AN INTELLIGENT MODEL FOR ALIGNING ACADEMIC PROGRAMS TO ESSENTIAL DIGITAL SKILLS IN THE LABOUR MARKET

DUNCAN K. NYALE

A thesis submitted to the Department of Computer Science and Information Technology in the School of Computing and Mathematics in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Information Systems of the Cooperative University of Kenya.

November, 2025

DECLARATION

Declaration by the candidate

This thesis is my original work and has not been presented for award of a degree in any other University or for any other award.

Signature: .... Date: ....20/11/2025.................................

DUNCAN K. NYALE

PISC01/7422/2021

Declaration by the supervisors

We confirm that the work reported in this thesis was carried out by the candidate under our supervision and has been submitted with our approval as university supervisors.

Signature: .... Date: ....20/11/2025.................................

Prof. Simon Karume, Department of Computer Science & Information Technology, School of Computing and Mathematics, Co-operative University of Kenya

Signature: .... Date: ....20/11/2025.................................

Dr. Andrew Kipkebut, Department of Computer Science & Information Technology, School of Science Engineering and Technology, Kabarak University

Signature: .... Date: .... 20/11/2025...............................

Dr. Fidelis Mukudi, Department of Mathematical Sciences, School of Computing and Mathematics, Co-operative University of Kenya

DEDICATION

I dedicate my study:

To GOD almighty without WHOM none of this would be possible.

To my parents for always being there for me and for always wishing the very best for me.

To my wife for her unwavering love, support and patience.

To my kids; may you grow up to enjoy the perfect will and destiny of GOD in your life.

To my brothers and sisters, for your unflinching encouragement.

God bless you all!

ACKNOWLEDGMENTS

I would like to thank my supervisors, Prof. Simon Karume, Dr. Andrew Kipkebut and Dr. Fidelis Mukudi for their guidance and support throughout the course of this research. In addition, I would also like to thank my friends, colleagues, the department faculty and staff at the School of Computing and Mathematics for the encouragement and making this a wonderful experience. I want to also offer my appreciation to those who were willing to participate in studies, without whom, this study would not have been possible. Finally, I would like to thank the Cooperative University of Kenya for granting me the scholarship to pursue this academic study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

AI -	Artificial Intelligence
AIED -	Artificial Intelligence in Education
AIEd -	Artificial Intelligence in Education
APEC -	Asia-Pacific Economic Cooperation
BoW -	Bag-of-Words
CAI -	Computer-Assisted Instruction
CBT -	Computer-Based Training
CMO -	Chief Marketing Officer
CPU -	Central Processing Unit
CS -	Computer Science
CV -	Cross-Validation
DSI -	Digital Skills Index / Curriculum Digital Skills Index
DSGI -	Digital Skills Gap Index
DSR -	Design Science Research
EIT -	European Institute of Innovation and Technology
EU -	European Union
GDP -	Gross Domestic Product
GPU -	Graphics Processing Unit
GridSearchCV -	Grid Search with Cross-Validation
HTML -	Hypertext Markup Language
IBM -	International Business Machines Corporation
ICT -	Information and Communications Technology
ILO -	International Labour Organization

IR 4.0 -	Industrial Revolution 4.0
IS -	Information Systems
IT -	Information Technology
IoT -	Internet of Things
ITU -	International Telecommunication Union
LMS -	Learning Management System
ML -	Machine Learning
MOOC -	Massive Open Online Courses
MSE -	Mean Squared Error
NER -	Named Entity Recognition
NLP -	Natural Language Processing
NLTK -	Natural Language Toolkit
O*NET -	Online Occupational Network
OECD -	Organization for Economic Co-operation and Development
PDF -	Portable Document Format
RAM -	Random Access Memory
regex -	Regular Expression
RQ -	Research Question
SEO -	Search Engine Optimization
spaCy -	Natural Language Processing Library
SQL -	Structured Query Language
STEM -	Science, Technology, Engineering and Mathematics
TF-IDF -	Term Frequency-Inverse Document Frequency
UNESCO -	United Nations Educational, Scientific and Cultural Organization
XAI -	Explainable Artificial Intelligence

CONCEPTUAL AND OPERATIONAL DEFINITION OF TERMS

Term	Conceptual Definition	Operational Definition
Model	A computer program that simulates or builds a prototype of a real-world feature, phenomenon, or event. This study uses a model to demonstrate and validate that developing a framework to match academic curricula to business needs is feasible.	The intelligent curriculum alignment model developed in this study, integrating web scraping, NLP, and machine learning to analyze and map digital skills.
Academic Programs	The core, required, and elective courses that lead to a degree or certificate.	The set of courses and modules offered by universities in Computer Science and Information Systems, as documented in official curriculum documents.
Digital Skills	A variety of skills for managing and accessing information via networks, digital devices, and communication applications.	Specific competencies extracted from curriculum documents and job postings, such as programming, data analysis, cybersecurity, and digital literacy.
Labour Market	Digital services industry, with the workers or labour providing the digital skills that employers demand.	The pool of job postings and skill requirements collected from technology recruiting platform (i.e., Dice.com) and analyzed for digital skill demands.
Employability	The attributes of a person that make that person able to gain and maintain employment.	The proportion of curriculum skills that match high-demand market skills, measured by the Curriculum Digital Skills Index (DSI).
Digital Enterprise	A company that has finished its digital transformation and fully integrated digital tools and technologies into all facets of its	Organizations that require advanced digital skills for their

	activities, from conception, realization to utilization.	operations, as reflected in job postings and skill requirements.
Digital Proficiency	The ability to use technology in order to complete a task. Digital proficiency enables people to effectively choose the right software or technology for completing a project or task.	The level of skill mastery demonstrated by graduates, measured by the presence and frequency of digital skills in curriculum documents.
Skills Gap	The disparity between the skills an employer expects their employees to have and the actual skills employees possess. This mismatch makes it challenging for employers to fill open positions.	The difference between the digital skills required by employers (from job postings) and those taught in academic programs (from curriculum documents).
Design Science Research (DSR)	A problem-solving methodology that emphasizes the creation, development, and rigorous evaluation of innovative artifacts designed to address real-world organizational and technical challenges.	The research approach used in this study, involving the development and validation of the intelligent curriculum alignment model.
Natural Language Processing (NLP)	A branch of artificial intelligence that enables computers to understand, interpret, and generate human language.	The pipeline of text processing techniques (tokenization, entity recognition, TF-IDF vectorization) used to extract digital skills from curriculum documents.
Decision Tree	A machine learning algorithm that uses a tree-like model of decisions and their possible consequences.	The supervised learning model trained to classify and map curriculum skills against market demands, using features such as skill frequency and entity count.
NoSQL Database	A type of database that stores and retrieves data in a non-relational format, suitable for handling large volumes of unstructured data.	The Chroma database used to store and manage the skillset database, including digital skills and their attributes.

Curriculum Digital Skills Index (DSI)	A quantitative metric indicating the relevance of curriculum digital skills to market demands.	The score calculated by the model, representing the proportion of curriculum skills that match high-demand market skills.
Prescriptive Analytics Report	A document that provides actionable recommendations for curriculum improvement based on data analysis.	The output generated by the model, suggesting specific courses or modules to address identified skill gaps.
Web Scraping	The process of automatically extracting data from websites.	The custom Python scripts used to collect job postings and skill requirements from Dice.com.
Machine Learning	A subset of artificial intelligence that enables computers to learn from data and make predictions or decisions without being explicitly programmed.	The algorithms (i.e., Decision Tree) used to classify and map digital skills, trained on labeled datasets.
Artifact	A tangible product or system created to solve a specific problem.	The intelligent curriculum alignment model, including its components and outputs.
Validation	The process of ensuring that a model or system performs as intended and produces reliable results.	The evaluation of the model using performance metrics (accuracy, precision, recall, F1-score, MSE) and cross-validation techniques.
Stakeholder	An individual or group with an interest in the outcomes of a project or research.	Academic institutions, industry partners, policymakers, and students affected by the curriculum alignment process.
Ethical Clearance	Approval from an ethics committee to conduct research involving human subjects or sensitive data.	The formal approval obtained from the university's ethics committee to collect and analyze curriculum and job posting data.

