ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

Machine Learning-Based Predictive Analysis on Electronic Billing Machines to Value Added Tax Revenues Growth

Patrick Mazimpaka[1], Prof. Sanja Michael (PhD)[2]
Master of Science in Information Technology, University of Kigali

Abstract:- This paper explores the application of predictive analysis in assessing the impact of Electronic Billing Machines (EBMs) on Value Added Tax (VAT) revenue growth in Rwanda. EBMs represent a significant innovation in VAT revenue collection, yet negative taxpayer perceptions and a lack of understanding impede their full potential. Without predictive analysis, accurately projecting the growth in VAT revenues due to EBM implementation remains a challenge for the Rwanda Revenue Authority.

Using machine learning techniques and historical data obtained from Rwanda Revenue Authority Annual reports, this research aims to develop a predictive system capable of forecasting the influence of EBMs on VAT revenue growth. Employing a descriptive survey design approach, secondary data analysis was conducted to gather information on VAT taxpayers, EBM users, VAT revenues, and other relevant variables.

The study reveals a notable increase in EBM adoption among VAT payers over a six-year period, indicating successful governmental efforts. Time series analysis demonstrates a positive correlation between EBM usage and significant increases in both VAT and total tax revenues. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model, specifically an ARIMA (1, 0, 0) model, is identified as suitable for forecasting VAT revenue growth due to its balance between accuracy and simplicity. The developed predictive system provides highly accurate forecasts of VAT revenue growth, facilitating informed fiscal policymaking and enhancing financial stability and revenue collection in Rwanda.

Keywords:- *Electronic Billing Machine (EBM), Value Added Tax Revenue (VAT), Revenue Growth, Forecasting, Predictive Model, Tax Payers, Rwanda Revenue Authority*.

I. INTRODUCTION

One of the best tools for increasing the public sector's performance potential, funding social insurance programs, and paying off public debt is taxation. The ability of a nation to levy additional taxes, both economically and administratively, determines how much money it can generate (Raymond, 2014). According to Kaldore (1963), a nation must raise tax revenue at a rate higher than the typical 10-15% found in developing nations in order to advance economically.

Since August,2013 all VAT-registered taxpayers in Rwanda are required by VAT Law No. 37/2012 of September 9, 2012, Article 24, to obtain and utilize Electronic Billing Machine to issue tax invoices to their clients on each transaction they complete. Tax crimes that violate the law have harsh penalties. With the support of this system, businesses will be assisted in maintaining accurate records, revenue protection will be improved, and sincere taxpayers would be shielded from unfair competition. (RRA, 2022).

However, despite their introduction, negative perceptions among taxpayers persist, primarily due to a lack of understanding regarding the specific influence of EBMs on VAT revenue generation. This lack of clarity poses a substantial challenge for the Rwanda Revenue Authority in accurately forecasting the potential growth in VAT revenues resulting from EBM utilization. In light of this, I have determined that machine learning models are required in order to evaluate and forecast the effects of employing EBM on future revenue growth.

Machine learning models are not 100% accurate, but with a good Machine Language model ranging between 95-99%+ accuracy on test data, you can have a great deal of confidence in the forecasts they produce. You need a lot of high-quality data that is precisely labeled with the output in order to train an ML model. To evaluate the model's accuracy on data that has already been labeled, a subset of the data, typically 10–20%, is taken before training. Once the model has been trained and tested, it can be used to analyze data with uncertain results to produce a forecast of future effects (people.ai, 2023).

A. Value Added Tax

Value-Added Tax (VAT) is a tax, which is payable on sales of goods or services within the country. The tax, in all cases, is ultimately payable by the final consumer of the good or service. Each party in the chain of supply (manufacturer, wholesaler and retailer) acts as a VAT collector. They collect VAT from their customer and include that VAT in their VAT return to Revenue. When returning the VAT collected, they can reclaim as appropriate, VAT which has been charged to them by their suppliers. (www.revenue.ie, 2023)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

B. Electronic Billing Machine

EBM is a digital tool and system made to track sales transactions, produce invoices, and make tax and value added compliance easier. It is also a particular system that the Rwanda income Authority (RRA) has put in place to improve tax compliance in the nation and streamline the processes for collecting income. Electronic billing machines (EBMs) give revenue authorities the ability to keep an eye on official business activities, perhaps improving VAT compliance. (IGC, 2023).

