FORECASTING EMERGING SKILL DEMANDS WITH MACHINE LEARNING TO INFORM CURRICULUM DEVELOPMENT IN UGANDA'S HIGHER EDUCATION

DENIS WANYAMA

J23MD19/216

A DISSERTATION SUBMITTED TO THE FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS OF THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS OF UGANDA CHRISTIAN UNIVERSITY

September,2025

UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

Abstract

Rapid technological advancement and evolving industry demands have widened the skill gap in Uganda's labor market. Higher education institutions often struggle to keep pace with these changes, leading to mismatches between graduate competencies and employer expectations. This study uses machine learning techniques to forecast emerging skill demands and inform the development of data-driven curriculum in Ugandan universities.

Drawing on more than one million job postings from 2021 to 2023, the research applies natural language processing (NLP), time series forecasting (ARIMA and Holt-Winters), and clustering algorithms to analyze labor market trends. Exploratory Data Analysis (EDA) revealed high-demand skills, while Holt-Winters outperformed ARIMA (MAE: 9.05 vs. 23.87), capturing the seasonal nature of skill fluctuations.

Key findings indicate a growing demand for roles such as interaction designers, network administrators, user experience professionals, and social media managers. In-demand technical skills include Python, Google Analytics, CSS, Tableau, AWS, and Sketch. The increasing emphasis on digital literacy and soft skills underscores the need for more flexible and adaptive curricula.

This study offers actionable recommendations for curriculum reform, including integrating technical skills, developing continuous learning pathways, and enhancing academic-industry collaboration. By applying machine learning to labor market analysis, the research equips universities, policymakers, and stakeholders with the information needed to align higher education with the demands of Uganda's evolving digital economy.

Keywords: Machine Learning, Labor Market Trends, Skill Forecasting, Higher Education Curriculum, Uganda, ARIMA, Holt-Winters, Data Science, skill mismatch.

Declaration

I, WANYAMA DENIS (B00142), solemnly declare that the research presented in this study, 'Forecasting Emerging Skill Demands with Machine Learning to Inform Curriculum Development in Uganda's Higher Education,' is my original work. All sources of information and data used in this research have been acknowledged and referenced in good faith. I have appropriately credited any assistance received during this study and confirm that I will not submit this research for any other degree or qualification at any other institution. The findings, conclusions, and recommendations presented in this study are based on analyzing and interpreting the data collected and reflect my understanding. I take full responsibility for the accuracy and integrity of this research. Furthermore, I declare that I will adhere to ethical issues throughout the process according to the principles of academic integrity and ethical research conduct.

Name: WANYAMA DENIS

Date: 30th April 2025

Signature...

Approval

This research report, 'Forecasting Emerging Skill Demands with Machine Learning to Inform Curriculum Development in Uganda's Higher Education,' has been reviewed and approved by:

Name: John B. Wabwire Habere

Date: September 25, 2025

Signature: Habere

Dedication

I dedicate this work to my family for their unwavering support, encouragement, and sacrifices that have fueled my academic journey. My parents, for their unconditional love and guidance, have helped me persevere, and my mentors, whose wisdom has inspired me to excel in research. Lastly, to all aspiring data scientists, educators, and policy makers working to bridge the gap between educational needs and industry demands, we must shape a future where higher education aligns with labor market requirements.

Acknowledgments

I extend my deepest gratitude to all those who contributed to this study, particularly Uganda Christian University and the Department of Computing and Technology, which provided the academic environment and resources.

I am particularly grateful to my supervisor, Mr. John B. Wabwire Habere, for his invaluable guidance, feedback, and unwavering support. I also acknowledge the mentors, lecturers, and colleagues who offered encouragement and technical assistance.

I appreciate the researchers and professionals whose work inspired this study. Their collective contributions made this research possible.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
ARIMA	AutoRegressive Integrated Moving Average
CBA	Competency-Based Approach
CSS	Cascading Style Sheets
CV	Curriculum Vitae
EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
GIS	Geographic Information System
GoU	Government of Uganda
HEIs	Higher Education Institutions
ICT	Information and Communication Technology
IoT	Internet of Things
IT	Information Technology
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MOES	Ministry of Education and Sports
MOGLSD	Ministry of Gender, Labour and Social Development
NCHE	National Council for Higher Education
NLP	Natural Language Processing
NPDIV	Fourth National Development Plan
OLS	Ordinary Least Squares
RMSE	Root Mean Square Error
SDGs	Sustainable Development Goals
TVET	Technical and Vocational Education and Training
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
WEF	World Economic Forum
NPDIII	Third National Development Plan

Glossary of Terms

Analytical Skills The ability to evaluate information, identify patterns, and solve complex problems using logical reasoning and evidence.

ARIMA Model A forecasting model combining Auto-Regressive (AR), Integrated (I), and Moving Average (MA) components to predict time series data.

Clustering A machine learning method used to group data into clusters based on similarity without prior labeling.

Curriculum Reform The process of updating academic content and delivery methods to align with the current labor market and industry demands.

Data-Driven An approach that bases decisions and insights on data analysis and empirical evidence rather than intuition.

Digital Skills Competencies needed to use digital tools and platforms, including programming, data visualization, and online collaboration.

Exploratory Data Analysis (EDA) A technique to summarize the main characteristics of data using visual and statistical methods before formal modeling.

Forecasting Predicting future trends or values based on historical data patterns.

Higher Education Institution (HEI) An accredited institution that provides tertiary-level education, such as a university or college.

Holt-Winters Model A time series forecasting method that accounts for trends, seasonality, and level in the data.

Job Market Trends Observable shifts in employment, job roles, and skills demand over time.

K-Means Clustering An unsupervised algorithm that partitions data into K distinct, non-overlapping clusters.

Labor Market The economic space where workers offer their services and employers demand labor.

Lifelong Learning The continuous, voluntary, and self-motivated pursuit of knowledge and skills throughout an individual's life.

Machine Learning A subfield of artificial intelligence focused on building algorithms that improve automatically through data exposure.

Natural Language Processing (NLP) A branch of AI that enables machines to understand, interpret, and generate human language.

Positivism A philosophical stance asserting that only empirical, observable, and quantifiable data leads to valid knowledge.

Skill Demand The level of need for specific skills in the labor market based on employer requirements.

Time Series Analysis A statistical method used to analyze sequences of data points over time to identify trends and patterns.

Workforce Planning A strategic process to ensure that the right number of people with the right skills are in the right place at the right time.

Chapter 1

General Introduction

1.1 Introduction

Rapid technological advancements and the proliferation of digital devices and services are shaping human development worldwide [1]. In Africa, for instance, an estimated 230 million jobs in Sub-Saharan Africa are expected to require digital skills by the end of the decade. Frontier technologies, including artificial intelligence, robotics, biotechnology and nanotechnology, accelerate this transformation, as demonstrated by the rapid development of coronavirus vaccines in 2020 [2]. However, significant downsides may emerge if these advances outpace society's ability to adapt. Automation and evolving skill requirements risk displacing jobs and exacerbating existing inequalities.

The stakes are even higher for developing countries such as Uganda. Without timely and practical training, the labor force can struggle to harness frontier technologies, hindering efforts to balance innovation with equity and achieve the Sustainable Development Goals. Higher education institutions must therefore play a pivotal role in designing up-to-date and relevant curricula. Mandated agencies, such as the National Council for Higher Education (NCHE) and the National Curriculum Development Center (NCDC), require practical forecasting tools to analyze labor market trends and emerging skill demands [3]. This proactive approach will enable universities to consistently align their programs with industry requirements, thereby enhancing the employability and global competitiveness of their graduates. Advancements in data science offer promising opportunities in this regard. Leveraging large datasets through machine learning techniques, such as Natural Language Processing (NLP), can enhance the accuracy of forecasting models and provide actionable insights into emerging skills and trends [4].

Despite these opportunities, Uganda's Big Data ecosystem still faces challenges, including a shortage of skilled data analytics personnel. This research proposal aims to investigate a data-driven approach to collecting and analyzing trends, forecasting emerging skill demands, and recommending the skills most sought after by employers [1]. In doing so, the study intends to contribute to the development of the curriculum in Ugandan higher education, thereby improving the relevance and quality of academic programs.

Subsequent chapters will cover the background, literature review, methodology, theoretical framework, data collection and analysis techniques, and ethical considerations to ensure the rigor and validity of the research findings.

1.2 Background of study

Higher education institutions in Uganda are increasingly challenged by the rapidly evolving demands of the digital age [5]. For example, the shift to Open Distance e-Learning during the COVID-19 pandemic exposed significant skills gaps between staff, which limited the effective adoption of new teaching methods [6]. Meanwhile, global trends driven by the digital revolution and rapid technological advancements have raised expectations for graduates to possess a broader skill set [7], resulting in mismatches between labor supply and demand, both globally and geographically, as well as industry-specific risks to the supply and demand of skills [8]. However, in Uganda, there is a substantial mismatch between the skills of graduates and those demanded by industry, contributing to unfilled vacancies and rising unemployment rates [9], [10].

Africa faces a critical shortage of skilled workers, further complicating efforts to recruit and retain qualified faculty for in-demand programs like data science and analytics [11] [12]. Moreover, unreliable data on emerging skills hampers timely curriculum updates [13]. Globally, CEOs have consistently ranked the skills gap as one of their top challenges, with research indicating that mismatches between graduate skills and job market requirements affect up to 40% of employers in the United States [14] [15].

To tackle these issues, higher education institutions in Uganda must adopt a data-driven approach to forecast emerging skill demands and analyze job market trends. Institutions can efficiently collect and analyze data from job postings and industry reports by leveraging machine learning techniques such as natural language processing (NLP). [16]. This strategy would enable continuous curriculum refinement to align educational offerings with industry's better needs, enhancing graduate employability and contributing to Uganda's broader economic development [17].

1.3 Problem statement

Uganda is undergoing a rapid digital transformation, as emphasized in the Third National Development Plan (NDP III) [18] and the Uganda Digital Transformation Roadmap [19]. However, higher education institutions have struggled to align their curricula with the evolving demands of the labor market [20]. This misalignment has contributed to a widening skills gap, with recent reports indicating that 83% of the Ugandan workforce lacks essential digital competencies, and graduate unemployment rates exceed national averages by up to 12% [9].

Traditional curriculum development practices, often based on expert judgment, static labor market reports, and outdated forecasting tools, are insufficient for capturing the dynamic and rapidly evolving nature of labor market needs [21]. As a result, many graduates enter the workforce with outdated skill sets, particularly in high-growth sectors such as information and communication technology (ICT), finance, manufacturing, and healthcare technology [22].

To bridge this gap, a data-driven approach that integrates machine learning and natural language processing (NLP) is needed to extract real-time insights from labor market data such as online job postings, industry reports, and alumni's employment outcomes. Techniques such as time series forecasting, skill clustering, and named entity recognition can help identify emerging skills and provide universities with actionable insights. This proactive approach supports evidence-based curriculum reform, enhances graduate employability, and contributes to Uganda's broader inclusive digital economic development goals [20].

1.4 Purpose of the Study

The Fourth Industrial Revolution has ushered in various emerging technologies [23]. Notably, artificial intelligence is transforming the global employment landscape. In Uganda, higher education institutions face the challenge of preparing graduates for jobs that may not yet exist, with skill demands continually evolving. Traditional methods of analyzing job market trends, such as online job postings and expert interviews, often lack comprehensiveness and foresight [24].

This study aims to develop data-driven methodologies for identifying emerging skill demands, informing curriculum development, and aligning education with future workforce needs.

1.4.1 Aim

This study aims to explore methods for analyzing trends and forecasting future skill demands using machine learning techniques and provide recommendations for circular development in Uganda.

1.4.2 Objectives

1. Collect and analyze labor market data to identify emerging skills and high-demand trends.
2. Apply and evaluate machine learning algorithms to accurately forecast future skills demands and the corresponding educational qualifications required by employers.
3. Generate actionable insights and recommendations that guide curriculum development, aligning educational programs with Uganda's future labor market demands.

1.4.3 Research Question

How can machine learning techniques be effectively applied to identify trends and forecast emerging skills demands across various industries to guide the development of higher education curricula in Uganda?

1.4.4 Hypothesis

Machine learning techniques can effectively analyze labor market trends and accurately forecast future skill demands, facilitating data-driven recommendations for aligning higher education curricula with industry needs and supporting circular development in Uganda.

1.5 Justification of the research

Technological advancements have significantly impacted employment patterns, raising concerns about job displacement and inequality, particularly in developing countries like Uganda. This study aims to bridge the gap between the skills offered by higher education and those demanded by industries by leveraging data-driven insights. Aligning educational offerings with market needs is vital for economic development, social progress, and achieving sustainable development goals.

1.6 Significance of the study

Addressing the skills gap between employer expectations and higher education training is vital for Uganda's socio-economic development. By employing machine learning techniques to forecast in-demand skills, this study aims to help institutions adapt curricula, preparing students for future careers. Collaboration with agencies like the National Council for Higher Education (NCHE) and the National Curriculum Development Centre (NCDC) ensures that the findings contribute to quality education, decent work, and infrastructure development, aligning with SDG 4 (Quality Education), SDG 8 (Decent Work and Economic Growth), and SDG 9 (Industry, Innovation, and Infrastructure) [25].

1.7 Scope and Limitations

1.7.1 Scope

The study analyzes job postings in Uganda from 2021 onwards to predict the data skills most employers demand, aiming to inform curriculum development in higher education.

1.7.2 Limitations

- Incomplete or inconsistent job postings may affect prediction accuracy.
- The reliability of Natural Language Processing (NLP) techniques may influence results.
- Due to differing labor market dynamics, the findings may not be applicable beyond Uganda.

