ASHESI UNIVERSITY

MACHINE LEARNING FOR PREDICTING ECONOMIC GROWTH

HOW CAN ZIMBABWE ACHIEVE VISION 2030?

UNDERGRADUATE THESIS

B.Sc. Management Information Systems

KUDAKWASHE GODKNOWS NDEBVUDZEMENE

2020

ASHESI UNIVERSITY

MACHINE LEARNING FOR PREDICTING ECONOMIC GROWTH

HOW CAN ZIMBABWE ACHIEVE VISION 2030?

UNDERGRADUATE THESIS

A thesis submitted to the Department of Computer Science and Information Systems, Ashesi University in partial fulfillment of the requirements for the award of Bachelor of Science degree in Management Information Systems

KUDAKWASHE GODKNOWS NDEBVUDZEMENE

MAY 2020

DECLARATION

I hereby declare that this undergraduate thesis is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:

…………………………………………………………………………………………...

Candidate's Name:

…………………………………………………………………………………………...

Date:

…………………………………………………………………………………………...

I hereby declare that the preparation and presentation of this undergraduate thesis were supervised in accordance with the guidelines on the supervision of undergraduate thesis laid down by Ashesi University.

Supervisor's Signature:

……………………………………………………………………………………………

Supervisor's Name:

……………………………………………………………………………………………

Date:

……………………………………………………………………………………………

Acknowledgement

I would like to dedicate this thesis to my late mother, who, through her diligent hard work, inspired me to pursue my education and reach this level. I wish to realize her hopes and dreams for me by submitting this thesis as a fulfillment towards attaining my university degree. Maybe Chakukura may your beautiful soul continue to rest in peace.

I would like to acknowledge my supervisor, Mr. Eric Ocran, for the guidance and direction to accomplish this research. I further recognize Prince Baah and David Sasu for their academic advice that helped me undertake this study. I am grateful for their unwavering support.

To the Nyikadzino family, I extend my gratitude for your gracious hospitality during the height of the COVID-19 Coronavirus fears in Ghana. With the global pandemic disrupting and threatening not only my university studies and the completion of this work but the very way of life as we know, you gave me a home to find refuge. To my family and friends, thank you for the words of encouragement that spurred me during both the challenging and good times.

Last but not least, I acknowledge God Almighty for being ever-present in my life, and indeed, during this undergraduate thesis.

Abstract

Though Machine Learning has been around for a while, it is still considered a new tool for economists and in its application to predicting economic growth. Studies that apply machine learning to predicting economic growth have found that the Random Forest algorithm is currently the best performing machine learning algorithm for predicting economic recessions and economic growth. However, besides studies evaluating the various machine learning algorithms, there is limited literature on the application of these techniques to help economists and policymakers solve problems. Developing African countries, like Zimbabwe, with their unique economic growth challenges, can harness the predictive qualities of this technology in development planning, setting, and achieving growth targets. In this thesis, I apply the random forest algorithm to make income predictions for Zimbabwe, in the country's hopes to attain upper-middle-income status by 2030.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ESAP	Economic Structural Adjustment Program
GDP	Gross Domestic Product
GNI	Gross National Income
GNU	Government of National Unity
IMF	International Monetary Fund
MA	Moving Average
MAE	Mean Absolute Error
MERP	Millennium Economic Recovery Programme
NDS1 & NDS2	National Development Strategy 1 & National Development Strategy 2
NERP	National Economic Recovery Programme
RMSE	Root Mean Squared Error
SADC	Southern African Development Community
STERP	Short Term Economic Recovery Programme
US$	United States dollar
WDIs	World Development Indicators
ZANU PF	Zimbabwe African National Union, Patriotic Front
ZIMASSET	Zimbabwe Agenda for Sustainable and Socio-Economic Transformation
ZIMPREST	Zimbabwe Programme for Economic and Social Transformation
ZW$	Zimbabwean dollar

Chapter 1: Introduction

The advent of Machine Learning has helped to provide further insight into previously complex problems. The use of this technology has improved problem-solving and decision-making, with its application in modeling and prediction helpful to answer difficult questions [15]. As a new tool in economics, machine learning can help improve policy decision-making. Researchers can project possible policy alternatives that allow policymakers to interrogate economic hypotheses. This will help clarify ambiguous policy positions and reduce the level of uncertainty involved in policy decision-making.

This study applies Machine Learning to examine the economic trajectory for Zimbabwe in the country's ambition to achieve Vision 2030 [13]. Under Vision 2030, the government hopes to grow the economy and attain upper-middle-income status in the year 2030. Currently, the government has prioritized facilitating economic stabilization through the Transitional Stabilisation Programme blueprint in implementation until December 2020 [24]. The Transitional Stabilization Program represents the 6^{th} out of 7 economic plans in the past ten years. Table 1.1 below highlights significant development inconsistencies that have perpetuated the continued economic decline for over 20 years. For the next decade till 2030, the development will be steered by two 5-year National Developmental Strategies (known as NDS1 & NDS2) that are still yet to be devised.

The study examines the significant factors that can drive Zimbabwe's income growth and steer the country towards attaining Vision 2030. Leveraging on Machine Learning's predictive capabilities, the study identifies significant development indicators that should take priority in the next two 5-year development plans in Zimbabwe's desire to attain Vision 2030.

Table 1.1: Economic Policies Timeline (2009 - 2019)

Date	Economic Policy	Duration
19 March 2009	Short Term Emergency Recovery Programme (STERP I)	2009
23 December 2009	Short Term Emergency Recovery Programme (STERP II)	2010 - 2012
1 July 2011	Medium Term Plan (MTP)	2011 - 2015
1 October 2013	Zimbabwe Agenda for Sustainable Socio-Economic Transformation (ZIMASSET)	2013 - 2018
19 April 2018	Vision 2030	2018 - 2030
5 October 2018	Transitional Stabilisation Programme (TSP)	2018 - 2020
June 2019	Zimbabwe National Industrial Development Policy (ZNIDP)	2019 - 2023

1.1: Background

In this section, I highlight the context of this research by recounting the economic history of Zimbabwe. The country inherited a prosperous economy from colonial Rhodesia when it gained independence in 1980 [23]. After independence, the country still enjoyed its privileged economic success and was known as the breadbasket of Africa due to a booming economy anchored by agriculture. However, a decade after independence, the country started to experience economic problems that encompass stunted growth, high unemployment, and high inflation. The following subsections detail the economic decline of Zimbabwe and the significant economic development plans employed.

1.1.1: Black Friday; The Beginning of Decline

14 November 1997 is a significant day that has been etched in the economic history of Zimbabwe. The day is dubbed Black Friday and is a watershed mark that represents the beginning of the decline of Zimbabwe's economy [7]. The exchange market of the Zimbabwean dollar opened the day with the local currency trading at 14 to 1 with the United States dollar but closed the day on 26 to 1. The local currency lost over 70% of its value on the day, and the ripple effects manifested as the Zimbabwean Stock Exchange crashed to drop 46% of the value of shares traded on that day [23, 35, 40].

Many studies have made inconclusive conclusions on the causes of Black Friday. However, as Munangagwa [3] indicates, the Black Friday seed was sowed from the failure of the International Monetary Fund's (IMF) Economic Structural Adjustment Program popularly known as ESAP. It was adopted to help the government guard against the socialistic public expenditure of the post-independence 1980s. From 1990 to 1995, ESAP intended to help the government transition to a more market-oriented economic structure that promotes growth that had stagnated from public expenditure crowding out private investment. Through liberalizing trade and fiscal reform, ESAP projected Gross Domestic Product (GDP) growth at 5% annually [3, 7, 23].

Though there were significant changes in liberalizing the economy, the restrictive fiscal policy failed to meet its objectives and had unintended outcomes of increasing unemployment and eroding incomes [3, 23, 40]. The 1992 drought significantly hampered ESAP, and the unbudgeted spending on food imports led to further increases in the budget deficit [7, 41]. After its 5-year tenure, ESAP's accelerated changes greatly affected the poor, who mostly ended up being unemployed.

When ESAP elapsed, it was succeeded by the Zimbabwe Programme for Economic and Social Transformation (ZIMPREST) that meant to pick up from where ESAP finished [7]. ZIMPREST intended to facilitate increased savings and investment to achieve higher growth targets [41]. However, in August 1997, the growing discontent of war veterans pushing for more post-independence social mobility pressured the government to pay out gratuities amounting to ZW$50,000 to appease the war veterans. Further, they were eligible for an approximate monthly pension of ZW$2,000 per month starting January 1998 [3, 23, 35, 40].

The war veterans were frustrated after being left impoverished under the reforms of ESAP and lack of welfare support thereafter. The gratuity payment was accomplished outside budget, and through monetization, because trade unions fiercely opposed increases in taxes and both developmental partners in World Bank and IMF withdrew support. The total gratuity payment accounted for 3% of GDP in 1997 and was retrogressive of efforts to reduce budget deficit intended under ZIMPREST [7].

In November 1997, the looming land reform program to compulsorily acquire white-owned commercial farms was the last straw [35]. Inadequate financing plans of this enormous undertaking and an already negative fiscal position exacerbated by the war veterans' gratuity payments led to a loss of confidence in the local currency and capital markets, and they crashed on 14 November 1997 [23, 35].

1.1.2: Plunge into Hyperinflation

In March 2000, the government launched Vision 2020. It was a long-term development plan to revive the economy and spearhead sustainable macroeconomic growth [40]. Among other Vision 2020 targets (see Appendix 1) included, doubling GDP between 1997 and 2020 and stabilizing inflation headlined the policy objectives [40]. However, the consequence of the land reform program rendered Vision 2020 implausible, and it fell off the radar of the policymakers as the basis of every government's short and medium-term policy and planning. The Millennium Economic Recovery Programme (MERP) of 2001 and the 2003 National Economic Revival (NERP) were implemented without a significant upturn in economic development indicators and contributed little to Vision 2020 [41].

However, all of these plans were overshadowed by the fast track land reform program that sought to correct the disfranchisement of majority-black citizens from arable land [23]. Despite its good intentions, the hastily implemented program saw a decline in agricultural productivity. Figure 1.1 below shows the downtrend in agriculture output as a percentage of GDP from over 15% in 2000 to nearly 12% in 2002.

The newly resettled black farmers lacked the expertise and equipment to take on commercial farming like the former white farmers [7]. With an agriculture-based economy, the loss of productivity was shared through the whole economy. The manufacturing industry heavily dependent on agricultural output for inputs recorded a drop in production of the sector's contribution to GDP falling from 13.3% in 2000 to 11.8% in 2001.

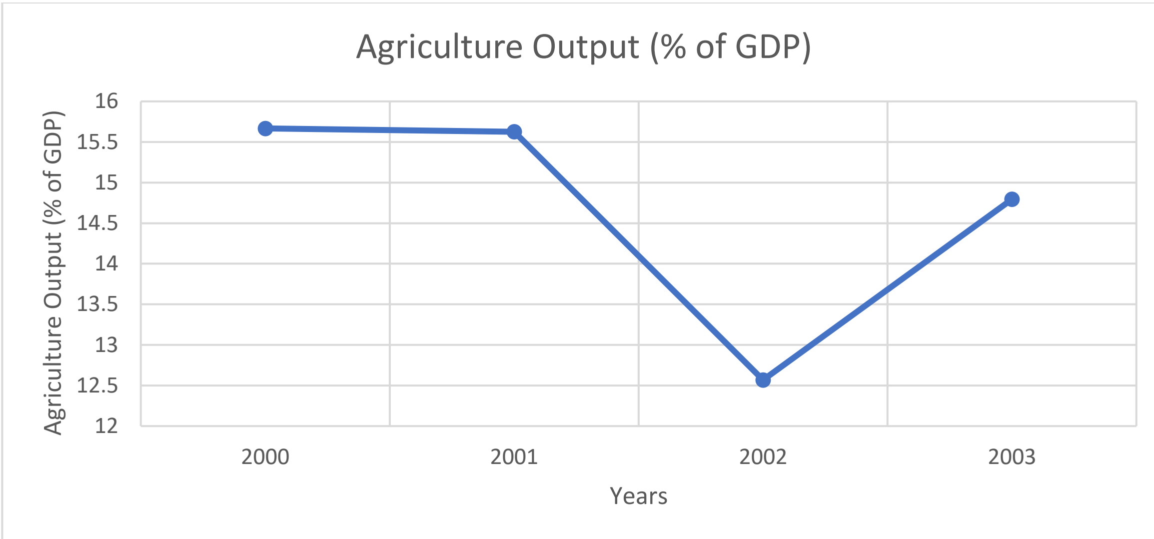

Figure 1.1: Agricultural output (2000-2003). Source: Own graph from World Bank figures

The loss of productivity in agriculture and manufacturing affected aggregate supply in the economy. Shrinking aggregate supply, ceteris paribus, cause price increases, and this started the price-wage spiral precipitated by a teetering exchange rate from the late 1990s [35]. In December 2003, Dr. Gideon Gono assumed the role of Governor of the Reserve Bank of Zimbabwe and declared inflation, which stood at 263% on year basis, as the number one enemy of the state [35]. He implemented a mildly successful tight monetary policy slowing down in the inflation in early 2004.

