International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Machine Learning Modelling of GDP Datasets of Botswana

R. Sivasamy[1], Moseki, K. K[2], Makatjane, K[3], K.Kelepile[4]

[12,3,4]Department Of Statistics, Faculty Of Social Sciences, University Of Botswana, Botswana

Abstract

Researchers use Principal Component Analysis (PCA) as a multivariate method to reduce the dimensionality of the dataset under consideration. It is a well-known fact that PCA transforms the original set of variables into a smaller set of linear combinations that account for most of the variance in the original set. The main objective of this paper is to investigate how exogenous variables X= (X_1, X_2, …, X_{14}) affect the gross domestic product (GDP) of Botswana data (Y) using different multivariate methods such as PCA, linear discriminatory analysis (LDA), factor analysis and other classifiers. Thus, we propose a machine learning project that works on datasets X and Vector(Y) and fits a machine learning model using this dataset representing Botswana GDP, which contains 65 rows (i.e., quarters between 01-01-2004 - 01-01-2020) and 15 columns. This data set Z= (X, Y) is treated as a matrix, which is the main data structure for our analysis. Effective principal components (PCs) associated with eigenvalues of (X^TX) in the variance-covariance matrix exceeding one are isolated, and the variable contribution of these PCs is summed by unsupervised learning algorithms. The X and Y datasets are then split into training subsets, X-train, Y-train, and test subsets, X-test and Y-test. Using supervised learning techniques, machine learning models are fitted and performance is monitored against test data, calculating the error between observed Y and predicted Y in each case. The results show 100% accuracy for some classifiers and more than 75% accuracy for some other classifiers.

Keywords: machine learning, dimensionality reduction, classifier analysis.

1. Introduction

A standard measure of the added value generated by the production of goods and services in a country over a certain period of time is called ‘Gross domestic product (GDP) ‘. All countries compile their data according to the 2008 System of National Accounts (SNA). As such, the GDP of a country measures the income earned from that production or the total amount spent on final goods and services (excluding imports).

Although GDP is the most important measure of economic activity, it does not provide an adequate measure of people's material well-being, for which there may be more suitable alternative indicators. This indicator is based on real GDP (also called GDP at constant prices or GDP by volume), which means that developments over time have been adjusted for price changes. The figures are also adjusted for seasonal effects. The indicator is available in different dimensions: percentage change compared to the previous quarter, percentage change compared to the same quarter last year and volume index (2015=100).

GDP of Botswana grew 5.8% in 2022 compared to last year. This rate is 60 -tenths of one percent less than the figure of 11.8% published in 2021.

The GDP figure in 2022 was $20,352 million, Botswana is number 116 in the ranking of GDP of the 196 countries that we publish. The absolute value of GDP in Botswana rose $1,584 million with respect to 2021.

The **GDP per capita of Botswana in 2022** was $7,863, $612 higher than in 2021, it was $7,251. To view the evolution of the GDP per capita, it is interesting to look back a few years and compare these data with those of 2012 when the GDP per capita in Botswana was $6,393.

If we rank the countries according to their GDP per capita, **Botswana** is in 87th position of the 196 countries whose GDP we publish.

1.1 Progression of the GDP in Botswana.

According to the World Bank, the Gross Domestic Product (GDP) of Botswana was worth 20.35 billion US dollars in 2022 [1]. This represents 0.01 percent of the world economy

Botswana has one of the highest GDPs per capita in sub-Saharan Africa, with a value of approximately $18,113 as of 2021 [2]. The country is also the world's largest producer of diamonds [2].

The economy of Botswana is heavily reliant on the mining industry, which accounts for an estimated **34.7%** of the country's GDP [1]. The mining sector is anchored by diamond mining, which is the largest contributor to the country's economy. Botswana is the largest exporter of diamonds in the world, and diamond mining contributes up to a third of the nation's GDP [12]. Debswana Company, a joint undertaking between De Beers and the government of Botswana, is responsible for all diamond mining in the country [1].

Agriculture accounted for the largest portion of Botswana's GDP at independence, but its contribution has since dropped to a mere **3%** [1]. The sector is severely limited by adverse climatic conditions, with a large part of the country situated in the Kalahari Desert. Corn, sorghum, millet, beans, and groundnuts are the main subsistence crops cultivated. A significant proportion of the sector's contribution to the economy also comes from livestock, with beef cattle rearing generating an estimated $34 million in 2010 [1].

Manufacturing accounts for **4%** of Botswana's GDP and includes diamond processing, food processing (predominantly beef), textiles, and mining [1].

(Granger, 1969) investigated causal relations through 'Econometric Models and Cross Spectral Methods'. (Fabrigar, Wegener, MacCallum, & Strahan, 1999) discussed various uses of exploratory factor analysis pertain to Psychological Research. (Gabriel, 1971) discussed principal component analysis for data sets with Biplot display of matrices. (Humphreys, L G; Igen, D. R, 1969) and (Humphreys & Montanelli, 1975) used parallel analysis and determined number of common factors. (Jennrich, 2001) explained the uses of orthogonal rotation through a novel procedure. (Kano, 1990) created noniterative estimation selecting an optimum number of factors in exploratory factor analysis. (Zwick & Velicer, 1986) focused Five Rules for Determining the Number of Components to retain in exploratory factor analysis. Most of the fundamentals related to univariate and applied multivariate statistical analysis can be found in (Johnson & Dean, 1992).

(Artis, Banerjee, & Marcellino, 2005) discussed time series methods and factor forecasts for UK. (Bai & Ng, 2002) and (Bai J. N., 2006.) determined several types of factor models to obtain factors and confidence intervals for forecasts. (Stock & Watson, 2002), and (Stock & Watson, 2012) used Macroeconomic forecasting and obtained diffusion indexes. For the purpose of dimension reduction under the many predictors' environment, a dynamic factor approach based on principal components regression was proposed in (Kim & Swanson, 2014).

(Pasini, 2017) investigated various aspects of principal component analysis for stock portfolio management. (Nakajimay & Sueishiz, 2020) studied forecasting of High-Dimensional Data related to the Japanese Macroeconomy. (Uematsu & Tanaka, 2019) employed penalized regression model to high-dimensional macroeconomic forecasting and variable selection jobs. (He, Yang, & Zhang.B, 2023) fitted robust PCA for high-dimensional data using characteristic transformation.