1.7.3 Conceptual Framework

The study integrates data science methodologies with the development of higher education curricula. Data science facilitates the collection and analysis of labor market data to identify emerging skill demands. Machine learning techniques enable systematic analysis of large datasets, providing insights that inform curriculum design. This data-driven approach ensures that educational programs align with evolving workforce needs, equipping graduates with relevant skills for the future.

Chapter 2

Literature Review

2.1 Theoretical Literature Review

In today's rapidly evolving job market, the demand for skills constantly shifts due to technological advancements, globalization, and changes in industry landscapes. Higher education institutions must ensure their curricula remain relevant and aligned with emerging skill demands to prepare students for the future workforce [26]. This literature review examined the trends, challenges, and opportunities in forecasting emerging skill demands and leveraging data-driven approaches, particularly machine learning techniques that predict future skills, informing curriculum development in Uganda's higher education [5].

2.1.1 Related Literature

Several studies have explored the relationship between emerging skills demands and higher education curriculum development.
The National Labor Force Survey 2021, conducted by the Uganda Bureau of Statistics (UBOS), highlighted a significant mismatch between the skills of the employed population and the job requirements [9]. Specifically, 43% of employed individuals required additional education for their current positions, indicating a gap between the educational qualifications of workers and the skill level demanded by employers. Additionally, despite most of the employed population reporting job satisfaction, only 27% were entitled to workplace benefits, suggesting a mismatch between the compensation and benefits received by workers and their expectations or needs. Moreover, the survey revealed that less than 10% of the employed population held secondary jobs, indicating a potential mismatch between the availability of job opportunities and individuals' desire or need for additional income sources. These findings underscore the need for targeted interventions to address skill mismatches, improve employment conditions, and promote inclusive economic growth in Uganda.

One effective strategy proposed is the adoption of data-driven approaches in curriculum development. This approach enables institutions to accurately identify emerging skill requirements and analyze job market trends. For example, the "Digital Work Digital Sovereignty at the Workplace" study examined university curricula in various countries, including Germany and Australia, highlighting the need for comprehensive updates to integrate data science and analytics skills demanded by modern industries [17].

Similar skills demand emerged when Natural Language Processing (NLP) was used in a study to bridge Namibia's Data Science Skills Gap [27]. Machine learning techniques can analyze trends and identify emerging skill demands. In the study "Predicting Students' Employability using Support Vector Machine" [28], Machine learning techniques were employed to analyze job postings and identify emerging skill demands in cybersecurity. The study found that machine learning techniques can provide accurate and up-to-date information on skills demands, which higher education institutions can use to revise their curricula [29].

Curriculum development in higher education is a complex process that involves various stakeholders, including educators, employers, and policymakers. According to the study titled "To Close the Skills Gap, Technology and Higher-Order Thinking Skills Must Go Hand in Hand" [30], the primary purpose of higher education is to prepare students for the workforce. This curriculum development should be aligned with the needs of the job market. For example, in recent years, there has been a growing demand for technology skills in many industries, including healthcare, finance, and Agriculture. This has led higher education institutions to revise their curricula, ensuring that students possess the skills necessary to succeed in the evolving job market.

2.1.2 Comparative Evaluation of Reviewed Research

The following table presents a comparative evaluation of the research studies reviewed, highlighting their methodologies, focus areas, and relevance to curriculum development:

Table 2.1: Comparative Evaluation of Reviewed Research

Research Study	Data Sources	Analytical Methods	Focus Area	Relevance to Curriculum Development
Generating Synthetic Data for Better Prediction Modeling in Skill Demand Forecasting [31]	Job advertisement sites	Web scraping, Natural Language Processing	Skill demand trends across various professional fields	Provides insights into current skill demands to inform curriculum updates
Importance of Skills Development for Ensuring Graduate [32]	Survey data	Statistical analysis	Alignment of graduate skills with employer expectations	Highlights the need for curriculum adjustments to meet industry requirements
Digital Technologies in Africa by the International Bank for Reconstruction and Development [33]	Survey data	Descriptive analysis	Digital skills gap in Africa	Emphasizes the necessity for integrating digital skills into curricula
Mapping the MIS Curriculum Based on Critical Knowledge and Skills [34]	Curriculum documents	Content analysis	Alignment of the MIS curriculum with industry needs	Guides curriculum revisions to include critical knowledge and skills
Utilizing Web Scraping and NLP to Better Inform Pedagogical Practice [35]	Job postings	Web scraping, NLP	Identification of in-demand skills in Kenya	Suggests curriculum enhancements based on identified skill demands
National Labor Force Survey 2021 [9]	Survey data	Statistical analysis	Skills mismatch in Uganda's labor force	Provides empirical evidence to drive curriculum reforms

Research Study	Data Sources	Analytical Methods	Focus Area	Relevance to Curriculum Development
Proposed Research	Job postings	Machine learning algorithms	Emerging skill trends in Uganda	Aims to align the curriculum with future skill demands through predictive modeling

2.1.3 Discussion

The reviewed literature consistently emphasizes the need for higher education institutions to adapt their curricula to the evolving nature of skill demands. Data-driven approaches, particularly machine learning and Natural Language Processing (NLP), have emerged as practical tools in identifying and forecasting these demands [29]. For instance, studies that utilize web scraping and NLP to analyze job postings have successfully identified the critical skills sought by employers, enabling institutions to tailor their programs accordingly. Moreover, machine learning models have been employed to predict future skill shortages, providing foresight to inform proactive curriculum development [36]. These methodologies enable the development of curricula that are responsive to current industry needs and anticipate emerging trends.

In the context of Uganda, empirical evidence points to a significant skills gap, particularly in digital and technological domains. The "Digital Technologies in Africa" report by the World Bank indicates that a substantial portion of the population lacks the necessary digital skills, hindering effective engagement with digital platforms and services [33]. This gap presents both a challenge and an opportunity for higher education institutions to revamp their curricula to include comprehensive digital literacy and advanced technological competencies. Integrating emerging technologies, such as Artificial Intelligence (AI), the Internet of Things (IoT), Robotics, Machine Learning (ML), and Deep Learning (DL), into academic programs is crucial for preparing students for the evolving job market [37].

2.1.4 Conclusion

Literature highlights the crucial role of data-driven approaches in aligning higher education curricula with the rapidly evolving demands of the job market. Machine learning and NLP techniques offer valuable tools for forecasting skill requirements, enabling institutions to adjust their programs proactively [38]. Addressing the digital skills gap through curriculum reform is essential in Uganda to meet the demands of a modern economy. Subsequent sections of this thesis will build upon these insights, proposing methodologies and strategies for implementing machine learning-driven curriculum development to bridge the skills gap and enhance graduate employability.

Chapter 3

Research Methodology

3.1 Research Design

This study employs a mixed-methods explanatory research design that combines quantitative and qualitative approaches within a data-driven framework. The positivist methodology emphasizes objectivity, empirical data analysis, and predictive modeling.

Quantitatively, the study utilizes machine learning algorithms (e.g., ARIMA, Holt-Winters, K-Means, Random Forest) on historical job posting data to forecast emerging skill demands. It uses Natural Language Processing (NLP) to extract structured information from unstructured job descriptions, including skills, qualifications, and job roles. These techniques support trend identification and forecast future skill requirements across various industries.

The study qualitatively interprets the results to understand underlying trends and sector-specific implications and to draw policy- and curriculum-related recommendations. This interpretive layer ensures the insights are data-driven and contextually relevant to Uganda's higher education system.

The data is sourced from Kaggle (`https://www.kaggle.com/datasets/ravindrasinghrana/job-description-dataset/data`), which aggregates job postings from global platforms. Combining data mining, text analytics, and time-series forecasting provides a comprehensive understanding of labor market dynamics.

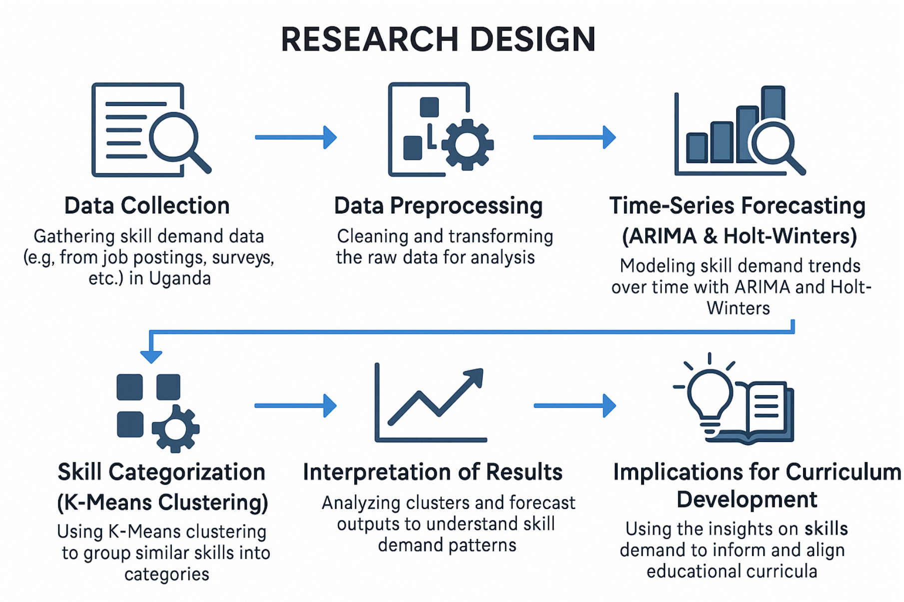

Figure 3.1: Research Design diagram

This diagram illustrates the overall research design employed in the study. It outlines the sequential stages, from data acquisition (using the global job postings dataset) to filtering for Uganda-specific records, followed by extracting Natural Language Processing (NLP) skills. The design incorporates Exploratory Data Analysis (EDA), K-Means clustering, and time series forecasting using ARIMA and Holt-Winters models. The process concludes with interpreting and translating findings into insights for curriculum development. The framework reflects a positivist, data-driven approach centered on structured empirical analysis and predictive modeling.

3.1.1 Philosophical Consideration

This study is grounded in the positivist research philosophy, which emphasizes the use of empirical, observable, and measurable evidence to understand real-world phenomena. Originating from the work of Auguste Comte, positivism posits that knowledge is derived from scientific inquiry, rather than from subjective belief or speculation. It upholds that valid knowledge is obtained through observation, experimentation, and quantitative reasoning.

Positivism is applied through the systematic analysis of structured job market data, including measurable variables such as skill frequency, job type, industry sector, and time of posting. By employing machine learning algorithms and statistical models, this research relies on objective, verifiable data to identify skill trends and forecast future demands [39].

Illustration: Positivism in Action A mathematical expression of forecasting based on time series modeling to help conceptualize this philosophy, consider the following:

Illustration: Positivism in Action

A mathematical expression of forecasting based on time series modeling to help conceptualize this philosophy, consider the following:

$$Y_t = \alpha + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t \tag{3.1}$$

Where:

- Y_t = Predicted demand for a skill at time t
- $X_{1,t}, X_{2,t}$ = Observed job market indicators (e.g., number of job postings, frequency of skills)
- α, β_1, β_2 = Coefficients estimated through regression
- ε_t = Error term (assumed to be random)

This formula illustrates how positivist research operationalizes variables into mathematical models to make predictions, a key aspect of the machine learning approach.

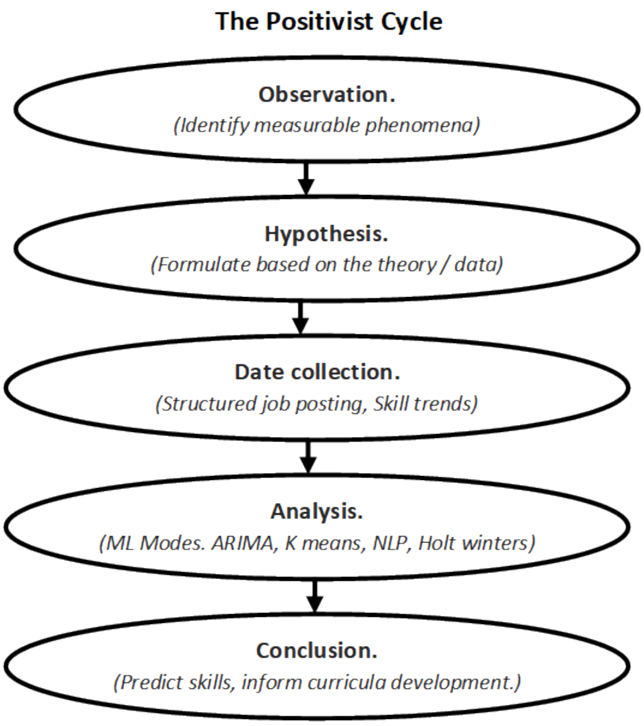

Figure 3.2: The Positivist Research Cycle

This diagram illustrates the positivist research approach employed in this study. The cycle begins with empirical observations of job market trends. It progresses through structured data collection, variable operationalization, and analytical processing, including Natural Language Processing (NLP) and statistical modeling (e.g., ARIMA, Holt-Winters, K-Means). This structured methodology tests hypotheses and transforms them into actionable insights that inform curriculum design and policy recommendations. The cycle reflects core tenets of positivism: reliance on observable, quantifiable evidence; use of scientific models; and the derivation of objective, reproducible conclusions from real-world data.

Alignment with Research Methodology

The positivist research philosophy closely aligns with the study's data-driven methodology, prioritizing objective, measurable, and quantifiable data. This alignment is evident in the systematic collection and analysis of labor market information through techniques such as web scraping, exploratory data analysis (EDA), and the application of machine learning algorithms.

By leveraging structured datasets such as job postings, skill frequency metrics, and industry trends, the study aligns with positivism's emphasis on empirical observation and scientific inquiry. Machine learning models, such as ARIMA, Holt-Winters, and K-Means, identify patterns and forecast future skill demands. These models rely on mathematical formulations, ensuring that results are replicable, verifiable, and unbiased.