From 2004 to 2007, in a bid to resuscitate supply and support the monetary policy targets, the governor initiated quasi-fiscal policies beyond the Central Bank's mandate [35]. The Farmers' Mechanization Program was one such initiative that sought to help the newly resettled black farmers with farming equipment such as tractors and inputs like fertilizers to increase agricultural output [3, 35]. Without support from development partner institutions such as the IMF and the World Bank, the governor resorted to financing the quasi-fiscal policies through monetization, which was retrogressive to curbing inflation [3, 7, 35].

In the meantime, the low agricultural output warranted increased government expenditure on food imports for immediate consumption. The Central bank utilized a quasi-fiscal initiative to try to reverse a negative current account that made foreign currency dearer and fueled its increased demand for importation. The Basic Commodities Supply Side Intervention (BACOSSI) facility was initiated to subsidize light industry manufacturers to provide low-cost necessities [35, 41]. Under the program, the central bank would finance manufacturers to produce essential goods that the citizens would buy as a hamper at a subsidized price of ZW$100 billion, which was only sufficient to buy a loaf of bread at the time [27].

The quasi-fiscal policies of the central bank led to unrelenting growth in the money supply that induced inflation at unprecedented levels [33]. The more money the Central Bank introduced into the economy through its fiscal interventionist programs, the more the currency lost its worth. For a decade, Zimbabwe threw caution to the wind and neglected Vision 2020 and economic planning. The total disregard of Vision 2020 in place of firefighting interventionist policies led the country to experience hyperinflation that peaked at a world record high year-on-year inflation rate of 89.7 sextillion percent in mid-November 2008 [7]. Instead of doubling, the economy drastically shrank with GDP halving between 1998 and 2008 as a result of massive deindustrialization that left industry capacity utilization at 10% [41].

1.1.3: STERPs from Hyperinflation

In February 2009, a new Government of National Unity (GNU) formed by the three most prominent political parties assumed responsibility of resuscitating the economy. The GNU introduced the Short Term Emergency Recovery Programme known as STERP. This policy framework meant to stabilize the economy, promote growth, and lay the foundation of a longer-term development plan [41]. Under STERP, the country had its first purple patch in terms of growth in over ten years post Black Friday by recording a 12% increase in GDP during 2009, as Figure 1.2 shows below. From 2009 to 2011, Zimbabwe was among the fastest-growing countries [34].

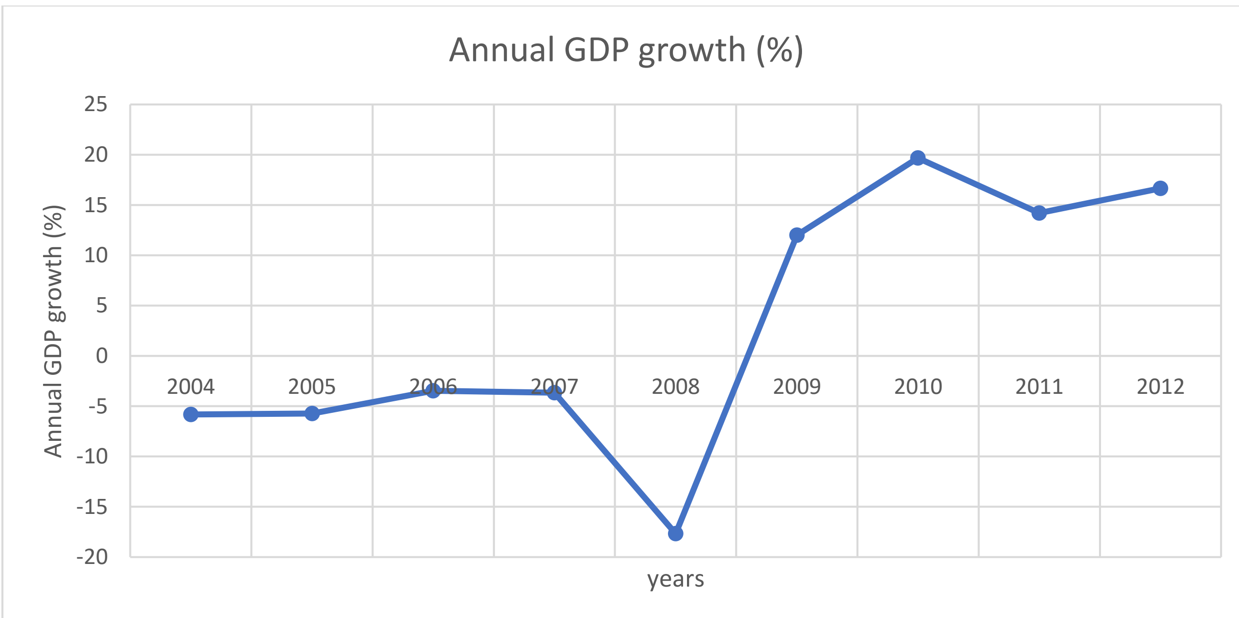

Figure 1.2: GDP growth (2004-2012). Source: Own graph from World Bank figures

The GNU government prioritized the formal adoption of the multi-currency regime [34]. It legalized the use of foreign currency that included the South African Rand and United States dollar as the medium of exchange in place of the obsolete Zimbabwean dollar. The multi-currency regime put a stop to hyperinflation and with year-on-year inflation averaging 4% between 2009 and 2013 [41]. Removal of quasi-fiscal activities by the central bank managed to reign in budget and current account deficits as demanded in the now infamous, '*We eat what we kill!*' mantra by then Finance minister, Tendai Biti.

The positive outlook encouraged investment and saw businesses reopen with capacity utilization rising to 60% as of December 2009 [34]. Backed by renewed confidence and support from the developmental partners, the country realized some resuscitation of the crucial sectors like health and education. STERP was a positive strategy due to its broad consultation in Zimbabwe and beyond. However, it was only a short-term measure to resuscitate an economy that was bereft of any form of life.

1.1.4: Depreciating ZIMASSET

In October 2013, the new ZANU PF led government launched the Zimbabwe Agenda for Sustainable Socio-Economic Transformation (ZIMASSET) to replace the Medium-Term Plan that had succeeded STERP at the beginning of the year [40]. The development plan intended to direct national development for the next five years until 2018. The government presented ZIMASSET as a sanctions-busting strategy that identified the '*illegal*' economic sanctions imposed by the United States of America (USA) and the European Union (EU) as the chief cause of Zimbabwe's economic downfall [12, 40, 41]. Thus, it was inwardly looking with the government earmarking the full exploitation and value addition of the country's human and natural resources to drive economic growth [12].

ZIMASSET was built around the four strategic clusters of Food Security and Nutrition, Social Services and Poverty Eradication, Infrastructure and Utilities, and Value Addition and Beneficiation [12]. These strategic clusters were to assist the government in prioritizing its programs for implementation with a view of realizing broad results that address the socio-economic challenges. ZIMASSET envisioned accelerating economic growth and wealth creation by projecting an average growth rate of 7.3% [40].

The success of ZIMASSET was dependent on the stability of the policy assumptions, (see Appendix 2), that did change to fail the plan. Liquidity challenges intensified in 2016, prompting the introduction of local proxy currency by the Reserve Bank of Zimbabwe called Bond Note. The prevalent use of the US dollar put pressure on prices and made Zimbabwean goods more expensive in the SADC region. This made Zimbabwean exports expensive while imports became cheaper, leading to more of the lucrative US dollar leaving the country. Some companies and individuals took advantage and externalized funds, thus intensifying flight of the liquid US dollar.

The high costs prevalent in the economy repelled the targeted increase in Foreign Direct Investment (FDI) into Zimbabwe. Due to the imbalanced inflow and outflow of the US dollar in the country, the multi-currency regime dominated by the US dollar virtually became a one currency regime of the local Bond Note currency. The impromptu change in government because of the military coup in November 2017 also contributed to the failure of ZIMASSET. However, this happened only a year before the policy would have elapsed. By that point, it was clear the inward-looking approaches like mobilizing local resources to liquidate US$6.1 Billion in external debt were not working [41].

Further evidence of ZIMASSET's inadequacies was delivered when Emmerson Mnangagwa took on the presidency on the interim bases in early 2018. He undertook a '*Zimbabwe is Open for Business*' initiative that informed his foreign policy [19]. This resulted in the reversal of the inward-looking approaches of ZIMASSET, such as the Economic and Indigenization Act, in a bid reach out to the International community for assistance with Zimbabwe's economic recovery.

1.1.5: A Decade of Action

2018 ushered in the Second Republic of Zimbabwe, and that required new thinking in economic development for the country. The appointment of Professor Mthuli Ncube, an Oxford graduate, and esteemed macroeconomic expert seemed to offer hope of incoming change from a history short-term policy thinking and policy inconsistency and inadequacy. Currently, there is no definitive development plan, with the country being steered on by the two-year Transitional Stabilization Plan that is due to expire in December 2020.

In a strategy document presented on 19 April 2018, the Government of Zimbabwe [13] spelled out Vision 2030 as rebuilding and transforming Zimbabwe to become an Upper-Middle Income Economy by 2030. Vision 2030 replaces Vision 2020, but it will likely turn out the same if there is inadequate planning and the policy inconsistencies persist. Aligning development policy with Vision 2030 will streamline efforts to achieve it and ensure its priority, unlike with Vision 2020. Vision 2030 timeline is convenient for the Second Republic since it is within the **Decade of Action** for achieving the United Nations Sustainable Development Goals (SDGs) [39]. Having informed policies based on Vision 2030 will allow the country to make up ground on achieving the SDGs. Growth of the Zimbabwean economy towards an upper-middle-income economy by 2030 can be the basis upon which Zimbabwe achieve the SDGs.

However, the current sentiment of Zimbabweans is very skeptical of Zimbabwe's ability to attain upper-middle-income status by the year 2030 [4]. There is strong resistance to the half-hearted austerity measures by the Minister of Finance that have further put the economy on the downward curve characterized by hyperinflation, eroded incomes, and liquidity challenges [4, 5, 21]. Though the government maintains that Vision 2030 is viable, the Zimbabwean citizenry is disillusioned with current government effort. It is running out of patience at the deterioration of standards of living sprouted by fears of previous hyperinflationary years. An independent assessment interrogating how Zimbabwe can achieve Vision 2030 can help bridge the impasse and hopefully propose development solutions that every Zimbabwean can get behind.

1.2: Problem Statement

From the documented economic history of Zimbabwe, we can identify that short-term policy thinking, policy inconsistency, and inadequacies appear to have hampered development. Though other factors such as corruption contributed to the economic meltdown in Zimbabwe, policymakers also share the blame [40]. Tendai Biti [34], serving as the Finance Minister in the GNU, reaffirms Beaubien [18] notion that Zimbabwe has been suffering from costly, incoherent, and inconsistent policies that need urgent addressing.

The total disregard of Vision 2020 and the constant change of policy and priorities has confused markets and eroded confidence in investors and developmental partners. Vision 2030 is a welcome return to long-term development planning. However, for it to succeed, the two 5-year national development strategies that are earmarked to support it should have identifiable priorities that can enable growth and transition to an upper-middle-income economy in the next ten years. To attain Vision 2030, we need to identify the key development indicators that should take priority in NDS1 & NDS2. This study proposes a data-backed policy formulation alternative to determine what should be contained in NDS1 & NDS2.