1.2 Machine Learning Methods to Botswana GDP

The machine learning method (MLM) has been playing remarkable advancements, producing data-driven insights and facilitating wise decision-making across various domains. Within the machine learning, principal component analysis (PCA), factor analysis, low rank approximations, and Markov switching methods emerge as sound techniques.

Basic principles of mathematics and concepts like Linear algebra are the foundation of Machine Learning and Deep Learning systems. So, we can say that Machine Learning creates a useful model with the help of both mathematical and statistical concepts. Statistics is an important concept to organize and integrate data in Machine Learning. Linear Algebra also helps to create better supervised models (such as Logistic Regression, Linear Regression, Decision Trees, Support Vector Machines (SVM)) as well as unsupervised Machine Learning algorithms such as PCA, Single Value Decomposition (SVD), Clustering etc.

There are the various graphical representations supported by Machine Learning projects that we can work on. It is remarked that parts of the given input dataset are trained based on their categories by classifiers provided by machine learning algorithms. Such classifiers also remove the errors from the trained data.

The primary objective is to study the correlation and covariance relationships that the variables in terms of microeconomic factors called exogenous features (X_1) Agriculture, (X_2) Mining, (X_3) Manufacturing, (X_4) Water and Electricity, (X_5) Construction, (X_6) Hotels and Restaurants , (X_7) Transport and Communication (X_8) Financial business and Services, (X_9) General Government, (X_{10}) Social personal and services, (X_{11}) value added , (X_{12}) Taxes on import, (X_{13}) other taxes on products, (X_{14}) Subsides and the response variable (15) Y=Total GDP.

We propose a machine learning project that works on the dataset $X=(X_1, X_2, \ldots, X_{14})$, and Vector(Y), and we fit the machine learning model using this dataset representing the GDP of Botswana containing 65 rows (i.e. quarters ranging from 01-01-2004 to 01-01-2020) and 15 columns. This dataset Z=(X,Y) is handled as a Matrix, which is a key data structure for our analysis.

This dataset is divided into input X and output (or target Y) for the supervised learning model, it represents a Matrix(X) and Vector(Y). The dataset X is alone used for the unsupervised learning case.

Generally, fitting a model for a large dataset is a most challenging tasks of machine learning. Moreover, a model built with irrelevant features is less accurate than a model built with relevant features. There are several methods in machine learning that automatically reduce the number of columns of a dataset, and these methods are known as Dimensionality reduction. The most commonly used dimensionality reductions method in machine learning is PCA. This technique makes projections of high-dimensional data for both visualizations and training models. PCA uses the matrix factorization method from linear algebra.

Section 2 discusses the best practices of handling the observed datasets such as correlation, scree plot and the principal components. In addition, classification of Total GDP by Linear Discriminant Analysis (LDA) is also explained. Further, several types of classifier analyses were used to classify the two divisions

N=below median and M=above median created over the Botswana GDP Dada. Section 3 provides a formal concluding remark.

2. PCA applied to GDP datasets of Botswana

We have applied a principal component analysis, (PCA), using Python to compute the eigenvalues, the eigenvectors loadings of the correlation matrix and the covariance matrix.

Table 1: Correlation Matrix for microeconomic variables

	Agricultur	Mining	facturing	lectricit	nstructio	staurer	municc	Service	vernme	services	added	import	roduct	ubsdie
Agriculture	1.00													
Mining	0.77	1.00												
Facturing	0.97	0.80	1.00											
Electricity	0.17	0.17	0.26	1.00										
constructi	0.95	0.79	0.99	0.33	1.00									
Restauren	0.93	0.80	0.97	0.38	0.99	1.00								
Communic	0.95	0.79	0.99	0.35	1.00	0.99	1.00							
Service	0.95	0.79	0.99	0.31	1.00	0.99	1.00	1.00						
Governme	0.95	0.78	0.99	0.32	0.99	0.98	0.99	1.00	1.00					
services	0.96	0.80	0.99	0.27	0.99	0.98	1.00	1.00	0.99	1.00				
added	0.94	0.85	0.98	0.33	0.99	0.99	0.99	0.99	0.99	0.99	1.00			
import	0.93	0.80	0.97	0.33	0.97	0.97	0.98	0.98	0.98	0.97	0.98	1.00		
Products	0.95	0.75	0.98	0.29	0.97	0.96	0.98	0.97	0.98	0.98	0.97	0.95	1.00	
Subsdies	-0.93	-0.76	-0.98	-0.38	-0.99	-0.98	-0.99	-0.98	-0.99	-0.98	-0.98	-0.97	-0.97	1.00

Inspecting though the correlation matrix, we find that most of the correlation coefficients of the macroeconomic variables except with electricity are greater than 0.5 and they show very strong positive linear correlation. Common factors are extracted as principal components of the entire set of predictor variables.

Scree Plot for PCA Explained: The main idea is to capture most of the variance of our data using a lower-dimensional space. A scree plot is a graphic that shows the explained variance per principal component. The measure of the plot can be the absolute value of the explained variance (eigenvalues). It's common in practice that the first few principal components explain the major amount of variance. Kaiser's rule is a commonly used method to select the number of components in a PCA. It's based on keeping the components with eigenvalues greater than 1.

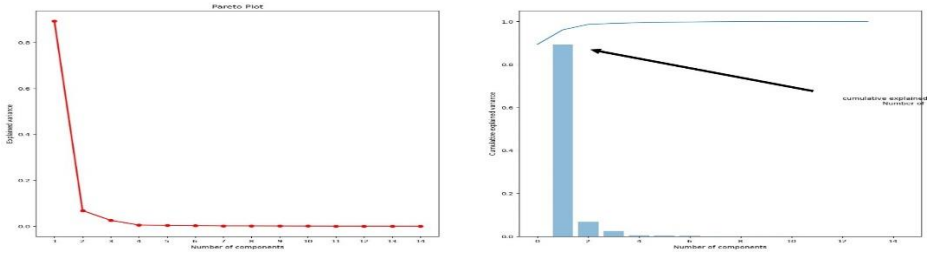

Figure 1: Scree Plot for PCA Explained

Eigen values:

array ([1.27016988e+01, 9.64883094e-01, 3.59404498e-01, 6.87006520e-02,
4.64051657e-02, 3.00618516e-02, 1.46071809e-02, 1.39676726e-02,
8.58397212e-03, 5.12146378e-03, 2.51514820e-03, 1.88828000e-03,
9.12270172e-04, 3.85122822e-11])

In terms of dimensionality reduction, we have found that factors 1 and 2 have an eigenvalue greater than 1. Specifically, factor 1 has a value of 12.7 and factor 2 has a value of 0.96 (almost 1).

