Faith Kabanda

N02425179W

Masters in Big Data

University of Science and Technology

Bulawayo, Zimbabwe.

TITLE

Short-Term Internet Penetration Trends in Zimbabwe (2020–2025): A Big Data Analysis of Usage Patterns and Their Role in Shaping the Economy Through Retail, Agriculture, and Healthcare Sectors.

Keywords: Internet Penetration, Big Data Analytics, Digital Inclusion.

Table of Contents

1. INTRODUCTION

1.1 Background

In today's digital economy, the internet has emerged as a powerful enabler of economic transformation driving business innovation, enhancing healthcare delivery, and boosting agricultural productivity (Allioui and Mourdi, 2023). In Zimbabwe, rising internet penetration has opened new avenues for growth across these sectors. However, high data costs, limited connectivity, and lack of digital tools remain significant obstacles. Small businesses struggle to scale due to expensive internet access, rural healthcare facilities face unreliable connections, and farmers without adequate connectivity are unable to leverage digital technologies that could enhance yields and market access.

Internet inclusion is essential for equitable economic growth. As Khan (2023) notes, SMEs benefit from digital platforms like WhatsApp Business and Facebook Marketplace, while mobile money solutions such as EcoCash enhance transaction efficiency. In healthcare, telemedicine and mHealth applications are improving patient outreach and record management (Rehman, Naz and Razzak, 2022). The agricultural sector also gains from digital advisory services and online marketplaces that support smarter farming decisions (Zengeya, Sambo and Mabika, 2021). Yet, barriers such as infrastructure gaps, affordability issues, and low digital literacy persist. As emphasized by Edquist, Goodridge and Haskel (2021), internet access is now a necessity for economic participation. This study investigates internet penetration and social media usage trends in Zimbabwe and focuses on usage of big data analytics and machine learning. The goal is to provide data-driven recommendations for improving internet access, affordability, and digital adoption across key economic sectors.

1.2 Problem Statement

Zimbabwe's economy depends on small and medium-sized enterprises (SMEs), healthcare, and agriculture, yet limited internet access and high data costs hinder growth, as highlighted by Allioui and Mourdi (2023). Internet penetration remains uneven, with urban areas experiencing high connectivity but lacking full penetration due to costly data, making access challenging for many. Meanwhile, rural regions have little to no connectivity due to inadequate network infrastructure, further widening the digital divide. This gap prevents businesses, healthcare providers, and farmers from fully utilizing digital tools, restricting economic progress (Edquist, Goodridge and Haskel, 2021).

SMEs that rely on digital platforms such as Facebook Marketplace, WhatsApp Business, and mobile transactions struggle due to expensive data and weak internet infrastructure, reducing their market reach and competitiveness (Khan, 2023). Similarly, digital healthcare solutions such as telemedicine and electronic health records remain underutilized due to limited internet connectivity, especially in rural clinics, restricting access to efficient medical services (Rehman, Naz and Razzak, 2022). In agriculture, farmers lack access to real-time market prices, weather forecasts, and online advisory services, leading to lower productivity and financial instability (Zengeya, Sambo and Mabika, 2021).

Addressing these challenges requires affordable internet expansion, stronger digital infrastructure, and digital literacy programs. As highlighted by Allioui and Mourdi (2023), leveraging digital solutions can drive financial growth and economic stability. This study examines Zimbabwe's internet penetration trends and their impact on key sectors, providing data-driven insights to bridge the digital divide and support economic development.

1.3 Aim and Questions

1.3.1 Aim

The aim of this study is to analyse and predict internet penetration trends in Zimbabwe using big data analytics, focusing on how improved internet access can enhance SMEs, healthcare, and agriculture, and in turn, contribute to economic growth through digital inclusion.

1.3.2 Research Questions

1. What have been the historical internet penetration trends in Zimbabwe, and how can big data analytics be used to analyse them?
2. What are the emerging social media usage patterns in Zimbabwe?
3. How can a predictive model be developed to estimate future internet penetration in Zimbabwe?
4. What data-driven strategies can be recommended to enhance digital inclusion and promote economic growth in Zimbabwe?

1.4 Significance of the Study

This study is crucial in demonstrating how improved internet access drives economic growth in Zimbabwe, particularly in retail, healthcare, and agriculture sectors. As highlighted by Edquist, Goodridge, and Haskel (2021), digital connectivity enhances productivity and economic expansion. However, high data costs and limited infrastructure hinder digital adoption. Despite the increasing reliance on digital tools, a significant gap remains. in Zimbabwe, where inadequate connectivity limits economic participation (Allioui and Mourdi, 2023). By analysing internet penetration trends and forecasting growth, this research provides insights into enhancing digital inclusion and economic participation.

In retail, SMEs rely on platforms like WhatsApp Business and Facebook Marketplace for market access and mobile money services such as EcoCash for transactions (Khan, 2023). Yet, high data costs and digital literacy gaps limit their potential (Allioui and Mourdi, 2023). Similarly, in healthcare, internet access enables telemedicine, digital records, and mobile health (mHealth) solutions, but weak connectivity in rural areas restricts their impact (Rehman, Naz, and Razzak, 2022). In agriculture, digital tools help farmers monitor weather, track market prices, and access expert guidance, but poor internet access limits their use (Zengeya, Sambo, and Mabika, 2021). Strengthening digital adoption across these sectors can boost economic stability, as emphasized by Edquist, Goodridge, and Haskel (2021). This study supports policy efforts to improve affordability, infrastructure, and digital literacy, ensuring that Zimbabwe's economy fully benefits from internet-driven innovations.

1.5 Definition of Key Terms

1. **Internet Penetration** – The percentage of the population with access to the internet (Edquist, Goodridge & Haskel, 2021).
2. **Big Data Analytics** – The process of analysing large datasets to uncover patterns, trends, and insights (Rehman, Naz & Razzak, 2022).
3. **Digital Inclusion** – Ensuring that individuals and businesses have access to affordable internet and digital tools (Khan, 2023).

1.6 Chapter Summary

This chapter introduced the study, highlighting how internet access impacts small businesses, healthcare, and agriculture in Zimbabwe. It outlined the problem of limited internet access and high costs, which restrict digital adoption and economic growth. The study aims to analyse internet penetration trends and predict future growth using big data analytics. The next chapter reviews existing literature on internet penetration, digital inclusion, and sector-specific impacts, identifying gaps that this research addresses.

2. LITERATURE REVIEW

2.1 Introduction

Internet penetration is a key driver of economic transformation, particularly in critical sectors such as retail, healthcare, and agriculture. In Zimbabwe, however, barriers like high data costs, limited infrastructure, and digital skill gaps continue to slow progress toward full digital inclusion. Although platforms like WhatsApp and Facebook are used for marketing and communication, unreliable connectivity limits their broader economic impact. To address these challenges, researchers have applied various machine learning techniques including decision trees, random forests, support vector machines (SVMs), k-nearest neighbours (KNN), deep learning, and linear regression to analyse internet usage trends, understand social media behaviour, and support evidence-based strategies for boosting digital access and sector-specific growth.

2.2 Machine Learning Techniques for Analyzing Internet Usage and Social Media Trends

Firstly, decision trees are widely used to analyse internet usage trends due to their clarity and ability to manage complex datasets. As highlighted by Hemal et al. (2024), decision trees can effectively model user behaviour and reveal patterns in digital access. An advantage is their interpretability, making results easy to understand for policymakers. However, they are prone to overfitting and sensitive to slight changes in data. A research gap exists in applying decision trees to Zimbabwean data, where digital growth is shaped by unique infrastructural and socio-economic conditions.

Secondly, random forests, an extension of decision trees, improve predictive accuracy by combining multiple tree models. Savci et al. (2022) demonstrated their effectiveness in classifying problematic internet use. These models are robust and reduce overfitting through ensemble learning. However, they demand high computational resources and are harder to interpret than single trees. Research is lacking on how random forests perform in tracking internet usage growth over time in African countries.

Thirdly, support vector machines (SVMs) are powerful for detecting patterns in social media behaviour and classifying user groups. As denoted by Sahoo and Gupta (2021), SVMs perform well on high-dimensional data and can identify subtle trends in online usage. Their main advantage is accuracy in binary classification tasks. Yet, SVMs are less interpretable and can be time-consuming to tune. There is limited evidence of SVM applications tailored to internet and social data from Zimbabwe or similar regions.

Fourthly, k-nearest neighbours (KNN) is a simple classification algorithm used to group users based on behavioural similarity. Savci et al. (2022) noted KNN’s effectiveness in identifying at-risk social media users. Its advantage lies in minimal training time and ease of implementation. However, it performs poorly with large datasets and is sensitive to irrelevant features. The method is underused in low-data environments like Zimbabwe, where it could serve as a practical alternative to complex models.

Fifthly, deep learning models such as neural networks are effective for analyzing vast amounts of unstructured social media data. As highlighted by Sahoo and Gupta (2021), these models excel at processing text, audio, and time-series content. Their strength lies in recognizing deep, non-obvious patterns in data. Nonetheless, they require extensive datasets and computing power, and their outputs are often difficult to interpret. A major gap exists in adapting these models for use in data-scarce African regions.