1.3: Objectives of the Study

This study utilizes the predictive capabilities of machine learning to determine the development indicators significant to increasing Zimbabwe's income. The identified development indicators should be considered priorities in NDS1 & NDS2 in support of Vision 2030. Though growth literature that correlates variables such as open trade with increased growth exists, it, however, does not inform how open trade matters to income growth than another variable like investment [17]. By using machine learning, the study focuses on the predictive power of variables and ranks them according to how they are important in increasing income growth. The high-ranking variables are considered priorities and inform the economic plans and income growth strategies policymakers should employ.

1.3.1: Research Questions

Within the context of Zimbabwe, the study applies machine learning to identify development indicators significant to growth, guided by the following questions of inquiry:

1. What are the development indicators significant for income growth in Zimbabwe?
2. What is the projected trend in annual income growth for Zimbabwe in the next ten years?
3. Can Zimbabwe achieve Vision 2030?

1.4: Significance of the Study

Having a parsimonious list of development indicators that take priority in economic planning is vital for Zimbabwe. The country is currently lagging in most development indicators, with the World Bank recently downgrading the country's status as a lower-middle-income country to a low-income country after only two months of upgrading into the former group [1]. Having a broad policy that tries to remedy all areas the economy is underperforming will spread thin government effort and the few resources available. Thus, a narrow policy framework to achieve a specific economic objective serves as a viable option towards development. Vision 2030 outlines the economic target for Zimbabwe for the next ten years, and it should be supported by development plans and policy framework geared towards achieving that objective.

Thus, this study is essential as an alternative approach to development planning and policy formulation by identifying the priorities of the strategies that should support Vision 2030. These priorities can be adopted to inform the two 5-year national development plans still in development [25]. Through conducting this study, I hope to advise the current government and policymakers on the development priorities needed to achieve Vision 2030. To the despondent Zimbabwean citizenry, I hope the study results can clarify the country's vision and give an objective analysis for them to evaluate Vision 2030. As a model of data backed economic planning and policy formulation, this study can help Developing African countries to embrace an alternative approach to development planning.

1.5: Scope of Study

This study is confined to exploring the significant development indicators to Gross National Income growth. The World Development Indicators (WDI), as defined by the World Bank, is a compilation of relevant, high-quality, and internationally comparable statistics about global development and the fight against poverty. The WDIs used in this study are only for Zimbabwe and have been extracted from the World Bank data bank.

The study will not compare the various machine learning techniques available. However, it adopts the use of one machine learning model, as recommended by current literature, as the best performing algorithm in predicting economic growth. Since this model is implemented using already defined libraries, the model is used as provided and will not require any optimization effort from the researcher.

1.6: Organization of Study

This thesis paper is organized into five chapters. Chapter one has covered the introduction, outlining the background of the study, the research problem, questions and objectives, the significance, scope, and organization of the study. Chapter two details the literature review, highlighting economic growth literature. It will also highlight research on machine learning algorithms in time series and economic analyses. The third chapter describes the methodology of the study, specifying the dataset, resources, and analytical steps undertaken to answer the research questions, and chapter four presents the analysis results of the investigation. Lastly, Chapter six summarizes the study and concludes by highlighting the study recommendations and limitations and suggestions for future work.

Chapter 2: Literature Review

There has been a lot of research on economic growth research, with many contributions citing Robert Solow and his growth model [31]. Most of the studies have focused on exploring empirical evidence of catch-up growth and the income convergence between developing and developed countries. Time series data or panel data is usually compared and analyzed to determine whether developing countries are catching up with developed countries based on a measure of per capita income [14, 29]. Other studies within growth literature focus on assessing the relationship of a specified economic or social variable such as investment or corruption, to economic growth. [10, 16, 30] These studies require the assumption that there exists a prior existing relationship between the economic variable and a measure of economic growth. Thus, growth debate has been dominated by discussions of whether economic growth stems from neoclassical theory or endogenous theory [2].

Neoclassical growth theory references the Solow growth model to explain the sources of economic growth. The Solow growth model is based on the productivity function that considers the knowledge in an economy, the capital stock, and the human capital as exogenous variables that contribute to creating economic output [14, 31]. These represent sources of growth that leads to an increase in output. Technology is represented by capital stock in the model as a crucial source of economic growth in neoclassical theory. However, capital stock experience diminishing return as more is acquired with time, thus allowing developing countries an opportunity to grow and catch-up with developed countries. However, without convergence evidence, the neoclassical theory is challenged to explain the differences in per capita income across the country. However, technical progress is considered significant to economic growth and recognizes a covariate of economic growth and income [14,29,31].

Endogenous growth theory suggests that economic growth can occur within an economic system without exogenous technical progress, savings, or population [26]. Prominent factors sources of growth in endogenous growth are human capital and investment [2]. The Endogenous theory does not recognize income convergence, citing that an increase in investment leads to constant or increasing returns [2, 26]. Hence, developing cannot catch up with developed countries. However, constant and increasing returns to investment suggest that the level of economic growth is limited by the social capabilities of a country such as weak institutions. This study is influenced by the endogenous growth theory to recognize the role of government input in delivering economic growth.

Salai-i-Martin and Artadi's [9] identifies investment, human capital, strong institutions, economic openness, limited public spending, and extended interval of peace as determinants of economic growth for the continent. However, from that list of determinants, we cannot identify which covariate of economic growth matters the most. We cannot prioritize one over the other and measure how much each variable compares against the other. We can employ machine learning to find the predictive power of the covariates of economic growth. Rather than rely on umbrella covariates, we can use machine learning to assess all possible sources of economic growth without any need for a prior theoretical framework that defines how a covariate related to economic growth [17].

Though machine learning has existed for longer, it is still considered a new tool for economists [15]. Economists are moving to adopt machine learning to handle the large economic datasets and model more complex economic phenomena. Machine Learning also intuitively presents results that are easier to interpret, especially for non-technical economists [17].

However, not all machine learning techniques have been successfully employed to support everyday decision making by policymakers and analysts. The random forest algorithm, a supervised machine learning technique, has been the most recommended machine learning algorithm for economic analysis.

Nyman and Ormerod's [28] compared a random forest regressor model against the Ordinary Least Squares (OLS), a commonly used regression technique in econometrics. Both were tested to predict economic recessions, and the random forest algorithm was found to possess the potential to give early warning signs of an impending downturn in economic activity. Bang et al. [17] employed the random forest algorithm, together with artificial neural networks and classification and regression trees to predict economic growth. The random forest proved to be the best performing algorithm after posting the lowest mean squared error in both training and testing samples of data. Thus, the random forest algorithm is currently the best performing algorithm in economic analyses. This is not surprising considering that the algorithm's performance in prediction is not isolated to economic projections. In a 2015 study to predict civil war onset by Muchlinskito et al. [8], the algorithm proved to have more predictive power than the logistic regression model. Fernandez-Delgado et al. [22] also establish that the random forest algorithm is the best performing machine learning technique for making predictions on time-series data. Thus, based on the literature recommendation of both economic analyses and non-economic analysis, this research will attempt to make predictions on income using the random forest algorithm to find the predictive determinants of economic and income growth for Zimbabwe.

Chapter 3: Methodology

This study aims to establish the development indicators significant to Zimbabwe's income growth ambitions. The central aim of this study is to identify economic levers to income using a machine learning algorithm as a data-backed alternative approach to development planning that presents Zimbabwe with increased chances of attaining upper-middle-income status. The following sections detail the dataset and data values used in this study, the machine learning algorithm used, and implementation resources employed, and the overall flow of analysis.

3.1: The World Development Indicators Dataset

The World Bank compiles the World Development Indicators (WDIs), which are high-quality statistics about global development and the fight against poverty [38]. These indicators span the following six data themes: Poverty and Inequality, People, Environment, Economy, States and Markets, and Global Links. Since WDIs are measures of development, I adopt them as levers for increasing income. I obtained a WDIs dataset for Zimbabwe, which contains over 1400 WDIs measures for 60 years spanning from 1960 to 2019. I will focus on considering all WDIs data themes and considering variables beyond the ones recognized in growth literature as covariates of economic growth. This will present me with the challenge of creating a practical list of WDIs that are levers of income growth from the whole dataset of all the WDIs under consideration.

3.1.1: Data Processing

After obtaining the dataset, I had to pre-process it before beginning analysis. One of the pre-processing tasks was to eliminate repeated measures. For example, GDP is measured in either local currency or US dollars. This duplication is observed for several other WDIs, such as Gross Domestic Savings. For the duplicated measures, I selected WDIs measured as a percentage of GDP or per capita. All monetary WDIs are in US dollars to easily standardize between the pre-hyperinflationary period of local currency use and post 2009 when the country adopted a multi-foreign currency regime that uses the US dollar as an overall denomination. The data cleaning task also involved getting rid of noise data comprising of WDIs fixed in nature. A WDI such as Land area (in square km), is not a likely lever that policymakers can influence. Other WDIs omitted included those that do not have any recorded data across all 60 years.

The other main pre-processing task was to convert all WDIs to lagged values. The lagged values are calculated as a 5-year moving average (MA) from the previous five years of data for each year's WDI measure. From empirical observation, Zimbabwean development plans have an average five-year tenure, which influenced the choice of using 5-year MAs to capture ample changes in each development cycle. The lagged values are meant to stabilize variance due to seasonal or cyclical fluctuations inherent within each development cycle. Thus, we also control for random shocks, such as a recession caused by random occurrences like the global outbreak of an infectious disease. Using the previous 5-year data also captures the persistence effects of past growth and convergence effects since development cycles are implemented in sequence, taking the output of the last cycle as input [17].

3.1.2: Handling Missing data

One of the significant challenges of conducting growth research on developing countries like Zimbabwe is that many variables have missing data for some considerable number of years. The lagged values only help impute missing data within the five next years of missing unlagged data. However, this problem is solved by assigning estimates to missing data. The missing lagged data for each WDI is replaced with the mean of lagged values of that WDI. Studies have found out that trained machine learning models with imputed data perform comparably, and often better than, those that omit variables with missing data [17].

3.1.3: Dependent and Independent Variables

The target variables of the study are Gross National Income per capita and annual Gross National Income per capita growth rate. Gross National Income (GNI) the sum of value added by all resident producers plus any product taxes (minus subsidies) not included in the valuation of output plus net receipts of primary income (compensation of employees and property income) from abroad [38]. GNI per capita is found by dividing GNI by the country's population, and GNI per capita growth is the observed change in GNI per capita over time.

The choice of GNI over GDP is informed by The World Bank, which classifies income groups by GNI per capita. However, since GDP is included in the calculation of GNI, GNI is a suitable proxy of economic output and income. Hence, Zimbabwe's income is measured using GNI to capture the economic output and income and obtain a comparable statistic to The World Bank's income groups. Further, The World Bank also cites that GNI closely correlates with other, nonmonetary measures such as life expectancy at birth, mortality rates of children, and enrollment rates in school [36]. This study hopes to assess how these unconventional covariates to income compare against the monetary measures in influencing GNI per capita growth rate.

The other 492 WDIs in the WDIs dataset are taken as the independent variables that predict the target variables. The challenge of this study is to discern the significant WDIs from all the independent variables and come up with a prudent list of WDIs that have empirically influenced the target variables. Appendix 3 summarizes the WDIs that make up the working dataset for this study.

3.1.4: Challenges of Monetary Variables

The target variables of GNI per capita and GNI per capita growth rate represent income values that have been adjusted for inflation. Thus, since the target variables are real values, there is a need to deflate all monetary indicators in nominal terms. Inflation is usually a significant component of apparent growth in any series measured in monetary values and needs to be controlled for in forecasting [30]. By converting the nominal variables to real terms, I can standardize the unit of analysis in real terms. Since the target variables are measured in constant 2010 US dollars, I have used 2010 as the base year to calculate a GDP deflator. It is the adopted price index for discounting the effects of inflation on the nominal monetary variables. Thus, after data processing, the data referred to in this study is in lagged values and real 2010 terms.

3.2: Random Forest Algorithm

This study will employ the Random Forest Regressor model as the machine learning tool for predicting GNI per capita growth. A random forest regressor is an Ensemble machine learning model based on the random forest algorithm. It is a supervised-learning machine learning technique that has proved to be the most reliable in economic projections.