Explained Variance Ratio:

array ([8.93306286e-01, 6.78599099e-02, 2.52767999e-02, 4.83169421e-03,

3.26366001e-03, 2.11424011e-03, 1.02731821e-03, 9.82341811e-04,
6.03707929e-04, 3.60190859e-04, 1.76889544e-04, 1.32802110e-04,
6.41596604e-05, 2.70855611e-12])

Thus, the factors that we will retain are two. Concerning, eigenvalues on explained variance ratio, we have found that the proportion for factor 1 is 89.33% and for factor 2 is 0.067% of the total variance. The first two components namely account for 89.4% of the total variation.

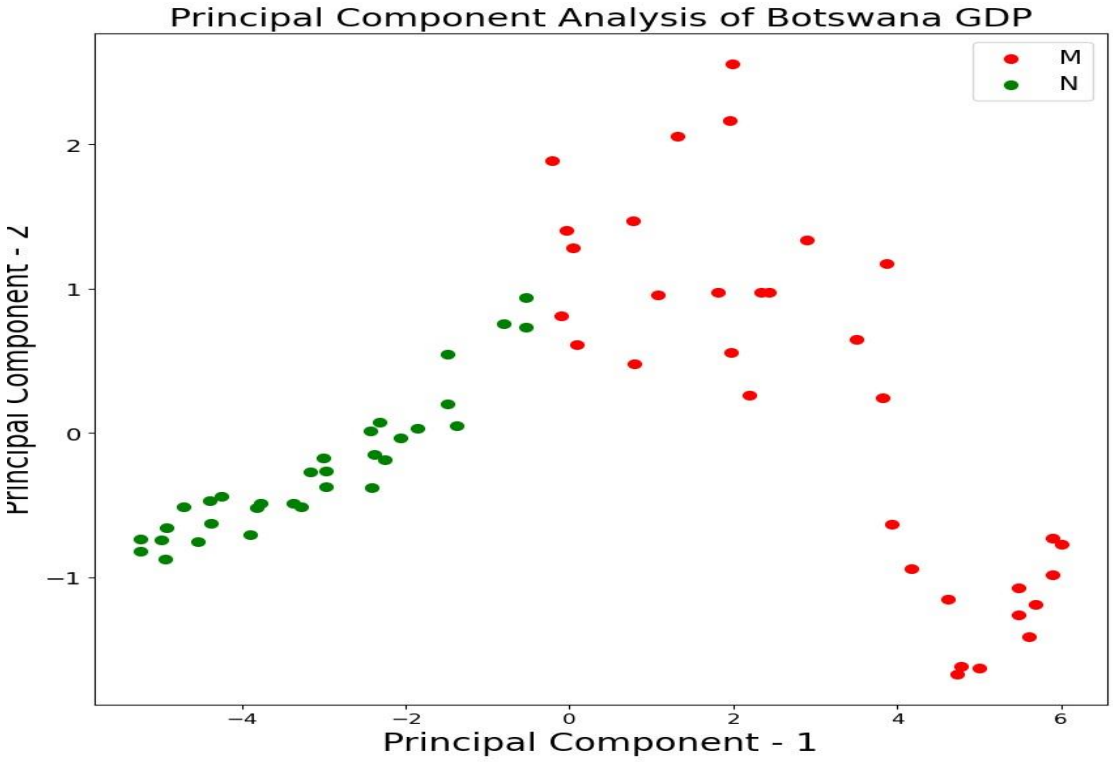

Figure 2: PC1 versus PC2

The total GDP vector Y is converted as string vector using the string 'N' for those values of the total GDP that fall below its median otherwise with string 'M' so that Y vector becomes a sequence of 65 symbols of 'N' and 'M'. Thus, the graph in Figure 2 shows the scatter dots of (PC1, PC2) in terms of red dots for 'M' and green dots for 'N'.

2.1 Classification of Total GDP by Linear Discriminant Analysis (LDA)

In the digital world of machine learning, Linear Discriminant Analysis (LDA) is a statistical method to determine the best separation between two or more classes. LDA requires a labelled training set of data points in order to learn the Linear Discriminant function. Once the Linear Discriminant function has been learned, it can then be used to predict the class label of new data points. So, with application of LDA, one can identify which class a particular data point of the target belongs to.

LDA is a supervised algorithm that aims to find the linear discriminants to represent the axes that maximize separation between different classes of data. Both LDA and PCA are used as dimensionality reduction techniques, where PCA is first followed by LDA.

The main purpose of LDA is to find the line (or plane) that best separates data points belonging to different classes. The key idea behind LDA is that the decision boundary should be chosen such that it maximizes the distance between the means of the two classes while simultaneously minimizing the variance within each classes data or within-class scatter. This criterion is known as the Fisher criterion and can be expressed as the formula for two classes. LDA algorithm works based on the following steps:

- The first step is to calculate the means and standard deviation of each feature.
- Within class scatter matrix and between class scatter matrix is calculated These matrices are then used to calculate the eigenvectors and eigenvalues.
- LDA chooses the k=2 eigenvectors with the largest eigenvalues to form a transformation matrix.
- LDA uses this transformation matrix to transform the data into a new space with k dimensions.
- Once the transformation matrix transforms the data into new space with k dimensions, LDA can then be used for classification or dimensionality reduction

