the longevity and success of companies in Nigeria's fast-food Industry and other related sectors. This study emphasizes the need of adopting data-driven approaches in corporate management in order to remain competitive and resilient in a continuously changing economy.

REFERENCES

Adeyemi, I. I., Isaac, O. A., & Olufemi, A. S. (2017). Strategic Management: A Policy To Enhance Sustainable Business Development In Small And Medium Scale Enterprises In Nigeria. *Archives of Business Research*, *5*(9). https://doi.org/10.14738/abr.59.3638

Aminu, S. A. (2018). Market Orientation and Small Medium Enterprises' (SMEs) Performance in Nigeria: A review. *Ilorin Journal of Marketing*, *3*(1), 122-132.

Boyapati, S. N., & Mummidi, R. (2020). Predicting sales using Machine Learning Techniques.

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Computer Science*, *200*, 993–1003. https://doi.org/10.1016/j.procs.2022.01.298

Iyortsuun, A. S. (2017). An empirical analysis of the effect of business incubation process on firm performance in Nigeria. *Journal of Small Business and Entrepreneurship*, *29*(6), 433–459. https://doi.org/10.1080/08276331.2017.1376265

Lee, W. I., Chen, C. W., Chen, K. H., Chen, T. H., & Liu, C. C. (2012). A comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods. *Journal of Marine Science and Technology*, *20*(2), 4.

Noor, T. H., Almars, A. M., Alwateer, M., Almaliki, M., Gad, I., & Atlam, E.

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.

(2022). SARIMA: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia. *Electronics*, *11*(23), 3986. https://doi.org/10.3390/electronics11233986

Nwokocha, V. C., Nwankwo, C. E., Nwosu, I. G., & Madu, I. A. (2020). An Appraisal of Production Subcontracting Toward Small and Medium Scale Enterprises Development in the Nigeria Industrial Sector: A Review Approach. *SAGE Open*, *10*(3), 215824402094100. https://doi.org/10.1177/2158244020941001

Ologbenla, P. (2021). Small and Medium Scale Enetrprises and Economic Performance in Nigeria: Acase of Chicken Republic Fast Food in Ekiti State. *Acta Universitatis Danubius. Œconomica*, *17*(3), 7-23.

Sakib, S. N. (2023). Restaurant sales prediction using machine learning. In *Handbook of Research on AI and Machine Learning Applications in Customer Support and Analytics* (pp. 202-226). IGI Global.

Singh, K., Booma, P. M., & Eaganathan, U. (2020). E-Commerce System for Sale Prediction Using Machine Learning Technique. *Journal of Physics: Conference Series*, *1712*(1), 012042. https://doi.org/10.1088/1742-6596/1712/1/012042

Tahir, F., & Inuwa, F. (2019). Empirical Investigation of the Factors Affecting Micro, Small and Medium Scale Enterprises Performance in Borno State, Nigeria. *International Business Research*, *12*(4), 30. https://doi.org/10.5539/ibr.v12n4p30

Corresponding author: Balogun, G. B.
balogungb@gmail.com
Department of Computer Science, University of Ilorin.