C. Application of EBM in Rwanda

- *Enhanced Tax Compliance:* EBMs are mandated by the Rwanda Revenue Authority (RRA) for use by registered taxpayers, primarily in the retail and service sectors. By using EBMs, businesses are required to record sales transactions electronically, ensuring accurate and transparent reporting of revenue. This reduces the possibility of under-reporting or manipulation of sales figures, leading to increased tax compliance and revenue collection.
- *Real-Time Data Reporting:* EBMs in Rwanda are connected to the RRA's central server through an online interface. This allows for real-time reporting of sales data to tax authorities. By transmitting data directly to the RRA, tax officials can monitor sales activities promptly, verify VAT collections, and identify any discrepancies or irregularities. Real-time data reporting facilitates efficient revenue monitoring and enforcement.
- *Simplified Tax Calculation and Invoicing:* EBMs simplify tax calculations for businesses by automatically applying the appropriate VAT rate to sales transactions. This reduces the likelihood of errors in VAT calculations and ensures accurate invoicing. The generated invoices contain all the necessary details, including the business's information, the buyer's details, the VAT amount charged, and other relevant transaction information. This simplification streamlines the invoicing process and promotes compliance with tax regulations.
- *Audit Trail and Compliance Verification:* EBMs maintain a digital record of all sales transactions, providing a reliable audit trail. Tax authorities can access this data during tax inspections or audits to verify the accuracy of VAT reporting. The availability of electronic records through EBMs simplifies compliance verification, enabling tax officials to cross-check the sales data with other financial records and identify any discrepancies or potential tax evasion.
- *Prevention of Revenue Leakages:* By implementing EBMs, the RRA can minimize revenue leakages resulting from under-reporting or manipulation of sales transactions. The electronic recording of sales data provides greater transparency, making it difficult for businesses to evade taxes. The use of EBMs contributes to a more efficient and effective revenue collection process, helping the government maximize its tax revenues. (RRA, 2019)

D. Significance

A study will make a significant contribution by identifying the contribution of Electronic Billing Machine to Value Added Tax revenues growth and forecasting VAT revenue growth as an impact of implementing EBM in Rwanda. When the government receives the funds required to finance its activities as intended, a tax administration is effective. Therefore, an administration is said to be efficient if it accomplishes its goal for the government and taxpayers at a fair, minimal cost.

The outcomes from this research will be used by Rwanda Revenue Authority for public awareness to convince the taxpayers (VAT registered) to use EBM instead of using punishments and penalties. It will use it also to make future plan related to Value Added Tax collection and management according to the results of prediction.

II. EMPIRICAL REVIEW

The introduction of VAT can be seen as one of the most significant developments in tax policy of recent decades (Simone, 2004). VAT adoption has increased from 47 countries in 1990 to over 140 today (IMF, 2011). Over this period, the VAT has rapidly grown to become the greatest portion of domestic tax collection. For instance, in 2011 VAT collections made up 24% of tax revenue in the Southern African Development Community and 36% of tax revenue within the East African Community (IMF, 2013). In 15 African nations, consumption taxes (VAT and excise levies) now contribute more than 36% of total revenue, more than any other tax. (ATAF, 2016).

According to Victor Steenbergen, Improving VAT compliance is one of the most critical issues for domestic revenue mobilization in developing countries for two reasons. First of all, since VAT collection tends to outpace all other tax revenue streams, even minor changes can have a large effect. A revenue authority's capacity to enforce tax compliance for all domestic tax types is strengthened by the paper trail created by the data generated to observe VAT liability.

In August 2013, Rwanda adopted a new law that stated that all businesses registered for VAT must provide customers, at each sale, a certified VAT receipt generated by Electronic Billing Machine, which contains a Sales Data Controller (SDC) with GPRS and a Certified Invoicing System (CIS) all working together. This must be purchased from a Rwanda Revenue Authority (RRA)-approved vendor and activated by the RRA. The findings of the study of Internal Growth Center in Rwanda showed that on average, the introduction of EBMs resulted in a VAT increase of 5.4 percent. This was relatively little, and much lower than expected by the Rwandan Revenue Authority.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

International Growth Center Rwanda was asked by the Rwanda Revenue Authority and Ministry of Finance to evaluate the impact of Electronic Billing Machines (EBMs), to help the Government of Rwanda understand if the expense of the machines is a worthwhile investment, how the EBMs can be made optimally cost-effective, and how they will increase revenues.