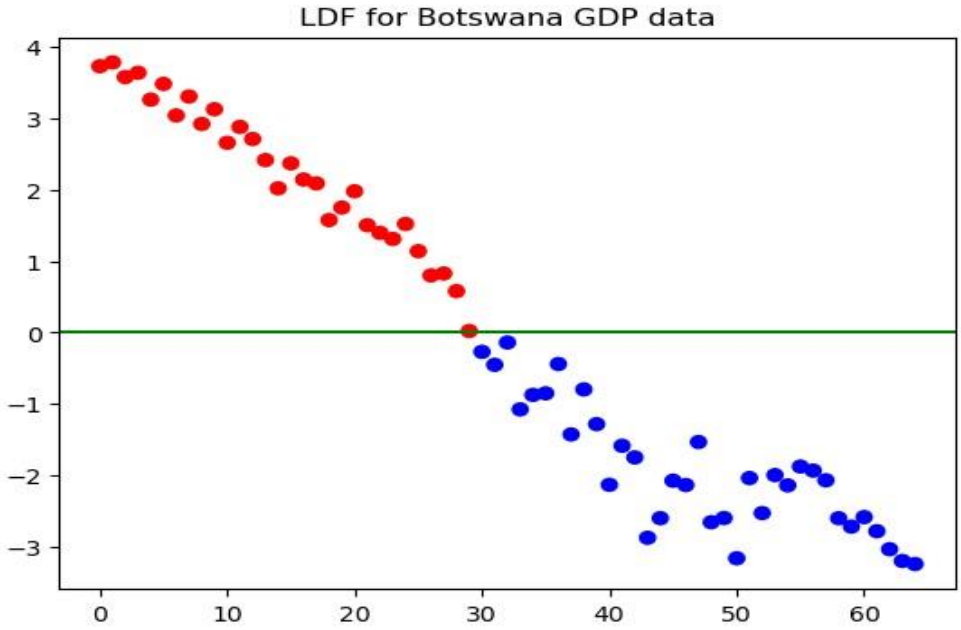

Figure 3: Separating line of LDF for Botswana GDP data in terms of red dots for 'M' and blue dots for 'N'.

In the Figure 3, data get projected on most appropriate separating line such that the distance between the means is maximized and the variance within the data from same class is minimized.

2.2 Classifier Analysis on Botswana GDP Dada

What is a classifier in machine learning? A classifier is any algorithm that sorts data into labelled classes, or categories of information. A classification model, is the end result of our classifier's machine learning. The model is trained using the classifier, so that the model, ultimately, classifies our data. There are both supervised and unsupervised classifiers.

Using the following command, we divide the Botswana GDP dataset into training datasets X_train, and Y_train and testing datasets X_test and Y_test:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=43)

We then used the following classifiers to analyse the Botswana GDP data set..

names = ["Logistic Regression", "Nearest Neighbors", "Linear SVM", "RBF SVM", "Gaussian Process",

based "Decision Tree", "Random Forest", "Neural Net", "AdaBoost",ing
"Naive Bayes", "QDA"]

Here, reported below is the observed result about the accuracy of classifying the given dada in the testing pairs(X_test, Y_test) based on the fitted model using the pairs(X_train, Y_train) by each of the classification model

Accuracy of Logistic Regression Classifier is:1.0
Accuracy of Nearest Neighbors Classifier is:1.0
Accuracy of Linear SVM Classifier is:1.0
Accuracy of RBF SVM Classifier is:0.47058823529411764
Accuracy of Gaussian Process Classifier is:0.5294117647058824
Accuracy of Decision Tree Classifier is:1.0
Accuracy of Random Forest Classifier is:1.0
Accuracy of Neural Net Classifier is:0.8235294117647058
Accuracy of AdaBoost Classifier is:1.0
Accuracy of Naive Bayes Classifier is:1.0
Accuracy of QDA Classifier is:1.0

3. Factor **Analytic Methods for subsets of Botswana GDP datasets**

This section considers a sub set on X_1= Agriculture, X_2= Mining, X_3=(Manu)factoring, X_4=Electricity and X_5=Construction of the GDP dataset analysed in the preceding sections. We analyse this sub set to exhibit more insights about the variance present in the data variables and contributions of individuals towards various dimensions using factor and PCA analysis.

Application of both factor and PCA resolves factor solutions and thus, her, we identify two components dim1 (first PC or PC1) and dim2 (second PC or PC2) to account for 95.05 % of variability of the data set X= (X_1, X_2, X_3, X_4, X_5). The outcome of the PCA analysis is shown in Figure 4.

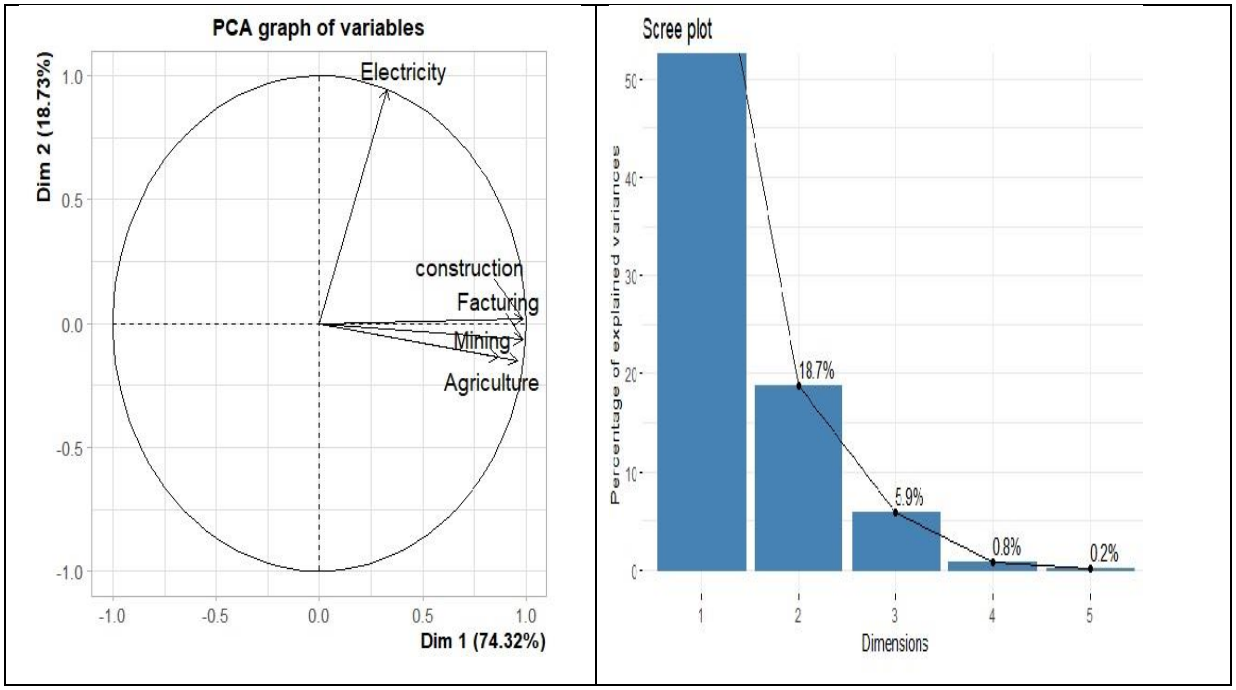

Figure 4: *The graph showing the outcome of PCA analysis in PC1 vs PC2 dimensions* ***(left) and the scree plot(right)***