Lastly, linear regression remains a valuable tool for modelling internet penetration trends and forecasting growth. As highlighted by Hemal et al. (2024), it clearly shows how variables like income or infrastructure investment relate to digital inclusion. Its key advantages are simplicity and transparency. However, it assumes a linear relationship between variables, which may not capture real-world complexity. Despite this, linear regression is ideal for Zimbabwe, where clear, cost-effective modelling is crucial for data-driven policy and economic planning.

2.3 The Role of Internet Penetration Across Key Sectors

Internet penetration has played a pivotal role in transforming Zimbabwe's retail, healthcare, and agricultural sectors, although affordability and infrastructural barriers continue to impede progress. In the retail sector, as denoted by Mandipa and Sibindi (2022), SMEs have leveraged digital platforms such as WhatsApp Business and Facebook Marketplace to expand their market reach. Similarly, the healthcare sector has made attempts to integrate electronic health records and telemedicine, though limited internet infrastructure has slowed implementation, as highlighted by Mhazo and Maponga (2022). In agriculture, digital advisory services and mobile-based climate tools have shown promise, but their reach remains limited due to poor rural connectivity, as emphasized by Mayoyo et al. (2023).

Despite these digital opportunities, widespread adoption remains a challenge across sectors. In retail, many businesses in rural areas still struggle with low digital literacy and limited ICT infrastructure, as denoted by Muzunze and Takavarasha (2022). In the healthcare sector, as highlighted by Madziwa (2021), frequent network disruptions and insufficient resources undermine digital service delivery. Likewise, in agriculture, farmers face difficulty accessing online platforms due to inadequate training and digital exclusion, as emphasized by Masere and Worth (2021). These persistent challenges demonstrate the need for locally tailored digital interventions and infrastructure development.

Cross-country comparisons reveal the benefits of strategic digital investment. As denoted by Matsongoni and Mutambara (2021), South Africa's support for SME digitalization has led to improved financial performance in the retail sector. In contrast, Zimbabwe lacks sufficient policy and financial support for sectors like healthcare, where providers continue to operate without reliable internet subsidies or telehealth incentives, as highlighted by Ncube and Songo (2024). Similarly, Kenya's successful implementation of mobile-based agricultural services has not been mirrored in Zimbabwe, due to poor infrastructure and inconsistent policy enforcement, as emphasized by Mayoyo et al. (2023).

To fully leverage the internet's transformative potential, Zimbabwe must strengthen access, affordability, and digital capabilities across retail, healthcare, and agriculture. As highlighted by Maphosa and Maphosa (2022), digital adoption among SMEs requires targeted training and financial support. In the healthcare sector, strategic investment in digital infrastructure can enhance access and outcomes, as denoted by Rehman, Naz and Razzak (2022). In agriculture, improving rural connectivity and digital literacy will be critical to maximizing productivity and resilience, as emphasized by Zengeya, Sambo and Mabika (2021).

2.4 Chapter Summary

This chapter explored the role of internet penetration in Zimbabwe's retail, healthcare, and agricultural sectors, highlighting key challenges and opportunities for digital adoption. The chapter also shed light on how researchers have applied various machine learning techniques including decision trees, random forests, support vector machines (SVMs), k-nearest neighbours (KNN), deep learning, and linear regression to analyse internet and social media usage trends. While internet access has enabled business growth, improved healthcare service delivery, and enhanced agricultural productivity, barriers such as high data costs, inadequate infrastructure, and low digital literacy continue to hinder progress. A global comparison underscores the potential benefits of strategic policy interventions in overcoming these challenges. Addressing these gaps through targeted digital inclusion initiatives and infrastructure development is essential for ensuring that Zimbabwe fully capitalizes on the economic opportunities presented by internet penetration.

CHAPTER 3 METHODOLOGY

3.1 Introduction

This study utilised big data analytics to analyse internet penetration trends in Zimbabwe, that can be impactful on the retail, healthcare, and agricultural sectors. The analysis incorporated data from Data Reportal Website that was collected from various reputable sources, all of which were cross-checked to ensure their accuracy. This approach allowed for a comprehensive and data-driven understanding of the factors influencing internet penetration and its impact on key sectors. The data collected spanned from 2020 to 2025, with a specific focus on the role of the internet in shaping Zimbabwe's economy. The aim was to use historical data and predictive modeling to forecast future trends in internet access and its implications for small businesses, healthcare providers, and farmers in Zimbabwe. As highlighted by Guzman et al. (2021), the combination of data from different sources allows for more accurate strategic forecasting, making the use of already collected data a practical and effective method for this analysis.

3.2 Data Collection and Sources

Data for this study was sourced from the Data Reportal Website, a collection of data from multiple platforms to ensure a holistic and accurate understanding of internet penetration trends in Zimbabwe (DataReportal, n.d.). The data was in the form of narrative text that was presented in various different formats such as bullet points, descriptive paragraphs, and quoted text, making it unstructured data, rather than being in a standardized, machine-readable structured data format like CSV or JSON. This made it challenging to process automatically, hence requiring manual extraction of relevant information needed information for analysis., The data was compiled by *We Are Social & Meltwater* and gathered from various primary data sources, with all datasets collected and verified through the GSMA Intelligence Website (GSMA Intelligence, n.d.). The data provided insights into national internet coverage, penetration rates, and service availability, as well as data on broadband service usage, customer growth, and regional access disparities. By cross-checking and consolidating these datasets, the researcher ensured that the analysis reflected the most current and accurate information available.

The selection of 2020 to 2025 as the timeframe for this study was significant. In 2020, Zimbabwe's Minister of ICT, Jenfan Muswere, emphasized that broadband connectivity was no longer a luxury but a fundamental right (Muswere, 2020). This marked a pivotal shift in policy, especially during the COVID-19 pandemic, which highlighted the crucial role of the internet in maintaining business operations, healthcare services, and agricultural productivity. The disruption caused by the pandemic underscored the necessity for reliable internet infrastructure, making this period an ideal window to assess both the immediate effects and long-term potential of internet penetration in Zimbabwe.

The data for this study was sourced from a range of reputable reports, focusing on the trends in digital penetration in Zimbabwe from 2020 to 2025. These reports were published by *We Are Social and Meltwater*, on the *Data Reportal Website*, which provided insights into the country's digital landscape, internet penetration, and user behavior. The reports from *Digital 2020: Zimbabwe* through *Digital 2025: Zimbabwe* offered a comprehensive analysis of internet use, social media trends, and digital adoption across various sectors. The reports accessed from the Data Portal Website include:

- **Digital 2020: Zimbabwe** (We Are Social & Meltwater, 2020)
- **Digital 2021: Zimbabwe** (We Are Social & Meltwater, 2021)
- **Digital 2022: Zimbabwe** (We Are Social & Meltwater, 2022)

- **Digital 2023: Zimbabwe** (We Are Social & Meltwater, 2023)
- **Digital 2024: Zimbabwe** (We Are Social & Meltwater, 2024)
- **Digital 2025: Zimbabwe** (We Are Social & Meltwater, 2025)

These datasets provided key insights into the evolving trends of internet access and digital infrastructure, which are crucial for analysing the impact of internet penetration on the retail, healthcare, and agricultural sectors in Zimbabwe.

3.3 Data Pre-processing and Analysis

Once the data was collected from the aforementioned source, Data Reportal Website, reports from 2020 to 2025 were manually converted into datasets in python as shown in the Python code in the Appendix section. These datasets were thoroughly cross-checked to ensure accuracy as mentioned above, which is a crucial step as denoted by Guzman et al. (2021). The Data Reportal reports for the selected years were read through carefully, one by one, in order to identify common features and extract useful data relevant to this study. This process helped in recognizing the key variables that could provide insights into internet penetration trends in Zimbabwe and their impact on the influential sectors such as the retail, healthcare, and agricultural sectors. The selected variables were: Total Population, Zimbabwe's Population By Age, Population Growth Rate, Urban vs. Rural Population, Internet Users, Social Media Users, Mobile Connections, and Gender Distribution. Zimbabwe's population by age was removed during feature engineering, as age was not considered useful for analyzing internet trends from 2020 to 2025. No missing data was found in the selected variables for this period.

Python was employed as it was seen by the researcher as the best fit tool, and the decision was supported by Igual, Seguí (2024). The following packages were used to analyse the data after pre-processing *`pandas, matplotlib.pyplot, numpy and sklearn.linear_model`*. Once the data was cleaned and processed, a forecast was generated using machine learning techniques, such as time series forecasting, to predict future trends in internet penetration. This analysis allowed for a data-driven approach to understanding the potential implications of increased internet access in Zimbabwe.

The following steps were taken during the analysis:

- Exploratory Data Analysis (EDA) involved visualization of tables and statistical summaries to compare trends per year at a glance and inform modelling choices.
- Linear regression was chosen for model development as a simple, interpretable algorithm well-suited for tracking changes over time.
- Evaluation used a domain-appropriate metric, Mean Absolute Error (MAE), to measure prediction accuracy.