Furthermore, using Natural Language Processing (NLP) to extract skill sets from unstructured job descriptions reflects a positivist commitment to transforming raw observations into structured, analyzable data. This reinforces the objectivity and validity of the research outcomes.

Overall, the positivist approach enhances the study's methodological rigor, ensuring that its conclusions, particularly those regarding curriculum development and skills forecasting, are based on evidence-based, reproducible findings suitable for informing policy and educational reform.

Justification for the Positivist Approach

The decision to adopt a positivist stance is based on its ability to provide objective, quantifiable insights into Uganda's evolving labor market. This approach facilitates the formulation of generalizable conclusions that inform curriculum development, ensuring alignment between higher education programs and industry needs [40].

3.2 Area of study

The study is conducted at the intersection of data science, higher education, and labor market analysis. It analyzes job postings to forecast skill demands and align workforce needs with educational programs. The study integrates multiple disciplines, including:

- Data Science provides methods for trend analysis and skill forecasting.
- Education Policy, which supports curriculum reform based on labor market needs.
- Labor Market Analytics addresses skill mismatches and employment trends in Uganda's economy.

3.3 Sources of Information

This study utilizes secondary sources to ensure a comprehensive and contextually grounded investigation. These sources were meticulously selected to provide a strong foundation for addressing the research questions and achieving the study's objectives. They encompass aggregated job posting platforms, academic literature, government publications, global labor market reports, and ethical and methodological guidelines for data-driven research. This diverse input enriches the analysis with both empirical data and scholarly insight, presenting significant potential for evidence-based recommendations and curriculum development:

Table 3.1: A summary of the primary sources used in this study highlights the diversity and relevance of each category in informing labor market analysis and curriculum development.

Source type	**Description**	**Example**
Job posting websites	Platforms that advertise vacancies and provide real-time insights into the skills, qualifications, and experience employers demand.	Online job advertising portals and other advertising websites
Labor market reports	Published studies and national or international reports on employment trends, workforce gaps, and sectoral skill demand.	Uganda Labour Market Reports, ILO Studies, and Research Institution Briefs
Academic literature	Peer-reviewed articles, conference proceedings, and books on curriculum development, digital skills, and higher education reforms.	Journal of Higher Education, International Conference, Educational Research Review

Table continues on next page

Source type	Description	Example
Data science tools and techniques	The modeling framework was informed by methodological guides and technical resources on machine learning, natural language processing (NLP), and data analysis.	Python libraries (scikit-learn, NLTK), textbooks, Coursera/edX courses
Government reports	Official strategic documents, national policies, and guidelines relevant to education, digital transformation, and labor development.	NCDC Curriculum Guidelines, NDP III, Digital Transformation Roadmap, UBOS, NDPs
International comparisons	Cross-national reports benchmark Uganda's skills ecosystem and higher education against global trends.	OECD, World Bank studies, Civil Society Organizations, and United Nations reports.
Ethical Guidelines	Information on ethical considerations and guidelines for research and website ethical guidelines in terms and conditions of the websites.	Institutional Ethics Policies, Terms of Use of Websites, Research Protocols

3.4 Population and Sampling Techniques

This study utilizes a non-probability purposive sampling technique, which is well-suited for research that relies on large-scale, heterogeneous secondary datasets. The target population comprises all job postings available in the dataset obtained from Kaggle, which includes 1,048,575 job entries collected between 2021 and 2023 from global job portals. The researcher applied filtering criteria to extract job postings relevant to Uganda, ensuring they were aligned with the Ugandan labor market context. These criteria included:

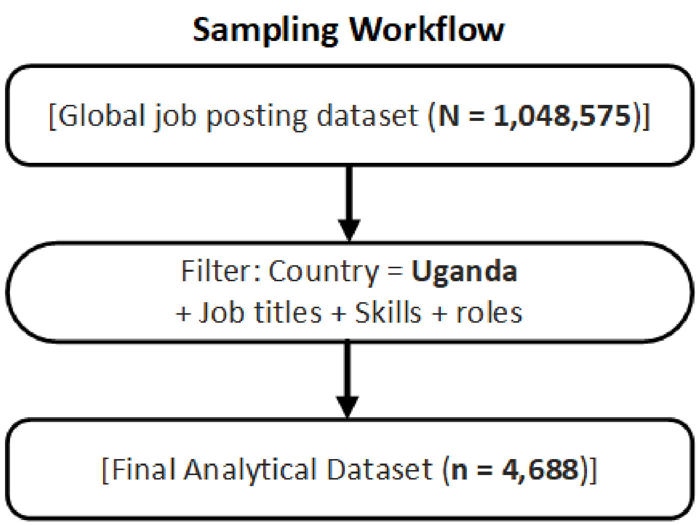

Figure 3.3: Sampling workflow illustrating the purposive selection of job postings based on relevance to Uganda's labor market.

After applying these conditions, a final sample of 4,688 job postings from Uganda was derived. This subset serves as the analytical dataset for exploring emerging skill demands and informing curriculum recommendations.

3.5 Variable definitions and measurements

This study focused on capturing the key aspects of skills demands and trends, as recommended in previous chapters. The analyzed variables represent the overarching categories of information, while the indicators are specific metrics or characteristics used to measure or assess each variable.

Table 3.2: Job Market Variables and Their Indicators

Variable	**Indicator**
Job Title	Frequency of specific job titles
Organization	Types of organizations hiring
Duty Station	Geographic locations of job postings
Reports to	Level of the position within the organizational hierarchy (e.g., reports to Senior Data Scientist or Head of Analytics)
About the Organization	Description of the hiring organization's mission, values, and industry focus
Skills	Frequency of specific skills (e.g., Python, R, SQL), diversity of demanded skills, and emerging trends over time
Educational Requirements	Required education level (e.g., bachelor's, master's), field of study, and certifications
Job Experience	Required experience level and specific job roles
Key Duties and Responsibilities	Main tasks and responsibilities for the job role
Qualification, Skills, & Experience	Academic requirements, technical skills, and professional experience
Job Location	Job's geographic location and possibility for remote work
Date Posted	Date the job posting was published
Job Summary	Overview of the job's purpose and expectations
Competencies	Desired soft skills and personal attributes
Employment Type	Type of contract (e.g., full-time, part-time)
Deadline	Expiration date of the job posting

These variables are analyzed using quantitative techniques, including:

I. **Skill Demand**: Measured by the frequency with which a skill appears in job postings.

II. **Industry Trends**: Assessed by distributing job postings across sectors.

III. **Geographical Demand**: Captures the regional distribution of job postings.

IV. **Educational Requirements**: Examines the qualifications required for job roles, including bachelor's degrees, master's degrees, and certifications.

V. **Emerging Skills**: Measured through year-over-year skill growth rates to identify high-demand competencies.

3.6 Data Collection

Secondary data were sourced from a Kaggle-hosted dataset that aggregates job postings from LinkedIn, Indeed, and Everjobs. The metadata accompanying the Kaggle dataset provided information about the period (2021-2023), geographical coverage, and occupational sectors.

3.6.1 Supplementary Data Sources

In addition to the Kaggle dataset, the study incorporates insights from Industry Reports, which validate quantitative findings and identify sector-specific demands. Government Labor Market Assessments include reports from the Ministry of Gender, Labor, and Social Development and the Uganda Bureau of Statistics, ensuring alignment with national workforce development priorities and policy frameworks.

3.6.2 Data Preprocessing

A comprehensive data preprocessing phase ensured the dataset was suitable for machine learning analysis and trend forecasting. One of the critical steps involved treating missing values, which were common across several fields, including job titles, skills, qualifications, and posting dates. Missing job titles, for instance, were considered vital for clustering tasks. Therefore, entries with null or blank values in this field were excluded from the dataset to preserve the analytical integrity.

In addition to addressing missing values, other preprocessing steps were applied to standardize the dataset's structure and enhance its interpretability. Column names were uniformly reformatted by converting all characters to lowercase and replacing spaces with underscores to ensure compatibility with programming libraries such as Pandas and scikit-learn. Skills, which were often listed as comma-separated strings, were split into individual elements to facilitate analysis of skill frequency, clustering, and forecasting. The job posting date field was converted into a consistent datetime format for time series modeling, allowing for proper indexing, resampling, and trend analysis over time.

Finally, feature engineering was performed to create additional variables that enriched the analysis. This included generating job clusters using K-Means algorithms based on extracted keywords from job descriptions and categorizing postings by roles, enabling skill and role analyses. These preprocessing techniques collectively ensured that the dataset was clean, consistent, and analytically robust, forming a solid foundation for the machine learning models used in the study.

3.7 Exploratory Data Analysis (EDA) and Skill Identification

Exploratory data analysis was conducted using descriptive statistics and visualizations to uncover patterns and trends. Word frequency analysis revealed the most common skills across job postings, while TF-IDF vectorization highlighted the most distinctive skills in specific sectors. Time-series graphs and bar plots were created using Matplotlib and Seaborn to illustrate trends in demand.

3.7.1 Overview of Tools and Libraries Used

The analysis was implemented in Python. Standard libraries such as re, string, and collections were used for text manipulation. Pandas and NumPy handled data cleaning, aggregation, and statistical analysis. Nltk was used for stop word removal and tokenization, while TfidfVectorizer quantified the relevance of skills across job descriptions.

For clustering, KMeans from sklearn. A cluster analysis was performed, and the optimal number of clusters was determined using the silhouette score. For forecasting, the study applied ARIMA from stats models and Holt-Winters Exponential Smoothing, with auto arima from pmdarima optimizing model parameters. Visualization libraries, matplotlib, pyplot, seaborn, and WordCloud, were employed to create informative graphics.

3.7.2 Skill Trend Identification and Visualization

Bar charts highlighted the top 10 skills, while WordCloud offered a visual overview of high-frequency terms. Time-series plots illustrated the growing demand for digital skills, such as AI and cloud computing. Cluster analysis categorized skills into occupational groups, including cybersecurity, marketing, software development, and project management.

3.8 Machine Learning Models for Forecasting Skill Demand

This section explores the machine learning approaches to forecasting skill demands in Uganda's labor market. Machine learning (ML) is a class of algorithms that allow systems to learn patterns from data and make predictions or decisions without being explicitly programmed [41][41]. Machine learning (ML) is generally classified into two categories: supervised learning, where the model is trained on labeled data (e.g., ARIMA for time series forecasting), and unsupervised learning, which detects patterns in unlabeled data (e.g., K-Means for clustering).

This study employed supervised learning, specifically ARIMA (Autoregressive Integrated Moving Average) and Holt-Winters Exponential Smoothing, to predict future skill demands based on historical trends. Unsupervised learning was implemented using the K-Means clustering algorithm to identify patterns and group similar skills across job postings.

3.8.1 ARIMA (Autoregressive Integrated Moving Average) Modeling Approach

The Auto Regressive Integrated Moving Average (ARIMA) model is a widely used time series forecasting technique that combines three key components to model and predict future values based on historical trends:

- **Auto Regression (AR)**: The model uses the dependency between an observation and several lagged observations. For example, the number of job postings in each month may be influenced by the number of postings in previous months.
- **Integration (I)**: This step involves differencing the data to achieve stationarity, i.e., making the data's mean and variance constant over time. It helps in removing trends or seasonality that can distort the forecasting process.
- **Moving Average (MA)**: The model captures the relationship between an observation and the residual errors from a moving average model applied to lagged observations.

The general ARIMA model is denoted as ARIMA (p, d, q), where:

- p is the number of lag observations in the model (AR term),
- d is the degree of differencing (to remove trend and make the data stationary),
- q is the size of the moving average window (MA term).

Application in This Study

This study employed ARIMA to analyze monthly job posting data for Uganda from 2021 to 2023, aiming to forecast future demand for skills. The following steps were undertaken:

The ARIMA model was applied using the following steps:

- **Stationarity Check**: The job demand time series was tested for stationarity using visual plots and statistical tests, such as the Augmented Dickey-Fuller test.
- **Differencing**: First-order differencing ($d = 1$) was applied to remove any trend, making the series stationary.
- **Model Selection**: Various combinations of (p, d, q) were evaluated based on AIC (Akaike Information Criterion) values. The best-performing model was chosen based on goodness of fit and predictive performance.
- **Forecasting**: The selected ARIMA model predicted monthly skill demand through 2024, offering a trend-focused projection.

While ARIMA effectively captured long-term trends, it was less capable of detecting short-term seasonal fluctuations compared to the Holt-Winters method. Nevertheless, its use was instrumental in verifying the general direction of labor market changes and helped cross-validate the results from other models.

The general mathematical form of the ARIMA(p, d, q) model is:

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.2}$$

Where:

- Y_t – The differenced (stationary) value of the time series at time t
- ϕ and θ – The model coefficients
- ε_t – The error term (white noise) at time t
- c – Constant term

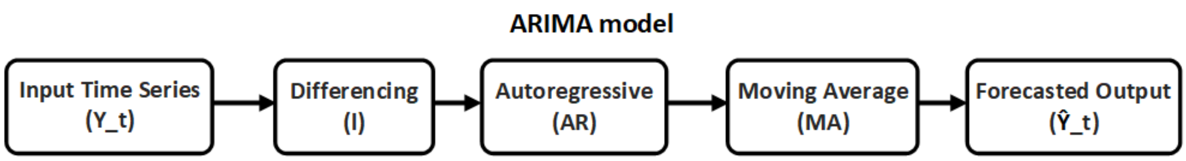

Figure 3.4: ARIMA forecasting model architecture showing the integration of differencing, autoregressive lags, and moving average smoothing.