ABSTRACT

This thesis addresses the critical challenge of persistent misalignment between academic digital skills offerings and evolving industry demands, worsened by the absence of a comprehensive, standardized framework for alignment. The research's primary objective was to develop an intelligent curriculum alignment model that systematically analyzes university curricula, identifies embedded digital competencies, and maps these to dynamic labour market digital skill requirements. Methodologically, the model integrates advanced Artificial Intelligence techniques: web scraping was employed to dynamically acquire real-time labor market data, resulting in a skillset database containing 9,077 unique digital skills extracted from 46,514 job postings; Natural Language Processing (NLP) facilitated detailed analysis of curriculum documents to extract relevant digital skills; and a supervised Decision Tree machine learning algorithm was trained to classify and map curriculum skills against market demands, offering prescriptive recommendations for curriculum improvement. The developed model's performance was rigorously validated on an independent test set, achieving a high overall accuracy of 88%, with an exceptional recall of 99% for high-frequency, in-demand skills, and balanced F1-scores between 0.88 and 0.89 across skill categories. These results demonstrate the model's reliability in capturing critical digital competencies aligned with industry needs. Theoretically, this research advances scholarship by presenting a novel hybrid AI framework for continuous curriculum adaptation, and the creation of a dynamic, data-driven digital skills ontology that moves beyond static taxonomies prevalent in prior studies. Practically, it offers academic institutions actionable intelligence through the Curriculum Digital Skills Index and Prescriptive Analytics Reports, enabling evidence-based curriculum development that enhances graduate employability, fosters academia-industry collaboration, and supports national digital transformation strategies. The study addresses methodological scope by detailing data acquisition, preprocessing, model construction, and evaluation processes, while acknowledging limitations including geographic focus, ethical considerations in AI deployment, labour market data constraints, and institutional adoption pace. Recommendations for future research emphasize enhancing model capabilities with broader datasets, soft skills inclusion, ethical AI integration, and longitudinal impact studies. Overall, this thesis represents a paradigm shift from manual, reactive curriculum review to a proactive, AI-driven, and continuously evolving approach. It equips educational institutions with a scientifically validated toolset to close the digital skills gap, thus ensuring higher education's relevance and responsiveness to the fast-changing digital economy.

CHAPTER ONE
INTRODUCTION

1.1 Background to the Study

The demand for workers with advanced digital abilities is rising as a result of the quick development of technology and the digital transformation of industries (ILO, 2021). Academic programs must adapt to the needs of the digital economy in order to guarantee that graduates have the skills needed to meet the changing demands of the job market. (Rovira et al., 2021). This research aims to develop a comprehensive model that facilitates the integration of digital skills into academic curricula, ultimately enhancing graduates' employability. The digital enterprise refers to the integration of technology into various aspects of business operations. As a result, employers place a high value on graduates with excellent digital abilities because they can assist businesses in navigating the complexity of the digital ecosystem (Kraus et al., 2022).

According to Astapciks, (2023), the digital enterprise is rapidly transforming the way businesses operate. Businesses must be able to use digital technology to boost their production, efficiency, and customer service if they want to stay competitive. The skills necessary for employment are also being significantly impacted by this transition. Employers are searching for graduates with strong digital capabilities in today's digital economy. The digital abilities that graduates possess and the talents that employers are looking for are diverging more and more, according to Moldoveanu et al., (2022) this is due to a number of factors, including: academic programs find it challenging to stay up to date with the most recent developments due to the quick speed of technological change; the mismatch between academic curricula and the demands of the digital economy and the absence of funding for investing in digital skills training at academic institutions

1.1.1 Aligning Academic Programs to the Digital Enterprise

Academic institutions must take a methodical approach to matching their curriculum with the demands of the digital economy in order to close the digital skills gap (Moldoveanu et al., 2022). This entails integrating digital competences across disciplines, redesigning the curriculum, developing the faculty, and working with industry partners (Dutta et al., 2020; EIT, 2022; Bhorat et al., 2023). The proposed model will provide a structured methodology for universities to implement these changes effectively.

Digital skills encompass a broad range of competencies related to utilizing digital technologies effectively. Employers prioritize applicants with digital literacy, data analysis, programming, information management, and cybersecurity abilities more and more in the current employment market (ITU, 2020; OECD, 2021a). Studies, however, indicate that there is a large disparity between graduates' skill sets and what employers in the digital era are looking for (Bhorat et al., 2023; Mamabolo & Myres, 2020). Universities must therefore modify their academic programs to close this gap and encourage the growth of graduates' digital skills. By integrating digital enterprise into academic programs, graduates will have the necessary knowledge and abilities to succeed in the employment market. According to studies, graduates with good digital abilities have better chances of finding employment and have higher earning potential (Kraus et al., 2022; Fraillon et al., 2020). Employers also appreciate graduates who can adopt new technology and support organizational efforts to shift digitally (Astapciks 2023; ITU, 2020). As a result, improving graduates' digital abilities through a curriculum that is in line with industry needs can greatly improve their employability.

1.1.2 Levels of Digital Proficiency

a) Basic Level Digital Skills

Basic digital skills are the fundamental abilities needed to effectively use and manage digital technology, which are essential for participation in the digital world. These skills include a particular set of proficiencies, such as Internet and Web Browsing, which is the capacity to view websites, do information searches, and traverse web pages (OECD, 2021a; Techopedia, 2020). Another key skill is Email Communication, which involves the use of email clients or web-based platforms to create, send, receive, attach files, manage contacts, and manage emails (IBM, n.d; Fraillon et al., 2020). Furthermore, basic digital skills encompass Word Processing, defined as proficiency in using word processing software to create, edit, and format documents (Bashay, 2020). Lastly, File Management is necessary, which is the basic knowledge of organizing, creating, and managing files and folders on computers or cloud storage platforms (Digital Unite, 2021).

b) Intermediate Level Digital Skills

The intermediate level of digital competency expands upon fundamental skills, allowing users to handle more difficult tasks and utilize a wider array of digital tools. These skills are crucial for boosting productivity and communication across various environments. They include several key areas: Social Media Management involves having working knowledge of how to use social networking sites like Facebook, X (formerly Twitter), and Instagram for personal or business goals, encompassing creating posts, interacting with others, and understanding privacy settings (ITU, 2020). Multimedia Editing requires proficiency in editing and enhancing image, audio, and video files using appropriate software (GCFGlobal, 2019). Furthermore, Digital/Online Collaboration is the capacity to work effectively with others online using platforms and methods such as virtual meeting software and cloud-based document sharing (Flowtrace, 2023; United Nations, 2020). Another vital intermediate skill is Data Security and Privacy, which is the knowledge of best practices for safeguarding personal information online, comprehending privacy settings, and recognizing common cybersecurity threats (Zahiroh, 2020). Finally, Data Analysis involves the ability to collect, interpret, and analyze data using spreadsheet software or dedicated data analysis tools (Volodymyr, 2022).

c) Advanced Level Digital Skills

This level describes an advanced level of competence and knowledge in using digital technologies. These sophisticated skills are often tailored to specific sectors or businesses, enabling individuals to utilize technology for complex tasks and specialized problem-solving. This includes several specialized areas: Programming and Coding involves possessing the necessary skills to create software, websites, or mobile applications (Bhorat et al., 2023; ITU, 2020). Cybersecurity is expertise in implementing measures to protect digital systems and networks from security threats while ensuring data privacy (Cabello, 2023). Data Visualization requires proficiency in using tools like Microsoft Excel, Tableau, or Python to analyze and interpret data, create visualizations, and ultimately draw actionable insights (Barany, 2022; Chapman, 2019). Finally, this advanced tier also covers Artificial Intelligence and Machine Learning, which is the understanding of the principles and applications of AI and machine learning algorithms, including niche areas such as natural language processing and computer vision (Hasan et al., 2023, Goyal et al., 2021).

1.2 Statement of the Problem

There is a critical gap between digital skills taught in academic programs and those demanded by the labour market, worsened by the absence of effective models to align curricula with evolving digital enterprise needs.

Kenya Digital Masterplan 2022-2032 highlights several observable gaps, including non-alignment to industry needs, poor manpower planning for digital skills, low ICT literacy among citizens, businesses, and employees, and lack of awareness of digital skills among citizens

(Ministry of ICT, Innovation and Youth Affairs, 2019). A World Bank report noted the need to address the gap between employability and the tertiary education system in African nations, such as Kenya. Additionally, the report showed that identical courses specialize in distinct competences at several universities in Kenya, highlighting the need to map curricular to market needs and synchronize competencies in all related courses (World Bank, 2019). Another study found out that most African nations, including Kenya, are dealing with the issue of the increasing gap between employability and the tertiary education system (ILO, 2021).