3.4 Ethical Considerations in Data Collection and Analysis

Ethical considerations in this study were paramount, especially given the sensitive nature of the data collected and the reliance on publicly available reports. First, the data used in this study was collected from reputable and authorized sources, ensuring that it complied with relevant data protection regulations. Moreover, while the study relied on secondary data obtained from these sources, ethical concerns regarding the use of publicly available data were considered. It was essential to acknowledge that online data collection, especially when aggregating and analysing public reports, required a level of transparency to ensure that the rights of data subjects were not violated (Krotov and Johnson, 2023). In line with ethical research practices, the data was used strictly for the purpose of this analysis, with no individual-level data or personal information disclosed. The potential risks of misusing such data for commercial or non-research purposes were carefully mitigated, and appropriate citations were included to give credit to the original data providers. In conclusion, this study's methodology upheld the highest ethical standards in data collection and analysis, ensuring that the research was both responsible and credible.

3.5 Chapter Summary

This chapter outlined the methodology employed in this study, which aimed to analyse internet penetration trends in Zimbabwe and their potential impact on the retail, healthcare, and agricultural sectors. The study utilized big data analytics, incorporating data from multiple reputable sources, including the *Data Reportal* Website, which hosted reports from *We Are Social and Meltwater*, spanning the years 2020 to 2025. These datasets, which were cross-checked for accuracy, provided insights into Zimbabwe's digital landscape, internet coverage, and user behavior through data on *GSMA Intelligence* Website. Data processing involved using Python for data analysis, and forecasting, employing packages like `pandas, matplotlib.pyplot, numpy and sklearn.linear_model`. The study also employed machine learning techniques to predict future trends in internet penetration. Ethical considerations were prioritized, ensuring compliance with data protection regulations and transparency in data usage. The research aimed to generate a comprehensive, data-driven understanding of internet access's role in shaping the Zimbabwean economy.

4. FINDINGS AND RESULTS

4.1 Demographics

Table 1: Demographics

Indicator	2020	2021	2022	2023	2024	2025
Total Population	Not specified	14.98 million	15.21 million	16.49 million	16.84 million	16.80 million
Population Growth Rate	Not specified	+1.5% from 2020	+1.6% from 2021	+2.1% from 2022	+2.2% from 2023	+1.9% from 2024
Gender Distribution	Not specified	52.3% female, 47.7% male	52.3% female, 47.7% male	52.8% female, 47.2% male	52.7% female, 47.3% male	52.3% female, 47.7% male
Urban vs. Rural Population	Not specified	32.3% urban, 67.7% rural	32.5% urban, 67.5% rural	32.5% urban, 67.5% rural	32.7% urban, 67.3% rural	32.8% urban, 67.2% rural
Internet Users	4.81 million (33%)	5.01 million (33.4%)	4.65 million (30.6%)	5.74 million (34.8%)	5.48 million (32.6%)	6.45 million (38.4%)

Social Media Users	980 thousand (6.6%)	1.30 million (8.7%)	1.55 million (10.2%)	1.50 million (9.1%)	2.05 million (12.2%)	2.10 million (12.5%)
Mobile Connections	12.18 million (83%)	14.76 million (98.5%)	Not specified	14.08 million (85.4%)	13.88 million (82.4%)	15.20 million (90.6%)

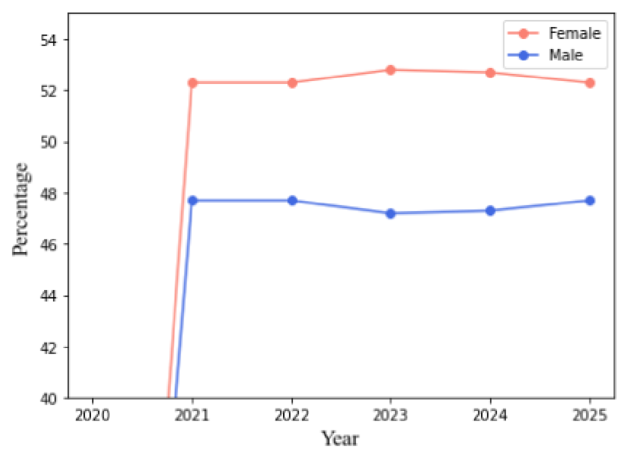

Figure 1: Gender Distribution

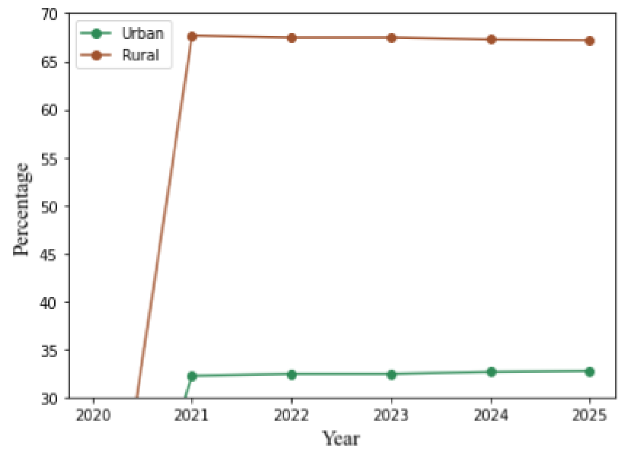

Figure 2: Urban vs. Rural Population

The demographic data of Zimbabwe from 2020 to 2025, as shown in table 1 as well as Figure 1 and Figure 2 above, show a steady increase in population, growing from approximately 14.98 million in 2021 to 16.80 million in 2025. The gender distribution remains relatively stable, with around 52-53% female and 47-48% male. The urban population gradually increases from 32.3% in 2021 to 32.8% in 2025, while the rural population slightly declines. Internet penetration fluctuates, with users rising from 4.81 million (33%) in 2020 to 6.45 million (38.4%) in 2025. Social media usage also increases, growing from 980,000 users (6.6%) in 2020 to 2.10 million (12.5%) in 2025. Mobile connections vary over the years, reaching a peak of 14.76 million (98.5%) in 2021 before stabilizing at 15.20 million (90.6%) in 2025. The median age remains low, ranging from 18.1 to 19.0 years across the years.

4.2 Internet Usage

Table 2: Internet Usage in Zimbabwe 2020 to 2025

Year	Total Internet Users (millions)	Annual Increase in Internet Users	Annual Growth Rate (%)	Internet Penetration Rate (%)
2020	4.81	413 000	9.4	33.0
2021	5.01	203 000	4.2	33.4
2022	4.65	265 000	6.0	30.6

2023	5.74	117 000	2.1	34.8
2024	5.48	115 000	2.2	32.6
2025	6.45	120 000	1.9	38.4

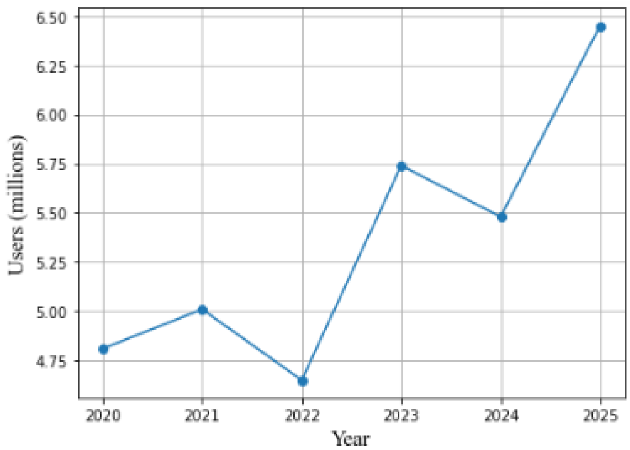

Figure 3: Total Internet Users (2020-2025)

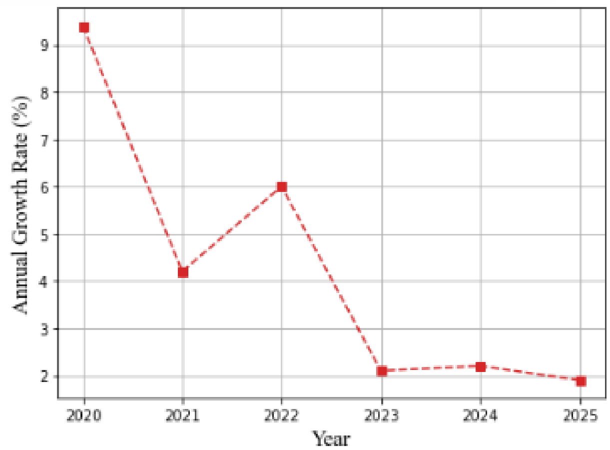

Figure 4: Annual Growth Rate In Internet Users (2020-2025)

From 2020 to 2025, internet usage trends in Zimbabwe revealed significant fluctuations in user numbers, growth rates, and penetration levels. Starting with 4.81 million users in 2020 (9.4% growth, 33% penetration), usage rose to 5.01 million in 2021, though growth slowed to 4.2%. In 2022, users dropped to 4.65 million, despite a 6.0% growth rate and a reduced 30.6% penetration. A rebound in 2023 brought usage to 5.74 million with only 2.1% growth and 34.8% penetration. In 2024, user numbers dipped slightly to 5.48 million (2.2% growth, 32.6% penetration). By 2025, users increased significantly to 6.45 million, with the highest penetration rate of 38.4% but the lowest growth rate of 1.9%, indicating shifting dynamics in internet adoption, accessibility, and infrastructure.