This diagram illustrates how ARIMA integrates differencing (I), autoregression (AR), and moving average (MA) components to generate forecasts from a time series. Each stage processes the series incrementally to smooth trends and predict future values.

3.8.2 Holt-Winters Exponential Smoothing

This study employed the Holt-Winters additive exponential smoothing model to forecast future demand for skills in Uganda's labor market. This model is particularly effective for time series data showing trend and seasonal patterns with constant magnitude, which aligns with the observed characteristics of the monthly job postings data (2021–2023). Unlike the multiplicative variant, the additive model assumes that seasonal effects remain constant over time and do not change with the level of the trend.

The Holt-Winters additive model consists of three components:

- **Level** – the smoothed estimate of the series at time t
- **Trend** – the estimated change in the level from one period to the next
- **Seasonality** – the estimated seasonal effect at time t

The Holt-Winters equations are:

$$\text{Level:} \quad \ell_t = \alpha(Y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{3.3}$$

$$\text{Trend:} \quad b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \tag{3.4}$$

$$\text{Seasonality:} \quad s_t = \gamma(Y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \tag{3.5}$$

$$\text{Forecast:} \quad \hat{Y}_{t+h} = \ell_t + hb_t + s_{t+h-m(k+1)} \tag{3.6}$$

Where:

- Y_t is the observed value at time t
- $\hat{Y}_{t+h}$ is the forecast h periods ahead
- α, β, γ are smoothing parameters for level, trend, and seasonality
- m is the number of periods in a season (e.g., 12 for monthly data)
- k is the integer part of $(h-1)/m$

Application for Job Forecasting in This Study
In this study, ARIMA was applied with the following approach:

- First, the job postings time series was made stationary through differencing ($d = 1$).
- The ARIMA(1,1,1) model was found to be optimal, meaning the current value depends on one lagged observation (AR), one level of differencing (I), and one lagged error term (MA).
- The model generated monthly forecasts of skill demand through 2024, emphasizing the trend rather than seasonal variations.

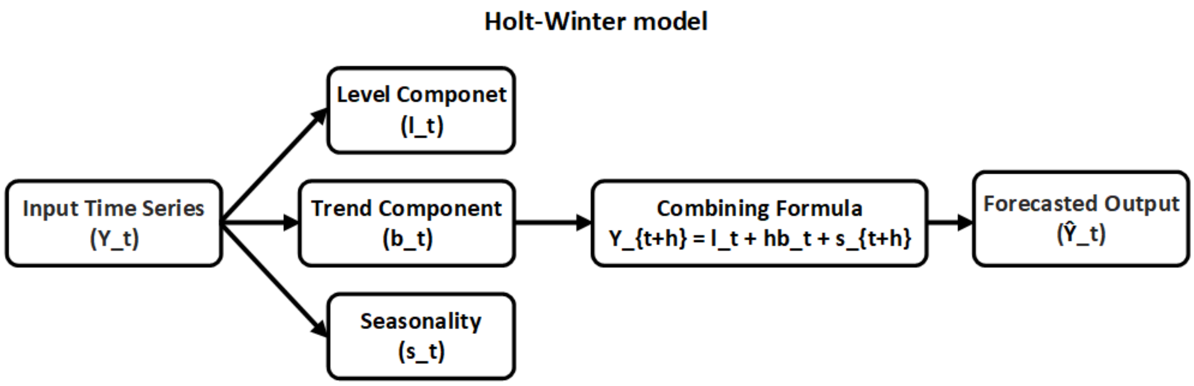

Figure 3.5: Holt-Winters Additive Model for Forecasting Skill Demand.

This diagram visualizes the components of the Holt-Winters additive model used to forecast monthly job postings in Uganda. The model decomposes the time series into three components: level (ℓ_t), trend (b_t), and seasonality (s_t), assuming constant seasonal effects over time.

It was applied to structured job posting data from 2021, 2022, and 2023 to predict skill demand for 2023 and 2024. The model captures both the upward trend in hiring and recurring seasonal peaks, providing data-driven insights to inform education and workforce planning.

3.8.3 Forecasting Process

The forecasting process followed a structured pipeline:

1. **Data Aggregation**: Time series data was generated by aggregating job postings by skill and date.
2. **Data Splitting**: The dataset was split into training and testing sets to validate model accuracy.
3. **Model Application**: Both ARIMA and Holt-Winters models were applied.
4. **Performance Comparison**: The models were evaluated to determine their forecasting accuracy.

3.8.4 Model Evaluation and Validation

Accuracy Metrics

Model accuracy was evaluated using:

- **Mean Absolute Error (MAE)**: Measures the average magnitude of errors in a set of predictions, without considering their direction.
- **Mean Squared Error (MSE)**: Measures the average squared difference between the actual and predicted values.

These metrics offer interpretable, quantitative evaluations of model accuracy. MAE represents average absolute deviation from actual values, while MSE penalizes more significant deviations more heavily. Lower values indicate better forecast performance.

External Validation

The reliability of the forecasting models was supported by:

- Cross-checking results with industry workforce reports.
- Benchmarking against global skill demand trends, ensuring external relevance.

3.9 Actionable Insights for Curriculum Development

The forecasted results identified key skill gaps by comparing high-demand skills with course offerings from major universities. The analysis revealed that most institutions still emphasize traditional IT and business subjects, with limited incorporation of high-demand skills, such as AI, cloud computing, and data science. These insights support recommendations for curriculum reform in Ugandan higher education institutions.

3.10 Methodological Constraints

While the study employs a robust data-driven approach, it is essential to acknowledge several methodological constraints that may impact the validity, generalizability, and scalability of the findings. Transparently addressing these challenges enhances the credibility and scientific rigor of the research.

Table 3.3: Key Methodological Constraints

Constraint	**Description**
Data Availability	Only 4,688 Ugandan job postings were extracted from a Kaggle dataset of over 1,048,575 entries. This limited availability may affect representativeness.
Data Quality	Variability in job posting formats across platforms led to data inconsistencies, necessitating extensive cleaning and standardization.
Bias	Online job platforms primarily represent urban and formal sector opportunities, potentially underrepresenting the demand from rural and informal sectors.
Technological Constraints	Changes in website structures or access restrictions (e.g., CAPTCHA, API limits) may disrupt web scraping and reduce the accuracy and completeness of the data.
Legal and Ethical Considerations	Compliance with ethical guidelines and terms of use was observed. The data was anonymized, but restrictions affected data reuse and interpretation.
Text Data Complexity	Ambiguity and redundancy in job description text complicated NLP preprocessing despite the use of TF-IDF and Named Entity Recognition (NER) tools.
Resource Constraints	Time and computational limits restricted the implementation of advanced models, such as LSTM or BERT, which could enhance prediction performance.

By documenting these methodological constraints, the study demonstrates transparency and lays the foundation for future research improvements, particularly in expanding data sources and adopting more advanced computational models. This approach aligns with best practices in qualitative and quantitative research, as advocated by [42], who emphasize addressing methodological limitations to enhance research validity and reliability.

Where possible, mitigation strategies were considered to manage these constraints. For instance, manual data cleaning improved data quality by cross-validating Kaggle-derived job listings against regional trends and selecting computationally efficient models that aligned with available resources. A more detailed account of mitigation approaches is provided in **Appendix A**.

3.11 Justification of Methodological Choices

The selection of analytical techniques in this study, specifically K-Means clustering for skill categorization and ARIMA and Holt-Winters models for time series forecasting, was guided by the nature of the data, the research objectives, and the need for interpretability and robustness.

3.11.1 K-Means Clustering

K-Means was selected for its simplicity, scalability, and capability to identify natural groupings in skill demand data. Given the unlabelled nature of the job posting dataset, K-Means was effective in uncovering hidden structures and recurring patterns in frequently co-occurring skills.

Alternative clustering methods, such as DBSCAN or hierarchical clustering, were considered; however, K-Means was favored due to its computational efficiency and clear interpretability, which are crucial for curriculum development applications.

3.11.2 Holt-Winters and ARIMA for Forecasting

Holt-Winters (Triple Exponential Smoothing) was chosen due to its effectiveness in handling seasonality and trend components in time series data. It is particularly suitable for relatively short-term forecasting with intense seasonal cycles, as observed in job posting trends.

ARIMA, in contrast, was selected for its robustness in capturing autoregressive and moving average components. It served as a baseline linear model for comparison with the Holt-Winters method. Both models are interpretable, easy to validate, and perform well on small to medium-sized datasets with missing periods, such as the dataset used in this study (which had a gap for 2021).

3.11.3 Why Advanced Models Like LSTM Were Not Chosen

While Long Short-Term Memory (LSTM) neural networks and other deep learning models, such as Prophet (Facebook's forecasting tool) or Transformer-based time series models, have demonstrated high predictive accuracy in complex forecasting tasks, they were not selected in this study for the following reasons:

- **Data Limitations:** Deep learning models generally require large, continuous, and high-frequency datasets for training. The job posting dataset used had missing periods and relatively short historical coverage (2021–2023), which limits the effectiveness of LSTM.
- **Interpretability:** A key goal of this study is to support curriculum development and policy planning. Traditional statistical models, such as ARIMA and Holt-Winters, are more accessible to non-technical stakeholders, including educators, government officials, and curriculum developers, for easier interpretation. Although LSTM is powerful, it operates more like a black box, making it less suitable for transparent decision-making.
- **Computational Cost:** LSTM models are resource-intensive to train and tune. Given the study's focus on reproducibility, scalability, and accessibility for institutions with limited computational resources, classical time series models were considered more practical.
- **Model Robustness and Transparency:** The selected models provide explainable forecasting processes, including precise diagnostics (e.g., autocorrelation plots, error terms, seasonal smoothing), which are essential for fostering trust in the analysis outcomes.

Chapter 4

Data Analysis, Presentation and Interpretation of Results

4.1 Introduction

This chapter applies machine learning techniques to forecast emerging skill demands across industries in Uganda. The primary objective is to generate actionable insights to inform curriculum reforms in higher education. The study addresses the central research question: *How can machine learning techniques effectively forecast skill demands and guide curriculum development in Uganda?* This chapter also builds upon the methodological models and libraries described in Chapter 3.

The underlying hypothesis posits that machine learning can produce accurate, data-driven predictions of future skill requirements. This chapter analyzes job postings, sector-specific trends, and frequency distributions to identify key digital and technical competencies increasingly shaping Uganda's labor market.

High-demand skills such as Python, Tableau, CSS, Java, AWS, and Google Analytics are particularly relevant in software development, data analytics, and IT security. In addition, sector-specific skills indicate a demand for AutoCAD in engineering, financial planning in business, and cybersecurity knowledge in IT roles. Skills in digital content, automation, and UX/UI design also reflect the interdisciplinary nature of today's evolving job market. These findings establish a foundation for aligning higher education curricula with the dynamics of the labor market.

4.2 Dataset

The core dataset was obtained from Kaggle (`https://www.kaggle.com`), a reputable platform for sharing datasets used in data science research. The specific dataset is titled "*Worldwide Job Postings*", compiled by [Name of Uploader if known], and made available under open-use terms. This dataset aggregates over 1,048,575 job postings from global online job platforms, including:

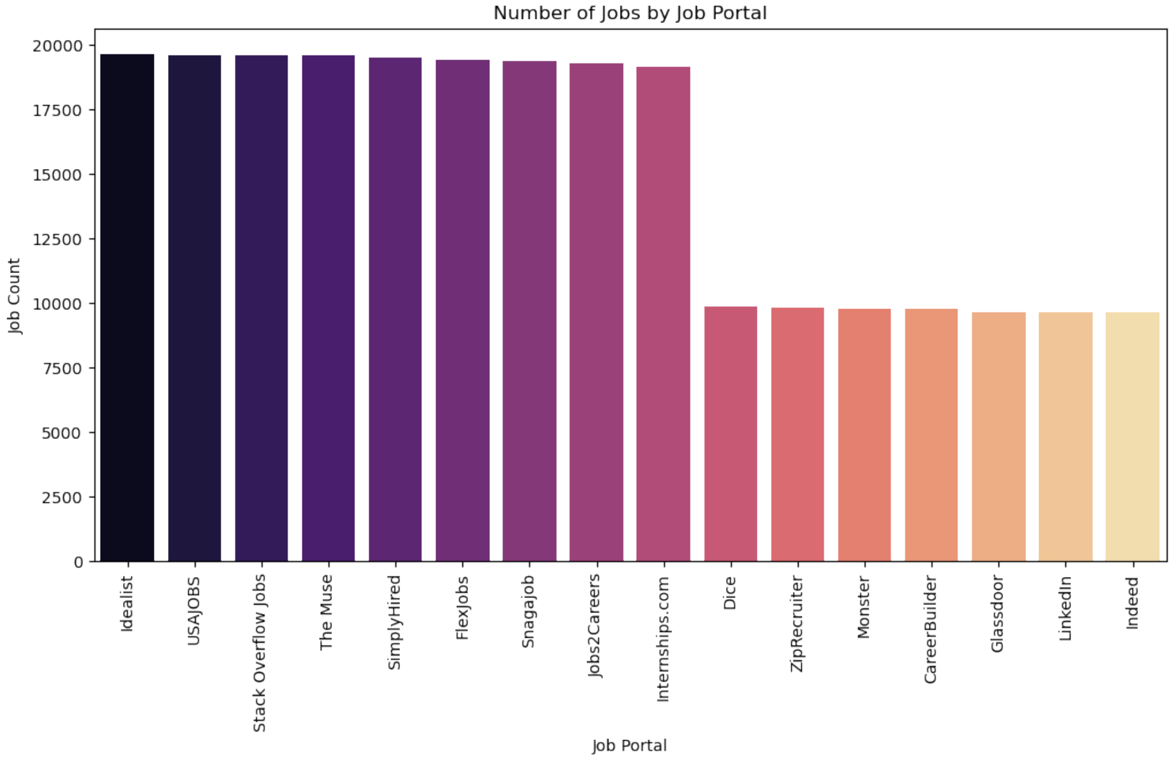

Figure 4.1: Number of jobs by portal

The figure illustrates the number of job postings sourced from various online job portals. Idealist, USAJOBS, and Stack Overflow Jobs are the leading platforms in terms of job count, reflecting their dominance in publishing job vacancies across various sectors and regions. This insight is vital for understanding where employers and job seekers are most active online.