3.1 Criteria for extracting number of factors

Extracting multiple factors from a large data set is like focusing a microscope. The researcher can examine several factor structures and thus choose the best representation from the given data. It should be noted that no precise quantitative basis for deciding the number of factors to be extracted has been developed.

Latent root criterion (LRC): This technique is simple to either component analysis or common factor analysis. Only factors having eigen values (=latent roots) greater than 1 are considered significant.

Percentage of variance criterion (PVC): It is used to achieve a specified cumulative percentage of total variance extracted by the successive factors. As there is no absolute threshold is mandatory, in the social sciences it is a common practice to recommend a factor solution that accounts for 60 percent of the total variance as sufficient amount.

The total amount of variance an original variable share with all other factors included in the analysis is known as communality. So, one can select enough factors to achieve a prespecified communality for each variable as an alternative criterion or as the degree of explanation.

Sree test Criterion:

The scree test is applied to identify the optimum number of factors that can be extracted before the unique variance begins to dominate the common variance structure. One such scree test is derived for feeding the data set X and the corresponding results are plotted for 5 factors in Figure 4. This plot shows the latent roots against the number of factors in their order of extraction. Starting from the first factor, the plot slopes steeply downward initially and then slowly become a horizontal line. In this case, the first two factors would quality based on either the LRC or PVC.

An important tool in interpreting factors is factor rotation. The simplest case is orthogonal rotation, in which the axes are maintained at 90 degrees. The correlation between a variable and a principal component (PC) is used as the coordinate of the variable on the PC. The representation of variables differs from the plot of the observations: the observations are represented by their projections, but the variables are represented by their correlations

In Figure 4, the original variables are plotted with unrotated axes (or reference axes) and then they are turned about the origin until some other position or 90 degree is reached. From visual inspection of Figure 4(l, we observe that 'X_4=Electricity' is independent of 'X_1=Agriculture'. Since, X_1= Agriculture, X_2= Mining, X_3=(Manu)factoring, and X_5=Construction are all highly correlated variables, they form the first principal component (PC1) accounting for 74.32% of variation and X4 forms the second principal component PC2 with 18.37 % of variation.

Quality of representation: The quality of representation of the variables on factor map is called **cos2** (square cosine, squared coordinates). We can visualize the cos2 of variables on all the dimensions as in Fig 5. If the objective is to select factors that discriminate among the subgroups of a sample, we must extract additional information. Figure 5 displays the contribution plot which shows the amount of contribution towards each dimension from each variable of X

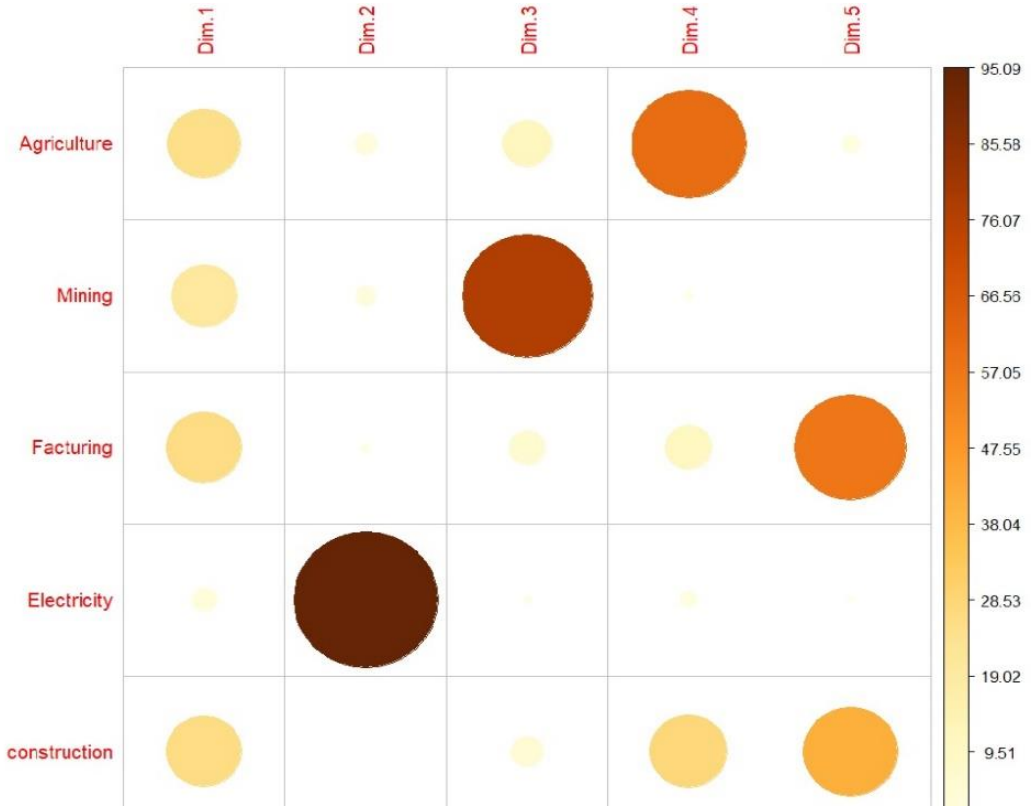

Figure 5: Contribution of each variable towards dimensions or components.