The Fourth Industrial Revolution has given rise to new and growing digital skills that are rising in demand in the employment market, according to a research by the World Economic Forum, (2023). However, academic programs take time to adjust to these changes, causing a mismatch between the graduates' existing skills and the skills that companies are looking for. Since it stifles innovation and economic growth, this mismatch has been recognized as a serious problem for both graduates and companies.

This problem has been addressed in a number of studies by putting up various strategies for matching academic curricula with developing digital skills. These methods' reliance on manual procedures and arbitrary evaluations, however, frequently imposes limitations on their effectiveness. For instance, a study by Hol et al., (2023) suggested a model for matching academic programs with market demands, however the strategy depended on expert judgment and surveys, which might be biased and subjective.

The demand for graduates possessing advanced digital skills has witnessed a significant increase (ITU, 2021). Nevertheless, numerous academic programs are currently facing challenges in aligning their curriculum with the requirements of the digital enterprise (Harry, 2021). This lack of alignment poses a hindrance to the employability of graduates and constrains their potential for success in the job market (Ma'dan et al, 2020). Existing literature emphasizes the vital role of digital skills in augmenting graduates' employability and enabling them to thrive in the digital economy (ITU, 2021). However, there is a notable scarcity of comprehensive models that offer guidance on effectively aligning academic programs with the demands of the digital enterprise (Moldoveanu et al., 2022). This research therefore focusses on the lack of alignment of curriculums to the digital skills requirements of the labour market.

1.3 Objectives of the Study

1.3.1 Broad Objective

The research aim was to develop an intelligent curriculum alignment model capable of analyzing university curricula to accurately identify prevailing digital skills and effectively map them to the essential digital skill requirements of the labour market.

1.3.2 Specific Objectives

i. To examine the current state of digital skills in academic programs and explore the gaps between industry demands and graduates' competencies.
ii. To design an intelligent curriculum alignment model for aligning academic programs to the digital enterprise.
iii. To implement a prototype of the intelligent model for aligning academic programs to the digital enterprise.
iv. To validate the curriculum alignment model.

1.4 Research Questions

i. What is the current state of digital skills in academic programs and what are the gaps between industry demands and graduates' competencies?
ii. How can an intelligent curriculum alignment model be effectively designed?
iii. How can an intelligent curriculum alignment prototype be developed to efficiently align academic programs with the essential needs and demands of the digital enterprise?
iv. How valid and reliable is the intelligent curriculum alignment model?

1.5 Justification for the Study

This study is justified by the critical and persistent gap between the digital skills offered through academic programs and the competencies demanded by the labor market in the era of rapid technological advancement. Existing curricula often lag behind industry needs, limiting graduate employability and hindering economic growth. There is a clear need for a dynamic, data-driven framework that can align university curricula with evolving digital skill requirements. By leveraging artificial intelligence to analyze real-time labor market data and academic content, this research addresses a pressing challenge faced by educational institutions worldwide. The study aims to fill this gap by developing an intelligent curriculum alignment model, thereby contributing to more relevant educational programs and ultimately supporting workforce readiness in the digital age.

1.6 Significance of the study

The study's significance lies in its potential to transform how academic institutions respond to labor market demands. By integrating AI techniques such as machine learning and natural language processing, the proposed model offers a scalable, adaptable solution for ongoing curriculum improvement. This can enhance graduate employability by ensuring students acquire skills that are current and in demand. Additionally, the framework facilitates stronger academia-industry collaboration by providing actionable insights based on labor market analytics. The research also supports broader national and global digital transformation goals by preparing a workforce equipped for the challenges and opportunities of the digital economy. Overall, the study advances theoretical knowledge in AI-driven educational alignment while delivering practical tools for policymakers, educators, and industry stakeholders.

1.7 Scope of the Study

The scope of this study is confined to the evaluation of digital skill levels in academic programs offered by institutions of higher learning. It encompasses programs within selected universities and colleges and focuses on the alignment of these programs with the requirements of the contemporary labour market. The boundaries of the study are limited to digital skills relevant to current employment trends on a global scale, excluding other skill domains or non-digital skill assessments. The investigation is geographically and institutionally bounded to the chosen sample of higher education institutions.

1.8 Expected Outcomes of the Study

The expected results of this study were designed to deliver three specific, actionable outcomes. First, the study aimed to produce a skillset database detailing the essential digital skills currently in demand within the labor market. Second, it sought the identification of digital skill gaps existing within academic curricula, followed by suggestions for new courses and programs to effectively address these deficiencies. Finally, the research expected to create an intelligent model capable of accurately aligning academic curricula with the essential digital skills required by the labor market.

1.9 Limitations and Delimitations of the Study

1.9.1 Limitations

The study's geographic focus may limit the generalizability of the findings to other regions with different labor market dynamics and educational systems.

Data quality and availability from labor market sources may impact the accuracy and comprehensiveness of the skill analysis.

The AI model's recommendations depend on the quality and representativeness of input data, which might not capture all emerging or niche skills.

Ethical considerations regarding AI use, including bias in data and algorithmic decisions, may affect the fairness and acceptance of the model.

The study is constrained by the pace at which academic institutions can implement recommended curriculum changes, which may lag behind rapid labor market shifts.

1.9.2 Delimitations

The research focuses explicitly on digital skills within university curricula, excluding broader non-digital competencies or vocational training programs.

The model targets alignment between academic programs and digital labor market demands, not addressing other factors influencing employability.

The study prioritizes machine learning techniques such as Decision Trees, NLP, and web scraping for data collection and analysis, without exploring alternative AI methods.

The labor market data used is drawn primarily from online job postings, which may omit informal or emerging employment trends.

The research timeframe and data collection are limited to a defined period, reflecting a snapshot rather than continuous real-time labor market changes.

1.10 Assumptions of the Study

The labor market rapidly evolves due to technological advances and new business models, requiring academic programs to adapt continuously.

There is a persistent mismatch between academic curricula and industry needs because of limited industry insight, insufficient funding, and slow curriculum updates.

AI can bridge the skills gap by analyzing labor market data to identify emerging digital skills and recommend curriculum changes.

The intelligent model needs continuous refinement through ongoing data analysis and collaboration with industry stakeholders to remain relevant.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

The need for a trained workforce with growing digital skills has become crucial for long-term economic growth and organizational success in the rapidly evolving digital ecosystem. Educational institutions must make sure that their academic programs align with the constantly shifting needs of the labour market as new technologies grow and shape various industries. In order to bridge the gap between academia and business, this methodical literature review delves into the research and advancements around the application of intelligent models for the alignment of academic curricula to the dynamic digital skills desired by companies. By examining the existing body of knowledge, important insights are obtained into the efficacy and potential of using intelligent algorithms to optimize curriculum design, anticipate industry expectations, and create a workforce that is adaptable and prepared to survive in the digital era. This review aims to foster a more adaptable, responsive, and harmonious approach to addressing the challenges presented by the labour market's quickly changing demands for digital skills in order to advance decision-making about education policy and the ongoing dialogue between academia, industry stakeholders, and policymakers.

2.2 Digital Skill Categories

Digital skills are essential in today's rapidly evolving technological landscape. They encompass a wide range of abilities and knowledge required to effectively navigate and utilize digital tools and technologies. From basic computer literacy to advanced coding and data analysis, digital skills are increasingly becoming a necessity in various industries.

2.2.1 Basic Digital Literacy Skills

To use digital technology efficiently, one needs to have a basic understanding of them. These abilities include knowledge of the fundamentals of using a computer, managing files, using the internet, communicating via email, and using word processing, spreadsheets, and presentation tools effectively. These abilities are necessary for people to function effectively in the digital environment and carry out daily duties (OECD, 2021a; Fraillon et al., 2020).

2.2.2 Information Literacy and Digital Research Skills

The ability to evaluate, analyze, and use digital information is essential given the wealth of information available online. Conducting efficient web searches, assessing sources for authenticity and dependability, and critically analyzing information are all part of information literacy abilities. People with effective digital research abilities can efficiently collect, arrange, and synthesize data from a variety of digital sources to aid in decision-making and problem-solving (Bashay, 2020; ITU, 2020).

2.2.3 Cybersecurity and Online Safety Skills

Cybersecurity and online safety knowledge are crucial in a society that is becoming more digital and linked. These abilities include being aware of typical internet dangers, utilizing secure passwords, and maintaining safe online habits. People who are skilled in cybersecurity can recognize potential security breaches and take appropriate action, reduce risks, and protect digital assets (Cabello, 2023; Zahiroh, 2020).