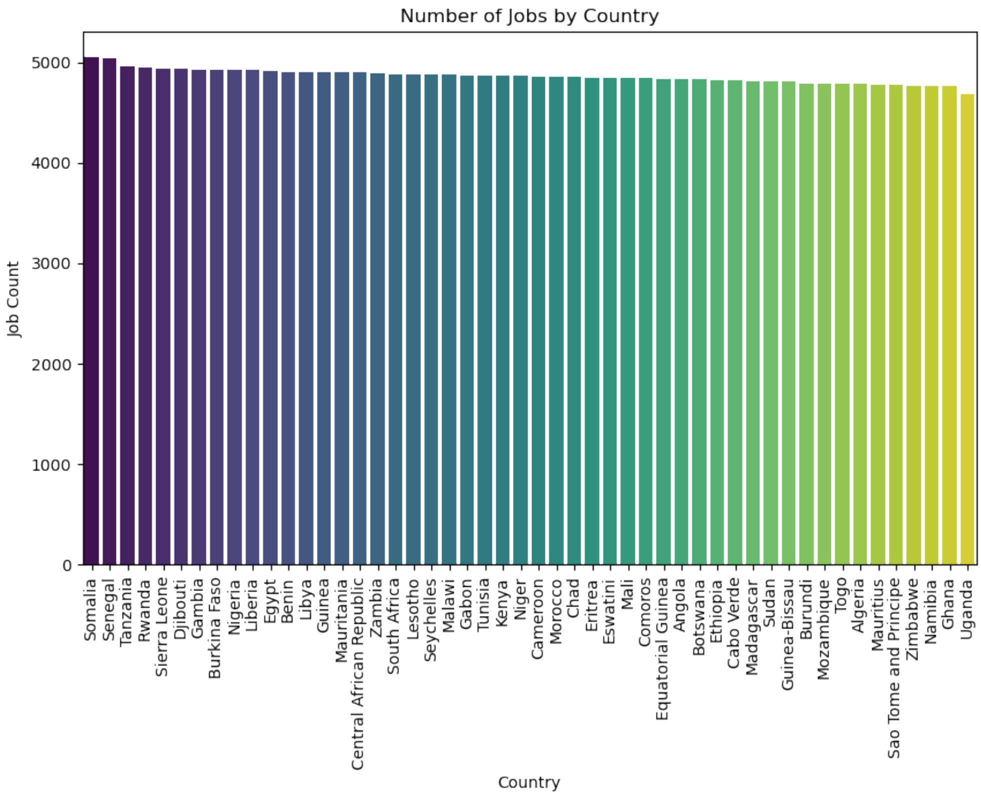

Figure 4.2: Number of Jobs by Country

This bar chart presents the distribution of job postings across African countries. Somalia, Senegal, and Tanzania show the highest number of job listings, indicating higher visibility of the digital job market or posting activity. Despite being a focus country, Uganda ranks among the lowest in job postings, suggesting either a lack of online listings or limited portal coverage.

4.3 Exploratory Data Analysis (EDA) and Skill Demand Trends

4.3.1 General Insights

A breakdown of qualifications revealed that university degrees were the most required, emphasizing the importance of formal education in the Ugandan labor market.

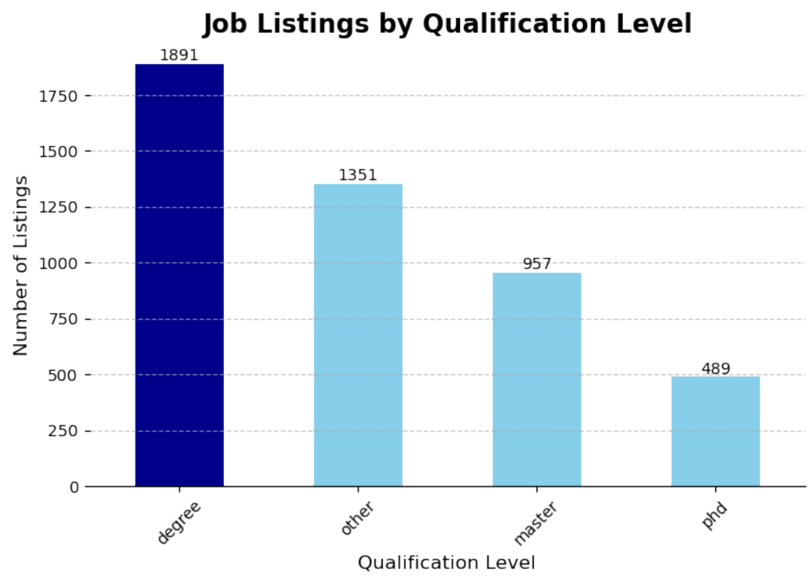

Figure 4.3: Most common job qualification.

Internships were the most frequent type of job, suggesting a strong demand for entry-level talent. Full-time and temporary contracts featured prominently, while part-time and contract-based roles indicated flexibility in the job market.

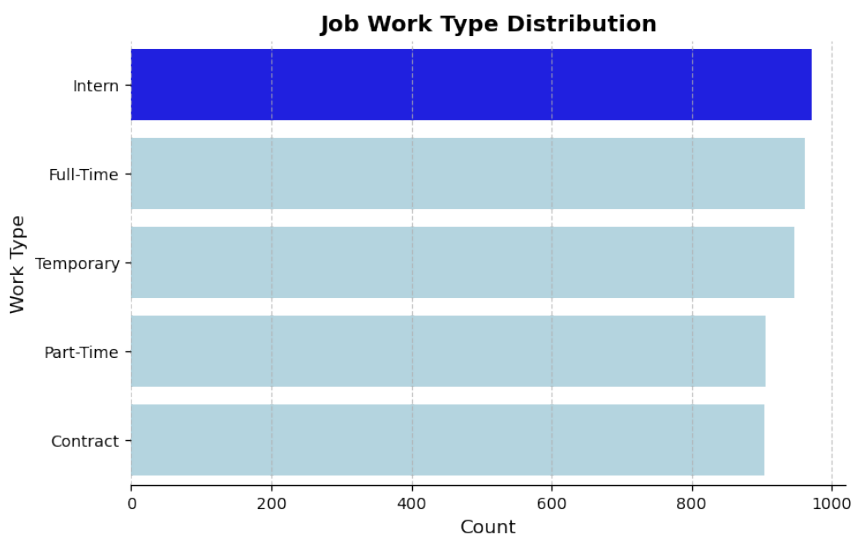

Figure 4.4: The distribution of job types.

Analysis revealed irregular posting patterns: limited data for 2021 (only September-December), full year data for 2022, and partial coverage for 2023 (January-September). These inconsistencies impacted the performance of forecasting models, particularly ARIMA, which is sensitive to gaps and irregular cycles in time series data. The Holt-Winters model, benefiting from its ability to smooth and adapt to seasonal trends, performed relatively better despite data constraints.

This observation emphasizes the importance of selecting time series models that can accommodate incomplete datasets and seasonality when forecasting skill demands in developing labor markets.

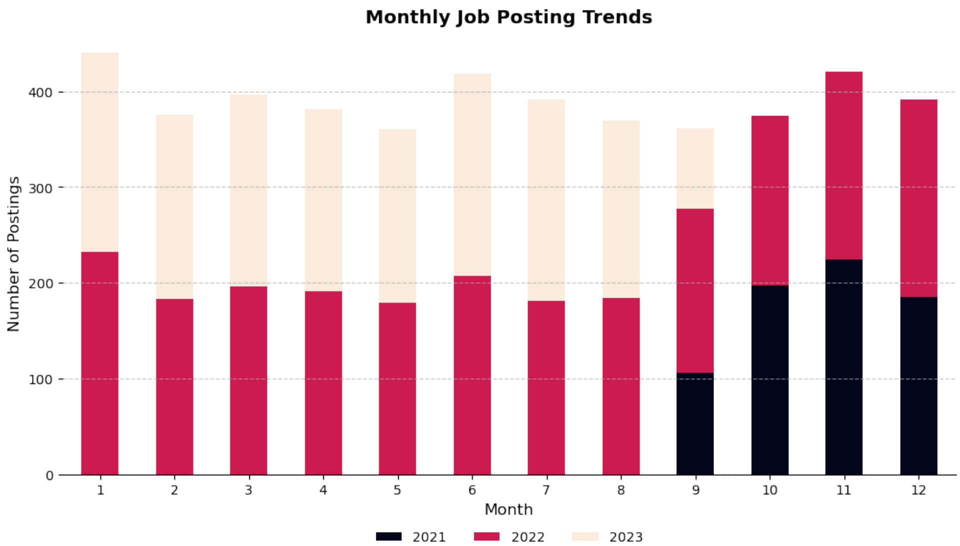

Figure 4.5: Number of job postings for each month by year.

Job postings frequently mention benefits such as health insurance, incentive programs, flexible work, wellness programs, and stock options, reflecting a growing emphasis on employee well-being and work-life balance.

Figure 4.6: Word cloud of the most common job benefits.

4.4 Most in-demand skills in Uganda

The study used anomaly detection techniques to identify spikes in the demand for skills such as Interaction Design, Network Administration, Data Analysis, and Executive Assistance. These roles represent evolving market priorities.

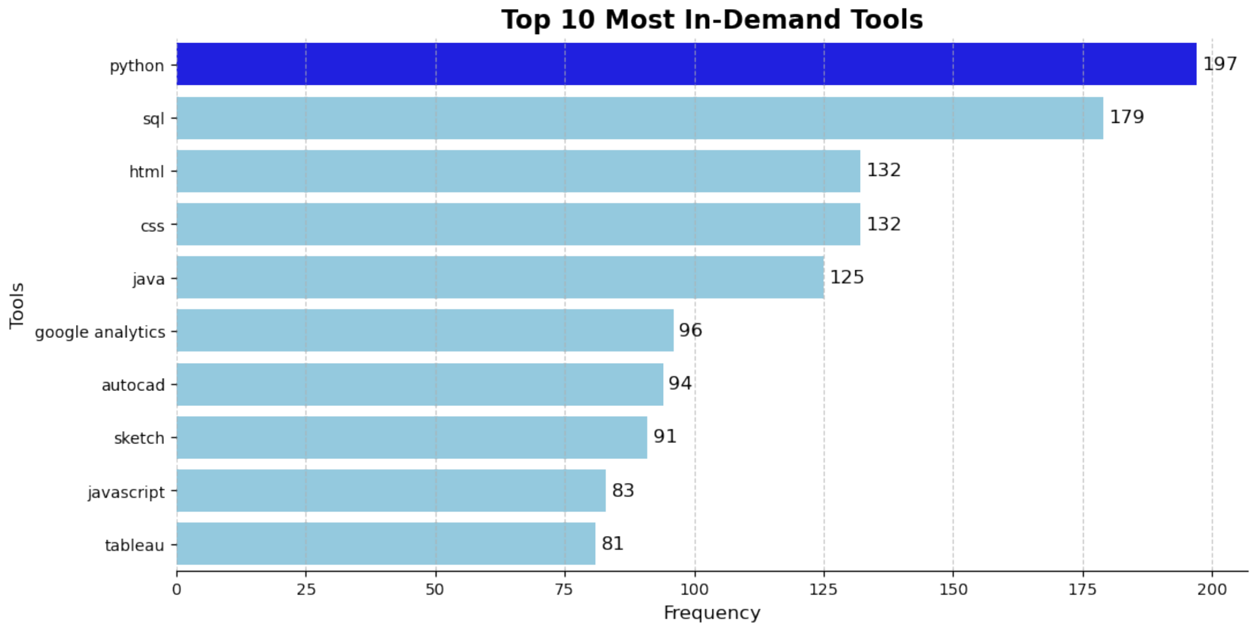

Figure 4.7: The most frequently mentioned tools

The analysis reveals that the Interaction Designer role is the most in-depth, emerging as the highest-ranked position across job postings. This is followed by the Network Administrator, which also shows significant demand, reflecting the ongoing importance of IT infrastructure management. In third place, the Data Analyst role is highly sought after, underscoring the growing need for data-driven insights in various industries. These findings highlight a clear trend towards roles centered around user experience design, network management, and data analytics in the current job market.