The study finds that adoption has spread quickly and reached 77.8% of tax-paying firms, but was slowing. The average impact of EBM on firms' VAT payments was estimated at 5.4%. Estimated impacts vary substantially by sector and size. The mystery shopper study found that - at least for low-cost goods in Kigali retail stores EBM utilization is low, but utilization increased substantially when consumers requested a receipt – and with utilization revenues increased. Taken together, the findings suggest that future strategies could appropriately be focused on concentrating the expansion of EBM coverage on specific sectors where both adoption rates are low and potential impacts are highest, while building on existing policies to strengthen firms' incentives to report transactions through EBMs.

The study was summarized for the Minister of Finance and delivered in detail to the executive management of the RRA, and the Chief Economist and Minister of MINECOFIN. The RRA requested that the International Growth Center undertake a more in-depth study to analyze variations in EBM compliance to guide tax auditing and compliance monitoring (IGCenter, 2014).

III. METHODOLOGY

A. Data Collection

During this study, the secondary data analysis will be conducted where the researcher will use collected from Rwanda Revenue Authority annual reports from 2015/16 fiscal year to 2020/21 fiscal. Within Rwanda Revenue annual reports that will be used by researcher, he will focus on the number of Value added tax payers, electronic billing machine users, collected value added taxes, contribution of revenues to the total collected revenues in percentage in each fiscal and come up with data set that will facilitate him to asses and examine the impact of Electronic Machine to the Revenue growth by comparing the years. The researcher will also forecast the future Value Added Taxes and predict the revenues growth based on the increase and usage of Electronic Billing Machine with respect to the findings among the fiscal years.

B. Research Design

The research design for this study encompasses both descriptive survey and predictive modeling approaches to forecast the growth of value-added tax (VAT) revenues and assess the influence of Electronic Billing Machines (EBMs) on VAT revenues, specifically focusing on EBM users across various fiscal years. This design integrates the exploration of current conditions (descriptive survey) with the development of predictive models to anticipate future outcomes, facilitating a comprehensive analysis of the relationship between EBM utilization and VAT revenue generation.

C. Study Population and Sample Size

For this study, the population of interest comprises taxpayers in Rwanda, specifically those registered for Value Added Tax (VAT) as documented by the Rwanda Revenue Authority for each fiscal year spanning from 2015 to 2021. The study will primarily focus on this group to analyze their characteristics and behaviors in relation to VAT compliance and the utilization of Electronic Billing Machines (EBMs).

Due to the use of secondary data from Rwanda Revenue Authority reports, the target population is not limited to VAT Registered tax payers using Electronic Billing Machines (EBMs). Instead, it includes all EBM users and VAT tax payers fall under the purview of RRA's reporting.

Analysis used the available secondary data from Rwanda Revenue Authority reports to understand the impact of EBMs on VAT revenues across the entire population of VAT-registered tax payers in Rwanda and EBM users. The analysis involved machine learning models for predicting and assessing the relationship between EBM usage and VAT revenue growth.

Table 1: Population and Sample Data

Financial Year	All tax payers	VAT payers	EBM users	Non-EBM Users	VAT Revenues (Rwf Bn)	Total Tax Revenues (Rwf Bn)
2015-2016	136334	16,047	11,434	4,613	323.2	986.7
2016-2017	212320	16,898	14,518	2,380	352	1,086.50
2017-2018	172006	22,296	17,681	4,615	405.7	1,234.10
2018-2019	195425	24,035	19,516	4,519	458.7	1,399.50
2019-2020	231463	28,248	23,885	4,363	491.5	1,494.80
2020-2021	238522	33,535	32,848	687	531.4	1,635.80

D. Data Analysis

- *Time Series*

Time series forecasting is one of the most applied data science techniques in business, finance, supply chain management, production and inventory planning. Many prediction problems involve a time component and thus require extrapolation of time series data, or time series forecasting. Time series forecasting is also an important area of machine learning (ML) and can be cast as a supervised learning problem. Time series forecasting means to forecast or to predict the future value over a period of time. It entails developing models based on previous data and applying them to make observations and guide future strategic decisions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

The future is forecast or estimated based on what has already happened. Time series adds a time order dependence between observations. This dependence is both a constraint and a structure that provides a source of additional information. Before we discuss time series forecasting methods, let's define time series forecasting more closely.

There are two main types of decomposition which are decomposition based on rates of change and decomposition based on predictability.