Four variables X_1, X_2, X_3 and X_5 make contributions, more or less the same amount, to dim1. The electricity i.e., X_2, alone forms the dim 2 axis. The variable X_1(Agriculture) contributes more in dim 4. More amount of share from X_2=mining is contributed to dim 3. Among X_3= Manufacturing, and X_5=Construction, each contribute a good amount to dim 5. Hence for it is suggested that best statistical analysis for drawing wise decision making can be based on all these dimensions.

Representation of Sample Points: The results, for individuals can be extracted using the function get_pca_ind() [*factoextra* package]. Similarly to the get_pca_var(), the function get_pca_ind() provides a list of matrices containing all the results for the individuals (coordinates, correlation between individuals and axes, squared cosine and contributions) We can plot chats according the cos2 of the corresponding individuals. We now draw a graph in Figure 6 showing the scatter points all 65 sample in the two-dimensional plot:

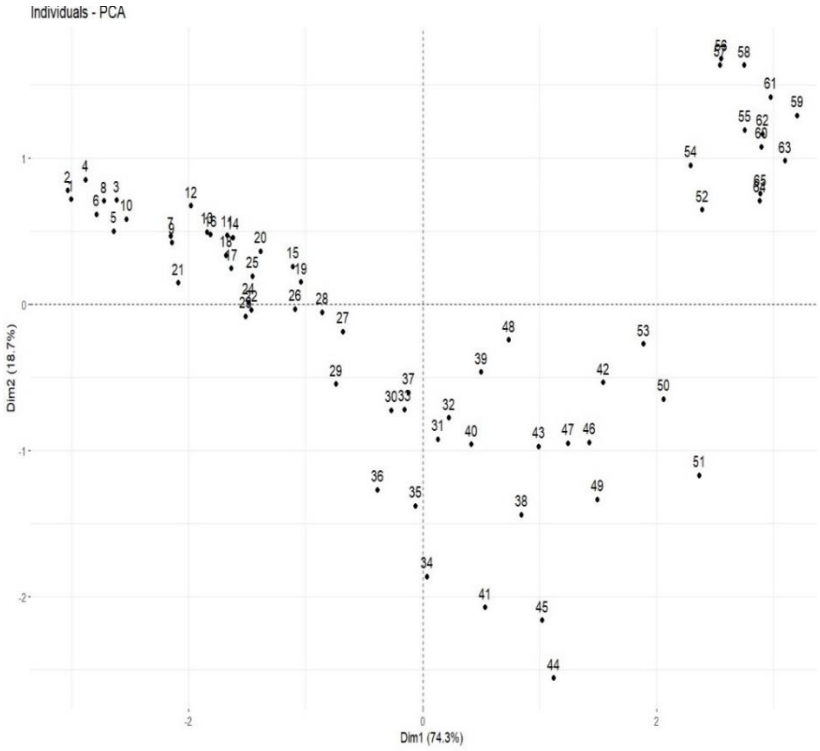

Figure 6: Two-dimensional Scatter plot showing the positions of all sample 65 sample points

From Figure 6 we observe that all 65 sample points scatter over all four quadrants. The cluster points are more in the second and the fourth quadrants.

4. Discussion and Conclusion

The proposed analysis of this article is just an example of how PCA is applied across different fields. Its versatility makes it a valuable tool for exploratory data analysis and feature extraction. Our contributions focused on how machine learning compliments statistical techniques. In particular, we employed the machine learning to find a lower-dimensional representation of the data while preserving its most important features. and aids further understanding of the PCA, LDA and factor analysis.

Those two Principal components PC1 and PC2 that were isolated based on the explained variation via an eigen-decomposition of the covariance matrix for the Botswana's GDP data. It is thus essentially linear combinations of the original data capturing most of the variance in the data and uncorrelated data living in the reduced PCA space such that the first component explains the most variance in the data and the second component explaining less.

The paper discusses how machine learning techniques can be applied to Botswana's GDP to extract the best principal components to help track domestic product performance in African countries, including Botswana. It compares the predictive power of different classifiers with predicting the actual GDP of Botswana and thus provides a good understanding of how machine learning models can be used to predict the GDP of other countries.

References

1. Artis, M., Banerjee, A., & Marcellino, M. (2005). Factor forecasts for the UK. *Journal of Forecasting* , 24, 279-298.

2. Bai, J. N. (2006.). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4), 1133-1150.
3. Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* , 70(1), 191-221.
4. Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, 4 (3), pp.272 – 299.
5. Gabriel, K. (1971). T he Biplot – Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, 58, pp.453 – 467.
6. Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross Spectral Methods. *Econometrica*, 37, pp.424 – 438. .
7. He, L., Yang, Y., & Zhang.B. (2023). Robust PCA for high-dimensional data based on characteristic transformation. *Australian and New Zealand Journal of Statistics*, 65, Issue2, 127-151, https://doi.org/10.1111/anzs.12385.
8. Humphreys, L G; Igen, D. R. (1969). Note on a Criterion for the Number of Common actors. *Educational and Psychological Measurement*, ,29, pp.571 - 578.
9. Humphreys, L. G., & Montanelli, J. (1975). An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors. *Mutivariate Behavioral Research*, 10, pp.193 – 206.
10. Jennrich, R. (2001). A simple General procedure for Orthogonal Rotation. *Psychometrika*, 66 (2), pp.289 – 306. .
11. Johnson, R. A., & Dean, W. (1992). *Applied Multivariate Statistical Analysis.* Third Edition, Upper Saddle River, New Jersey: Prentice – Hall, Inc.
12. Kano, Y. (1990). Noniterative estimation and the choice of the number of factors in exploratory factor analysis. . *Psychometrika*, 55 (2), pp.277 – 291. .
13. Kim, H., & Swanson, N. (2014). Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. *Journal of Econometrics* , 178(2), 352-367.
14. Nakajimay, Y., & Sueishiz, N. (2020). Forecasting the Japanese Macroeconomy Using High-Dimensional Data. ttps://ssrn.com/abstract=3403414 or http://dx.doi.org/10.2139/ssrn.3403414.
15. Pasini, G. (2017). Principal component analysis for stock portfolio management. *.nternational Journal of Pure and Applied Mathematics*, 115(1):153–167. https://doi.org/10.12732/ijpam.v115i1.12.
16. Stock, J., & Watson, M. (2002). Macroeconomic forecastig using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2), 147-162.
17. Stock, J., & Watson, M. (2012). Generalizaed shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics*, 30(4), 481-493.
18. Uematsu, Y., & Tanaka.S. (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *Econometrics Journal* ,, volume 22 , p. 34 - 56 .
19. Zwick, W. R., & Velicer, W. (1986). Factors Influencing Five Rules for Determining the Number of Components to Retain. *Psychological Bulletin*, 99 (3), pp.432 – 442.