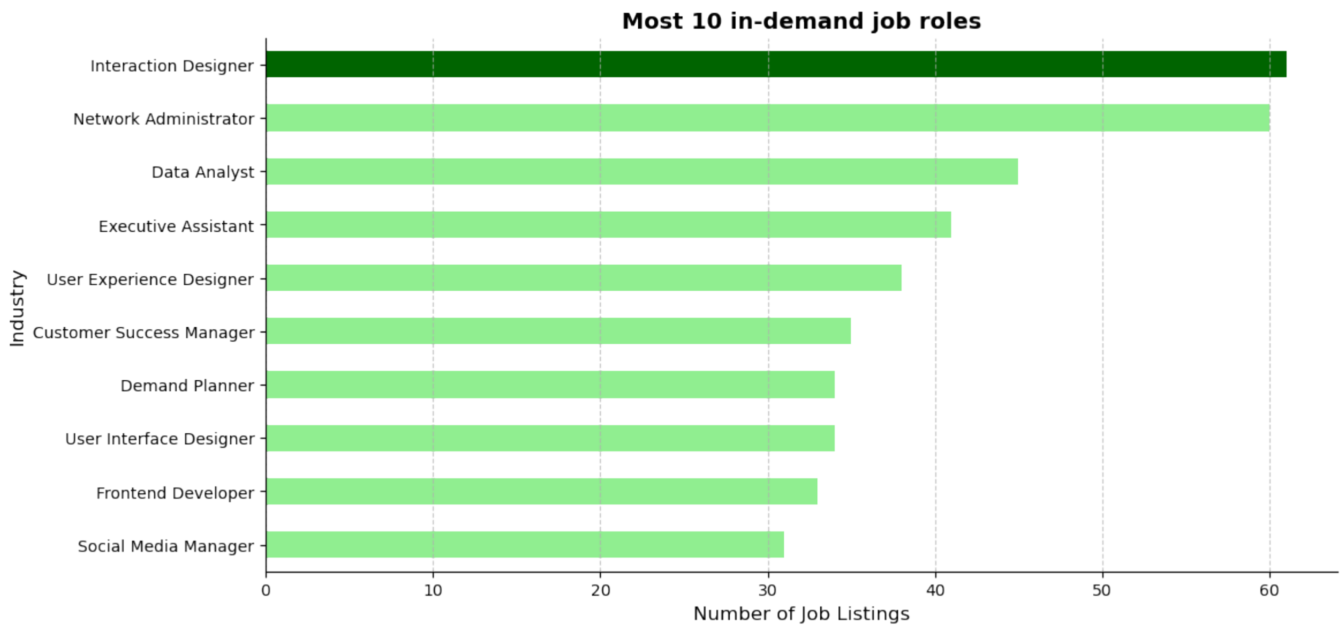

Figure 4.8: Ranking of the Most In-Demand Job Roles.

4.4.1 K-Means Clustering of Skill Demand

We employed the K-means clustering algorithm to identify patterns and groupings in the skill demand data, as justified in Section 3.8. This unsupervised machine learning technique partitions the data into a predefined number of clusters, ensuring that data points within each cluster are more similar to one another than to those in other clusters. K-Means' choice was motivated by its effectiveness in revealing inherent structures in multidimensional data and its interpretability for grouping related skills. According to the methodology (Section 3.8), the optimal number of clusters was determined using the elbow method, which examines the within-cluster sum of squares, and further validated by silhouette scores to ensure well-separated clusters. This approach guarantees that each cluster represents a meaningful grouping of skill characteristics.

Table 4.1: Clusters of Industry-Specific Skills

Cluster ID	Top Words	Industry
8	design, user, principles, prototyping, wire-framing	UX/UI Design
3	management, communication, planning, design, development	Project Management
6	security, network, management, database, troubleshooting	IT
0	pediatric, care, patient, assessment, treatment	Healthcare
2	data, analysis, research, legal, communication	Data Science / Research
4	management, communication, coordination, vendor, planning	Supply Chain / Operations
1	sales, management, relationships, negotiation, product	Sales / Business Development
5	marketing, content, analytics, seo, social	Digital Marketing
7	test, testing, development, automation, tracking	Software Development

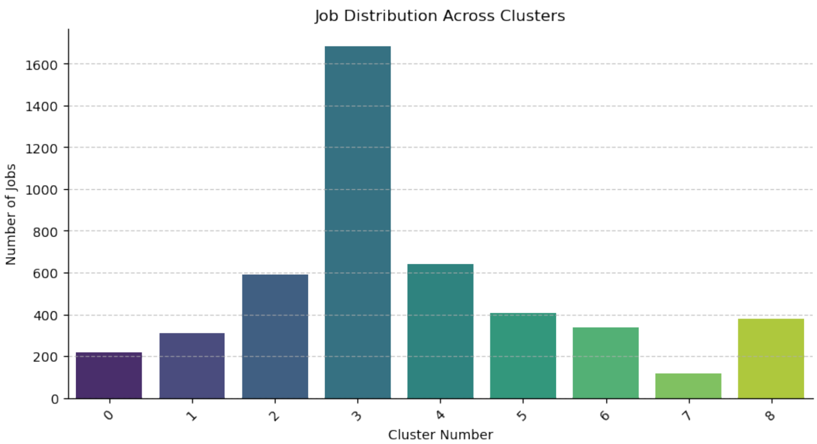

Figure 4.9: K-Means Clustering of Uganda's Skill Demand.

This figure visualizes the distribution of skill demand data across the clusters identified by the K-Means algorithm. Each cluster is depicted in a distinct color, illustrating how the skill demand observations are grouped. The clustering is based on similarities in skill demand patterns (e.g., frequencies or co-occurrence of specific skills), revealing distinct groupings of related skill sets in the job market.

After experimenting with different values of K, the analysis settled on K = 3 clusters, the point at which additional clusters provided diminishing returns in explaining variance. The K-Means algorithm grouped the skill demand data into three distinct clusters, each characterized by a particular profile of skills or job categories. The characteristics of the identified clusters are summarized below (and detailed in Table 4.1)

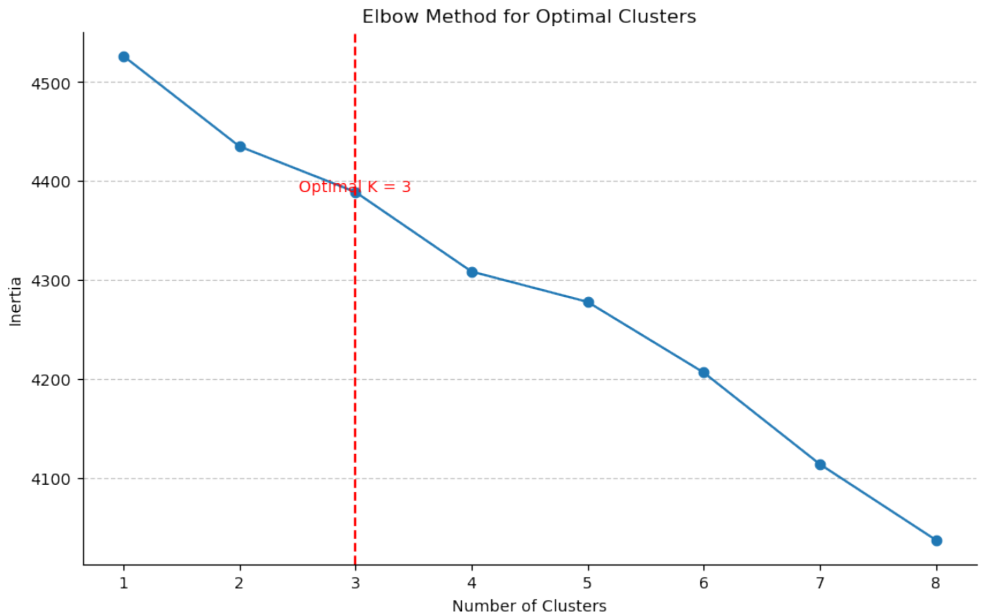

Figure 4.10: Elbow Plot Showing the Optimal Number of Clusters (K = 3).

4.5 Time Series Forecasting of Skill Demand

To complement the cluster analysis, a time series forecast was conducted to understand trends in overall skill demand and predict future needs. The forecasting involved two approaches: an ARIMA model and the Holt-Winters exponential smoothing model, applied to the historical dataset of job postings from 2021 to 2023. Section 3.8 of the methodology provides the detailed rationale for using these models: ARIMA was chosen for its strength in modeling temporal dependencies and trends in data, while Holt-Winters was selected to explicitly capture seasonality, which is expected in monthly job demand data. Using both models, we aimed to cross-validate the forecasts and ensure the robustness of the conclusions.

4.5.1 Time series analysis for skill forecasting

To understand and anticipate emerging labor market needs, this study applied time series forecasting techniques to project future trends in skill demand. Forecasting models are crucial for enabling higher education institutions and policymakers to proactively align training programs with anticipated changes in the job market. Two models were selected and compared: the ARIMA (Auto Regressive Integrated Moving Average) model and the Holt-Winters Exponential Smoothing model.

Historical job posting data from 2021 to mid-2023 served as the basis for training the model. The forecasts were then generated to cover the period from late 2023 through 2024. The results, visualized in the forecast graph, show each model's predictive performance and distinct characteristics.

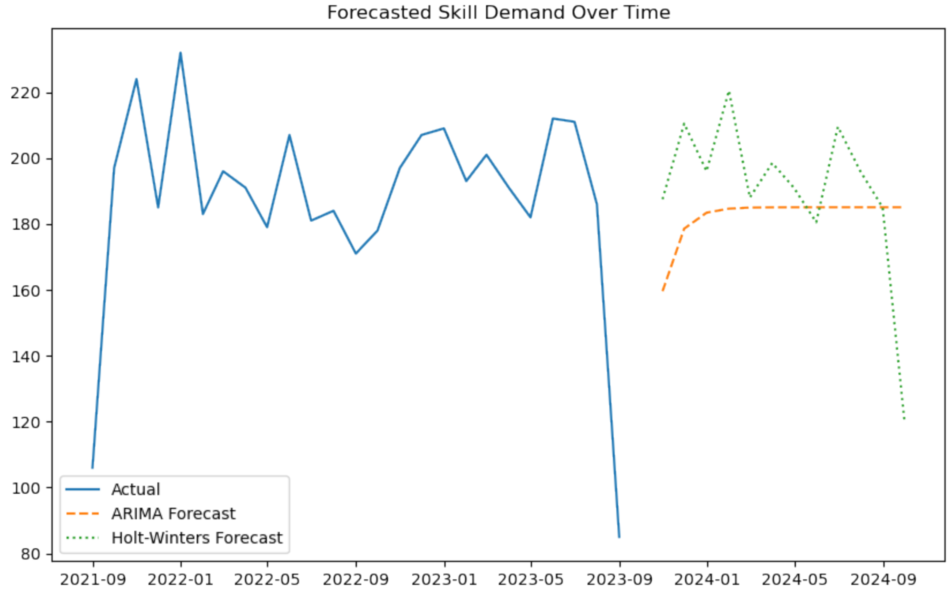

Figure 4.11: Comparative line graph showing ARIMA vs. Holt-Winters forecasts.

ARIMA Model Forecast

The ARIMA model is a widely used statistical forecasting method that combines autoregressive terms (AR), differencing (I), and moving average terms (MA) to model time-dependent structures in data. It is particularly effective in capturing trends and linear dependencies in time series data that do not exhibit strong seasonality.

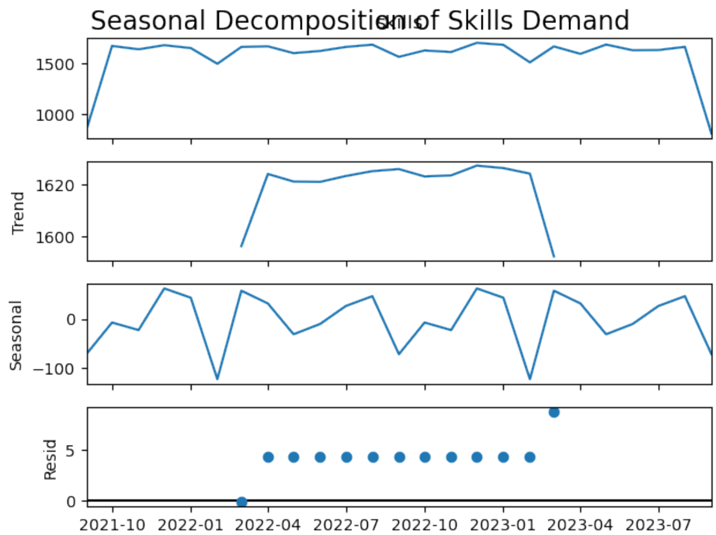

Figure 4.12: Seasonal Decomposition of Skills Demand

This multi-panel chart shows the seasonal decomposition of skill demand using additive decomposition. The top panel displays the original data, the estimated trend, seasonal effects, and residuals. The decomposition reveals distinct seasonal fluctuations and a stable underlying trend, offering valuable insights for curriculum planning and labor market forecasting.

In this study, the ARIMA model produced a relatively stable and smooth forecast for skill demand over the projected period. As shown in the graph (orange dashed line), the model predicts that skill demand will remain consistent, with only marginal changes throughout 2024. While this stability benefits long-term planning, the model does not capture the fluctuations and seasonal effects in historical data. This limitation suggests that the ARIMA model may underrepresent cyclical hiring behaviors or sector-specific demand surges.

Despite this, the ARIMA forecast provides valuable insight into the overall trend in skill demand, making it useful for strategic, long-range curriculum planning where volatility is less of a concern.

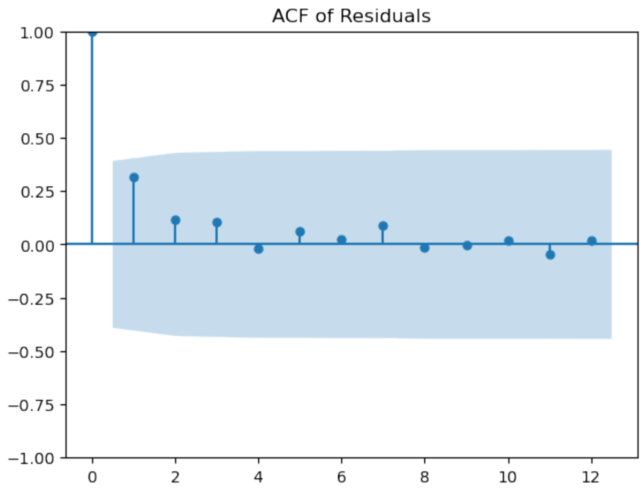

Figure 4.13: Autocorrelation Function (ACF) of Residuals

This plot visualizes the autocorrelation of residuals from the fitted ARIMA model. The absence of significant spikes outside the confidence bounds suggests that the residuals are not autocorrelated, supporting the model's adequacy in capturing the temporal structure of the data. This is a key diagnostic for validating the model's performance.