- *Decomposition Based on Rates of Change:* is an important time series analysis technique, especially for seasonal adjustment. It seeks to construct, from an observed time series, a number of component series (that could be used to reconstruct the original by additions or multiplications) where each of these has a certain characteristic or type of behavior. If data shows some seasonality (e.g. daily, weekly, quarterly, yearly) it may be useful to decompose the original time series into the sum of three components: **Y(t) = S(t) + T(t) + R(t)**

Where **S(t)** is the seasonal component, **T(t)** is the trend-cycle component, and **R(t)** is the remainder component.

- *Decomposition Based on Predictability:* The theory of time series analysis makes use of the idea of decomposing a time series into deterministic and non-deterministic components (or predictable and unpredictable components). In statistics, Wold's decomposition or the Wold representation theorem, named after Herman Wold, says that every covariance-stationary time series can be written as the sum of two time series, one deterministic and one stochastic. Where **Yt** is the time series being considered, **Et** is an uncorrelated sequence which is the innovation process to the process — that is, a white noise process that is input to the linear filter **{bj} b** is the possibly infinite vector of moving average weights (coefficients or parameters), **nt** is a deterministic time series, such as one represented by a sine wave.

➢ *ARMA and SARMA Models*

Autoregressive Moving Average (ARMA) and Seasonal Autoregressive Moving Average (SARMA) models are used in time series analysis to capture the underlying patterns and relationships within a dataset. ARMA models extract features related to the autoregressive component, which capture the linear relationship between an observation and a certain number of lagged observations. This component reflects how the current value of the time series is influenced by its past values. SARMA models include features related to the seasonal autoregressive component, which captures the influence of seasonal patterns on the time series data. This component accounts for periodic fluctuations that occur at regular intervals, such as daily, monthly, or yearly patterns. It also extracts features related to the seasonal moving average component, which models the relationship between an observation and a linear combination of past error terms with seasonal effects.