4.3 Social Media Usage

Table 3: Social Media Usage in Zimbabwe 2020 to 2025

Year	Total Social Media Users (millions)	Annual Change in Social Media Users	Annual Growth Rate (%)	Social Media Penetration Rate (%)
2020	0.98	-135 000	-12.0	6.6
2021	1.30	320 000	33.0	8.7
2022	1.55	250 000	19.2	10.2
2023	1.50	-50 000	-3.2	9.1
2024	2.05	550 000	36.6	12.2
2025	2.10	50 000	2.4	12.5

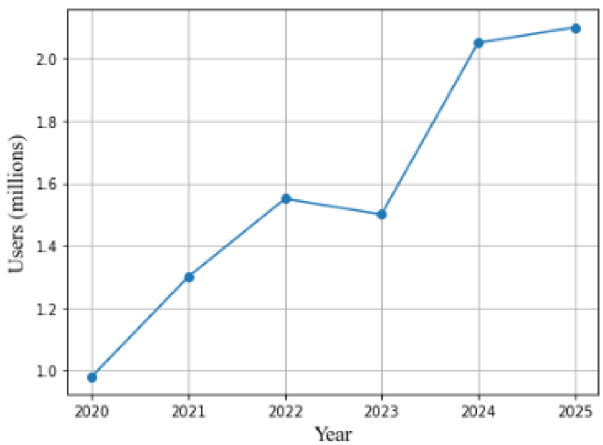

Figure 5: Total Social Media Users (2020-2025)

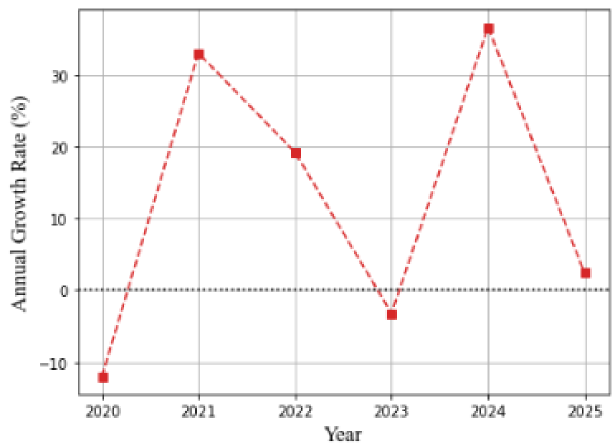

Figure 6: Annual Growth Rate of Social Media Users (2020-2025)

Between 2020 and 2025, social media usage in Zimbabwe experienced notable fluctuations in user numbers, growth rates, and penetration. In 2020, user numbers fell to 980,000 after a drop of 135,000, resulting in a -12% growth rate and 6.6% penetration. Adoption surged by 2022, reaching 1.55 million users with a 250,000 annual increase, a 19.2% growth rate, and 10.2% penetration. However, in 2023, users declined slightly to 1.50 million, with a -3.2% growth rate and 9.1% penetration. A strong rebound occurred in 2024 as users rose to 2.05 million (36.6% growth, 12.2% penetration). By 2025, growth slowed to 2.4% with 2.10 million users and 12.5% penetration, suggesting a period of stabilization after a sharp rise in adoption.

4.4 Summary Statistics

Table 4: Summary Statistics for Internet Users from 2020 to 2025

Metric	Total Internet Users (millions)	Annual Increase in Internet Users	Annual Growth Rate (%)	Internet Penetration Rate (%)
Count	6.000000	6.000000	6.000000	6.000000
Mean	5.356667	205 500	4.300000	33.800000
Standard Deviation	0.674675	118 267	2.963781	2.632109
Minimum	4.650000	115 000	1.900000	30.600000
25th Percentile (Q1)	4.860000	117 750	2.125000	32.700000
Median (50%)	5.245000	161 500	3.200000	33.200000
75th Percentile (Q3)	5.675000	249 500	5.550000	34.450000
Maximum	6.450000	413 000	9.400000	38.400000

The table 4 above presents the statistical summary of internet usage trends in Zimbabwe from 2020 to 2025 and represents the final stage of the exploratory data analysis (EDA) in this research. Key insights include an average of 5.36 million total internet users over the six-year period, with a mean annual increase of 205,500 users. The internet penetration rate averaged 33.8%, showing modest variability (standard deviation of 2.63%) and reaching a peak of 38.4% in 2025. The annual growth rate in user numbers exhibited a wider range, from a low of 1.9% to a high of 9.4%, indicating periods of both stagnation and rapid adoption. These insights provide a foundational understanding of internet trends in Zimbabwe, serving as a baseline for deeper analysis on its socio-economic impacts.

Table 5: Social Media Usage in Zimbabwe from 2020 to 2025

Metric	Total Social Media Users (millions)	Annual Change in Social Media Users	Annual Growth Rate (%)	Social Media Penetration Rate (%)
Count	6.000000	6.000000	6.000000	6.000000
Mean	1.580000	164 167	12.666667	9.883333
Standard Deviation	0.432897	256 484	19.972648	2.240908
Minimum	0.980000	-135 000	-12.000000	6.600000
25th Percentile (Q1)	1.350000	-25 000	-1.800000	8.800000
Median (50%)	1.525000	150 000	10.800000	9.650000
75th Percentile (Q3)	1.925000	302 500	29.550000	11.700000
Maximum	2.100000	550 000	36.600000	12.500000

The table 5 above provides a detailed statistical summary of Social Media Usage in Zimbabwe from 2020 to 2025, marking the final stage of the exploratory data analysis (EDA) in this segment of the research. Over the six-year period, Zimbabwe saw an average of 1.58 million social media users, with considerable variability as shown by a standard deviation of 0.43 million. The annual change in users was highly volatile, ranging from a decrease of 135,000 users in 2020 to an increase of 550,000 in 2024, reflecting a fluctuating digital engagement landscape. Similarly, the annual growth rate ranged from -12.0% to 36.6%, while social media penetration steadily increased, averaging 9.88% and peaking at 12.5% in 2025. These insights highlight both the growth potential and instability of social media adoption, providing crucial context for analyzing its influence on Zimbabwean society and digital transformation.

5. DISCUSSION OF RESULTS

5.1 Demographics

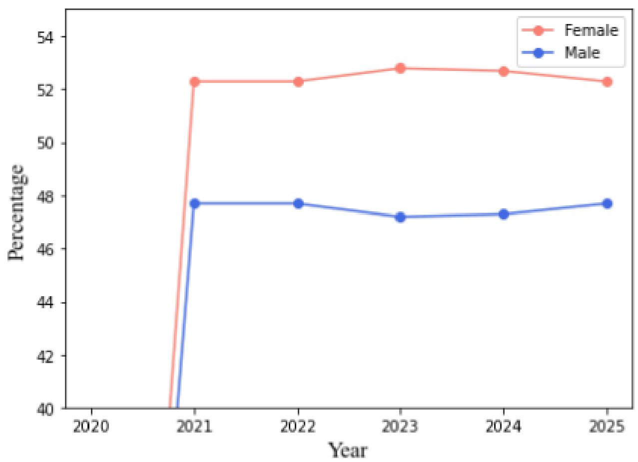

Figure 1: Gender Distribution

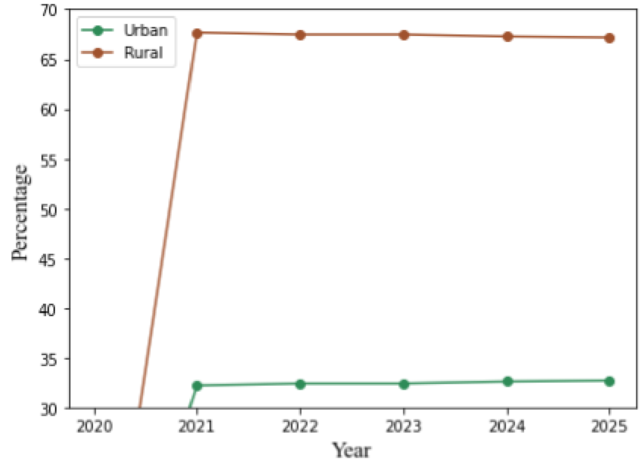

Figure 2: Urban vs. Rural Population

The demographic data for the year 2020 (Population Growth Rate, Population Growth Rate, Gender Distribution and Urban vs. Rural Population) are missing and this is shown in table 1 in chapter 4.1, as well as in figure 1 and 2 above, which suggests the presence of missing data. Based on the classification of missing data, as discussed by Carpenter and Smuk (2021), this appears to be Missing Not At Random (MNAR). This classification is justified because the absence of data for 2020 could be due to a systematic issue, such as the lack of data collection processes during that year or external disruptions affecting data availability. Unlike Missing At Random (MAR), where the probability of missingness depends on observed variables, MNAR suggests that the missing values themselves are related to unobserved factors, such as infrastructural limitations or data privacy concerns that may have prevented data reporting.