1. Appendix

						Trade									
			Manu	Water		Hotels	Transpo	Financia	General	social pe	Value	taxes or	other taxes on		
	Agricultu	Mining	Facturin	Electricit	constr	Restauren	Commu	Service	Governme	services	added	import	Product	Subsdies	Total GDP
Time	Agricultu	Mining	Facturin	Electricit	constr	Restauren	Commu	Service	Governme	services	added	import	Product	Subsdies	Total GDP
2004Q1	185.5	2721.9	569.5	183.1	605	1261.8	346.5	1261.5	1429.6	509.3	9073.3	549.5	681.8	-51.3	10253.3
Q2	317.6	1348.6	582.2	194.3	586	1294.1	337	1253.5	1522.9	513.2	7949.9	768.1	412.6	-57.4	9073.2
Q3	265.1	3647.8	571.9	208.5	608	1279.8	375.6	1389.7	1602.8	537	10486.1	719.3	326.1	-54.7	11476.8
Q4	181.6	3083.2	611.2	220.9	578	1280.5	390.5	1425.8	1596.4	550.7	9918.3	710	666.3	-61.1	11233.4
2005Q1	222.2	4040.3	606.6	160.4	577	1237.7	404.7	1405.4	1615.8	588.7	10858.3	798.4	744.7	-58.3	12343.1
Q2	254.2	2939	620.9	174	595	1323.6	416.5	1427.3	1693.5	610	10054.2	718.5	419.1	-65.2	11126.7
Q3	288.6	5341.3	634.7	183.7	615	1405.6	492.5	1518.6	1852.6	661.1	12993.9	646.9	609	-62.2	14187.7
Q4	162.9	3784.5	649.8	196.7	631	1532	511.4	1567.3	2033.6	684.7	11753.8	1011.1	399.3	-69.5	13094.7
2006Q1	350.2	4483.6	715.1	170.4	665	1676.4	524.4	1648.8	1723.6	718.3	12675.4	766.1	626.8	-66.3	14002
Q2	258.4	3463.4	722.3	177.9	686	1700	550	1649	2001.7	725.2	11933.6	848.6	501.7	-74	13209.9
Q3	355.1	5746.3	812.6	205.4	804	1789.6	569.2	1734.1	2028.3	756.8	14801.2	951.5	533.6	-70.6	16215.7
Q4	247	5325.6	873.1	231.4	684	1866.3	606.9	1643.5	2064.9	738.7	14281.4	862.6	614.2	-78.9	15679.3
2007Q1	319.9	4938.4	938.7	195.3	775	1927	603.9	1897.3	2021.5	761.2	14378	814.2	617	-75.3	15733.9
Q2	478.6	3760.5	1092.6	193.6	859	2037.4	669.8	1907.7	2251.5	789.7	14040.8	1145	672.7	-84.1	15774.4
Q3	456.6	6165.2	1100.1	183.8	899	2035.8	698.4	1974.1	2299.5	854.7	16666.9	1160.1	767.4	-80.2	18514.2
Q4	249.6	4703.6	1075.2	178.8	1011	2243.4	725.1	1873.9	2314.7	850.2	15225.9	1171.3	822.6	-89.6	17130.1
2008Q1	416	4966	940.5	150.8	854	2269.1	779.2	2032.1	2326.4	900.4	15634.3	1130.1	805.7	-85.5	17484.6
Q2	452.7	3584.5	1123.3	159	918	2452.9	806.5	2171.4	2782.2	906	15356.7	1171.4	854.4	-95.5	17286.9
Q3	558.8	5395.9	1132.1	161.1	948	2583.8	844.7	2514.3	2864.1	1003	18006	1194.7	1039.3	-91.1	20148.9
Q4	460	4697.1	1083	183.2	1007	2741.2	873.5	2598.9	2926.9	1037.9	17608.6	1205.1	1088.6	-101.8	19800.4
2009Q1	428.6	1950.5	1100.1	83.1	1061	2725.5	909.6	2374.6	2881.6	1123.8	14637.9	1007.4	1047.9	-97.1	16596
Q2	535.8	3272.8	1202.2	79.2	1138	2882.6	950.4	2365.8	3093	1122.8	16642.5	1089	1108.7	-108.5	18731.5
Q3	603.5	2519.6	1147.9	65.2	1200	2841.3	1032	2377.3	3156.2	1191.7	16134	1155.4	1286	-103.5	18471.9
Q4	503.1	3466.7	1212.1	89.6	1097	2713.4	1102	2577.1	3185.5	1154.9	17101	1283.6	1394.1	-115.7	19663
2010Q1	490.8	3242.2	1277.1	128.2	1153	2943	1041	2703.1	3136.9	1261.5	17376.6	1082.3	1257.7	-98.9	19617.7
Q2	595	3759.1	1357.3	101.1	1151	3267.1	1080	2881.7	3179	1302.7	18673.9	1049.8	1301.7	-110.5	20914.9
Q3	580.3	5085.7	1412.3	85.5	1351	3496.6	1165	2911.5	3338.1	1355.8	20782.1	1102.3	1327.7	-105.4	23106.7
Q4	495.2	4573.8	1501.6	97.1	1401	3377.5	1185	3113.3	3723.4	1322.8	20790.5	1190.5	1364.9	-117.8	23228.1
2011Q1	499.9	5824.8	1362.8	-0.1	1425	3821.3	1201	3162.6	3021	1328.4	21646.7	897.4	1433.1	-114.3	23862.8
Q2	667.8	6113.8	1426.5	-11.2	1473	3716.6	1235	3352.9	3725.1	1441.8	23141.1	1412.1	1473.4	-102.2	25924.5
Q3	716.1	6598.8	1585	-39.6	1638	3839.1	1325	3624.8	3677.8	1492.5	24457.6	1412	1201.6	-113.9	26957.3
Q4	752.2	5991.6	1699.8	-7.9	1712	4218.3	1353	3838.5	4353.7	1618.4	25529.1	1333.9	1490.2	-117.9	28235.3