Nyman and Ormerod's [28] study shows that a random forest regression model predicts economic recessions better than the Ordinary Least Squares (OLS) regression model. Fernandez-Delgado [22] also establish that the random forest algorithm is the best performing machine learning technique for prediction using time series data. Bang et al. [17] concur, presenting evidence that the random forest algorithm performed comparatively best against other machine learning algorithms to predict economic growth.

The random forest algorithm makes use of decision trees by combining them into, as the name suggests, a forest of decision trees. With a given set of data, the algorithm randomly selects a set of rows from the dataset and uses it to make a decision tree. After making the regression calculations on the decision tree, the algorithm moves to randomly select another set of data and make up another decision tree. This process is repeated for a specific number of times corresponding to the stated number of decision trees for each model.

The algorithm will find the answer to its regression question by averaging the result of each decision tree in the model. The combining of decision trees to make a prediction eliminates the bias than a model that makes uses of a single decision tree. Thus, the bias of prediction is reduced when spread among multiple decision trees leading to a more accurate forecast. The following pseudo steps summarize the algorithm and construction of decision trees in a random forest model.

I. Randomly select $n \leq N$ observations from the learning sample.
II. At the "root" node of the tree, select $k \in K$ inputs from X where x is the dataset
III. Find the split in each variable selected in (ii) that minimizes the mean square error at that node and select the next split that achieves the minimal error.

IV. Repeat the random selection of inputs and optimal splits in (ii) and (iii) until some stopping criteria (minimum improvement, the minimum number of observations, or the maximum number of levels) is met.

Bang et al. [17]

3.2.1: Implementation Resources

The random forest regressor models are implemented using Scikit-learn, an open-source python machine learning library. The library was initially released 12 years ago and has developed to be a gold standard in machine learning by having extensive implementations of machine learning algorithms that include regression, clustering, and neural networks [11]. I have chosen to use this library considering the following advantages:

a. Easily accessible random forest regressor function included in the library's random forest regressor class.
b. Easily accessible metrics class to calculate model validation metrics
c. The library easily integrates with other python libraries used in loading, handling, manipulating, and visualizing data such as Pandas, NumPy, and Matplotlib

I will use the Scikit-learn library within Jupyter notebook, an environment that will allow combining both programming code and presentation of results within the same document.

3.2.2: Model Validation

A regression model is subject to overestimating or underestimating its predictions. Thus, it is objectively evaluated on the absolute error the model overestimates or underestimates its projections [42]. The absolute error gives us a measure of how, on average, a prediction is over or below the actual value. I will use the Mean Absolute Error (MAE) calculated using the metrics class in the Scikit-learn library to evaluate the regression model.

I will also evaluate the models using the Root Mean Squared Error (RMSE). Similar to MAE, RMSE also measures the differences between the predicted values and the values observed. However, the RMSE, as a measure of accuracy, quantifies the prediction error in the model. I calculate the RMSE by taking the square root of the Mean Squared Error (MSE), which can also be obtained using the metrics class. I will use both MAE and RMSE, to judge the quality of my models with lower values for each error metric preferred.

3.2.3: Optimization

The Scikit-learn is a focused machine learning library that is clean and highly optimized. Thus, the models I will use will require no optimization effort. However, to guard against the tendency of the random forest regressor to overfit for some dataset, I will refine the models by increasing the number of decision trees. The more decision trees I employ, the better the models learn more from data and improve on predictions.

3.3.: Flow of Analysis

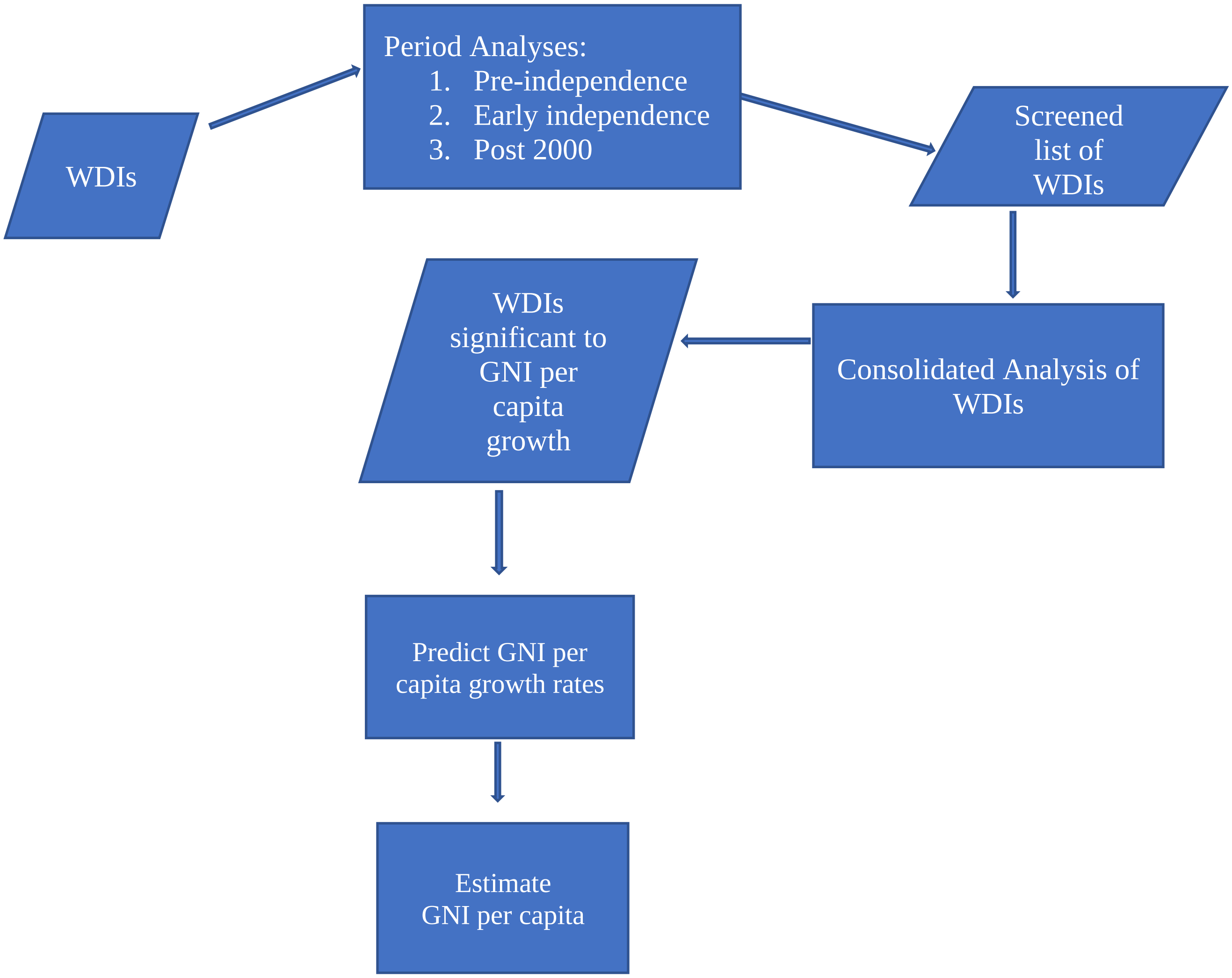

Figure 3.1: Analysis Flowchart

The flow chart in Figure 3.1 summarizes the analytical steps of this study. The challenge is to eliminate WDIs from the working dataset based on each WDIs influence on GNI per capita growth. The significant WDIs will be adopted as predictor WDIs for GNI per capita growth and to forecast the per capita income growth rates for Zimbabwe in the next decade till 2030. The analytical steps to accomplish this are outlined as follows:

3.3.1: Period Analyses

Period analyses investigate the WDIs significant to GNI per capita growth for Zimbabwe in three parts; before independence in 1980, the first two decades of freedom, and the period post-2000 to 2019. These are characteristically different periods in the economic history of Zimbabwe, that demand independent examination to find out what drove income changes in each period. Further, conducting the period analyses reveals common WDIs that are significant to GNI per capita growth across all three periods and could be argued are predictor WDIs of per capita income growth.

A random regressor model is built for each period, using corresponding WDI data for each period. The random regressor models will be trained to estimate GNI per capita growth for each period. However, rather than predictions, the analysis aims to find the significant WDI that explains GNI per capita growth rate during the time. From all three period models, we obtain a narrowed list of WDIs to further evaluate in an overall model.

3.3.2: Consolidated Model

The period analyses present a narrowed list of all WDIs that influence to GNI per capita growth from all periods as output. This list becomes the input for a consolidated random forest regressor model that further narrows down the WDIs list. The WDIs from the overall model will be used to make GNI per capita growth predictions. If the projections are accurate, the WDIs are adopted as predictor WDIs and approves the model to make per capita income growth predictions. The consolidated model analysis will help assess the significant WDIs that each period offers against those from other periods, consolidate the results of the initial period study, compile a list of predictor WDIs and evaluate a model to predict GNI per capita growth.

3.3.3: GNI per capita Projections

The consolidated analysis produces a fit model to predict GNI per capita growth. The model will be used to estimate the per capita income growth rate from predictor WDIs data. These growth rate predictions represent the annual increase in per capita income; hence we can use the growth rates to estimate annual per capita income.

Extending the lagged values of predictor WDIs beyond 2019 produces lagged values that can be used to forecast GNI per capita growth and GNI per capita for 2020 and beyond. The GNI per capita growth rate projection will help assess how long it will take to attain the minimum per capita income required of upper-middle-income economies. The per capita income forecast will reveal the current pace towards achieving an upper-middle-income economy status. The projections will evaluate how Zimbabwe is currently placed to achieve Vision 2030.

Chapter 4: Analysis and Results

This Chapter discusses the random forest regressor models used in the analytical steps to find the significant WDIs to GNI per capita growth and presents their results. The chapter follows the flow analysis presented in Chapter 3 and first presents the results of the period examination, followed by the consolidated model and, lastly, but not least, the GNI per capita prediction results.

4.1: Period Analyses Results

Table 4.1: Period Models' Performance summary

Metric		**Mean Absolute Error (MAE)**	**Root Mean Squared Error (RMSE)**	**R-squared score (%)**
Model Name	**Period Covered**			
Rhodesia	1965 - 1979	1.55	1.75	79.79
Independent	1980 - 1999	2.1	2.86	-114.99, 90.93*
Millennium	2000 - 2019	1.87	2.25	85.35

Table 4.1 above summarizes the performance of the random forest regressor models that were built during the periodic analyses. The Rhodesia model is trained and tested with data from 1965 to 1979. The Independent model uses 20 years of WDIs data for the early independent years, while the Millennium model uses post-2000 data up until 2019.

* R-squared score recorded during model training

4.1.1: Rhodesia Model

The model analyzes 15 years of WDIs data from 1965 to 1979 to assess the WDIs significant to GNI per capita growth in late colonial Zimbabwe, then known as Rhodesia. This period's insights can provide a foundation for policymakers. As a way of getting back to basics, we can discern what worked then and assess if it is applicable in present-day Zimbabwe.

Running four iterations of the Rhodesia model shows that the model improves on initial results between 100 and 200 decision trees. The *GridSearchCV()* function calculates that the model optimizes results from 130 decision trees. The rhodesia05 iteration, with 130 decision trees, as the best performing iteration, offers relatively optimal MAE and RMSE results. The model posts a MAE of 1.55, improving on an initial MAE by 0.14. It also returns a lower RMSE of 1.75 as compared to 1.91. Thus, the model likely overestimates or underestimates predictions by a MAE of 1.5 and RMSE of 1.75. The model's R-squared score of 79.79%, shows that the WDIs explain approximately 80% of the total variation in GNI per capita growth predictions of rhodesia05. This establishes a reasonably strong relationship between the WDI features and our target variable.

After deciding on the optimal model iteration, I used the *feature_importances()* function from a random forest regressor model to identify the independent variables that are most important in explaining the target variable. These variables represent the WDIs that were significant to GNI per capita growth for Zimbabwe before independence. Table 4.2 shows the high ranking WDIs according to their importance score. Net primary income ranks highest with an importance score of 0.086. Adolescent fertility rate ranks second with 0.069 with Agricultural land, coming close in third with a score of 0.064.