The trends observed in the figures align with broader discussions on digital accessibility and demographic shifts. As highlighted by Murthy et al. (2021), disparities in digital infrastructure can influence the availability and collection of demographic data, particularly in rural areas. The dominance of the rural population in Figure 2 suggests potential challenges in digital data collection, reinforcing arguments made by Wang et al. (2021) on the uneven evolution of data-driven technologies like IoT. Additionally, as denoted by Auxier and Anderson (2021), social media and digital platforms play a crucial role in demographic data collection, which may explain why certain population segments, particularly those with limited digital access, remain underrepresented. A slight increase in the urban population, accompanied by a nearly insignificant decline in the rural population, reflects common urbanization trends where major shifts had already occurred in previous years. This is consistent with the discussion by Allioui and Mourdi (2023), who emphasized that digital transformation significantly impacts financial and social inclusivity, reinforcing the gradual yet steady movement toward urban digital integration.

5.2 Internet Usage

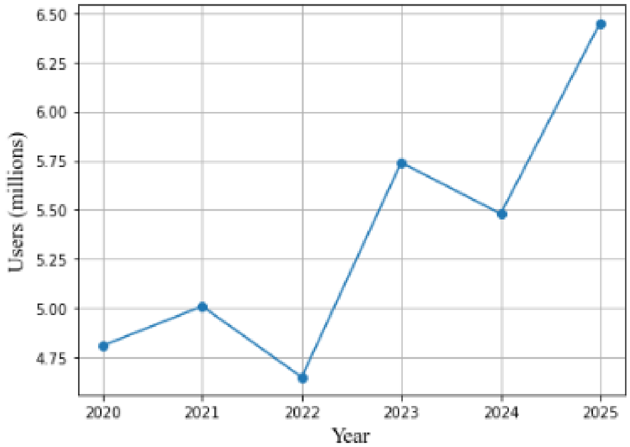

Figure 3: Total Internet Users (2020-2025)

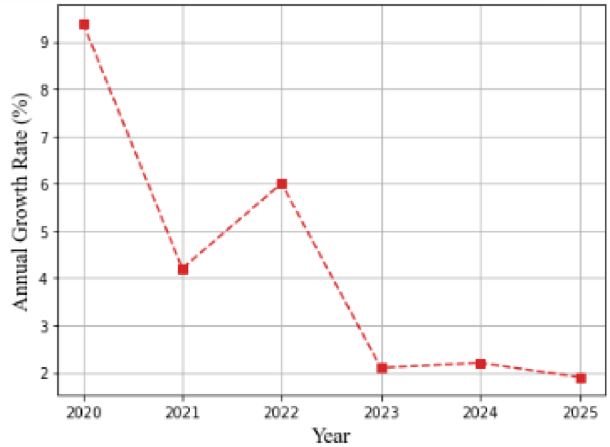

Figure 4: Annual Growth Rate In Internet Users (2020-2025)

The fluctuations in internet growth rates observed in Zimbabwe between 2020 and 2025, as shown in figure 3 and 4 above, indicate shifts in digital accessibility and infrastructure development. Growth rates play a crucial role in measuring digital expansion, as highlighted by Murthy et al. (2021), who argue that disparities in internet adoption reflect broader digital divide concerns. The growth rate started high at 9.4% in 2020, then dropped to 4.2% in 2021, and continued to decline to 6.0% in 2022, 2.1% in 2023, 2.2% in 2024, and finally 1.9% in 2025. This declining trend suggests a saturation effect, where initial rapid adoption slows as internet penetration reaches a larger proportion of the population. This trend aligns with findings by Edquist et al. (2021), who emphasized that while the Internet of Things (IoT) contributes to economic expansion, its adoption follows an S-curve where growth stabilizes after early adoption phases. While penetration increased to 38.4% in 2025, the diminishing growth rate suggests that external factors such as affordability, infrastructure, and policy limitations may be impeding further acceleration.

Additionally, the variation in penetration rates reflects the evolving role of internet accessibility in economic and social development. As denoted by Khan (2023), digital penetration is a key factor in the expansion of e-commerce and e-business, particularly for small and medium enterprises (SMEs), which form a significant portion of Zimbabwe's economy. The penetration rate rose from 33.0% in 2020 to 33.4% in 2021, before dropping to 30.6% in 2022. It then rebounded to 34.8% in 2023, before slightly declining again to 32.6% in 2024 and finally increasing to 38.4% in 2025. The temporary decline in penetration rates in 2022 and 2024 suggests potential disruptions in service availability or affordability issues, which can hinder consistent growth. However, the subsequent rise in 2025 suggests efforts to improve digital infrastructure or policy interventions that enhanced accessibility. The fluctuations emphasize the challenges identified by Maphosa and Maphosa (2022), who pointed out that infrastructure and affordability remain significant barriers to sustained ICT adoption in Zimbabwe.

The steady yet fluctuating adoption of internet services in Zimbabwe also aligns with global trends in big data utilization and digital transformation. As highlighted by Krotov and Johnson (2023), internet growth directly influences data availability, shaping business strategies and economic policies. Similarly, Rehman et al. (2022) argue that leveraging big data analytics in sectors like healthcare and finance depends on stable internet penetration. The observed trends in Zimbabwe suggest a transitional phase where digital adoption is progressing, albeit with fluctuations due to infrastructural, economic, and policy-related factors. The long-term sustainability of internet growth will likely depend on addressing these challenges, ensuring that penetration rates remain stable while growth rates do not decline to levels that may indicate stagnation.

5.3 Social Media Usage

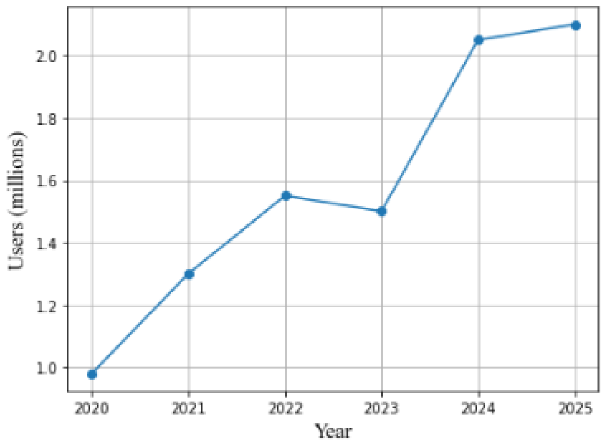

Figure 5: Total Social Media Users (2020-2025)

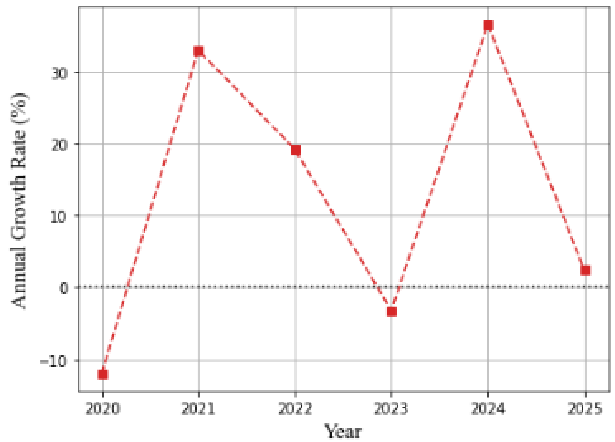

Figure 6: Annual Growth Rate of Social Media Users (2020-2025)

The fluctuations in social media usage in Zimbabwe between 2020 and 2025, as shown in figure 5 and 6 above, highlight significant shifts in digital engagement and accessibility. The growth rate initially showed a sharp decline at -12% in 2021, reflecting a temporary disengagement from social media platforms, likely due to economic or accessibility challenges. However, by 2022, the strong recovery at 19.2% demonstrated renewed interest, aligning with global digital expansion trends. As highlighted by Auxier and Anderson (2021), social media adoption often follows cyclical patterns influenced by changes in user behavior, platform popularity, and internet availability. This fluctuation is further supported by Krotov and Johnson (2023), who argue that big web data trends shape how users interact with digital platforms, influencing adoption patterns. The penetration rate followed a similar trajectory, beginning at 6.6% in 2020, rising to 10.2% in 2022, and reflecting an overall expansion of digital engagement despite short-term setbacks.

The temporary decline in 2023, where the growth rate fell to -3.2%, suggests factors such as platform fatigue, affordability concerns, or internet instability. However, the sharp increase in 2024, where the growth rate peaked at 36.6%, indicates a resurgence in social media adoption, likely driven by enhanced access, affordability improvements, or shifting digital trends. This aligns with Khan (2023), who emphasizes that social media penetration plays a crucial role in e-commerce and digital marketing, particularly for SMEs in developing economies. The penetration rate increased to 12.2%, reinforcing the idea that social media adoption is influenced by broader digital infrastructure trends. Similarly, Maphosa and Maphosa (2022) highlight the role of ICT advancements in increasing digital connectivity, which could explain the observed growth.