2.2.4 Coding and Programming Skills

In today's digital economy, coding and programming talents are highly prized. People who are proficient in programming languages like Python, Java, and JavaScript can make websites, build software programs, and automate operations. In the disciplines of software development, web development, data analysis, and artificial intelligence (AI), these abilities are very important (Bhorat et al., 2023; ITU, 2020).

2.2.5 Data Analysis and Interpretation Skills

The capacity to evaluate and comprehend data has grown essential as firms depend more and more on data-driven decision-making. Data analysis abilities involve gathering, purifying, and converting data, employing statistical methods, and successfully communicating insights through the use of data visualization tools. Marketing, finance, and healthcare are just a few of the sectors that require proficiency with tools like Microsoft Excel, SQL, and data visualization platforms like Tableau or Power BI (EIT, 2022; ITU, 2020).

2.2.6 Digital Marketing Skills

As businesses look to improve their online presence and communicate with clients effectively, digital marketing skills are in great demand. These abilities include social media marketing, SEO, content development, email marketing, and online advertising. For marketing initiatives to be effective, users must be proficient in digital marketing platforms and tools including Google Analytics, social media management tools, and email marketing software (Hartman, 2020).

2.2.7 Artificial Intelligence (AI) and Machine Learning Skills

The importance of AI and machine learning skills has grown as a result of their increasing integration in numerous businesses. Individuals that are skilled in deep learning, natural language processing (NLP), neural networks, and machine learning algorithms may create AI-powered solutions, automate procedures, and extract knowledge from huge datasets (Botvinick et al., 2020; ITU, 2020).

2.2.8 Cloud Computing

This includes the capability of storing, managing, and processing data on remote servers. Additionally, it entails the capability of managing and deploying software applications on cloud infrastructures like Microsoft Azure and Amazon Web Services (Amazon, 2023).

2.2.9 Internet of Things (IoT)

This involves the ability to create and oversee connected devices that can communicate with other devices online. It also includes the capacity to develop and deploy IoT solutions for multiple sectors, including manufacturing and healthcare (IBM, 2023; ITU, 2020).

2.2.10 Virtual and Augmented Reality

Utilizing virtual reality and augmented reality technologies entails the capacity to develop immersive experiences. Additionally, it has the ability to develop programs for a variety of industries, such as gaming, education, and healthcare (Rauschnabel et al., 2022; Unity Technologies, 2019).

2.3 Digital Skills Landscape

2.3.1 Global Perspective

The Digital Skills Gap Index (DSGI), developed by Wiley Research (2021), ranks 134 economies based on factors including employer needs and employee skills, the gender gap in STEM fields, and academic research output. It aims to narrow the digital skills gap. It challenges leaders in business, education, and public policy to overcome the difficulties of bridging the digital skills gap. The Index revealed that problem-solving is the most important soft talent for the 21st century, followed by algorithms and data analytics. However, nearly half of respondents believe workers are incapable of handling data ethically. The largest digital skills gap among APEC economies is in education and training, with 60.6% of respondents stating that available programs don't match demands. A significant gender disparity in STEM disciplines is acknowledged. Major obstacles to reducing the gap include lack of digital knowledge and opposition to up- and re-skilling.

2.3.2 Regional Perspective

The African Growth Initiative at Brookings, in partnership with the Development Policy Research Initiative, published a report titled Digitalization and digital skills gaps in Africa: An empirical profile, produced by Bhorat et al. (2023). According to the paper, while Africa is quickly digitizing, there is a major digital skills gap. Major insights are that Africa's digitalization gap index is 0.42, with only 25% possessing essential digital skills, particularly in rural areas. Women are more affected, with only 20% having the necessary abilities. Only 36% accept digital payments and 8% conduct online transactions.

Additionally, an earlier paper titled Competing in a Digital Age: The development of IT Skills and Jobs in Kenya and Uganda by Mercy Corps and Moringa School (2020) highlighted the significant skills gap in the IT sector in Kenya and Uganda. The article emphasizes Kenya's shortage of IT specialists despite the possibility of digital opportunities. A significant factor in Kenya's lack of IT skills is the quality of education offered in state colleges. Through innovation centers, specialized training facilities, online platforms, and universities like Strathmore, this problem is being addressed. Employers, however, continue to have trouble selecting the best people and effectively using them. Additionally, the study notes that by 2022, the market for technology-related higher education is anticipated to be worth $51.54 million, and 95,000 IT workers are anticipated by the year's end. The gap is a result of things like underinvestment in IT education, a mismatch between skills learned in school and those needed in the IT industry, and a lack of access to digital infrastructure in rural areas. The study suggests policy changes to solve this, including funding IT education, curriculum reform, funding for digital infrastructure, and fostering an environment that encourages entrepreneurship.

2.3.3 Local Perspective

The ICT talent cultivation for Kenya's digital economy is a whitepaper developed by UNESCO and Huawei. (2022) to support Kenyan policymakers in developing a country-level strategy for digital talent cultivation, analyzing the need, status, and proposing a new approach to meet challenges and meet the digital economy's demands. The paper predicts that the digital economy is to increase the global GDP from 15.5% in 2016 to 25% in 2026, with 50-55% of Kenyan occupations requiring digital skills by 2030. Kenya's digital economy is expected to contribute 9.24% of its GDP by 2025, compared to Africa's 5.2%. However, the growing ICT gap is due to low enrollment in intermediate and advanced courses.

Additionally, the top 20 digital skills in need in Kenya, according to a 2019 World Bank assessment, are in the following fields: software, ICT, programming, Java, Python, Script, JavaScript, SQL, data, network, servers, backend, frontend, cloud computing, cloud, HTML, agile (World Bank, 2019).

2.4 Related Literature

2.4.1 Current State of Digital Skills in Academic Programs and the Gaps between Industry Demands and Graduates' Competencies

In order to meet industry needs, academic institutions are expected to equip their graduates with the digital skills that are increasingly important in today's workforce. This assessment of the literature seeks to determine the discrepancies between graduates' competences and industry demands by looking at the level of digital skills in academic programs today.

a) Technology Integration in Academic Programs

The state of digital skills in academic programs has been the subject of several studies in recent years. According to these findings, there is a considerable difference between the digital abilities that are taught in academic programs and those that companies want. According to a World Economic Forum report, 60% of businesses feel that graduates do not have the necessary digital skills (World Economic Forum, 2023). According to another report by the Society for Human Resource Management (2021), just 70% of companies can locate competent individuals who possess the necessary digital skills.

In recent years, academic program technology integration has drawn the attention of researchers. In order to increase students' digital literacy and improve their employability prospects, higher education institutions have reportedly increasingly included digital tools and platforms into their curricula contend that many academic programs still lack thorough ways to guarantee students obtain the requisite digital skills demanded by the labour market despite such attempts (Mn et al., 2020).

The digital proficiency of graduates in diverse sectors has also been the subject of numerous studies. While graduates of education programs have some digital capabilities, González et al., (2021) discovered that there were still gaps in their capacity to use digital technologies for teaching and learning. While some graduates in the hotel business had great digital abilities, others lacked competency in things like data analysis and social media marketing, according to another study by Putra et al., (2022) that looked at the digital competencies of graduates in this sector.

b) Industry Demands for Digital Skills

Digital abilities are among the top ten skills needed for occupations in the future, according to a World Economic Forum survey (World Economic Forum, 2023). However, according to a McKinsey and Company poll, only 16% of CEOs think that their staff members have the requisite digital skills to keep up with the evolving business landscape (McKinsey and Company, 2021). This demonstrates how crucial academic institutions are to equipping graduates with digital skills. The discrepancies between industry demands and graduates' digital skill competencies have been shown in numerous studies. According to a study by the European Commission, there is a considerable disparity between the graduates' skills and the digital skills that businesses demand (European Commission, 2022). In a similar vein, an article in the Financial Times indicated that while just 15% of businesses feel school graduates have the required abilities, 80% of firms believe that digital skills are vital (Colback, 2023).