Holt-Winters Forecast

The Holt-Winters Exponential Smoothing model, also known as Triple Exponential Smoothing, extends basic exponential smoothing by incorporating trend and seasonality components. This makes it particularly suitable for time series data where skill demand exhibits cyclical or seasonal patterns, such as academic intake seasons or industry-specific hiring waves.

The Holt-Winters forecast (green dotted line in the graph) more accurately reflects the periodic fluctuations observed in historical data. It projects alternating peaks and troughs throughout the forecast horizon, indicating that demand for specific skills will fluctuate cyclically rather than remaining constant.

This model's sensitivity to seasonality makes it a strong candidate for short- to medium-term planning, especially in sectors that experience frequent shifts in labor demand. However, the higher volatility in its forecast may also reflect overfitting if the seasonal components are not well-calibrated.

Overall, the Holt-Winters model demonstrates greater responsiveness to actual labor market behavior, which is particularly important for designing agile and modular training programs that adapt to real-time needs.

4.6 Model Performance and Evaluation

The models were evaluated using three standard error metrics:

- **Mean Absolute Error (MAE)** – measures the average magnitude of errors.
- **Root Mean Squared Error (RMSE)** – provides a measure of error magnitude that penalizes large deviations.
- **Mean Absolute Percentage Error (MAPE)** – assesses the relative accuracy of the forecasts in percentage terms.

Table 4.2: Forecast Error Metrics for ARIMA and Holt-Winters Models

Model	**MAE**	**RMSE**	**MAPE (%)**
ARIMA model	19.99	32.87	15.72%
Holt-Winters model	9.05	13.11	6.53%

This table compares the performance of the ARIMA and Holt-Winters forecasting models based on three evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The Holt-Winters model achieved lower error values across all metrics, indicating superior forecasting accuracy compared to the ARIMA model.

The results indicate that the Holt-Winters model outperformed the ARIMA model across all three evaluation metrics. Its lower MAE and RMSE suggest it provided more accurate predictions with smaller average and squared errors. Furthermore, the significantly lower MAPE indicates that the Holt-Winters model achieved greater relative accuracy in forecasting demand for skills.

The Holt-Winters model's superior performance can be attributed to its ability to capture seasonal variations and trend components inherent in the job market data. On the other hand, while the ARIMA model effectively captured the general trend, it struggled to represent the cyclical patterns in hiring that characterize skill demand over time.

These findings support the selection of the Holt-Winters model as the preferred approach for forecasting skill demand in dynamic labor markets, particularly when seasonality plays a significant role. Its predictive accuracy can help inform data-driven decision-making in curriculum development, workforce planning, and policy formulation.

4.7 Forecasted Skills for the Next 12 Months

The top five skills expected to grow in demand based on the Holt-Winters model are shown below:

Table 4.3: Forecasted Demand for Selected Digital Skills

Month	Python	Tableau	CSS	Java	Google Analytics
2024-01-31	510	230	120	410	160
2024-02-29	500	225	115	400	155
2024-03-31	490	220	110	390	150
2024-04-30	480	215	105	380	145

These forecasts provide actionable insights for university curriculum committees and policy bodies seeking to align educational offerings with projected labor market demands.

Chapter 5

Discussion of Results

This chapter interprets the analytical findings of Chapter 4 in the context of the Ugandan labor market and the education system. It aims to derive actionable insights from K-Means clustering and time series forecasting (ARIMA and Holt-Winters), focusing on their implications for curriculum design, workforce planning, and policy development.

The discussion aligns with the research objectives: interpreting skill clusters to reveal labor market structures, analyzing forecasts to identify future demand trends, and linking findings to education reforms. It compares Uganda's trends with global shifts to assess readiness for the digital economy.

This chapter supports evidence-based strategies to align education with evolving labor market needs by combining data-driven results with contextual analysis.

5.1 Interpretation of Forecast Results

The "Forecasted Skill Demand Over Time" chart illustrates the historical trend and predictive performance of two time-series models, ARIMA and Holt-Winters, on skill demand data from September 2021 through projected values into 2024.

The blue line represents the observed skill demand from late 2021 to mid-2023. The series exhibits visible seasonality and fluctuations, characterized by a general upward trend, followed by periods of volatility and a sharp decline toward the end of the historical period. This sudden drop could indicate missing data or an external disruption, such as changes in data collection or labor market shocks.

The orange dashed line shows the forecast produced by the ARIMA model. This model captures the overall level of demand but smooths out fluctuations, resulting in a relatively stable estimate. While it provides a conservative prediction, it fails to reflect the seasonal variations evident in the actual data.

In contrast, the green dotted line, representing the Holt-Winters forecast, closely aligns with the seasonal dynamics observed in the historical data. This model captures periodic peaks and troughs more effectively, suggesting better responsiveness to underlying seasonal skill-demand patterns. However, the volatility in the Holt-Winters forecast can also amplify noise if not properly tuned.

Overall, the Holt-Winters model appears more suitable for short-term forecasts where seasonality is prominent. In contrast, the ARIMA model offers a more stable and accurate estimation of long-term trends. The differences between the models underscore the importance of aligning the forecasting technique with the specific characteristics of the data and the intended use of the forecast, whether for planning, curriculum reform, or policy intervention.

5.2 Discussion of Findings

The findings from both the clustering and time series forecasting analyses present a compelling case for urgent and informed curriculum reform in Uganda's higher education system. This study reveals the skill sets in demand nationwide and how these demands evolve, aligning with local development goals and global labor market trends.

The cluster analysis identified three broad categories of skills:

1. Technology and data-oriented skills
2. Business, finance, and soft skills
3. Sector-specific technical expertise

These clusters highlight the multifaceted nature of Uganda's labor market, where emerging digital skills intersect with enduring needs in traditional sectors, including agriculture, healthcare, and education. Notably, the prominence of digital and analytical skills echoes the transformation of global labor markets, as described in the World Economic Forum's *Future of Jobs Report* (2023), affirming that Uganda is not isolated in its shift toward a digital economy.

The time series forecasts, particularly the Holt-Winters model, provide an additional layer of insight, illustrating that skill demand is increasing and exhibiting seasonal fluctuations. This cyclical nature of demand emphasizes the importance of timing in curriculum design—such as aligning graduation cycles with peak hiring periods and introducing modular, fast-paced training programs during academic off-seasons. The growing demand trajectory also necessitates a strategic scaling of academic program capacity, particularly in high-demand fields such as Information and Communication Technology (ICT), data analytics, and healthcare.

Evaluation of model performance revealed that Holt-Winters significantly outperformed ARIMA in predictive accuracy (MAE: 9.05 vs. ARIMA's 23.87), highlighting its ability to capture trends and seasonality in skill demand. This suggests that traditional trend-based models may be insufficient for planning purposes where periodic shifts in hiring patterns are critical.

The implications of these findings extend to multiple stakeholder levels:

- **Educators and curriculum developers:** There is a clear directive to integrate high-demand skills, primarily digital and soft skills, into academic programs across disciplines.
- **Institutional leaders:** The data support investing in faculty development, industry partnerships, and infrastructure to support experiential learning and certification pathways.
- **Policymakers (e.g., NCHE):** The findings offer a framework for updating accreditation standards and incentivizing programs that directly address labor market intelligence.

Additionally, the results highlight gaps where specific skills—especially emerging or niche ones such as cybersecurity, green energy, and digital marketing—may be underrepresented in academic offerings. These insights equip curriculum planners with foresight to embed forward-looking content into existing programs, thus enhancing the relevance and adaptability of graduates.

Importantly, Uganda's demand trends align with global patterns, particularly in technology, business, and resilience-related skills [43]. This alignment validates domestic reforms and positions Uganda to benefit from cross-border employment opportunities and digital service exports, provided its graduates are adequately trained. However, the analysis also reflects Uganda's unique context, including the critical need for skills in agribusiness, healthcare, and renewable energy, which must be prioritized to drive national development.

In conclusion, this study's findings present a data-driven rationale for dynamic and context-sensitive reforms in higher education. They underscore the need for a skills-based, industry-linked, and future-ready curriculum strategy that strikes a balance between global competitiveness and local relevance. These insights form the basis for the strategic recommendations outlined in Chapter 6.

Chapter 6

Conclusions and Recommendations

6.1 Introduction

This chapter presents the study's strategic conclusions and actionable recommendations, grounded in the data-driven insights derived from cluster analysis and time series forecasting of skill demand in Uganda. Building on the empirical findings in Chapter 4, which identified three major skill clusters: Technology/Data, Business/Soft Skills, and Sector-Specific Skills, this chapter outlines the implications for curriculum reform, policy interventions, and long-term workforce development. These recommendations guide higher education institutions and policymakers toward ensuring that Uganda's graduates are equipped with relevant, future-ready competencies. This approach aligns with Uganda's Fourth Development Plan (NDPIV) [44], Vision 2040 [45], and SDG 4 (Quality Education) [46], emphasizing the importance of proactive, evidence-based educational reform.

6.2 Summary of Key Findings

The study demonstrated that machine learning models, particularly the Holt-Winters forecasting method, effectively identify labor market patterns and forecast future demand for skills. Compared to ARIMA, Holt-Winters more accurately captures seasonality and fluctuations in skill needs. The forecast results indicate a steady upward trend in demand for digital and technical skills, including artificial intelligence, cloud computing, and data analytics, alongside a growing emphasis on soft skills and specialized sector knowledge.

The cluster analysis categorized skill requirements into three distinct groups:

1. **Technology/Data-Oriented Skills**
2. **Business/Soft Skills**
3. **Sector-Specific Technical Skills** (e.g., agriculture, healthcare, and renewable energy)

These findings underscore the evolving, multidimensional nature of Uganda's labor market and highlight the critical need for curricula that combine technical expertise, soft skills, and sector-specific knowledge [47].

6.3 Recommendations

6.3.1 Recommendations for Higher Education Institutions

i. **Curriculum Reform:** Institutions should integrate emerging technical skills such as artificial intelligence (AI), cyber security, data analytics, and cloud computing into their core programs to better prepare students for the evolving workforce.

ii. **Promote Data-Driven Teaching:** Faculties should adopt data analytics and visualization tools across disciplines to foster data literacy and interdisciplinary collaboration.

iii. **Strengthen Academia-Industry Linkages:** Establish long-term partnerships with employers to co-develop curricula, conduct certification boot camps, and organize internships or guest lectures [48].

iv. **Support Lifelong Learning:** Introduce flexible modular certification programs for graduates and professionals seeking career progression or retraining [49].

6.3.2 Policy Recommendations for the Ugandan Government

i. **Invest in STEM and Digital Education:** Increase budget allocations for science and digital skills training at tertiary institutions.

ii. **Establish a National Skills Intelligence Platform:** Develop a real-time labor market analytics dashboard to monitor trends and guide evidence-based education policy planning.

iii. **Support Up-skilling Initiatives:** Fund re-skilling and up-skilling programs targeting youth and mid-career professionals to enhance their employability.

iv. **Promote Inclusive Access:** Address gender and regional disparities in access to digital skills training to foster equity and national development.

6.3.3 Limitations and Future Research Directions

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

- **Data Coverage:** The study relied on formal job postings, potentially excluding informal sector dynamics.
- **Forecasting Constraints:** The clustering and forecasting models depend on historical data and may not capture sudden economic or technological disruptions.

Future research should consider:

1. Integrating informal sector and freelance job data.
2. Exploring advanced forecasting models, including hybrid AI approaches.
3. Refining skill cluster taxonomies into more granular, actionable categories.
4. Assessing the impact of educational reforms on the employability of graduates.
5. Conducting comparative studies across regions and countries to identify best practices.

6.4 Conclusion

This study demonstrates the significant role that machine learning techniques can play in forecasting labor market demand for skills. By leveraging these tools, the study identified key trends crucial for aligning higher education curricula with the evolving needs of the workforce.

Digital and technical skills will continue to dominate various sectors, requiring both immediate and long-term reforms to educational policy and practice. The recommendations in this chapter offer a roadmap for higher education institutions, the Ugandan government, and industry players to collaborate in creating an education system that meets the demands of a rapidly changing job market through the strategic use of big data [50]. If implemented, these insights can help Uganda bridge its skills gap, ensure its workforce is future-ready, and accelerate sustainable economic growth and social development.

References

[1] A. Al-Maskari, T. Al Riyami, and S. Ghnimi, “Factors affecting students’ preparedness for the fourth industrial revolution in higher education institutions,” *Journal of Applied Research in Higher Education*, 2022.

[2] United Nations, “Technology and innovation report 2021: Catching technological waves, innovation with equity,” 2021.

[3] U. Bhandari, “Chasing quality tvet curricula,” *Journal of Technical and Vocational Education and Training (TVET)*, vol. 19, no. 1, pp. 36–47, 2025.

[4] A. Khaldibekova, “Osobennosti pravovogo regulirovaniya dogovorov okazaniya uslug po kommercheskomu ispol’zovaniyu big data: Big data kak ob’ekt grazhdanskikh prav,” *Nauka i nauchnyi potentsial: Osnova ustoichivogo innovatsionnogo razvitiya obshchestva*, vol. 1, no. 1, pp. 135–143, 2022.

[5] J. Brasse, “Identification of future skills using data-driven methods: A systematic literature review and directions for future research,” in *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)*, (Maui, Hawaii), pp. 5248–5257, HICSS, 2024. Ulm University.

[6] National Council for Higher Education (NCHE), “National council for higher education minimum standards for open distance and e-learning,” Tech. Rep. August, NCHE, 2019. [Online].