I employed the *SelectFromModel()* function from Scikit-learn's feature selection class. This function used the random forest regressor to identify indicators with feature importance greater than a threshold feature importance score of 0.01. The threshold is an arbitrary value chosen to narrow down the features based on predictive power. Table 4.2: shows the top 27 indicators, all with an importance score higher than 0.01, that is, more than 1% influence on GNI per capita growth, are noted for further analysis.

Table 4.2: WDIs Significant to GNI per capita growth (1965 – 1979)

Ranking	**WDI**	**Importance Score**
1.	Net primary income (BoP, US$)	0.086
2.	Adolescent fertility rate (births per 1,000 women ages 15-19)	0.069
3.	Agricultural land (sq. km)	0.064
4.	Arms imports (SIPRI trend indicator values)	0.056
5.	Multilateral debt service (TDS, US$)	0.047
6.	Cereal yield (kg per hectare) Importance: 0.03	0.03
7.	Debt service (PPG and IMF only, % of exports of goods, services, and primary income)	0.028
8.	School enrollment, secondary (gross), gender parity index (GPI)	0.028
9.	Insurance and financial services (% of service imports, BoP)	0.027
10.	Net migration	0.027
11.	Fossil fuel energy consumption (% of total)	0.026
12.	Cereal production (metric tons)	0.023
13.	Foreign direct investment, net (BoP, US$)	0.022
14.	Manufacturing, value added (% of GDP)	0.021
15.	Rural population growth (annual %)	0.021
16.	Secondary education, general pupils (% female)	0.021
17.	Primary income receipts (BoP, US$)	0.02
18.	Service imports (BoP, US$)	0.019
19.	School enrollment, primary (gross), gender parity index (GPI)	0.016
20.	Survival to age 65, female (% of cohort) Importance: 0.016	0.016
21.	External debt stocks, public and publicly guaranteed (PPG) (DOD, US$)	0.015
22.	Total greenhouse gas emissions (kt of CO2 equivalent)	0.15
23.	Energy imports, net (% of energy use)	0.14
24.	External debt stocks, total (DOD, US$)	0.14
25.	Multilateral debt service (% of public and publicly guaranteed debt service)	0.14
26.	PPG, All sources (NFL, US$)	0.014
27.	Secondary education, teachers	0.012

4.1.2: Independent Model

This model extends the period analysis of development indicators significant to GNI per capita growth from the Rhodesia model to the early independent years in Zimbabwe. As a period of the new Zimbabwean Republic, there was an intentional focus on policies to grow and share the economic gains addressing the inequalities experienced during Rhodesia.

The data analyzed in this model spans 20 years from 1980 to 1999 and contains 306 WDIs from 490WDIs under review. After running the first iteration of the Independent model with 50 decision trees, the model records a MAE of 2.14 and RMSE of 2.9. This poor performance is further confirmed by the model's negative R-squared score of -133.54%. Thus, our model predictions are worse off than just taking the mean of our target variable as a prediction during this period. This highlights the model's failure to learn and follow the trend of the data.

I ran more iterations of the Independent model, increasing the number of decision trees each time, but it did not result in significant improvement to the model. Based on that evidence, I can conclude that this model cannot provide a reliable GNI per capita growth prediction. Its failure to learn and follow the trend can be attributed to the many structural changes that dominated the early independence years. The effect is seen in Figure 4.1, which shows the high fluctuation of the actual GNI per capita growth rate from 1980 to 1999. Even though I have controlled for variation using the lagged values, there are pronounced changes between negative growth brought forward from the late 1970s and the positive growth of the early 1980s.

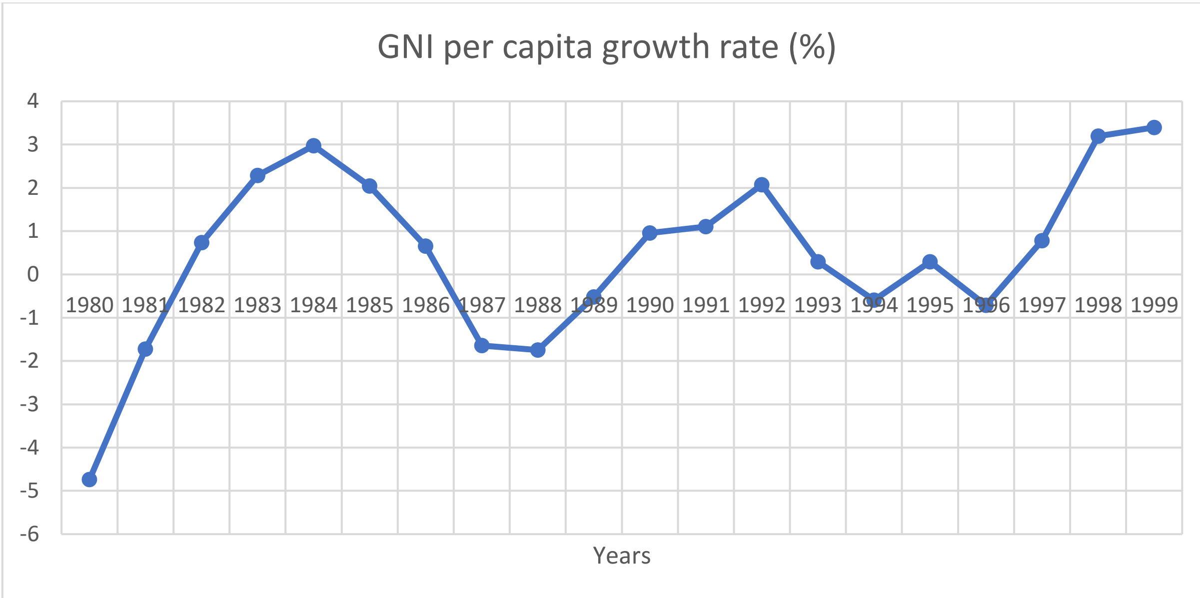

Figure 4.1: GNI per capita growth rate (1980-1999).

Besides the structural changes, the 20 years represent an inflection period characterized by economic booms in the 1980s and economic decline of the 1990s. Thus, to improve results, it can be better to dissolve this period and divide its data between the Rhodesia model and the Millennium model.

However, though the model is incapable of making predictions, it can still be useful to identify features that influence the target variable. I will use the R-squared score measured on the training dataset, to attest that, though the model performs poorly on predictions, it is still valid enough to return feature importance. I select independent04 iteration with 165 decision trees and a slightly lower MAE of 2.1 and RMSE of 2.86. The iteration returns a training R-squared score of 90.93%, which shows that during training, 91% of total variations in the target were attributed to the WDI features. This shows a strong relationship between the explanatory variables and the target variable is established when building the model. Table 4.3 presents the WDIs significant to GNI per capita growth for Zimbabwe from 1980 to 1999.

Table 4.3: WDIs Significant to GNI per capita growth (1980 – 1999)

Ranking	**WDI**	**Importance Score**
1.	Manufacturing, value added (annual % growth)	0.182
2.	Total reserves in months of imports	0.148
3.	Services, value added (annual % growth)	0.091
4.	Pupil-teacher ratio, secondary	0.044
5.	Portfolio investment, net (BoP, US$)	0.03
6.	Changes in inventories (US$)	0.028
7.	Cereal yield (kg per hectare)	0.022
8.	Export unit value index (2000 = 100)	0.02
9.	Computer, communications, and other services (% of commercial service exports)	0.019
10.	Use of IMF credit (DOD, US$)	0.019
11.	Inflation, GDP deflator (annual %) Importance: 0.017	0.017
12.	Total debt service (% of exports of goods, services, and primary income)	0.017
13.	Real interest rate (%)	0.016
14.	Lending interest rate (%)	0.015
15.	Computer, communications, and other services (% of commercial service imports)	0.014
16.	External balance on goods and services (% of GDP)	0.011
17.	Insurance and financial services (% of service imports, BoP)	0.011

Out of the 306 WDIs analyzed in the independent model, 17 registered an importance score higher than 0.01. Manufacturing, value added (annual % growth) ranked first and is the most significant WDI that explains GNI per capita growth during this period by registering a score of over 0.18., that is, over 18% influence on the target variable. Soon after independence in 1980, the country was well placed to experience growth in manufacturing capacity recording an average of 3.1%. The steady increase in manufacturing capacity reflects the then government’s intentional focus on growing the economy after years of war, and it is only natural that the increasing manufacturing capacity translated to GNI per capita growth. Thus, the huge importance score of Manufacturing, value added (annual % growth) to GNI per capita growth, is by no means improbable.

4.1.3 Millennium model

This model identifies the WDIs important to GNI per capita growth rate for Zimbabwe in the years 2000 to 2019. In the economic history of Zimbabwe, this period has been touted as lost decades with the 2000-2009 decade characterized by pronounced economic decline, hyperinflation, and erosion of incomes. The analysis of indicators significant to the increases in GNI per capita growth rate is essential to find the policy levers were available to policymakers during the turbulent economic period.

The data analyzed spans 20 years from 2000 to 2019, comprising of 485 WDIs with recorded data. The analysis steps from the Rhodesia and Independent models are repeated for this model as well. The selected iteration of the Millennium model was built with 200 decision trees and reported the relatively low MAE of 1.87 and RMSE of 2.25. The millennium02 also posted an R-squared score of 85.35%. As the best performing iteration, it is used to determine the feature importance. The significant WDIs to GNI per capita growth in Zimbabwe post the year 2000, as identified by an important score higher than 0.01, are listed in Table 4.4 below.

Table 4.4: WDIs Significant to GNI per capita growth (2000 – 2019)

Ranking	WDI	Importance Score
1.	Net official flows from UN agencies, All Agencies (US$)	0.062
2.	Immunization, DPT (% of children ages 12-23 months)	0.036
3.	Death rate, crude (per 1,000 people)	0.034
4.	Services, value added (% of GDP)	0.032
5.	Immunization, measles (% of children ages 12-23 months)	0.023
6.	Broad money (% of GDP)	0.022
7.	Claims on central government, etc. (% GDP)	0.021
8.	Service imports (BoP, US$)	0.021
9.	Nurses and midwives (per 1,000 people)	0.018
10.	Manufacturing, value added (annual % growth)	0.017
11.	Land under cereal production (hectares)	0.016
12.	Food imports (% of merchandise imports)	0.015

Ranking	WDI	Importance Score
13.	Goods exports (BoP, US$)	0.015
14.	Imports of goods and services (annual % growth)	0.015
15.	Agricultural raw materials imports (% of merchandise imports)	0.014
16.	Probability of dying at age 5-14 years (per 1,000 children age 5)	0.014
17.	Depositors with commercial banks (per 1,000 adults)	0.013
18.	Prevalence of stunting, height for age (% of children under 5)	0.013
19.	Prevalence of undernourishment (% of population)	0.013
20.	Borrowers from commercial banks (per 1,000 adults)	0.011
21.	Contributing family workers, total (% of total employment) (modeled ILO estimate)	0.011
22.	External debt stocks, private nonguaranteed (PNG) (DOD, US$)	0.011
23.	Personal remittances, paid (US$)	0.011
24.	Ratio of female to male labor force participation rate (%) (modeled ILO estimate)	0.011

Net official flows from UN agencies, All Agencies (US$), which is the aggregate disbursements of grants and loans from UN-affiliated agencies, ranks first with a score of 0.062. The other top-ranked indicators important in determining GNI per capita growth during this period include immunization of Diphtheria and Measles in children and the Death rate, which represent unconventional factors that influenced GNI per capita growth.

4.2: The Consolidated Model

From the period analyses, across the three periods, a total of 64 distinct WDIs are determined as significant to GNI per capita growth rate. From an original list of 492, the 64 WDIs make up the narrowed list that is 13% of the former. Out of the 64, 4 have been found significant in more than one period. These are Cereal yield (kg per hectare), Insurance and financial services (% of service imports, BoP), Manufacturing, value added (annual % growth), and Service imports (BoP, US$). It can be argued that since these WDIs are important to GNI per capita growth in more than one period, they should be considered predictor WDIs.