Industry demands for digital skills have changed dramatically in recent years. Plekhanov et al., (2022) draw attention to the rising demand for expertise in a variety of fields, including cybersecurity, artificial intelligence, and data analytics. Future Skills Centre, (2023) also stress the expanding demand for graduates skilled in social media management and digital marketing due to the expansion of online enterprises and marketing techniques.

c) Assessment of Graduates' Digital Competencies

A report by Asia Development Bank, (2023) revealed that a significant portion of graduates lacked critical digital skills, creating difficulties in the recruiting process. In contrast, Mn et al., (2020) used a systematic review to evaluate graduates' digital competencies and discovered that while some students excelled, a sizable portion had trouble using cutting-edge technological tools and apps. The rapid technological change is one of the causes of this disparity. In the next five years, 90% of employment will require some level of digital capabilities, and 50% of those jobs would require advanced digital skills, according to a study Li, (2022). To keep up with the shifting demands of the industry, education institutions must be quick to update their curricula.

The absence of cooperation between academic institutions and industry is another factor contributing to the gap. Only 23% of educators work with business partners to build curricula, according to a survey by IBM (Columbus, n.d.). Because of this lack of cooperation, academic institutions could not be up to date on current market trends and requirements.

d) Faculty Readiness and Training

One of the most important factors in bridging the gap between industry demands and graduate competencies is the faculty members' preparedness to teach students digital skills. According to Evans (2021), faculty development and support in incorporating technology into teaching approaches are essential to ensuring that students gain relevant and useful digital skills. Reddy et al. (2023) contend that many academics struggle to adopt new technology, which limits their capacity to impart digital skills to students in an efficient manner. Another study by Zan et al. (2021) discovered that business school faculty members' degrees of digital literacy varied, with some lacking expertise in using digital tools for teaching and research.

Lack of emphasis on digital skills in academic programs is one potential cause of the digital skills gaps between graduates and faculty members. According to a study by Zhao et al. (2021), certain academic programs have integrated digital skills into their curricula to a greater or lesser level. This demonstrates the necessity for academic programs to prioritize teaching digital skills more in order to ensure that graduates are ready for the demands of the workplace. Lack of coordination between academic curricula and industry expectations is another possible cause of the digital skills gap. According to a research by ITU (2021), university programs do not adequately teach the digital skills that employers demand. This shows that in order to make sure that graduates are equipped with the skills needed for success in the profession, academic programs need to collaborate more closely with industry partners.

e) Strategies for closing the Misalignment

Experiential learning possibilities have been advocated in various research as a way to close the knowledge gap between academic curricula and market demands. Incorporating experiential learning activities into an information management course, for instance, was found to assist students build digital abilities like data analysis and visualization.

Academic institutions have launched many measures to overcome these disparities. The Loughborough University London for instance, has created a digital skills program that instructs students (Loughborough University London, 2023). The University of Waterloo has created a program that incorporates digital skills throughout all undergraduate degrees in a similar manner (University of Waterloo, 2021).

2.4.2 Developing an Intelligent Model for Aligning Academic Programs to the Digital Enterprise

Education has been profoundly impacted by the digital transformation, which makes it necessary to integrate academic programs with the needs of the contemporary workforce. An exceptional opportunity to create a model that successfully matches academic curricula with the needs of the digital economy is provided by artificial intelligence (AI). This section examines pertinent studies and methods for faculty development, industry partnership, and curriculum design using AI as the main component. AI is able to analyze data, spot trends, and create recommendations, which aids academic institutions in creating curricula that caters to the needs of the digital sector, cultivating faculty with the necessary skills, and establishing connections with industrial partners.

a) Curriculum Design Using AI

Researchers have recently paid a lot of attention to the creation of an artificially intelligent model for curriculum design. According to a study by Kim et al. (2022) a model can be used to design a customized curriculum for each student by using machine learning algorithms to assess data from numerous sources, such as student feedback, learning outcomes, and market trends. The model enhanced student involvement and curriculum satisfaction, according to the study. Similar to this, Thongprasit and Wannapiroon, (2022) suggested an intelligent curriculum design system that creates tailored learning paths for pupils using data analytics and machine learning algorithms. To create a customized learning experience, the system takes into account a number of variables, including the student's history, hobbies, and professional aspirations. The method enhanced student performance and curriculum satisfaction, according to the study.

A crucial component of integrating academic programs with the digital industry is curriculum design. AI can help teachers comprehend market trends and integrate cutting-edge technologies into the curriculum. According to Sarker. (2021), machine learning algorithms can be used to examine industry data and pinpoint the crucial competencies needed by the digital enterprise. Academic institutions can successfully adapt their curricula to satisfy these demands by utilizing AI. A study by Crompton, (2023) employed AI to create a personalized learning system for learners. The technology was able to recognize the strengths and weaknesses of each learner and offer them individualized learning exercises.

Another investigation by Zhou et al. (2021) employed AI to create a Massive open online courses (MOOC) curriculum. The curriculum was created to respond to the demands and interests of the students.

Table 2.1 Summary of Research Studies on Application of AI in Academic Programs

S. No	Source	Focus	Methodology	Outcome
1	Kim et al. (2022)	Customized curriculum using an AI model	Machine learning algorithms to assess student data, enhance student involvement and satisfaction	Enhanced student involvement and curriculum satisfaction
2	Thongprasit and Wannapiroon, (2022)	Tailored learning paths using AI system	Data analytics and machine learning algorithms to personalize learning based on student history, hobbies, and aspirations. Improve student performance and satisfaction	Enhanced student performance and curriculum satisfaction
3	Sarker, (2021)	Integrating academic programs with digital industry	Machine learning to analyze industry data and identify crucial competencies. Adapt curriculum to meet industry demands	Successful adaptation of curricula to meet industry demands
4	Crompton, (2023)	Personalized learning system using AI	AI to identify learner strengths and weaknesses, provide individualized learning exercises	Individualized learning experiences for learners
5	Zhou et al., (2021)	AI in massive open online courses (MOOC) curriculum	AI to create a curriculum responsive to student needs and interests	Customized MOOC curriculum responding to student needs

The findings of this research suggest that AI has the potential to greatly enhance the alignment between academic curricula and the digital industry's needs. However, several challenges must be resolved before AI can be fully implemented in this educational setting. These difficulties include: the requirement for reliable student learning data; the need for efficient data analysis algorithms alongside the need to guarantee that AI systems are both impartial and fair; and finally, the requirement to create practical methods for incorporating AI into the curriculum.

b) Faculty Development and AI Integration

Researchers have paid a lot of attention to the creation of an artificially intelligent model for faculty growth. An intelligent faculty development system was developed by Swiecki et al. (2022) in their study. It employs machine learning algorithms to assess data from many sources, such as student feedback and faculty performance, to pinpoint areas that need improvement and offer tailored training programs. According to the study, the method enhanced instructor performance and program satisfaction.

Similar to this, Hooda et al. (2022) proposed an intelligent faculty assessment system that assesses teacher performance and offers tailored feedback using data analytics and machine learning algorithms. The approach provides a thorough evaluation by taking into account a number of variables, such as teaching efficacy, research production, and service to the institution. According to the study, the approach enhanced instructor performance and evaluation process satisfaction.

Faculty members that are knowledgeable with AI and its applications are needed to integrate the technology into academic curricula. Ng et al, (2023) look into how faculty development programs might improve instructors' AI literacy. Their study highlights the necessity for specialized training and resources to enable professors to successfully use AI tools into their teaching approaches. AI can also help in faculty development. For instance, Seo et al.'s (2021) work used AI to design a customized academic coaching system. The use of technology allowed instructors to receive feedback on their methods of training and help them improve their delivery. A faculty development program for MOOC teachers was created using AI in another work by Zhou et al. (2021). The application was created to assist teachers in producing high-quality, engaging MOOCs.

c) Industry Collaboration and AI Models

Researchers have studied the creation of an artificially intelligent model for industry collaboration. According to a study by Halder and Saha. (2020), an intelligent model for industry collaboration can be created by using machine learning algorithms to analyze data from various sources, such as industry trends and student performance, in order to pinpoint potential areas of cooperation and create business alliances between academia and industry. The model boosted industrial engagement and academic collaboration, according to the report.

Similar to this, Pantanowitz et al. (2022) suggested an intelligent industry partnership system that leverages data analytics and machine learning algorithms to find possible industry partners and create tailored collaboration programs. The technology considers a variety of elements, such as academic capabilities and industry needs, to deliver a customized cooperation experience. According to the report, the system enhanced industrial engagement and academic collaboration. Academic programs must continue to be in line with the digital enterprise, which requires collaboration between academia and industry. A model powered by AI can let the two of them communicate and operate together. Chaudhry and Kazim (2021) suggest an AI-powered platform that links educational institutions with business partners in order to pinpoint developing workforce demands and jointly create pertinent academic curricula.