2012Q1	728	4617.1	1601.4	-23.3	1831	4211.2	1509	3790.4	3681.4	1609.6	23555.9	1235	1368.9	-115.8	26044.1
Q2	796	5524	1627.7	-265.7	1859	4173.6	1565	4072.5	4145.2	1687.8	25184.9	1267.2	1390.5	-116.6	27726
Q3	808.9	4443.1	1668.5	-167.5	1879	4258.9	1617	4235	4268.4	1709.6	24721	1431	1509	-119.2	27541.9
Q4	628.4	4702.3	1625.6	-164.7	1896	4253	1634	4369.9	4828.7	1743.1	25516.2	1627.2	1536.4	-121.4	28558.5
2013Q1	649.3	4670.5	1725.8	-4.6	1954	4822	1662	4296.6	3981.8	1742.6	25500.4	1580.3	1468.5	-124.4	28424.8
Q2	780.6	7933.6	1784.4	-126.1	1989	5006.7	1677	4377.4	4346.5	1833.7	29603	1511	1472.5	-125.6	32461
Q3	704	6124	1865.3	63.5	2000	5451.6	1750	4424.3	4422.3	1864.2	28668.7	1571.7	1557.5	-132.3	31665.4
Q4	743	5515.3	1909.6	-56	2055	5931	1817	4618.7	4980.4	1903.6	29417.6	1740.1	1584.8	-135.4	32607.2
2014Q1	721.2	7270.3	1784.5	-295	2140	6378.4	1924	4637.1	4717	1982.8	31260	1759.4	1636.5	-141.5	34514.5
Q2	774.2	9402	1898.4	110.9	2205	6553.9	1922	4734.7	4875.5	2020.5	34497.3	1714.3	1654.3	-140	37725.9
Q3	762.1	7289.1	1986.3	-25.5	2161	6674.8	1954	4905.1	5055.7	2053.3	32816.8	1788.3	1694	-143.4	36155.7
Q4	787.8	8441.8	2071.1	-371.9	2184	6575.3	1987	5007.4	5297.8	2081.9	34062.3	1845.4	1706.6	-141.8	37472.5
2015Q1	771.8	7620	2044.9	-296.6	2331	6132.6	2022	5063.6	5376.3	2118.8	33184.7	1836.3	1690.8	-148.9	36562.8
Q2	812.5	7646.1	2102.8	-4.2	2421	5747.6	2089	5324.4	5562.5	2139.5	33841.6	1887.1	1768	-153.4	37343.2
Q3	799.4	6937	2131.5	-18.1	2403	5777.7	2199	5476.1	5783.2	2175.4	33664.2	1932.7	1857.1	-158	37296
Q4	831.3	3759.1	2162.5	99.8	2452	6053.5	2248	5653.4	5769.1	2219.2	31248.1	1885.5	1889.3	-158.7	34864.1
2016Q1	849.7	7567.4	2128.8	-91.2	2526	6947.6	2253	5591.8	5901.7	2254.4	35929.4	2015.3	1862.4	-164.5	39642.6
Q2	888	8763.6	2171.1	95.4	2613	8007.8	2370	5741.1	5969.6	2289.1	38908.7	1988.4	1952.9	-169.4	42680.6
Q3	900.5	9785.3	2268.3	-6.2	2708	8059.4	2483	5940.5	6085.7	2333.9	40558.4	2008.7	2004.7	-174.5	44397.3
Q4	857.9	8796.3	2291.2	397	2729	7982.3	2536	6075.4	5969.9	2381.9	40016.4	2041.9	1962	-176.7	43843.6
2017Q1	860.6	7768.1	2155.8	162	2732	8934.3	2576	6120.8	6092.9	2413.1	39815.4	2099	1921.8	-178.1	43658.1
Q2	886.2	7773.5	2266.7	454.6	2816	8388.7	2612	6172.5	6492.1	2455.6	40317.5	2157.2	2031.3	-179.4	44326.6
Q3	912.3	8607.3	2384.3	531.2	2950	8179.8	2695	6287.9	6583.9	2499.4	41630.4	2197.4	2229.3	-180.8	45876.4
Q4	924.7	7112.8	2407.6	622.4	2990	9556.3	2741	6468.4	6670.7	2531.3	42024.9	2202.7	2196.1	-182.2	46241.5
2018Q1	932.2	7336.1	2309.3	614.8	3005	8703.9	2816	6611.4	6714.8	2565.7	41608.6	2306.3	2173.5	-187.7	45900.7
Q2	957.1	7418.1	2404	621	3110	9385.9	2833	6659.5	6842.1	2600.9	42832	2257.6	2110.6	-193.3	47006.9
Q3	964.3	8912.8	2515.4	568.5	3215	8952.5	2915	6734	6944.6	2648.8	44371	2266.9	2396.2	-199.1	48835.1
Q4	940.7	7747.5	2566	496.9	3244	9610.3	2945	6884.3	7036.9	2686.8	44158.7	2282.5	2386.5	-205.1	48622.6
2019Q1	950	8202.7	2426.9	581.9	3256	9298.7	2966	6968.9	7122.4	2715.3	44487.9	2391.1	2346.3	-211.2	49014.1
Q2	962.1	7347.1	2565.2	513.6	3320	9978.1	2980	7106.7	7230	2744.7	44747.9	2364.2	2282.4	-217.6	49177
Q3	976.2	7649.1	2661	479.8	3404	9569.5	3072	7199.2	7279	2777.7	45066.9	2272.8	2471.1	-222.1	49588.8
Q4	957	6881.6	2699.5	401.3	3443	9919.5	3112	7338.3	7388.4	2808.5	44949.6	2295.4	2467.2	-223.7	49488.5
2020Q1	887.8	7840.8	2586.8	415.4	3462	9911.2	3145	7485.6	7563.6	2836.5	46134.7	2328	2486.5	-222.5	50726.8

2. Conflict of Interest

There is no conflict of interest with anybody/any organization.

3. Acknowledgement

ALL four authors thank the institutions of the University of Botswana for providing us with the infrastructure without financial sources.