However, I test this hypothesis by building an overall model, named Consolidated model, that evaluates the narrowed list of WDIs. The Consolidated model analyzes the 64 WDIs further to examine their importance to GNI per capita growth rate. The model extends the building and training procedures from the period analyses. The selected iteration of the Consolidated model reported a relatively low MAE of 1.5 and an RMSE of 1.94. The model posted an R-squared score of 83.05%. Like the models from period analyses, I use this iteration to calculate the feature importance.

Table 4.5: Overall WDIs Significant to GDP per capita growth (1965 – 2019)

Ranking	**WDI**	**Importance Score**
1.	Services, value added (annual % growth)	0.32
2.	Manufacturing, value added (annual % growth)	0.247
3.	Death rate, crude (per 1,000 people)	0.093
4.	Survival to age 65, female (% of cohort)	0.023
5.	Primary income receipts (BoP, US$)	0.017
6.	Imports of goods and services (annual % growth)	0.016
7.	Depositors with commercial banks (per 1,000 adults)	0.015
8.	Borrowers from commercial banks (per 1,000 adults)	0.014
9.	Multilateral debt service (% of public and publicly guaranteed debt service)	0.014
10.	External debt stocks, private nonguaranteed (PNG) (DOD, US$)	0.011

Table 4.5 identifies the WDIs that are the best predictors of GNI per capita growth for Zimbabwe in overall across periods. From the four WDIs that are significant to GNI per capita growth across more than one period, annual growth in Manufacturing still relatively significant GNI per capita growth. Together with the other nine, they supersede the former four WDIs. However, before adopting the ten as predictor WDIs, I build a beta model of the Consolidated model that only uses these ten variables to predict the target. This will help evaluate if the other 54 WDIs can be omitted from the model when required to make predictions.

The Consolidated Beta (hereafter referred to as Beta) model replicates the Consolidated model by using the same 1900 decision trees. The Beta model reports a lower MAE of 1.46 and RMSE of 1.88 than the Consolidated model. The model posted a slightly higher R-squared score of 84.04%. This confirms that we do not need to collect data for the other omitted 54 WDIs to predict GNI per capita growth since the same prediction performance is achieved using only the ten high ranked WDIs as predictor variables.

I also use the Beta model to calculate the feature importance selected features, and nearly the same predictor ranking is maintained with a few changes. Private nonguaranteed external debt, that is, long-term external obligations of private debtors that are not guaranteed for repayment by a public entity, rises in ranks from tenth to fourth as its influence on GNI per capita growth increases from 0.011 to 0.049. Multilateral debt service, that is, repayment of principal and interest to the World Bank, regional development banks, and other multilateral agencies, calculated as a percentage of public and publicly guaranteed debt repayments increase in importance from 0.014 to 0.043. Figure 4.2 below plots the predictor WDIs against their respective importance scores from the Beta model.

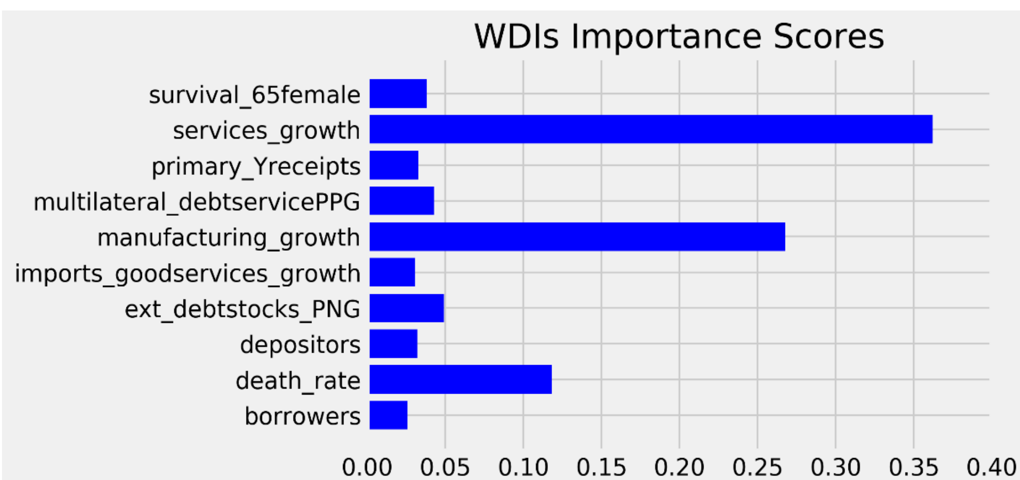

Figure 4.2: Predictor WDIs importance scores, Consolidated Beta model

Further, I evaluated the predictor WDIs and the Beta model in predicting GNI per capita growth by testing its prediction results against the actual data. The test data is the actual GNI per capita growth rate of random 14 years that is set aside when building the Beta model. The corresponding WDIs data for the test data is fed into the Beta model to make the predictions. Figure 4.3 shows how the Beta model projections compare against actual data.

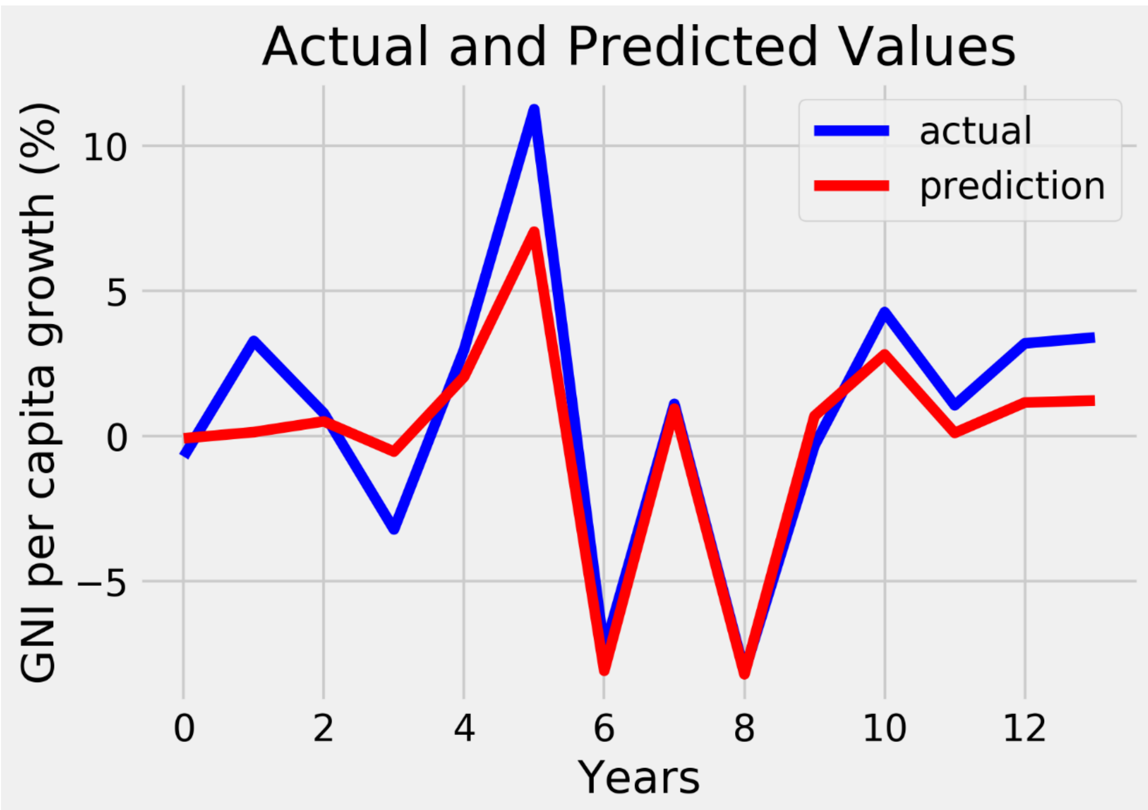

Figure 4.3: Actual and predicted GNI per capita growth rate

From Figure 4.3, the predictions (red line) coincide well with the actual GNI per capita growth (blue line). The red line closely follows the trend of the real data and does well to estimate the actual. This plot further validates the predictor WDIs influence on GNI per capita growth. It also appraises the performance of the Beta model and highlights how it is a good fit to make per capita income growth predictions for Zimbabwe.

5.1: GNI per capita Prediction

For the current 2020 fiscal year, The World Bank classifies upper-middle-income economies as those with a GNI per capita between US$3, 996, and US$12, 375 [37]. As of 2018, Zimbabwe had a GNI per capita of US$1, 317. This presents Zimbabwe with a challenge to grow its per capita income by a herculean average of 20.3% every year in the next ten years, assuming the lower limit to the upper-middle-income status remained unchanged.

To understand how likely the country can achieve Vision 2030, I first determine the current position. Based on the current trend and trajectory, I estimate the GNI per capita growth rates for each year until 2030. Extending the lagged values beyond 2019 provides 5-year moving averages estimates of the predictor WDIs and GNI per capita growth. These are fed into the Beta model to make GNI per capita growth predictions needed to estimate the GNI per capita for the next ten years. The following plot in Figure 5.1 illustrates the prediction results.

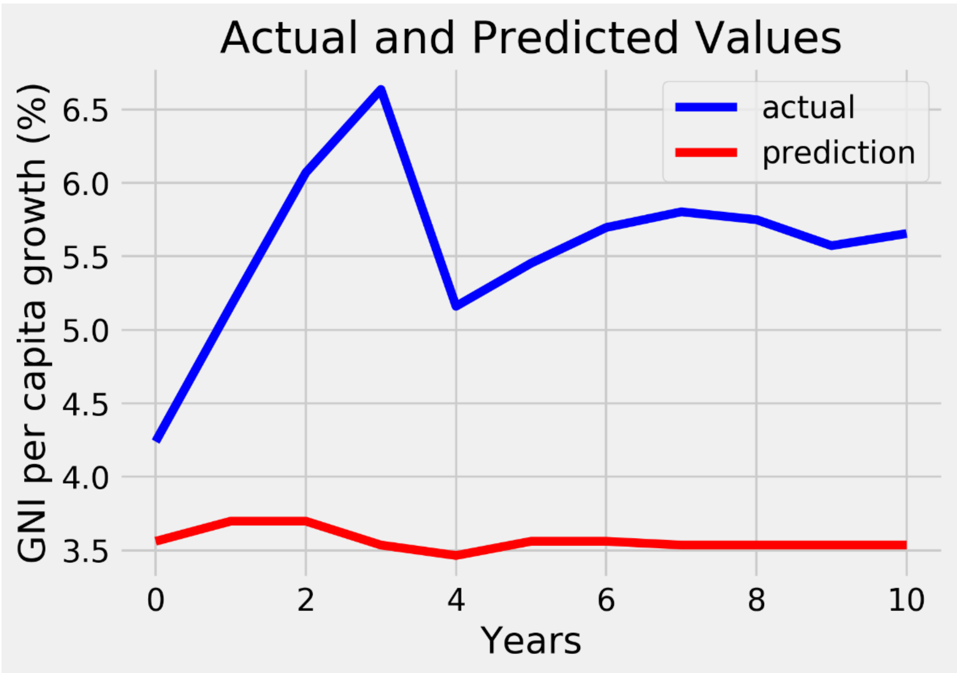

Figure 4.4: Actual (5-year MAs) and projected GNI per capita growth rate (2020 – 2030)

The plot above shows the 5-year MA estimates representing actual data and the Beta model prediction of the GNI per capita growth rate. The MA estimated real data correlates around 5.5% because of a 14% increase in per capita income in 2018. This is an outlier that was likely caused by the rebasing of the economy in October 2018 [20]. The plot in Figure 4.5 shows the same data with the 2018 outlier omitted in calculating data for the subsequent years.