By 2025, the growth rate dropped significantly to 2.4%, while the penetration rate slightly increased to 12.5%, suggesting that adoption had begun to stabilize. This trend aligns with Edquist et al. (2021), who argue that internet-based adoption curves follow a pattern of rapid early growth followed by slower stabilization. The slowing growth rate suggests that social media platforms may have reached a saturation point in Zimbabwe, where most potential users had already adopted the technology. Additionally, as denoted by Rehman et al. (2022), social media engagement trends are increasingly shaped by data analytics, content availability, and platform dynamics, meaning that future growth may depend on factors beyond basic access. The overall trends indicate that while social media adoption in Zimbabwe has expanded significantly, sustaining long-term growth will likely require improvements in internet infrastructure, affordability, and digital literacy.

5.4 Forecast – Linear Regression

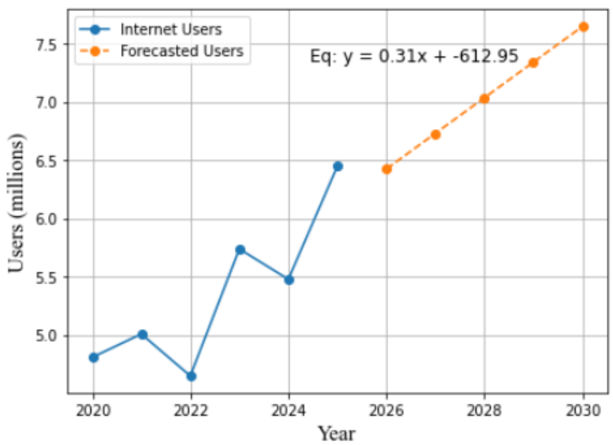

Figure 9: Total Internet Users (2020-2030)

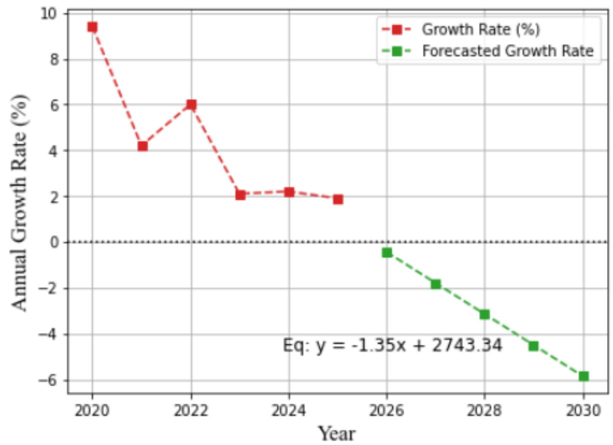

Figure 10: Annual Growth Rate in Internet Users (2020-2030)

The forecast results in figure 9 and 10 above, suggest that Zimbabwe's internet user base will continue growing, but at a declining rate, as indicated by Figure 10's regression equation $y=-1.35x+2743.34$. The trend aligns with Murthy et al. (2021), who highlight that internet growth often slows as adoption reaches a larger share of the population. The fluctuation in growth rates from 2020 to 2025, starting at 9.4% in 2020 and dropping to 1.9% by 2025, supports the idea of a saturation effect, where early rapid adoption slows due to affordability constraints, infrastructure limitations, and digital divide concerns. This pattern is also consistent with Edquist et al. (2021), who argue that technological adoption follows an S-curve, with initial expansion giving way to stabilization. The projected decline in growth rate post-2025 suggests that, without significant policy interventions, Zimbabwe could face stagnation in digital expansion, potentially limiting the economic benefits of increased internet connectivity.

If economic conditions, pricing, and infrastructure development remain unchanged, Zimbabwe could reach a plateau in internet adoption, reducing its impact on key industries such as retail, healthcare, and agriculture. The variations in penetration rates, which rose to 38.4% in 2025 despite prior declines, suggest external factors like affordability and service disruptions influence accessibility. As noted by Khan (2023), digital penetration is critical for expanding e-commerce and SME growth, sectors that are essential to Zimbabwe's economy. However, the declining growth rate suggests that further digital transformation may be hindered without improved infrastructure, affordability, and policy support. To avoid stagnation, efforts should focus on making internet services more accessible and affordable, ensuring continued digital inclusion and economic benefits in the long run.

5.5 Evaluation Metrics

The results presented in the image show the Mean Absolute Error (MAE) for both the Internet Users and the Annual Growth Rate models:

```
MAE for Internet Users: 0.296 million
MAE for Growth Rate: 1.241%
```

Figure 11: Mean Absolute Error (MAE)

As given by Figure 11 above, an MAE of 0.296 million means that, on average, the model's predictions for the total number of internet users are off by about 296,000 users per year. Given the scale of users (in millions), this is a reasonably low error, indicating that the model is fairly accurate for user forecasting (Hodson, 2022). An MAE of 1.241% for the growth rate suggests that the predicted annual internet growth rate deviates from actual values by an average of 1.24 percentage points. Considering that growth rates in the dataset range from around 1.9% to 9.4%, this is a moderate error acceptable for high-level forecasting but potentially sensitive for precise planning. These MAE values suggest that the linear regression models perform acceptably well for both predicting internet user counts and growth rates, especially in a trend analysis context. However, for long-term strategic decisions, it might be beneficial to consider non-linear models or incorporate more variables to reduce prediction error further.

5.6 Limitations

While this study provides valuable insights, several limitations must be acknowledged:

- **Limited Scope:** The study focused on national trends, potentially overlooking regional disparities. A more localized approach could reveal specific challenges faced by different communities.
- **Manual Data Extraction:** Relevant data had to be manually extracted from unstructured formats, which was time-consuming and limited the time available for other critical aspects of the research.
- **Data Availability:** Some useful variables such as device types or income-based access were unavailable. Including such data would have allowed for a more detailed analysis of digital access patterns.

Recognizing these limitations provides a foundation for refining future research methodologies and ensuring more comprehensive analyses.

5.7 Unexpected Outcomes

Several unexpected outcomes emerged during the study:

- **Internet Fluctuations:** Internet usage was expected to follow a steady upward trend, but unexpected drops were observed in 2022 and 2024. These declines were not anticipated and require further investigation to understand possible causes such as affordability issues, infrastructure challenges, or policy changes.
- **Social Media Dip:** In contrast to internet usage, which dipped in 2022 and 2024, social media usage showed a separate decline in 2023. This decoupling suggests that different factors influence general internet access and specific platform usage.
- **Mobile Connection Plateau:** Despite population growth, mobile connections did not increase proportionally, suggesting saturation or shifting usage patterns that warrant deeper analysis.

5.8 Societal/Ethical Implications

The study also raised several societal and ethical considerations:

- **Data Collection Limitations:** The data used was collected by external entities for broad reporting purposes, not specifically for research, which may affect its accuracy or relevance.
- **Privacy Concerns:** Relying on third-party digital data raises concerns about user consent and privacy, especially when individuals are unaware of how their data may be reused.
- **Misinterpretation Risk:** There is a risk of misinterpreting or overgeneralizing patterns from such data, potentially leading to policy decisions that do not reflect the lived realities of underrepresented groups.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusion

This study successfully addressed the first three research questions by analyzing historical internet penetration trends, social media usage patterns, and developing a predictive model for Zimbabwe's digital landscape. Firstly, the historical trends revealed that internet penetration rose from 33.0% in 2020 to 38.4% in 2025, despite fluctuations such as dips in 2022 (30.6%) and 2024 (32.6%). This analysis, using big data techniques, demonstrated how infrastructure and affordability challenges impact adoption. Secondly, emerging social media usage patterns showed notable volatility growth rates ranged from -12% in 2021 to 36.6% in 2024, with overall penetration rising from 6.6% to 12.5% over the five-year period. Thirdly, a linear regression model was developed to forecast internet user trends, yielding a Mean Absolute Error (MAE) of 0.296 million users and 1.241% for growth rate, highlighting reasonable predictive accuracy for high-level planning. These findings offered a foundational understanding of Zimbabwe's digital evolution. The next section addresses the fourth research question by providing data-driven strategies to enhance digital inclusion and promote sustainable economic growth.

6.2 Recommendations

To enhance digital accessibility and sustainable growth, the following recommendations are proposed:

- **Improve Digital Infrastructure:** Expanding rural broadband and investing in reliable, affordable internet through public-private partnerships can reduce the digital divide and ensure sustainability.
- **Promote Affordable Internet Access:** Reducing costs through subsidies or flexible pricing can boost adoption in low-income areas and help SMEs grow via digital platforms.
- **Enhance Digital Literacy Programs:** Digital skills training and awareness campaigns, especially in underserved areas, can boost engagement and empower key sectors like farming, healthcare, and small business.