➢ *Flowchart of the Predictive System*

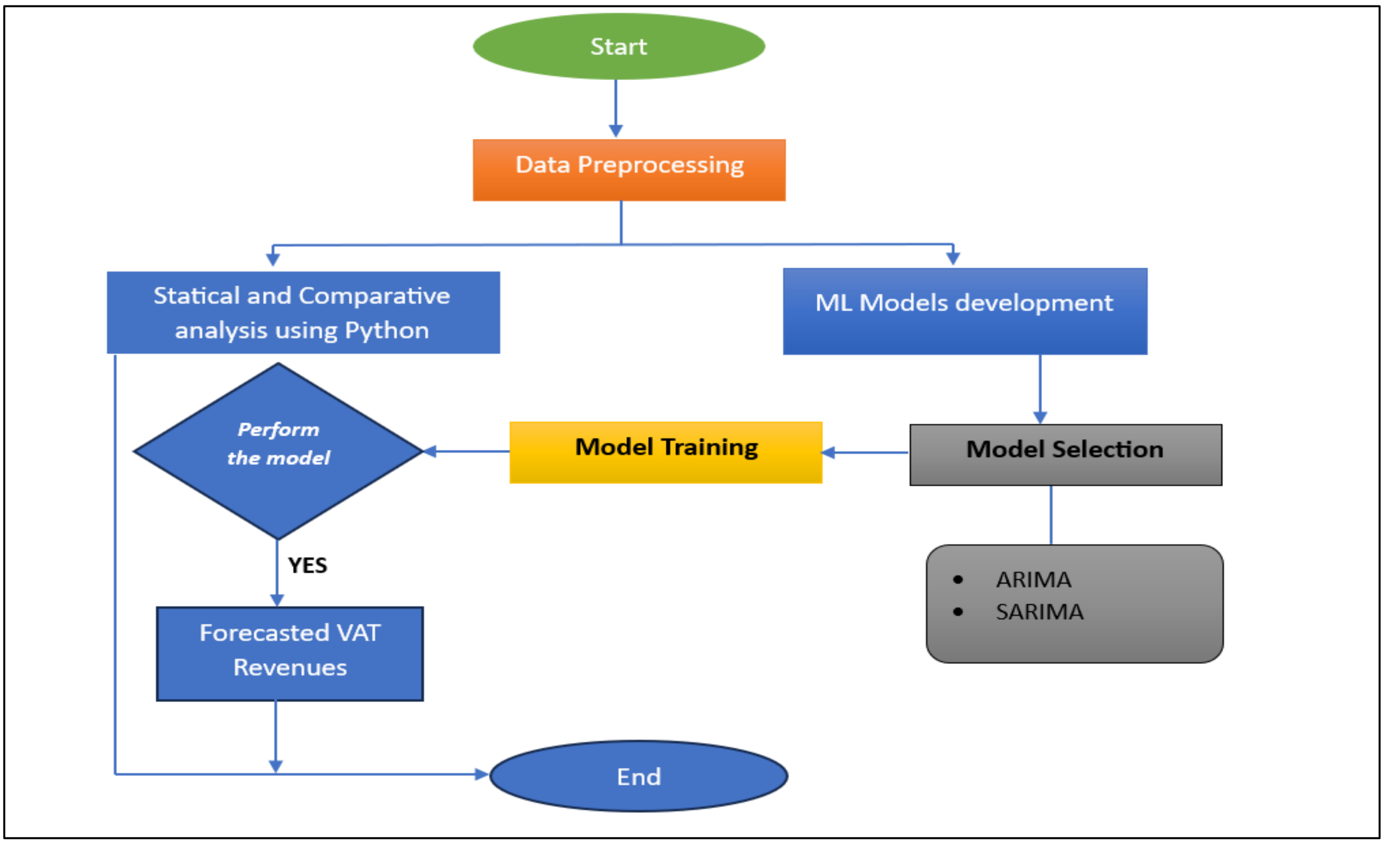

Fig 1: Flowchart of Predictive System

Volume 9, Issue 3, March – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

IV. RESULTS AND FINDINGS

A. Examining effect of EBMs on the Growth of VAT Revenues in Rwanda

The statistical analysis revealed that the rate of EBM users among VAT payers was on **71.3%** in the fiscal year of 2015/2016. Due to the efforts made by the government of Rwanda, the number of EBM users among VAT payers increased to **98%** which means that the VAT payers have successful adopted the government's wish of using electronic billing machine.