AI can be employed to promote industry cooperation. A study by Bagai and Mane, (2023) propose to employed AI to pair students with business mentors. The algorithm was able to recognize pupils with an interest in particular businesses and pair them with mentors who had firsthand knowledge of such fields. Another platform for industry-academia collaboration was developed using AI in another study by Natarajan and Szeto, (2023). The platform was created to make it easier for business partners to connect with and work with academic researchers.

d) AI-Driven Adaptive Learning Systems

Artificial intelligence (AI)-driven adaptive learning systems have drawn interest recently because of their potential to customize education based on the needs of specific students and market demands. In their in-depth analysis of AI-driven adaptive learning platforms, Kim et al. (2022) demonstrate how well these platforms can match academic curricula with the constantly altering demands of the digital economy.

The creation of an AI model presents numerous prospects for integrating academic curricula with the digital economy. The opportunity to tailor students' educational experiences is one of the main opportunities. AI can be used to examine student performance and interest data to determine each student's best learning paths (Tapalova, et al, 2022). Another chance is the potential for automating duties like grading and counseling. This can free up faculty time so they can work on projects that are more imaginative and strategic. (Darling-Hammond et al, 2019). AI can also be utilized to give staff and students feedback. Both students and teachers can benefit from this input in terms of learning improvement (Ojajuni et al, 2021)

The body of knowledge on AI use in higher education has some shortcomings. The scarcity of study on how AI affects student learning is one of the main gaps. Understanding how AI can be applied to raise student learning results is crucial. (Seo et al,2021)

e) Challenges and Ethical Considerations

Challenges and ethical issues must be considered while developing an AI model for integrating academic programs with the digital enterprise. The ethical ramifications of employing AI in curriculum design are examined by Kousa and Niemi, (2023), who also emphasize the significance of taking privacy, bias, and openness into account when putting AI-driven solutions into practice. The absence of data is one of the main issues in developing an AI model for integrating academic programs with the digital enterprise. Data on the skills and abilities required in the digital enterprise are required in order to design an effective model. But getting this information is frequently a challenge (OECD, 2021b).

The ethical ramifications of deploying AI in higher education provide another difficulty. For instance, there is worry that AI might be applied to discriminate against particular student groups. Before AI is widely deployed in higher education, it is crucial to solve these ethical questions. Before AI is widely deployed in higher education, it is crucial to solve these ethical questions (Nguyen et al, 2022).

2.4.3 Implementing an Intelligent Model in Enhancing Graduates' Digital Skills and Employability

The job market is evolving due to technology advancements, requiring a workforce with digital skills to meet industry needs. Educational institutions are implementing models to improve graduates' digital skills and employability. This literature review aims to evaluate the effectiveness of these models and identify key factors contributing to their success.

a) Digital Skills needed in the Workforce

Digital skills are important because they cover a broad range of competencies that allow people to use and benefit from digital technologies for a variety of work-related tasks. These abilities are useful for jobs involving technology, but they are also now expected in a wide range of fields and occupations. Digital skills enable workers to adapt to shifting work settings and maintain their competitiveness in the job market, whether in communication and cooperation, data analysis, or problem-solving (ILO, 2021). The kinds of abilities needed in the contemporary workforce have been profoundly changed by the digital revolution. The necessity for higher education institutions to modify their curricula in response to the increased need for digital skills across industries was underlined in a research by Feijao et al. (2021).

One of the most important digital skills needed in the workplace is a basic understanding of computers. This involves having a fundamental understanding of software applications, operating systems, and troubleshooting methods. According to a Pew Research Center research, 73% of American adults own a desktop or laptop computer, compared to 91% who own smartphones (Anderson and Perrin, 2021). As a result, it is crucial for people to possess a solid foundation in fundamental computer skills in order to employ these tools successfully for work-related tasks. Knowledge of managing social media is another crucial digital skill. For businesses to connect with clients and reach their target market, social media has become an essential tool. 3.78 billion people use social media globally as of right now, up 5.9% from the prior year, according to a Hootsuite report (Hootsuite, 2021). Strong social media management abilities are therefore increasingly sought after in the workforce.

The ability to analyze and interpret data is also becoming more and more crucial, in addition to having a working knowledge of computers and managing social media. Businesses are seeking personnel who can efficiently gather, analyze, and interpret data in order to make educated decisions in light of the growth of big data. In fact, according to IBM research, demand for data analysts would increase by 28% by 2022 (Columbus, n.d.). As a result, in today's work environment, those with great data analysis abilities are in high demand.

b) Implementing Models for Enhancing Graduates' Digital Skills

Higher education institutions have started putting models for digital skill development into place in response to the demand for graduates with strong digital skills. In a study by Barboutidis and Stiakakis. (2023) that examined the impact of a digital skills model at Vocational Training Institutes, the employability of graduates improved. An Akojie and Haynes. (2023) study found that using a digital skills model greatly improves graduates' employability. The study used evidence-based data to strengthen academic programs' relevance for careers, match curriculum topics with business needs, get students ready for the workforce, and boost degree-related employment placement rates. The model sought to give graduates the digital skills essential to compete on the labour market. According to the study, graduates who successfully completed the model had a better likelihood of finding job than those who did not.

Similar to this, Khan et al.'s (2021b) study assessed how well a digital competency model improved graduates' employability and digital abilities in Malaysia. The study indicated that those who had better levels of digital skills and were more employable. The study found that the incorporation of the model into the curriculum, feedback from stakeholders in the industry, and the utilization of practical exercises were the main elements that contributed to the model's success.

c) The Role of Internships and Industry Collaboration

In order to bridge the knowledge gained in the classroom with the practical skills needed in the workplace, internships and cooperation with industry stakeholders are essential. In a research Hecker and Loprest, (2019) did on the effects of industry partnerships within the context of a digital skills model, they found that graduates' level of job ready had significantly increased.

Case studies are crucial for assessing how well digital skills models improve graduates' employability. Case studies give graduates actual experience applying digital skills to address real-world issues, according to a study by Raji et al. (2023). According to the study, graduates who successfully completed a digital skills model that included case studies were more employable than those who did not.

d) Assessing the Impact on Employability

Employability results must be evaluated to determine a digital skills model's efficacy. Graduates who successfully completed a digital skills program had greater employment rates than non-participants, according to a longitudinal study by Ezell. (2021).

The usefulness of a digital skills model in improving graduates' employability in China was assessed in a different study by Zhang et al. (2023). According to the survey, graduates who successfully completed the model had a better probability of finding jobs in fields that need digital abilities. The use of real-world scenarios, cooperation with industry partners, and incorporation of the model into the curriculum were highlighted by the study as the main elements contributing to the model's success.

e) Feedback from Industry Stakeholders

When assessing whether graduates' digital skills are in line with the needs of the job market, industry stakeholders' perspectives are extremely important. Carlisle et al. (2023) interviewed managers and business executives in the European tourism industry to get comments on the skills of graduates who had gone through a digital skills model. The feedback offered insightful suggestions for continued program improvement.

Evaluation of the efficacy of digital skills models in boosting graduates' employability requires input from industry stakeholders. A study by Suarta and Suwintana. (2021) found that input from industry stakeholders is essential for determining the skills needed in the labour market and creating models that satisfy these needs. According to the study, graduates who successfully completed a digital skills model created with input from industry stakeholders were more employable than those who did not.

2.4.4 Validating an intelligent Curriculum Alignment Model

Curriculum development and education have been completely transformed by artificial intelligence (AI). It is crucial to make sure that these models are validated and adhere to the relevant standards given the growing use of AI in numerous educational applications. In order to ensure that learning objectives, instructional methodologies, and assessment techniques are all coordinated, curriculum alignment is a crucial component of education. However, creating successful curriculum similarity can be difficult, especially in tough school environments. The goal of this review section is to investigate current initiatives to confirm the efficacy of these models by looking at pertinent studies.

a) AI Model Validation in Education

AI models in education aim to improve data-driven decision-making, adaptive assessment, and personalized learning. Strong validation techniques are crucial to the successful integration of these models. Validation aids in the informed adoption and implementation of AI-powered educational technology by teachers, administrators, and policymakers (Tapalova et al., 2022).

Machine learning (ML) is one of the AI models that is most frequently utilized in education. Large-scale data analysis and prediction are both done using machine learning (ML) methods. Alhothali et al. (2022) created a model based on machine learning to predict student success in online courses. A big online learning platform's data was used to validate the model, and the findings demonstrated that the ML model could predict student performance with an accuracy rate of more than 80%. Another element of AI in education is the use of natural language processing (NLP). NLP is used to analyze and comprehend human language for numerous educational applications, such as language learning and automated grading systems. Ramesh and Sanampudi (2022) developed an automated essay grading system based on NLP. The model was validated using a dataset of more than 10,000 essays, and the results showed that the NLP model could score essays very accurately.