6.3 Gaps for Future Research

Further research can address the following areas to build on the findings of this study:

- **Impact of Digital Policies:** Examining how government regulations and policies affect internet adoption and digital accessibility will provide critical insights for policymakers. Investigating the effectiveness of current policies can help shape more inclusive digital expansion strategies.
- **Evolving Role of Social-Media:** Investigating how social media influences economic activities, such as e-commerce and digital marketing, can shed light on new growth opportunities. Understanding user engagement trends can also inform businesses on how to maximize social media platforms for revenue generation.

Future research in these areas will contribute to a deeper understanding of Zimbabwe's digital landscape, supporting data-driven decision-making and strategic planning. Addressing these research gaps can help bridge the digital divide and enhance economic growth through informed policy and technological advancements.

End of Paper: 20 Pages

7. REFERENCES

Abendin, S. and Duan, P., 2021. International trade and economic growth in Africa: The role of the digital economy. Cogent economics & finance, 9(1), p.1911767.

Allioui, H. And Mourdi, Y., 2023. Exploring the full potentials of IoT for better financial growth and stability: A comprehensive survey. Sensors, 23(19), p.8015.

Auxier, B. and Anderson, M., 2021. Social media use in 2021. Pew Research Center, 1(1), pp.1-4.

Blessed, M., Eric, M., Tanaka, M. and Susan, M., 2022. The impact of ICT adoption decision-making: Case of Zimbabwean SME owner-managers. International Journal of Research and Innovation in Applied Science, 7(6), pp.55-58.

Brodny, J. and Tutak, M., 2022. Digitalization of small and medium-sized enterprises and economic growth: Evidence for the EU-27 countries. Journal of Open Innovation: Technology, Market, and Complexity, 8(2), p.67.

Carpenter, J.R. and Smuk, M., 2021. Missing data: A statistical framework for practice. Biometrical Journal, 63(5), pp.915-947.

Chilunjika, A., Chilunjika, S.R. and Uwizeyimana, D., 2024. Implementing e-Health initiatives in Zimbabwe's public health sector. Journal of Economic Development, Environment and People, 13(1), pp.55-66.

DataReportal (n.d.) Digital in Zimbabwe. Available at: https://datareportal.com/digital-in-zimbabwe (Accessed: 10 February 2025).

Dudhat, A. and Agarwal, V., 2023. Indonesia's digital economy's development. IAIC Transactions on Sustainable Digital Innovation (ITSDI), 4(2), pp.109-118.

Edquist, H., Goodridge, P. And Haskel, J., 2021. The Internet of Things and economic growth in a panel of countries. Economics of Innovation and New Technology, 30(3), pp.262-283.

GSMA Intelligence (n.d.) GSMA Intelligence. Available at: https://www.gsmaintelligence.com/ (Accessed: 10 February 2025).

Guzman, L.A., Arellana, J., Cantillo-García, V. and Ortúzar, J.D.D., 2021. Revisiting the benefits of combining data of a different nature: Strategic forecasting of new mode alternatives. Journal of Advanced Transportation, 2021(1), p.6672961.

Hemal, S.H., Khan, M.A.R., Ahammad, I., Rahman, M., Khan, M.A.S.D. and Ejaz, S., 2024. Predicting the impact of internet usage on students' academic performance using machine learning techniques in Bangladesh perspective. Social Network Analysis and Mining, 14(1), p.66.

Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, 2022, pp.1-10.

Igual, L. and Seguí, S., 2024. Introduction to data science. In Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications (pp. 1-4). Cham: Springer International Publishing.

Jiao, S. and Sun, Q., 2021. Digital economic development and its impact on economic growth in China: Research based on the prespective of sustainability. Sustainability, 13(18), p.10245.

Khan, S.A., 2023. E-marketing, e-commerce, e-business, and internet of things: an overview of terms in the context of small and medium enterprises (SMEs). Global applications of the internet of things in digital marketing, pp.332-348.

Krotov, V. and Johnson, L., 2023. Big web data: Challenges related to data, technology, legality, and ethics. Business Horizons, 66(4), pp.481-491.

Kurniawati, M.A., 2022. Analysis of the impact of information communication technology on economic growth: empirical evidence from Asian countries. Journal of Asian Business and Economic Studies, 29(1), pp.2-18.

Madziwa, P.K., 2021. Effects of network disruptions on health service delivery among private practicing physicians in Bulawayo Metropolitan Province, Zimbabwe. IJHISI, 14(2), pp.17-21.

Mandipa, G. and Sibindi, A.B., 2022. Financial performance and working capital management practices in the retail sector: empirical evidence from South Africa. Risks, 10(3), p.63.

Maphosa, V. and Maphosa, M., 2022. Factors influencing the adoption of ICT for remote work among Zimbabwean SMEs: A case study of Bulawayo Metropolitan province. International Journal of Advanced and Applied Sciences, 9(3), pp.150-158.

Masere, T.P. and Worth, S., 2021. Influence of public agricultural extension on technology adoption by small-scale farmers in Zimbabwe. South African Journal of Agricultural Extension, 49(2), pp.25-42.

Mashavira, N., Nyoni, N.D., Mathibe, M.S. and Chada, L., 2023. Work-life balance in the Zimbabwe Retail Sector: Testing a job-engagement and job-satisfaction model. Acta Commercii, 23(1), pp.1-10.

Matsongoni, H. and Mutambara, E., 2021. Assessing the role of informal manufacturing SMEs and effectiveness of government economic programmes to enhance their performance in Zimbabwe. International Journal of Entrepreneurship, 25.

Mayoyo, A., Chapoto, A., Matchaya, G., Aheeyar, M., Chiwunze, G., Ebrahim, G., Ajayi, O.C., Afun-Ogidan, O., Fakudze, B. and Kasoma-Pele, W., 2023. Digital climate adaptation in agriculture profile for Zimbabwe. Colombo, Rotterdam, and Abidjan: International Water Management Institute, Global Centre on Adaptation, and the African Development Bank.

Mhazo, A.T. and Maponga, C.C., 2022. The political economy of health financing reforms in Zimbabwe: a scoping review. International journal for equity in health, 21(1), p.42.

Murthy, K.B., Kalsie, A. and Shankar, R., 2021. Digital economy in a global perspective: Is there a digital divide?. Transnational Corporations Review, 13(1), pp.1-15.

Muzunze, N. and Takavarasha, S., 2022, November. Investigating The Challenges of Adopting Data Analytics In Zimbabwe's Retail Sector. In 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT) (pp. 1-6). IEEE.

Ncube, J.N. and Songo, S., 2024. Tele-health Services in the Post-COVID-19 Pandemic Era: The Zimbabwean Context. Zimbabwe Journal of Health Sciences, 4(1), pp.106-119.

Ponde-Mutsvedu, L. and Chirongoma, S., 2022. Tele-evangelism, tele-health and cyberbullying in the wake of the outbreak of COVID-19 in Zimbabwe. In Religion and the COVID-19 pandemic in Southern Africa (pp. 103-114). Routledge.

Rehman, A., Naz, S. and Razzak, I., 2022. Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. Multimedia Systems, 28(4), pp.1339-1371.

Romero-López, M., Pichardo, C., De Hoces, I. and García-Berbén, T., 2021. Problematic internet use among university students and its relationship with social skills. Brain sciences, 11(10), p.1301.

Sahoo, S.R. and Gupta, B.B., 2021. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, p.106983.

Savci, M., Tekin, A. and Elhai, J.D., 2022. Prediction of problematic social media use (PSU) using machine learning approaches. Current Psychology, 41(5), pp.2755-2764.

Sharma, A., Pruthi, M. and Sageena, G., 2022. Adoption of telehealth technologies: an approach to improving healthcare system. Translational medicine communications, 7(1), p.20.

Tete, F., Chikoko, L. and Murendo, C., 2024. Factors influencing the resilience of rural agricultural and retail Micro, Small and Medium Enterprises in Midlands Province of Zimbabwe. Cogent Business & Management, 11(1), p.2404476.

Wang, J., Lim, M.K., Wang, C. and Tseng, M.L., 2021. The evolution of the Internet of Things (IoT) over the past 20 years. Computers & Industrial Engineering, 155, p.107174.

Watambwa, L. and Shilongo, D., 2021. Analysisian analysis of the impact of SME financing on economic growth in Zimbabwe (2015-2019). Available at SSRN 3779488.

Zengeya, T., Sambo, P. And Mabika, N., 2021. The adoption of the Internet of Things for smart agriculture in Zimbabwe. Computer Science & Information Technology, 2021, pp.99-106.