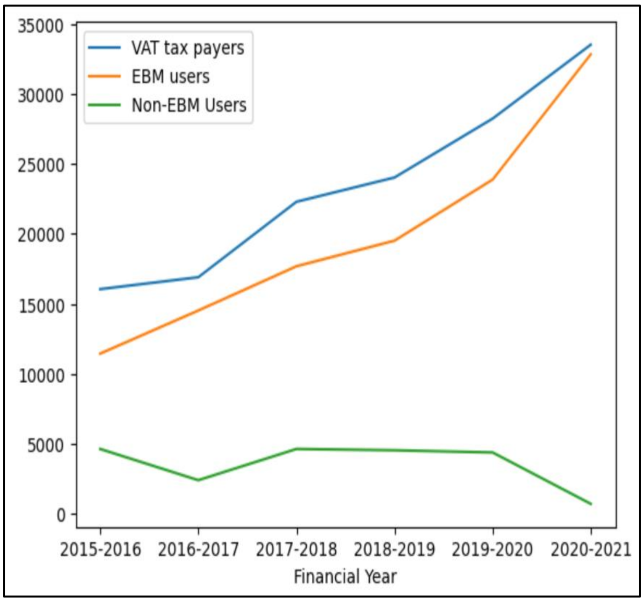

Fig 2: EBM Adoption within VAT Payers

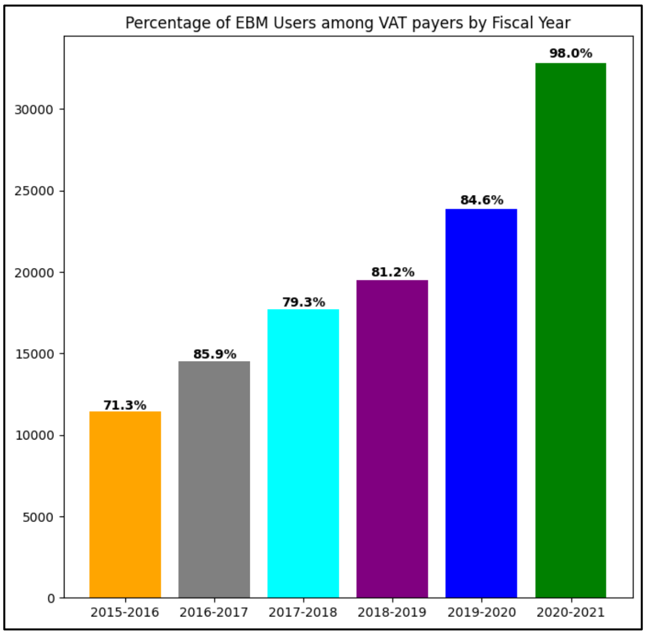

Fig 3: EBM users within VAT Payers

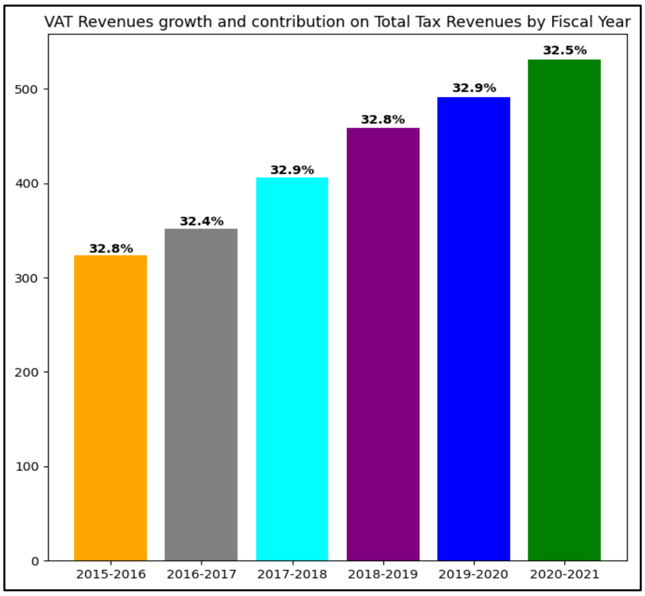

Fig 4: VAT Contribution to Tax Revenues Per Year

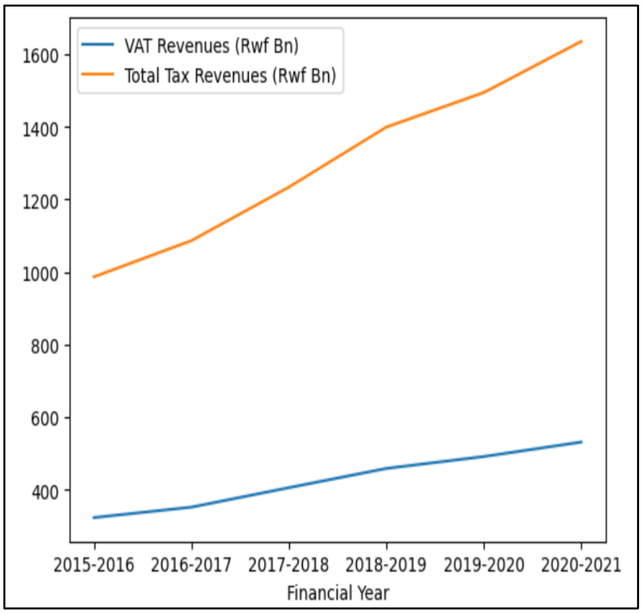

Fig 5: VAT Revenues Trends

Between 2015 and 2021, the trends in VAT revenue depicted a degree of fluctuation without a discernible upward or downward trajectory. VAT revenue emerged as a significant component of total tax revenue, albeit not constituting the majority portion. The graph illustrates the year-over-year growth in VAT revenue, exemplified by a 0.4% increase between fiscal years 2015-2016 and 2016-2017. Additionally, the percentage of total tax revenue attributed to VAT, with fiscal year 2020-2021 recording VAT revenue contributing 32.9% to the overall tax revenue collection

.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

B. Building and Valuating the Accuracy and Reliability of the Predictive Model

```
from statsmodels.tsa.stattools import adfuller
test_result=adfuller(vat_df['VAT'])
ef adfuller_test(vat):
    result=adfuller(vat)
    labels = ['ADF Test Statistic','p-value','#Lags Used','Number of Observations Used']
    for value,label in zip(result,labels):
        print(label+' : '+str(value) )
    if result[1] <= 0.05:
        print("strong evidence against the null hypothesis(Ho), reject the null hypothesis. Data has no unit root and is stationary")
    else:
        print("weak evidence against null hypothesis, time series has a unit root, indicating it is stationary ")
adfuller_test(vat_df['VAT'])
```

```
ADF Test Statistic : -0.11127555604939436
p-value : 0.9483037988479045
#Lags Used : 0
Number of Observations Used : 5
weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary
```

```
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(vat_df.VAT, order=(1,0,0))
fitted = model.fit()
fitted.summary()
```