Another crucial aspect of AI in education is its capacity to provide students with feedback in real-time. Cevikbas and Kaiser (2022) developed an AI-based method to provide real-time feedback to students during online conversations. The system was put to the test using data from an online course, and the results showed that the AI model could accurately provide feedback to students in real-time.

b) Challenges of AI-Based Model Validation

One of the challenges with validating AI models in education is the lack of standards in data collection and analysis. This is because it might be difficult to create a consistent dataset because educational data is typically diverse, unstructured, and complex. According to a recent study by Guan et al. (2020), the lack of consistency in educational data is a significant barrier to the validation of AI models in education.

Another challenge is the demand for interpretability and explicability in AI models. Understanding the decision-making process of an AI model is essential to ensuring that it complies to educational values. However, as AI models are occasionally referred to as "black boxes," it could be challenging to comprehend how they arrive at judgments. This issue is brought up by Khosravi et al. (2021), who argue that interpretability and explainability are crucial for the validation of AI models in education. Additionally, it's critical to guarantee that AI models are unbiased and inclusive. AI models have the potential to reinforce racial and gender inequalities that already present in educational settings. This may result in the discriminatory treatment of particular student groups, which is against the fundamental tenets of education. Assuring inclusion and fairness in AI models is a big difficulty in education, according to a recent study by Nguyen et al. (2022).

There's also a requirement for ongoing analysis and development of AI models. Since educational institutions are dynamic, the information utilized to train AI models may eventually become old. To guarantee that AI models stay applicable and efficient, it is necessary to continuously evaluate and enhance them. Llego, (2023), who contends that ongoing evaluation and development are essential for the validation of AI models in education, draws attention to this problem.

Collaboration is also required amongst the various parties participating in the validation of AI models in education. This group comprises academics, researchers, decision-makers, and technologists. To guarantee that AI models adhere to educational ideals and suit the interests of various stakeholders, collaboration is crucial. Collaboration is essential for the validation of AI models in education, claims a recent study by Kim et al. (2022).

2.5 Research Gaps

2.5.1 Current State of Digital Skills in Academic Programs and the Gaps between Industry Demands and Graduates' Competencies

The causes of the discrepancies between graduates' digital skills and industry-specific skill sets have not been thoroughly studied. Although the literature already in existence highlights the importance of digital skills, pedagogical techniques, curriculum design, and opportunities for experiential learning have not been fully investigated in relation to the conversion of academic courses into employable proficiencies. To narrow the gap, more study is required. Also, the precise digital skills that are most in demand in the workforce are not covered by the studies. The reports only offer a broad picture of the digital skills required for the workforce. However, it's critical to pinpoint the precise digital talents that are most in demand across a range of sectors and professions. The development of academic programs that better prepare graduates for the workforce can be aided by this data.

2.5.2 Developing an Intelligent Model for Aligning Academic Programs to the Digital Enterprise

A knowledge gap exists about the opportunities and problems associated with the practical application of AI models across a variety of educational institutions, despite the growing desire in doing so. There is a lack of thorough empirical research that explore practical barriers, faculty attitudes, student outcomes, and organizational dynamics, despite the fact that existing literature offers insights into theoretical elements. To effectively integrate academic programs with the digital enterprise and consider contextual distinctions, this gap must be filled before building and implementing AI models.

Employers' contributions to the creation of AI models for academic programs are not considered by the study. Employers may offer useful insights on the digital skills required in the workforce, and they can collaborate with academics to create AI models that are in line with the requirements of the workforce. The ethical ramifications of utilizing AI in academic programs, such as potential discrimination against particular groups or the collecting and storage of personal data without consent, must be considered.

The way that AI will affect employment in the future is also crucial, since it is likely to automate jobs in the years to come, since a result, academic programs that teach students how to be creative, critical thinkers, and problem solvers will be necessary.

2.5.3 Implementing an Intelligent Model in Enhancing Graduates' Digital Skills and Employability

There is still a lack of empirical studies that thoroughly evaluate the varied impact of particular models on different dimensions of digital skills (e.g., technical proficiency, digital literacy, adaptability), and their subsequent influence on various aspects of employability (e.g., job placement, career advancement), despite the growing body of literature examining the relationship between educational models and graduates' digital skills and employability.

There is also a gap in the evaluation of a model's efficacy in improving graduates' digital skills and employability. The model's long-term effects on graduates' professional growth and career trajectories, including job performance, career promotion, and all-around employability need to be looked at. The studies also disregard a model's long-term effects on various student demographics, assuming that all students benefit equally from it. However, it's possible that some student types will benefit from the model more than others. Studies should also consider the model's cost-effectiveness into account, performing a cost-benefit analysis to see if the advantages outweigh the costs.

2.5.4 Validating an intelligent Curriculum Alignment Model

There is a significant gap in the study literature about the empirical validation of the Artificial Intelligent Curriculum Alignment Models, despite the growing interest in the application of artificial intelligence (AI) in education. Despite the fact that many researchers have looked at the potential advantages of employing AI for curriculum design and alignment, less focus has been placed on systematically evaluating the efficacy, accuracy, and usefulness of the models in actual educational contexts. Therefore, there is a need for in-depth empirical research that examines how much the model improves curricular alignment, raises learning outcomes, and gives teachers useful information to help them make the best possible instructional decisions.

A further area which requires attention is the model's adaptability and generalizability to various academic fields and institutions. The model's effectiveness at coordinating curriculum with the demands of the digital enterprise might be specifically investigated in a number of scenarios, including various academic disciplines, program types, and institutional settings.

Additionally, it might look at how the model might be modified to fit the particular requirements and traits of different educational courses and institutions.

Table 2.2 Summary of Research Gaps

S. No	Focus	Source	Key Findings	Research Gap(s)
1	Technology Integration in Academic Programs	World Economic Forum (2023), Mn et al. (2020), González et al., (2021), Putra et al. (2022)	- Discrepancy between digital skills taught in academic programs and those demanded by companies - Lack of thorough methods in academic programs to ensure students acquire requisite digital skills - Graduates' digital proficiency varies across sectors	- Need for studies on effective methods to integrate digital tools and platforms into academic curricula. - Examination of barriers hindering thorough implementation of digital skill acquisition methods. - Lack of thorough methods to ensure students acquire necessary digital skills
2	Industry Demands for Digital Skills	World Economic Forum (2023), European Commission (2022), Colback, (2023), Plekhanov et al. (2022)	- Businesses perceive a gap in digital skills among graduates - Discrepancy between graduates' skills and industry demands. - Emerging demand for expertise in fields like cybersecurity and artificial intelligence	- Further exploration of specific digital skill gaps across industries. - Investigation into evolving digital skill demands and their implications on academic curricula.

3	Assessment of Graduates' Digital Competencies	Asia Development Bank (2023), Mn et al. (2020), Li, (2022)	- Significant portion of graduates lack critical digital - Variation in digital competencies among graduates	- Lack of Studies examining the root causes of digital skill deficiencies among graduates. - Longitudinal research on the effectiveness of academic programs in addressing digital competency gaps.
4	Faculty Readiness and Training	Evans (2021), Zan et al. (2021)	- Importance of faculty readiness in teaching digital skills - Varied levels of digital literacy among faculty members	- Exploration of effective strategies for faculty development in digital skill instruction. - Investigation into the impact of faculty digital literacy on student outcomes.