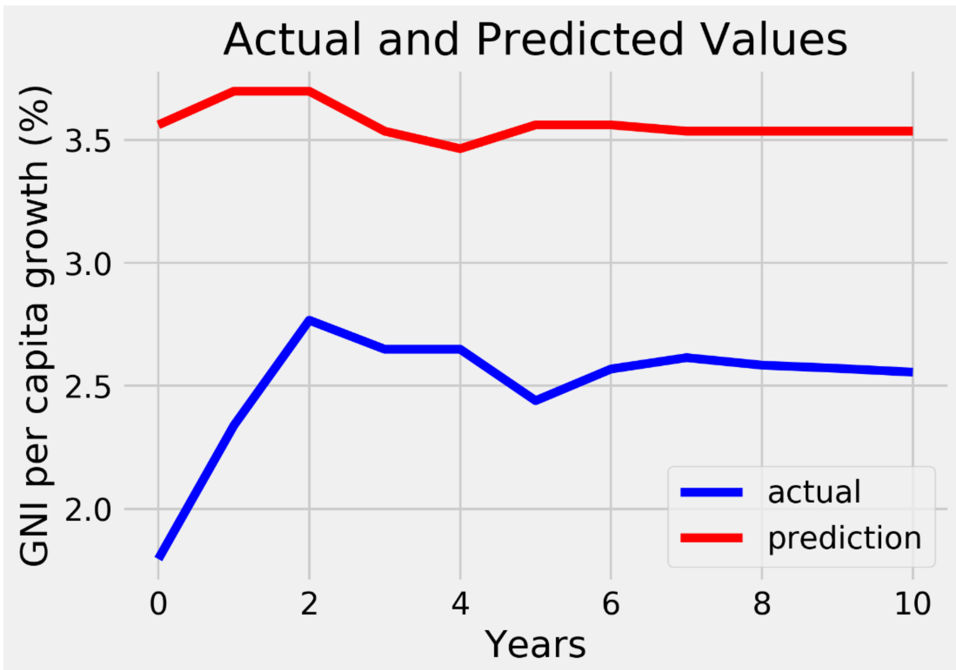

Figure 4.5: Adjusted Actual (5-year MAs) and projected GNI per capita growth (2020-2030)

Omitting the outlier adjusted the actual trend below the predicted trend, peaking at 3.5% for year 2. Though the prediction values are higher than actual, we can identify the similarities in the pattern between the two estimates. Due to the compromised actual data, the prediction values of GNI per capita growth were used to estimate per capita income. However, these predictions values are also limited by bias in MA estimates of predictor WDIs resulting in forecasts coinciding around a mean of 3.56%.

Table 4.6 forecasts Zimbabwe will likely attain a GNI per capita of approximately US$1996 by 2030. This projection is US$2000 less than the required minimum of US$3996 to achieve upper-middle-income status. Going by this trend, it will likely take Zimbabwe approximately 30 years starting from 2021 to reach a per capita income of US$3996 based on a mean per capita income growth rate of 3.56%. This makes Vision 2030 unattainable within the current timeline and policy thrust.

Table 4.6: GNI per capita and GNI per capita growth predictions (2020 – 2030)

Year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
GNI per capita growth (%)	3.56	3.7	3.7	3.53	3.46	3.56	3.56	3.53	3.53	3.53	3.53
GNI per capita (US$)	1406	1458	1512	1566	1620	1678	1737	1799	1862	1928	1996

Chapter 5: Recommendations and Conclusions

5.1: Summary

This research examined the World Development Indicators that are important in predicting the per capita Gross National Income growth rate for Zimbabwe. The study used the feature selection capabilities of a random forest regression model to identify the WDIs that have the most significant influence on per capita income growth. After two series of feature selection, ten WDIs with an importance score greater than 1% were examined as candidate predictor WDIs.

Another random forest model was built to make GNI per capita growth predictions using the candidate predictor WDIs. The model recorded a mean absolute error of 1.46 that makes it likely to underestimate or overestimate its GNI per capita growth predictions by 1.46%. However, the model can capably predict GNI per capita growth, as shown by the plot of predicted values that coincide with the actual values. The model proved fit to make per capita income growth predictions, and this allows us to adopt the ten WDIs summarized in Table 5.1 as predictor WDIs. This identifies the WDIs as the development indicators that are significant to and have the most influence in predicting GNI per capita growth.

Future projections of GNI per capita growth confirms the current sentiment that Zimbabwe cannot achieve upper-middle-income economy status in the next ten years. Based on the prevailing trend, GNI per capita is predicted to grow on an average of 3.56% from 2020 to 2030. This projects that, at this rate, it will take 30 years to increase Zimbabwe's per capita income to US$3996, which is the current minimum required to attain upper-middle-income economy status. Thus, Vision 2030, is untenable and currently beyond Zimbabwe.

Table 5.1: Predictor WDIs for per capita income growth

Ranking	Predictor WDI	Importance Score
1.	Annual growth in the Service sector	0.362
2.	Annual growth in the Manufacturing sector	0.267
3.	Death rate	0.118
4.	Private and nonguaranteed external debt	0.049
5.	Multilateral debt repayments	0.043
6.	Women surviving to age 65	0.038
7.	Primary income receipts	0.033
8.	Depositors with commercial banks (per 1,000 adults)	0.032
9.	Annual growth in imports of goods and services	0.030
10.	Borrowers from commercial banks (per 1,000 adults)	0.026

However, to achieve the income targets of Vision 2030, under the projected 30 years, policymakers should aim to effect policy that increases per capita income growth. According to the Beta model, the predictor WDIs, represent possible strategies to increase GNI per capita growth and, by extension, income. I propose NDS1 & NDS2 should focus effort to increase annual growth in the manufacturing and service sectors, respectively. The two industries have a combined 63% influence on GNI per capita growth rate and are empirically significant to inducing proportionate increases in per capita income growth. Further, they are within immediate policy reach than the other predictor WDIs such as death rate.

5.2: NDS1 Recommendations

NDS1 should aim to promote annual growth in the manufacturing sector. Though this indicator ranks second, the current predicament dictates it takes priority over the service sector. The Confederation of Zimbabwean Industries (CZI) [6] reported a 36.4% capacity utilization for the manufacturing industry in 2019. It is expected to fall this year due to challenges imposed on the sector by the infectious spread of Coronavirus. Thus, NDS1 should support the sector to recover lost capacity that likely spurs to growth in per capita income and help the country take the first steps towards upper middle-income economic status.

5.3: NDS2 Recommendations

Though Zimbabwe is an agriculture-based economy, the discovery of annual growth in the service sector as the most significant WDI on GNI per capita growth suggests that the country should diversify the economy by growing its services sector. This demands NDS2 to focus on service sector activities because a growing service sector capacity induces the highest proportionate increase in per capita income. Besides creating a conducive environment of already established service sector companies, NDS2 should also encourage service sector start-ups that diversify the economy and drive its growth. The future of per capita income growth for Zimbabwe is in the service sector, and NDS2 should make that future reality for Zimbabwe.

5.4: Limitations

The study is limited to the use of the WDIs data set from the World Bank. Though comprehensive, this is not an exhaustive data set of variables that may influence GNI per capita growth. The data does not include some of the institutional measures that can be adopted as social capability measures for economic growth, as suggested in growth literature.

The study also uses a single machine learning algorithm based on the recommendation of past research. As the accuracy in machine learning algorithms improves, this algorithm is susceptible to being overtaken as the best machine learning approach in predicting economic values. Further, the use of libraries in the implementation of the machine learning algorithm makes the analysis subject to the unfortunate error and bias found in the used libraries.

5.5: Suggestions for Future work

This section offers suggested extensions to this study. The first proposal is to utilize the extensive data beyond the World Development Indicators. Though they are comprehensive statistics, they do not contain other variables that might influence economic activity. Institutional variables such as democracy and corruption indexes represent interesting variables to assess as predictor variables of income and economic variables.

An improvement of the random forest models used in this study might also uncover explored results. Optimizing the varied parameters together with combining and comparing different methods of feature selection can lead to improvements in the predictions of growth in per capita income using random forest algorithms.

References

[1] Batanai Matsika. 2019. World Bank downgrades Zim to a low-income country. Video. (18 October 2019). Retrieved April 18, 2020 from https://www.cnbcafrica.com/videos/2019/10/18/world-bank-downgrades-zim-to-a-low-income-country/

[2] Bennett McCallum. 1996. Neoclassical vs. endogenous growth analysis: An overview. National Bureau of Economic Research, Cambridge, MA. DOI: https://doi.org/10.3386/w5844

[3] Chidochashe L Munangagwa. 2009. The economic decline of Zimbabwe. *Gettysburg Economic Review* 3, 1 (2009), 22 pages.

[4] Costa Nkomo. 2019. ED's Vision 2030 a "pipe dream." NewZimbabwe.com. Retrieved from https://www.newzimbabwe.com/eds-vision-2030-a-pipe-dream/

[5] Crecey Kuyedzwa. 2019. It's official: Hyperinflation has returned to Zimbabwe. Fin24. Retrieved from https://www.fin24.com/Economy/Africa/its-official-hyperinflation-has-returned-to-zimbabwe-20191012

[6] Crecey Kuyedzwa. 2020. Zim industry capacity utilisation could fall to 27% in 2020. Fin24. Retrieved from https://www.fin24.com/Economy/Africa/zim-industry-capacity-utilisation-could-fall-to-27-in-2020-20200216

[7] Daniel Makina. 2010. Historical perspective on Zimbabwe's economic performance: A tale of five lost decades. *Journal of Developing Societies* 26, 1 (March 2010), 99–123. DOI: https://doi.org/10.1177/0169796X1002600105

[8] David Muchlinski, David Siroky, Jingrui He, and Matthew Kocher. 2010. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. Retrieved from https://www.academia.edu/17202831/Comparing_Random_Forest_with_Logistic_Regression_for_Predicting_Class-Imbalanced_Civil_War_Onset_Data

[9] Elsa Artadi and Xavier Sala-i-Martin. 2003. The Economic Tragedy of the XXth Century: Growth in Africa. National Bureau of Economic Research, Cambridge, MA. DOI: https://doi.org/10.3386/w9865

[10] Esman Morekwa Nyamongo, Roseline N. Misati, Leonard Kipyegon, and Lydia Ndirangu. 2012. Remittances, financial development, and economic growth in Africa. *Journal of Economics and Business* 64, 3 (May 2012), 240–260. DOI: https://doi.org/10.1016/j.jeconbus.2012.01.001

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, (2011), 2825–2830.

[12] Government of Zimbabwe. 2013. Zimbabwe Agenda for Sustainable Socio-Economic Transformation - ZIMASSET. Retrieved from http://www.veritaszim.net/node/930

[13] Government of Zimbabwe. 2018. Towards an upper-middle income economy by 2030. Retrieved from http://www.veritaszim.net/sites/veritas_d/files/GoZ%20Presentation%20DC%20-%2019-4-2018.pdf

[14] Gregory N. Mankiw, David Romer, and David N Weil. A Contribution to the Empirics of Economic Growth. *Q. J. Econ* (1992), 31 pages.

[15] Hal R. Varian. 2014. Big data: New tricks for Econometrics. *Journal of Economic Perspectives* 28, 2 (May 2014), 3–28. DOI: https://doi.org/10.1257/jep.28.2.3

[16] Jaleel Ahmad and Andy C. C. Kwan. 1991. Causality between exports and economic growth: Empirical evidence from Africa. *Economics Letters* 37, 3 (November 1991), 243–248. DOI: https://doi.org/10.1016/0165-1765(91)90218-A

[17] James Bang, Tinni Sen, and Atin Basuchoudhary. 2015. *Machine-learning techniques in economics. New tools for predicting economic growth*. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-69014-8

[18] Jason Beaubien. 2006. Government Policies lead to collapse of Zimbabwe Economy. NPR.org. Retrieved from https://www.npr.org/templates/story/story.php?storyId=5446596

[19] Jefferson Ndimande and Knowledge Moyo. 2018. "Zimbabwe is open for business": Zimbabwe's foreign policy trajectory under Emmerson Mnangagwa. *Afro Asian Journal of Social Sciences* 9, 2 (September 2018), 25 pages.

[20] MacDonald Dzirutwe. 2018. Zimbabwe rebases data, boosting GDP numbers by 40 percent. Reuters. Retrieved from https://uk.reuters.com/article/uk-zimbabwe-economy-idUKKCN1MF1G8

[21] MacDonald Dzirutwe. 2020. Zimbabwe March inflation jumps to 676.39% y/y - Zimstat. Reuters. Retrieved from https://af.reuters.com/article/zimbabweNews/idAFL5N2C248H

[22] Manuel Fernandez-Delgado, Eva Cernadas, Senen Barro, and Dinani Amorim. 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15, (2014), 49 pages.

[23] Michael R. Vusani. 2014. *The economic downfall of Zimbabwe from 1980 to 2008*. Africa University, Zimbabwe.