SARIMAX Results

Dep. Variable:	VAT	No. Observations:	6
Model:	ARIMA(1, 0, 0)	Log Likelihood	-31.757
Date:	Tue, 12 Sep 2023	AIC	69.515
Time:	10:36:24	BIC	68.890
Sample:	06-30-2016	HQIC	67.014
	- 06-30-2021		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	427.0883	80.029	5.337	0.000	270.234	583.942
ar.L1	0.8879	0.358	2.481	0.013	0.186	1.589
sigma2	1788.5383	3529.542	0.507	0.612	-5129.237	8706.314

Ljung-Box (L1) (Q):	0.43	Jarque-Bera (JB):	2.09
Prob(Q):	0.51	Prob(JB):	0.35
Heteroskedasticity (H):	1.37	Skew:	-1.43
Prob(H) (two-sided):	0.84	Kurtosis:	3.46

Fig 6: Results of Built Model

The SARIMA model results indicate that a straightforward autoregressive model with a first-order lag, denoted as ARIMA(1, 0, 0), is employed for forecasting the time series data. The log likelihood, though not directly interpretable on its own, suggests a relatively good fit of the model to the observed data. In terms of model selection criteria, both the AIC and BIC exhibit values of 69.515 and 68.890, respectively, which are lower than other competing models, indicating a favorable trade-off between model fit and complexity. Similarly, the HQIC value of 67.014, being lower, also points towards the suitability of this model. Overall, these findings suggest that the ARIMA(1, 0, 0) model is a reasonable choice for forecasting the time series data, given its ability to provide a good balance between accuracy and simplicity.

C. Forecasting Future VAT Revenue Growth

The developed model has been asked to predict and forecast VAT revenue growth within the next 5 years starting from 2021 financial/fiscal year and the following are the results.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

	VAT	vat First Difference	Seasonal First Difference	forecast
2016-06-30	323.2	0.0	0.0	NaN
2017-06-30	352.0	0.0	0.0	NaN
2018-06-30	405.7	0.0	0.0	NaN
2019-06-30	458.7	0.0	0.0	NaN
2020-06-30	491.5	0.0	0.0	NaN
2021-06-30	531.4	0.0	0.0	NaN
2021-06-30 00:00:00	NaN	NaN	NaN	653.100081
2022-06-30 00:00:00	NaN	NaN	NaN	681.900081
2023-06-30 00:00:00	NaN	NaN	NaN	735.600081
2024-06-30 00:00:00	NaN	NaN	NaN	788.600081
2025-06-30 00:00:00	NaN	NaN	NaN	821.400081
2026-06-30 00:00:00	NaN	NaN	NaN	983.000162