Appendix

Python Code Snippets

```
# Title: Statistical Summary of Internet Usage in Zimbabwe (2020-2025)

import pandas as pd

# Create the dataset
data = {
    #"Year": [2020, 2021, 2022, 2023, 2024, 2025],
    "Total Internet Users (millions)": [4.81, 5.01, 4.65, 5.74, 5.48, 6.45],
    "Annual Increase (thousands)": [413, 203, 265, 117, 115, 120],
    "Annual Growth Rate (%)": [9.4, 4.2, 6.0, 2.1, 2.2, 1.9],
    "Internet Penetration Rate (%)": [33.0, 33.4, 30.6, 34.8, 32.6, 38.4]
}

# Create DataFrame
df = pd.DataFrame(data)

# Display statistical summary
summary = df.describe()
print("Statistical Summary of Internet Usage in Zimbabwe (2020-2025):\n")
print(summary)
```

```
# Title: Statistical Summary of Social Media Usage in Zimbabwe (2020-2025)

import pandas as pd

# Create the dataset
data2 = {
    #"Year": [2020, 2021, 2022, 2023, 2024, 2025],
    "Total Social Media Users (millions)": [0.98, 1.30, 1.55, 1.50, 2.05, 2.10],
    "Annual Change in Social Media Users": [-135000, 320000, 250000, -50000, 550000, 50000],
    "Annual Growth Rate (%)": [-12.0, 33.0, 19.2, -3.2, 36.6, 2.4],
    "Social Media Penetration Rate (%)": [6.6, 8.7, 10.2, 9.1, 12.2, 12.5]
}

# Create DataFrame
df = pd.DataFrame(data2)


# Display statistical summary
summary = df.describe()
print("Social Media Usage in Zimbabwe (2020-2025) with Statistical Summary:\n")
print(summary)
```

```python
#demographics visuals

import matplotlib.pyplot as plt
import numpy as np

# Data
labels = ["2020", "2021", "2022", "2023", "2024", "2025"]
gender_female = [0, 52.3, 52.3, 52.8, 52.7, 52.3]
gender_male = [0, 47.7, 47.7, 47.2, 47.3, 47.7]

urban_population = [0, 32.3, 32.5, 32.5, 32.7, 32.8]
rural_population = [0, 67.7, 67.5, 67.5, 67.3, 67.2]

# Create subplots
fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# Gender Distribution Line Graph
axes[0].plot(labels, gender_female, label='Female', color='salmon', marker='o')
axes[0].plot(labels, gender_male, label='Male', color='royalblue', marker='o')
axes[0].set_ylim(40, 55)
axes[0].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[0].set_ylabel("Percentage", fontname="Times New Roman", fontsize=15)
axes[0].set_title("Figure 1: Gender Distribution", pad = 15, y = -0.3,
          fontname="Times New Roman", fontsize=15)
axes[0].legend()

# Urban vs. Rural Population Line Graph
axes[1].plot(labels, urban_population, label='Urban', color='seagreen', marker='o')
axes[1].plot(labels, rural_population, label='Rural', color='sienna', marker='o')
axes[1].set_ylim(30, 70)
axes[1].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[1].set_ylabel("Percentage", fontname="Times New Roman", fontsize=15)
axes[1].set_title("Figure 2: Urban vs. Rural Population", pad = 15, y = -0.3,
          fontname="Times New Roman", fontsize=15)
axes[1].legend()

plt.tight_layout()
plt.show()
```

```python
import matplotlib.pyplot as plt

# Data
years = ["2020", "2021", "2022", "2023", "2024", "2025"]
internet_users = [4.81, 5.01, 4.65, 5.74, 5.48, 6.45]  # in millions
annual_growth_rate = [9.4, 4.2, 6.0, 2.1, 2.2, 1.9]  # in %

# Create subplots
fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# Plot 1: Total Internet Users
axes[0].plot(years, internet_users, marker="o", linestyle="-", color="tab:blue", label="Internet Users")
axes[0].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[0].set_ylabel("Users (millions)", fontname="Times New Roman", fontsize=15)
axes[0].set_title("Figure 3: Total Internet Users (2020-2025)", pad=15, y=-0.3, fontname="Times New Roman", fontsize=15)
axes[0].grid(True)

# Plot 2: Annual Growth Rate
axes[1].plot(years, annual_growth_rate, marker="s", linestyle="--", color="tab:red", label="Annual Growth Rate")
axes[1].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[1].set_ylabel("Annual Growth Rate (%)", fontname="Times New Roman", fontsize=15)
axes[1].set_title("Figure 4: Annual Growth Rate In Internet Users (2020-2025)", pad=15, y=-0.3,
                  fontname="Times New Roman", fontsize=15)
axes[1].grid(True)

# Adjust layout and show the plots
plt.tight_layout()
plt.savefig("internet_usage_visuals.png", dpi=300, bbox_inches="tight")
plt.show()
```

```python
import matplotlib.pyplot as plt

# Data
years = ["2020", "2021", "2022", "2023", "2024", "2025"]
social_media_users = [0.98, 1.30, 1.55, 1.50, 2.05, 2.10]  # in millions
growth_rate = [-12, 33, 19.2, -3.2, 36.6, 2.4]  # in %

# Create subplots
fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# Plot 1: Social Media Users
axes[0].plot(years, social_media_users, marker="o", linestyle="-", color="tab:blue", label="Social Media Users")
axes[0].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[0].set_ylabel("Users (millions)", fontname="Times New Roman", fontsize=15)
axes[0].set_title("Figure 5: Total Social Media Users (2020-2025)", pad=15, y=-0.3, fontname="Times New Roman", fontsize=15)
axes[0].grid(True)

# Plot 2: Annual Growth Rate
axes[1].plot(years, growth_rate, marker="s", linestyle="--", color="tab:red", label="Growth Rate (%)")
axes[1].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[1].set_ylabel("Annual Growth Rate (%)", fontname="Times New Roman", fontsize=15)
axes[1].set_title("Figure 6: Annual Growth Rate of Social Media Users (2020-2025)", pad=15, y=-0.3,
                  fontname="Times New Roman", fontsize=15)
axes[1].axhline(0, color="black", linestyle="dotted")  # Reference line at 0%
axes[1].grid(True)

# Adjust layout and show the plots
plt.tight_layout()
plt.savefig("social_media_usage_visuals.png", dpi=300, bbox_inches="tight")
plt.show()
```

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
# Data
years = ["2020", "2021", "2022", "2023", "2024", "2025"]
years_int = np.array([int(year) for year in years]).reshape(-1, 1)
internet_users = [4.81, 5.01, 4.65, 5.74, 5.48, 6.45]
annual_growth_rate = [9.4, 4.2, 6.0, 2.1, 2.2, 1.9]

# Train a linear regression model for internet users
model_users = LinearRegression()
model_users.fit(years_int, internet_users)

# Train a linear regression model for growth rate
model_growth = LinearRegression()
model_growth.fit(years_int, annual_growth_rate)

# Get the equation for internet users
slope_users = model_users.coef_[0]
intercept_users = model_users.intercept_
equation_users = f"y = {slope_users:.2f}x + {intercept_users:.2f}"

# Get the equation for growth rate
slope_growth = model_growth.coef_[0]
intercept_growth = model_growth.intercept_
equation_growth = f"y = {slope_growth:.2f}x + {intercept_growth:.2f}"

# Forecast the next 5 years
future_years = np.array([2026, 2027, 2028, 2029, 2030]).reshape(-1, 1)
forecast_users = model_users.predict(future_years)
forecast_growth = model_growth.predict(future_years)

# Plot the original data and the forecast
fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# Plot 1: Total Internet Users with forecast
axes[0].plot(years_int, internet_users, marker="o", linestyle="-", color="tab:blue", label="Internet Users")
axes[0].plot(future_years, forecast_users, marker="o", linestyle="--", color="tab:orange", label="Forecasted Users")
axes[0].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[0].set_ylabel("Users (millions)", fontname="Times New Roman", fontsize=15)
axes[0].set_title("Figure 9: Total Internet Users (2020-2030)", pad=15, y=-0.3, fontname="Times New Roman", fontsize=15)
axes[0].grid(True)
axes[0].legend()
axes[0].text(0.45, 0.90, f"Eq: {equation_users}", transform=axes[0].transAxes, fontsize=12, verticalalignment="top")

# Plot 2: Annual Growth Rate with forecast
axes[1].plot(years_int, annual_growth_rate, marker="s", linestyle="--", color="tab:red", label="Growth Rate (%)")
axes[1].plot(future_years, forecast_growth, marker="s", linestyle="--", color="tab:green", label="Forecasted Growth Rate")
axes[1].set_xlabel("Year", fontname="Times New Roman", fontsize=15)
axes[1].set_ylabel("Annual Growth Rate (%)", fontname="Times New Roman", fontsize=15)
axes[1].set_title("Figure 10: Annual Growth Rate in Internet Users (2020-2030)", pad=15, y=-0.3,
                  fontname="Times New Roman", fontsize=15)
axes[1].axhline(0, color="black", linestyle="dotted")  # Reference line at 0%
axes[1].grid(True)
axes[1].legend()
axes[1].text(0.40, 0.15, f"Eq: {equation_growth}", transform=axes[1].transAxes, fontsize=12, verticalalignment="top")

# Adjust layout and show the plots
plt.tight_layout()
plt.savefig("internet_usage_forecast_with_equation.png", dpi=300, bbox_inches="tight")
plt.show()

# Print the equations
print("Equation for Internet Users:")
print(equation_users)
print("\nEquation for Growth Rate:")
print(equation_growth)
```

GitHub Link

https://github.com/faithkabanda/2025-SIDS5204-Big-Data-Science-Project.git