[24] Ministry of Finance and Economic Planning, Zimbabwe. 2019. Govt launches Transitional Stabilisation Programme. Retrieved from http://www.zimtreasury.gov.zw/index.php?option=com_content&view=article&id=131:govt-launches-transitional-stabilisation-programme&catid=83&Itemid=613

[25] Ministry of Finance and Economic Planning, Zimbabwe. 2020. Press statement on the commencement of preparations for the development of the first five-year national development strategy (2021-2025). Retrieved from http://www.zimtreasury.gov.zw/index.php?option=com_content&view=article&id=201:press-statement-on-the-commencement-of-preparations-for-the-development-of-the-first-five-year-national-development-strategy-2021-2025&catid=92&Itemid=762

[26] Nevin Cavusoglu and Edinaldo Tebaldi. 2006. Evaluating growth theories and their empirical support: An assessment of the convergence hypothesis. *Journal of Economic Methodology* 13, 1 (March 2006), 49–75. DOI: https://doi.org/10.1080/13501780600566396

[27] Pindula. 2017. BACCOSSI. Retrieved from https://www.pindula.co.zw/BACCOSSI

[28] Rickard Nyman and Ormerod Paul. 2016. *Predicting Economic Recessions Using Machine Learning Algorithms*. University College London, London.

[29] Robert J. Barro and Xavier Sala-i-Martin. 1992. Convergence. *Journal of Political Economy* 100, 2 (1992), 223–251

[30] Robert Nau. 2014. Principles and risks of forecasting. (September 2014). Retrieved April 5, 2020 from https://people.duke.edu/~rnau/Principles_and_risks_of_forecasting--Robert_Nau.pdf

[31] Robert Solow. 1956. The Solow Growth Model. Retrieved October 14, 2019, from http://www.pitt.edu/~mgahagan/Solow.htm

[32] Samuel Adams. 2009. Foreign direct investment, domestic investment, and economic growth in Sub-Saharan Africa. *Journal of Policy Modeling* 31, 6 (November 2009), 939–949. DOI: https://doi.org/10.1016/j.jpolmod.2009.03.003

[33] Sonia Munoz. 2006. Suppressed inflation and money demand in Zimbabwe. *IMF Working Paper* No. 06/15 (January 2006) 20 pages.

[34] Tendai Biti. 2015. Rebuilding Zimbabwe: Lessons from a coalition government. Retrieved from https://www.cgdev.org/sites/default/files/Tendai-Biti-Zimbabwe-Sept-2015.pdf

[35] Terrence Kairiza. 2012. *Unbundling Zimbabwe's journey to hyperinflation and official dollarization*. Ph.D. Dissertation. National Graduate Institute for Policy Studies (GRIPS), Tokyo

[36] The World Bank. Why use GNI per capita to classify economies into income groupings? Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/378831-why-use-gni-per-capita-to-classify-economies-into

[37] The World Bank. 2019. World Bank country and lending groups. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519

[38] The World Bank. 2020. World development indicators. Retrieved from http://datatopics.worldbank.org/world-development-indicators/

[39] United Nations. 2020. Decade of action. United Nations Sustainable Development Goals. Retrieved from https://www.un.org/sustainabledevelopment/decade-of-action/

[40] Vusumuzi Sibanda and Ranganayi Makwata. 2017. *Zimbabwe Post Independence Economic Policies: A Critical Review*. LAP LAMBERT Academic Publishing, Germany.

[41] Wellington G. Bonga. 2014. Economic policy analysis in Zimbabwe: A review of Zimbabwe economic policies: Special reference to Zimbabwe Agenda for Sustainable Socio-Economic Transformation (ZIMASSET). *SSRN Electronic Journal* (January 2014) 19 pages. DOI: https://doi.org/10.2139/ssrn.2384863

[42] Will Koehrsen. 2017. Random Forest in Python. Medium. (December 2017). Retrieved April 21, 2020 from https://towardsdatascience.com/random-forest-in-python-24d0893d51c0

Appendices

Appendix 1

<u>Vision 2020, Long term Development targets</u>

1. Doubling current GDP per capita in 23 years
2. Stabilize inflation to single digit
3. Achieve and maintain positive real interest rates
4. Reduction of budget deficit to manageable levels
5. A substantial decrease in the unemployment rate
6. Increase investment and national savings to at least 30%

To meet these targets, the following actions were to be pursued:

a. Efficient public sector resources management in which the government lived within its means, reduce the deficit, SEs reforms, enhance revenue generation.
b. Industrialization – promoting value addition to local raw materials, further processing of manufactured outputs.
c. Agriculture - target full commercialization and expand output ahead of inflation.

Appendix 2

ZIMASSET policy assumptions.

1. Improved liquidity and access to credit by key sectors of the economy such as agriculture; - this is only possible if we increase exports and value addition to our products.
2. Establishment of a Sovereign Wealth Fund; - no turn back is required in this case
3. Improved revenue collection from key sectors of the economy, such as mining.
4. Increased investment in Infrastructure such as energy and power development, roads, rail, aviation, telecommunication, water, and sanitation, through acceleration in the implementation of Public-Private Partnerships (PPPs) and other private sector-driven initiatives.
5. Increased Foreign Direct Investment (FDI) into Zimbabwe.
6. Establishment of Special Economic Zones.
7. Continued use of the multi-currency system.
8. Effective implementation of Value Addition policies and strategies.
9. Improved electricity and water supply.

Appendix 3

Target Variables		GNI per capita growth, GNI per capita
Independent Variables	**Economy**	Changes in inventories, Exports of goods and services, External balance on goods and services, Final consumption expenditure, General government final consumption expenditure, Gross capital formation, Gross fixed capital formation, Gross national expenditure, Households and NPISHs final consumption expenditure, Imports of goods and services, Agriculture, forestry, and fishing, value added, Chemicals (value added in manufacturing), Food, beverages and tobacco (value added in manufacturing), Industry (including construction, value added), Machinery and transport equipment (value added in manufacturing), Manufacturing (value added), Medium and high-tech Industry, Other manufacturing (value added), Services (value added), Textiles and clothing (value added in manufacturing), Adjusted savings, Coal rents, Forest rents, GNI, Gross domestic savings, Gross savings, Gross value added at basic prices, Inflation, Mineral rents Net primary income, Net secondary income, Total natural resources rents, Current account balance
	Environment	Agricultural irrigated land, Agricultural land (sq. km), Agricultural machinery, Arable land, Cereal production, Cereal yield, Crop production index, Fertilizer consumption, Food production index, Forest area, Land under cereal production, Livestock production index, Permanent cropland, Access to clean fuels and technologies for cooking, Access to electricity, Alternative and nuclear energy, Combustible renewables and waste, Electricity production from oil, gas and coal sources, Electricity production from renewable sources (excluding hydroelectric), Energy intensity level of primary energy, Fossil fuel energy consumption, Renewable electricity output, Renewable energy consumption, Annual freshwater withdrawals, Aquaculture production, Capture fisheries production, Renewable internal freshwater resources, Terrestrial protected areas, Total fisheries production
	Global links	Commercial banks and other lending, Debt service External debt stocks, IBRD loans and IDA credits, IFC (private nonguaranteed, NFL), IMF repurchases and charges (TDS), Multilateral debt service, Net ODA

	received per capita, PNG, commercial banks and other creditors (NFL), Present value of external debt, Public and publicly guaranteed debt service, Short-term debt, Total debt service, Use of IMF credit (DOD), Merchandise trade, Net barter terms of trade index, Personal transfers (receipts), Portfolio equity (net inflows), International migrant stock, Net migration, Refugee population by country or territory of asylum/origin, International tourism (expenditures/receipts), International tourism (number of arrivals/departures)
People	Adolescent fertility rate, Age dependency ratio, Birth rate (crude per 1000 people), Death rate (crude per 1,000 people), Female headed households, Fertility rate, Life expectancy at birth, Mortality rates, Population, Rural population, Sex ratio at birth (male births per female births), Survival to age 65 (female/male), Teenage mothers, Adjusted net enrollment rate (primary), Adolescents out of school, Children out of school, Educational attainment (primary/secondary/post-secondary/Bachelors), Expenditure on education (primary/secondary/tertiary), Government expenditure on education (total), Gross intake ratio in first grade of primary education, Literacy rate (adult/youth), Lower secondary completion rate, Net intake rate in grade 1, Over-age students (primary), Persistence to last grade of primary, Primary completion rate, Primary education pupils, Primary education teachers, Progression to secondary school, Pupil-teacher ratio (primary/secondary/tertiary),School enrollment (preprimary/primary/secondary/ tertiary), Secondary education pupils, Secondary education teachers, Secondary education vocational pupils, Tertiary education academic staff, Contributing family workers, Employers, Employment in agriculture/industry/services, Employment to population ratio, Informal employment, Labor force participation rates, Labor force, Part time employment, Ratio of female to male labor force participation rate, Self-employed, Share of youth not in education, employment or training, Unemployment, Vulnerable employment, Wage and salaried workers, Adults and children newly infected with HIV, Antiretroviral therapy coverage, Births attended by skilled health staff, Cause of death (communicable diseases and maternal, prenatal and nutrition conditions/injury/non-communicable diseases), Current health expenditure, Domestic general government health expenditure, Domestic private health expenditure per

	capita, External health expenditure, Hospital beds (per 1000 people), Immunization, Incidence of HIV/malaria/tuberculosis (per 100000 people), Lifetime risk of maternal death, Newborns protected against tetanus, Nurses and midwives (per 1000 people), Out-of-pocket expenditure per capita, People practicing open defecation, People using at least basic drinking water services/sanitation services/handwashing facilities, Physicians (per 1000 people), Pregnant women receiving prenatal care, Prevalence of HIV, Prevalence of overweight/stunting/underweight, wasting (% of children under 5),Probability of dying at age 5-14 years, Smoking prevalence, Tuberculosis case detection rate, Tuberculosis treatment success rate, UHC service coverage index, Women's share of population (aged 15+) living with HIV, Proportion of seats held by women in national parliaments,
Poverty and Inequality	Average transaction cost of sending remittances from a specific country, Average transaction cost of sending remittances to a specific country
States and Markets	Average number of visits or required meetings with tax officials (for affected firms), Average time to clear exports through customs (days), Business extent of disclosure index, Cost of business start-up procedures, Cost to export, Depth of credit information index, Ease of doing business score, Firms competing against unregistered firms, Firms expected to give gifts in meetings with tax officials, Firms experiencing electrical outages, Firms experiencing losses due to theft and vandalism, Firms formally registered when operations started, Firms offering formal training, Firms that do not report all sales for tax purposes, Firms that spend on R&D, Firms using banks to finance investment, Firms using banks to finance working capital, Firms visited or required meetings with tax officials, Firms with female participation in ownership, Firms with female top manager, Labor tax and contributions, Losses due to theft and vandalism, New business density, Other taxes payable by businesses, Power outages in firms in a typical month (number), Private credit bureau coverage, Procedures to build a warehouse (number), Procedures to register property (number) Profit tax, Start-up procedures to register a business (number), Strength of legal rights index, Tax payments, Time required to build a warehouse, Time required to enforce a contract, Time required to get electricity, Time

		required to obtain an operating license, Time required to register property, Time required to start a business, Time to export, border compliance, Time to export, documentary compliance, Time to import, border compliance, Time to import, documentary compliance, Time to prepare and pay taxes, Time to resolve insolvency, Total tax and contribution rate, Automated teller machines (ATMs), Borrowers from commercial banks, Commercial bank branches, Depositors with commercial banks, Listed domestic companies, Market capitalization of listed domestic companies, S&P Global Equity Indices, Stocks traded, Customs and other import duties (% of tax revenue) Expense, Interest payments, Net acquisition of financial assets, Net incurrence of liabilities, Net investment in nonfinancial assets, Net lending (+) / net borrowing (-), Other expense, Other taxes, Revenue, excluding grants, Social contributions, Subsidies and other transfers, Tax revenue, Armed forces personnel, Arms exports & imports, Military expenditure, Intentional homicides (per 100000 people), Air transport (freight), Air transport (passengers carried), Fixed broadband subscriptions, Fixed telephone subscriptions, Individuals using the Internet, Mobile cellular subscriptions, Secure Internet servers, Patent applications, Scientific and technical journal articles, Trademark applications, ICT goods exports