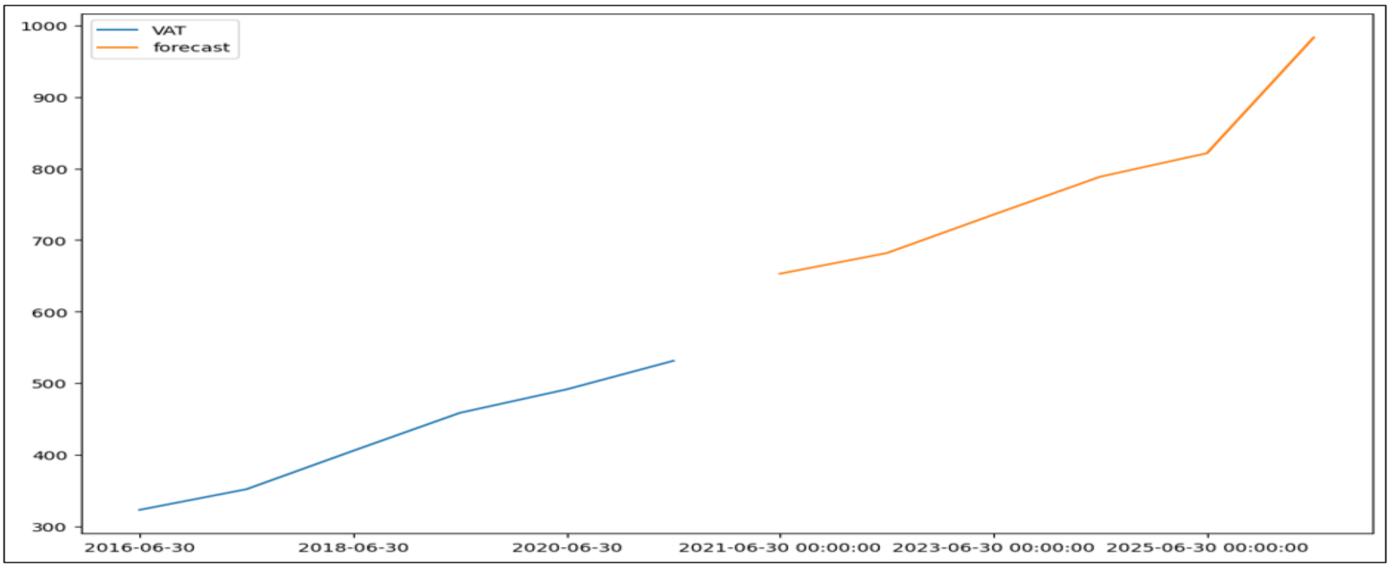

Fig 7: VAT Revenues Forecasting

The development of a predictive system utilizing machine learning models to create a forecasting model for future VAT revenue growth, taking into account the influence of Electronic Billing Machines (EBMs), has provided a powerful tool for revenue projection.

This predictive system showcases its efficacy by offering highly accurate forecasts of VAT revenue growth. It provides invaluable support for decision-makers in crafting informed fiscal policies and resource allocation strategies. The results demonstrate that the adoption and utilization of EBMs have a discernible and positive effect on VAT revenue growth, ultimately fostering enhanced financial stability and revenue collection in the designated context.

V. CONCLUSION

In conclusion, this study provides compelling insights into the impact of Electronic Billing Machines (EBMs) on Value Added Tax (VAT) revenue growth in Rwanda, aligning closely with the objectives set forth. The findings reveal a significant rise in EBM adoption among VAT payers over the studied six-year period, reflecting successful governmental efforts to promote their usage.

Through meticulous time series analysis, a clear and positive correlation emerges between EBM adoption and increased VAT and total tax revenues, affirming the objectives of examining the effect of EBMs on revenue growth. The SARIMA modeling results further validate the study's objectives by identifying an optimal forecasting model (ARIMA(1, 0, 0)) that effectively captures the observed data trends. Moreover, the development of a machine learning-based predictive system demonstrates the potential for accurate VAT revenue growth forecasts, fulfilling the objective of creating a predictive model capable of forecasting future revenue growth based on the influence of EBMs. In summary, this research underscores the efficacy of EBM adoption as a driver for revenue generation, contributing to improved financial stability and government revenue collection in Rwanda.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAR1503

REFERENCES

[1]. Raymond, O. &. (2014). The Impact of E-Taxation on Revenue Generation in Enugu, Nigeria. International Journal of Advanced Research, 1.

[2]. RRA. (2022, April). electronic_billing_machine_general_obligation.pdf. Retrieved from www.rra.gov.rw: https://www.rra.gov.rw/fileadmin/user_upload/electronic_billing_machine_general_obligation.pdf

[3]. people.ai. (2023, April). 6 Sales Forecasting Methodologies to Better Predict Revenenue. Retrieved from https://www.people.ai/: https://www.people.ai/blog/sales-forecast#:~:text=Machine%20learning%20in%20sales%20forecasting,and%20real%2Dtime%20performance%20data.

[4]. www.revenue.ie. (2023, February 09). what-is-vat. Retrieved from www.revenue.ie: https://www.revenue.ie/en/vat/what-is-vat/index.aspx

[5]. IGC. (2023, January). reaping-benefits-electronic-billing-machines. Retrieved from www.theigc.org: https://www.theigc.org/collections/reaping-benefits-electronic-billing-machines

[6]. RRA. (2019, November 19). explanation-of-ebms. Retrieved from tax-handbook.rra.gov.rw: https://tax-handbook.rra.gov.rw/handbook/explanation-of-ebms/.

[7]. Simone, K. a. (2004). Tax Policy in Developing Countries: Some. IMF.

[8]. IMF. (2011). Revenue Mobilisation in Developing Countries, Policy Paper. INTERNATIONAL MONETARY FUND

[9]. IGCenter. (2014, April). evaluation-electronic-billing-machines-effectiveness-tax-administration. Retrieved from theigc.org: https://www.theigc.org/collections/evaluation-electronic-billing-machines-effectiveness-tax-administration