[7] N. Zafar and J. Ahamed, “Emerging technologies for the management of covid-19: A review,” *Sustainable Operations and Computers*, vol. 3, no. January, pp. 249–257, 2022.

[8] World Economic Forum, “The global risks report 2025: 20th edition (insight report),” January 2025. Available online.

[9] Uganda Bureau of Statistics (UBOS), “The national labour force survey 2021 - main report,” 2021.

[10] G. Brunello and P. Wruuck, “Skill shortages and skill mismatch in europe: A review of the literature,” IZA Discussion Paper Series 12346, Institute of Labor Economics (IZA), 2019.

[11] World Bank Group, “Africa’s pulse: An analysis of issues shaping africa’s economic future,” Tech. Rep. Volume 16, World Bank Group, Office of the Chief Economist for the Africa Region, October 2017. Public Disclosure Authorized.

[12] J. Domingue, M. D’Aquin, E. Simperl, and A. Mikroyannidis, “The web of data: Bridging the skills gap,” *IEEE Intelligent Systems*, vol. 29, no. 1, pp. 70–74, 2014.

[13] D. Filmer and L. Fox, *Youth Employment in Sub-Saharan Africa.* Africa Development Series, Washington, DC: World Bank, 2014.

[14] M. Aring, “Report on skills gaps,” tech. rep., UNESCO, 2012. Background paper for the Education for All Global Monitoring Report 2012: Youth and Skills: Putting Education to Work.

[15] M. Laboissiere and M. Mourshed, “Closing the skills gap: Creating workforce-development programs that work for everyone,” *McKinsey*, no. February, pp. 1–6, 2017.

[16] O. O. Adepoju and C. O. Aigbavboa, “Assessing knowledge and skills gap for construction 4.0 in a developing economy,” *Journal of Public Affairs*, vol. 21, no. 3, p. 1, 2021.

[17] A. Shajek and E. A. Hartmann, eds., *New Digital Work: Digital Sovereignty at the Workplace*. Cham, Switzerland: Springer Nature Switzerland AG, 2023. Open Access under Creative Commons Attribution 4.0 International License.

[18] National Planning Authority, “Third national development plan (ndpiii),” National Development Plan NDPIII, Government of Uganda, 2020. Accessed July 2025.

[19] Ministry of ICT and National Guidance, Government of Uganda, “Digital transformation roadmap for uganda,” policy document, Ministry of ICT and National Guidance, Government of Uganda, 2023. Accessed July 2025.

[20] P. Namuliira, I. Kasirye, P. Okillong, B. Atwine, S. C. Nakkazi, P. Nakitende, and M. M. Kajumba, “Skills acquisition in uganda: Bridging the gap between education and employment in the era of digital transformation,” Tech. Rep. Research Series No. 164, Economic Policy Research Centre (EPRC), Kampala, Uganda, Sept. 2024. Copyright © Economic Policy Research Centre (EPRC).

[21] F.O.R.C. Review and Development Services, “Terms of reference for curriculum review & development services for selected vocational education,” 2019. Unpublished report.

[22] Organisation for Economic Co-operation and Development (OECD), “OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market,” tech. rep., OECD Publishing, Paris, 2023. OECD Employment Outlook ISSN 1013-0241 (print), ISSN 1999-1266 (online).

[23] Republic of Uganda National Task Force, “National Strategy on Fourth Industrial Revolution: A Report by the National Task Force,” policy report, Government of Uganda, May 2021. Accessed July 2025.

[24] UNICEF Uganda, “Skills for the future: Young people navigate the 21st century,” tech. rep., UNICEF Uganda, Kampala, Uganda, July 2024. Flagship report addressing youth skills in Uganda.

[25] United Nations Development Programme, *Sustainable Development Goals*. UNDP, 2015. Accessed July 2025.

[26] C. Howard, “Digital skills for youth employment: Fostering digital transformation for social impact,” *Include Knowledge Platform*, 2023. Accessed July 2025.

[27] T. N. Kashupi and E. Chikohora, “Using nlp to bridge data science skills gap in namibia: A survey,” in *2021 3rd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, IEEE, 2021. Accessed July 2025.

[28] C. D. Casuat, E. D. Festijo, and A. S. Alon, "Predicting students' employability using support vector machine: A smote-optimized machine learning system," *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 2101–2106, May 2020.

[29] Y. Deng, "A systematic review of application of machine learning in curriculum design among higher education," *Journal of Emerging Computer Technologies*, vol. 4, no. 1, pp. 15–24, 2024.

[30] M. Qiu, Y. Xu, and E. O. Omojokun, "To close the skills gap, technology and higher-order thinking skills must go hand in hand," *Journal of International Technology and Information Management*, vol. 29, no. 1, pp. 98–123, 2020.

[31] M. Backenköhler, A. Singla, F. Scherzinger, and V. Wolf, "Data-driven approach towards a personalized curriculum," in *Proceedings of the 11th International Conference on Educational Data Mining (EDM)*, pp. 246–251, 2018.

[32] M. J. Alam, K. Ogawa, and S. R. Bin Islam, "Importance of skills development for ensuring graduates' employability: The case of bangladesh," *Social Sciences*, vol. 11, no. 8, p. 360, 2022.

[33] T. Begazo, M. P. Blimpo, and M. A. Dutz, *Digital Africa: Technological Transformation for Jobs.* World Bank, 2023.

[34] S. J. Shim, "Mapping the mis curriculum based on is critical knowledge and skills: An empirical examination of is practitioners," *Regional Industry Review*, vol. 41, no. 1, pp. 289–313, 2019.

[35] S. Lunn, J. Zhu, and M. Ross, "Utilizing web scraping and natural language processing to better inform pedagogical practice," in *Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE)*, (Uppsala, Sweden), IEEE, 2020.

[36] M. Busari, "Predictive analytics for identifying skill gaps and future workforce requirements," *Unpublished manuscript*, March 2025. Available from the author at the University of Ibadan.

[37] UNESCO, "Harnessing the Era of Artificial Intelligence in Higher Education," tech. rep., UNESCO, 2023. [Online; accessed 2025-07-24].

[38] European Commission, Directorate-General for Employment, Social Affairs and Inclusion, "Future skills, career guidance and lifelong learning in public employment services," tech. rep., European Commission, Brussels, Belgium, 2022. Thematic paper. Directorate-General for Employment, Social Affairs and Inclusion, European Commission. Contact: Hilde Olsen, EMPL-PES-SECRETARIAT@ec.europa.eu.

[39] J. Maksimovi and J. Evtimov, "Positivism and post-positivism as the basis of quantitative research in pedagogy," *Research in Pedagogy*, vol. 13, no. 1, pp. 208–218, 2023.

[40] Y. S. Park, L. Konge, and A. R. Artino, "The positivism paradigm of research," *Academic Medicine*, vol. 95, no. 5, pp. 690–694, 2020.

[41] M. Kubat, "An introduction to machine learning," *An Introduction to Machine Learning*, no. December, pp. 1–458, 2021.

[42] C. T. K.-H. Stadtländer, "Qualitative, quantitative, and mixed-methods research," *Microbe Magazine*, vol. 4, no. 11, pp. 485–485, 2009.

[43] M. Odent, *The Future of Homo.* 2019.

[44] "Fourth national development plan (ndp iv) 2025/26–2029/30," tech. rep., Government of Uganda, 2025. Published March 2025.

[45] Ministry of Finance, Planning and Economic Development, Uganda, "Uganda vision 2040," Tech. Rep. 3, 2007.

[46] SDG Youth and S. Camp, "The 2030 agenda for sustainable development's 17 sustainable development goals (sdgs)," 2023.

[47] Ministry of Education and Sport, "Digital agenda for the education and sports sector," tech. rep., July 2020. Technical report.

[48] T. O. Pacho, "Transforming higher education in africa through university-industry collaboration," in *Proceedings of the 20th International Conference on Private Higher Education in Africa*, (Africa), Apr. 2022.

[49] UNESCO, *Making Lifelong Learning a Reality.* 2022.

[50] A. Frank and A. Lucky, "International digital organization for scientific research: Leveraging big data for effective educational management," Jan. 2025.

[51] J. Saldana, *The Coding Manual for Qualitative Researchers.* London, UK: SAGE Publications Ltd, 2009. First published 2009.

Appendices

Appendix A: Mitigation Strategies for Methodological Constraints.

This appendix outlines the measures taken to address the methodological constraints identified in Chapter Three. These strategies were implemented to ensure data reliability, analytical rigor, and ethical compliance in the study.

Table 6.1: Constraints and Mitigation Strategies

Constraint	**Mitigation Strategy**
Data Availability	The limited number of Uganda-specific job postings was supplemented with regional and international datasets from Kaggle. Metadata indicated that postings were aggregated from platforms such as LinkedIn and Indeed. Additionally, government and institutional labor reports were consulted to provide context for the trends.
Data Quality	A comprehensive data cleaning pipeline was employed, including removing duplicates, normalizing categorical fields, and consistent formatting. Text preprocessing involved tokenization, lowercasing, and removing stop words to prepare the data for machine learning models.
Bias	Potential urban and formal sector bias was mitigated by acknowledging the dataset's limitations and comparing the findings with the existing literature on informal labor trends in Uganda. Interpretive caution was applied in generalizing findings.
Technological Limitations	The study utilized pre-collected, static datasets from Kaggle, which were accompanied by clear documentation and version control, to minimize disruptions caused by dynamic web structures. This ensured reproducibility and transparency in data sourcing.
Legal & Ethical Concerns	The study adhered to guidelines for using secondary data, ensuring compliance with Kaggle's licensing terms and the privacy policies of the source platforms. All data was anonymized and used for academic purposes only.
Text Data Complexity	Natural Language Processing (NLP) methods, including TF-IDF vectorization and Named Entity Recognition (NER), were used to extract structured variables from unstructured job descriptions. This improved the quality and interpretability of the dataset.

Continued on next page

Constraint	Mitigation Strategy
Resource Constraints	The study prioritized models that strike a balance between accuracy and computational feasibility to maintain efficiency, such as ARIMA and Holt-Winters for forecasting, and K-Means for clustering. Feature selection was performed to reduce dimensionality and focus on variables with high impact.

These mitigation strategies addressed specific limitations during the analysis and enhanced the generalizability and credibility of the study's findings. This approach aligns with Creswell's [42] and Saldana's [51] recommendations for research practices, which emphasize methodological transparency and adaptive strategies in data-intensive research. The methodological constraints and their corresponding mitigation strategies outlined here will be revisited in Chapter Five to support the discussion on research limitations and proposed future directions. Researchers intending to replicate or extend this work are encouraged to consider and refine these approaches in light of evolving data availability and technological advancements. For further methodological guidance, researchers may refer to Creswell's principles of robust research design, Silverman's perspectives on data quality and ethical issues in qualitative research, and documentation from Kaggle on data sourcing and compliance.

UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

SCHOOL OF RESEARCH & POSTGRADUATE STUDIES
DISSERTATION CORRECTION COMPLIANCE FORM (POST VIVA FORM)

Date: **30th June 2025**

Name of Candidate: **Wanyama Denis** Reg.No: **J23MD19/216**

Title of Dissertation: **Forecasting Emerging Skill Demands with Machine Learning to Inform Curriculum Development in Uganda's Higher Education.**

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	The candidate should clearly show how each objective is achieved	Each objective has been explicitly addressed in the methodology (Chapter 3), results (Chapter 4), and discussion (Chapter 5). **Objective 1** (identify emerging skills) is achieved through data collection and EDA. **Objective 2** (forecasting skill demands) is achieved through ARIMA and Holt-Winters modelling. **Objective 3** (generate actionable insights) is achieved through curriculum recommendations.	Chapters 3–6 (page.14–37).
2	Clearly show the relationship between objectives and the methods used to achieve the objectives.	A direct link between objectives and methods is clarified in. **Chapter 3**. Clustering (Section 3.11.1, p.20) addresses emerging skill identification (Objective 1). Time series forecasting using ARIMA and Holt-Winters (Sections 3.8–3.8.4, pp.15–18) addresses predictive analysis (Objective 2). Recommendations (Section 3.9, p.19; Chapter 6, pp.36–37) address curriculum alignment (Objective 3).	Chapter 3 (page 15–20), Chapter 6 (page 36–37).
3	The literature should be always linked to the purpose and rationale of the study	The literature review has been revised to show clear connections to the research problem and rationale. For example, the **UBOS National Labor Force Survey (2021)** (Section 2.1.1, p.5) demonstrates the mismatch in skills and directly	Chapter 2 (pp.5-7), linked to Chapter 1 (pp.1-3).

		supports the study rationale in the Problem Statement (Section 1.3, p.2). Similarly, global and regional studies on data-driven curriculum forecasting (Table 2.1, p.6) are tied to the purpose of applying ML for skill prediction.	

S/N	COMMENTS BY INTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Chapters should be renamed as per the UCU research manual.	All chapters have been renamed and restructured to conform to the UCU research manual. For example, Chapter One is now “General Introduction,” Chapter Four is “Data Analysis, Presentation and Interpretation of Results,”, Chapter Five is “Discussion of Results, and Chapter Six is “Conclusions and Recommendations.”	Revised chapter titles in the Table of Contents and within the main text.
2			
3			

S/N	COMMENTS BY VIVA VOCE PANNEL	ACTION TAKEN	INDICATOR
1	Reports should be formatted/typeset using LaTeX.	The entire thesis has been reformatted and typeset in LaTeX, following departmental guidelines for structure, referencing, and formatting.	Final report submitted in LaTeX format
2			
3			

Candidate’s Name Signature

Wanyama Denis

Supervisor’s Name Signature

John B Wabwire Habere

NB: Post Viva compliance form is designed to capture all the corrections recommended by internal examiner (supervisor), external examiner and viva panel.