5	Strategies for Closing the Gap	Loughborough University London (2023), University of Waterloo (2021)	- Advocacy for experiential learning to bridge knowledge gap - Implementation of digital skills programs by academic institutions	- Assessment of the effectiveness of experiential learning in enhancing digital skill acquisition. - Comparative analysis of various institutional programs aimed at addressing digital skill gaps. -Lack of comprehensive strategies to bridge the digital skills gap
6	Curriculum Design Using AI	Kim et al. (2022), Thongprasit and Wannapiroon (2022), Sarker (2021), Crompton (2023), Zhou et al. (2021)	- AI-driven models enhance curriculum customization, student performance, and satisfaction. - Integration of AI helps identify industry demands and adapt curricula accordingly.	- Limited exploration on long-term effectiveness of AI-driven curriculum design. - Need for research on scalability and transferability of AI-driven models across different educational settings.
7	Faculty Development and AI Integration	Swiecki et al. (2022), Hooda et al. (2022), Ng et al. (2023), Seo et al. (2021), Zhou et al. (2021)	- AI-powered faculty development enhances instructor performance and satisfaction. - Training programs improve AI literacy among faculty members. - Customized AI-driven	- Insufficient investigation into long-term impact of AI integration on faculty development. - Need for standardized evaluation metrics

			systems aid in instructional improvement. - AI facilitates creation of high-quality MOOCs.	for assessing effectiveness of AI-driven faculty development programs.
8	Industry Collaboration and AI Models	Halder and Saha (2020), Pantanowitz et al. (2022), Chaudhry and Kazim (2021), Bagai and Mane (2023), Natarajan and Szeto (2023)	- AI-driven models improve collaboration between academia and industry. - Customized partnership programs enhance engagement and collaboration. - AI platforms facilitate mentorship and collaboration. - Streamlined platforms improve industry-academia interactions.	- Lack of studies assessing long-term sustainability and adaptability of AI models for industry collaboration. - Need for research on ethical implications and inclusivity in AI-driven industry collaboration models.
9	AI-Driven Adaptive Learning Systems	Kim et al. (2022), Tapalova et al. (2022), Darling-Hammond et al. (2019), Ojajuni et al. (2021), Seo et al. (2021)	- AI-based adaptive learning systems personalize education and automate tasks, improving learning outcomes and feedback mechanisms. - Challenges include the scarcity of research on AI's impact on student learning and the ethical considerations of AI deployment.	- Lack of comprehensive studies on AI's long-term impact on student learning outcomes. - Ethical implications such as bias and privacy require further exploration.
10	Challenges and Ethical Considerations	Kousa and Niemi (2023),	- Ethical considerations include privacy, bias,	- Need for more research on

		OECD (2021b), Nguyen et al. (2022)	and fairness in AI-driven curriculum design and deployment. - Challenges include data scarcity and ethical implications in AI applications.	mitigating biases and ensuring fairness in AI-driven education. - Further exploration of ethical models and guidelines for AI deployment in education is necessary.
11	Validating Artificial Intelligent Curriculum Alignment Models	Tapalova et al. (2022), Alhothali et al. (2022), Ramesh and Sanampudi (2022), Cevikbas and Kaiser (2022)	- AI models improve decision-making, assessment, and personalized learning. - Challenges include data standardization and model interpretability.	- Lack of standardized methods for data collection and analysis in educational AI models. - Interpretability and fairness of AI models require more attention for validation.
12	Effectiveness of a Model in Enhancing Graduates' Digital Skills and Employability	Feijao et al. (2021), Anderson and Perrin (2021), Hootsuite (2021), Columbus (n.d.), Barboutidis and Stiakakis (2023), Akojie and Haynes (2023), Khan et al. (2021b), Hecker	- Digital skills models improve graduates' employability and address industry demands. - Internships and industry collaboration play a crucial role in bridging the gap between academia and industry.	- Limited exploration on the scalability of digital skills models across different educational contexts and industries. - Need for standardized evaluation metrics to assess the long-term impact of digital skills models on

	and Loprest (2019), Raji et al. (2023), Ezell (2021), Zhang et al. (2023), Carlisle et al. (2023), Suarta and Suwintana (2021)		graduates' employability

2.6 Theoretical Framework

This study is underpinned by several theoretical frameworks that guide the development and application of the intelligent curriculum alignment model. The primary frameworks include the Digital Competence Framework (DigComp 2.2), the Technological Literacy Framework, and Competency-Based Curriculum Theory. These frameworks collectively provide the conceptual foundation for aligning academic programs with essential digital skills in the labour market.

2.6.1 Digital Competence Framework (DigComp 2.2)

The DigComp 2.2 framework, developed by the European Commission, provides a comprehensive structure for defining and assessing digital competencies. It identifies five key areas: information and data literacy, communication and collaboration, digital content creation, safety, and problem-solving (Vuorikari et al., 2022). This framework was applied in the study to ensure that the digital skills extracted from labour market data and mapped to academic curricula are aligned with internationally recognized standards. By using DigComp 2.2, the research ensures that the identified skills are relevant, comprehensive, and benchmarked against global best practices (Vuorikari et al., 2022).

2.6.2 Technological Literacy Framework

The Technological Literacy Framework emphasizes the knowledge and skills needed to use, manage, evaluate, and understand technology. It goes beyond mere proficiency with tools, advocating for a deeper comprehension of how technology affects society and the individual (ITEA, 2007). This framework was applied to ensure that the curriculum alignment model not only identifies technical skills but also considers the broader implications of technology use, including ethical considerations, critical thinking, and creative problem-solving. The framework guided the inclusion of both foundational and advanced digital skills in the model, ensuring a holistic approach to curriculum alignment (ITEA, 2007).

2.6.3 Competency-Based Curriculum Theory

Competency-Based Curriculum Theory focuses on clearly defined competencies that prepare professionals for the workforce. This theory was applied to ensure that the curriculum alignment model is not only comprehensive but also practical and sustainable. By using this theory, the research ensures that the identified skills are directly linked to the competencies required by employers, making the curriculum more relevant and effective (Kim, 2015). The theory also supports the customization of the model to different academic disciplines and institutional contexts, enhancing its applicability and impact (Kim, 2015).

2.6.4 Application in the Study

These theoretical frameworks were applied throughout the research process. The DigComp 2.2 framework guided the identification and categorization of digital skills, ensuring alignment with global standards. The Technological Literacy Framework informed the inclusion of both technical and broader digital competencies, ensuring a holistic approach. Competency-Based Curriculum Theory ensured that the identified skills are directly linked to workforce requirements, making the curriculum more relevant and effective. Together, these frameworks provide a robust foundation for the development and validation of the intelligent curriculum alignment model, ensuring that it is both theoretically sound and practically applicable.

2.6.5 Influence of Theoretical Frameworks on the Conceptual Framework

The development of the conceptual framework in this study was rooted in the constructs derived from the underpinning theoretical frameworks: The Digital Competence Framework (DigComp 2.2), the Technological Literacy Framework, and Competency-Based Curriculum Theory. The linkages between these theoretical constructs and the conceptual framework components were explicitly established to ensure alignment between theory and practical investigation.

The Digital Competence Framework provided the foundational classification of digital skills, identifying core domains such as information literacy, communication, content creation, safety, and problem-solving. These domains translated directly into key constructs within the conceptual framework, shaping the dimensions of digital skill assessment within academic curricula. This enabled the conceptualization of curriculum content as multidimensional digital competencies, facilitating a comprehensive mapping process against labour market demands.

The Technological Literacy Framework enriched the conceptual framework by introducing constructs related to critical understanding and ethical use of technology, beyond technical proficiency. This inclusion ensured the model accounted for not only skill acquisition but also the contextual and reflective competencies essential for holistic digital literacy, thereby broadening the scope of curriculum evaluation to include ethical, cognitive, and creative aspects.

Furthermore, Competency-Based Curriculum Theory grounded the conceptual framework in the practical alignment of academic offerings with workforce requirements. It emphasized the identification of clearly defined competencies as measurable outcomes, which informed the framework's focus on skill relevancy, employability, and curriculum responsiveness. This linkage ensured that the conceptual framework operationalized competencies as dynamic constructs shaped by labour market evolution and institutional adaptability.

The table 3 below briefly illustrates these direct linkages between the theoretical constructs and the corresponding elements within the conceptual framework:

Table 2.3 Theoretical Constructs and Conceptual Framework Components Linkages

Theoretical Framework	Core Constructs	Influence on Conceptual Framework	Conceptual Framework Components
Digital Competence Framework (DigComp 2.2)	Info/data literacy, communication, content creation, safety, problem-solving	Defined scope and categories of digital skills to be assessed in curricula	Digital proficiency domains; skill category dimensions
Technological Literacy Framework	Ethical technology use, critical thinking, creative problem-solving	Broadened curriculum content beyond skills to include cognitive and ethical competencies	Holistic digital literacy; reflective and ethical competencies
Competency-Based Curriculum Theory	Clearly defined competencies linked to workforce demands	Ensured curriculum outcomes are aligned with measurable	Employability focus; competency relevancy and

	employment competencies	mapping

2.7 Conceptual Framework

This conceptual framework illustrates an intelligent curriculum alignment model aimed at bridging the gap between academic programs and the evolving digital skill requirements of the labor market.

Variables

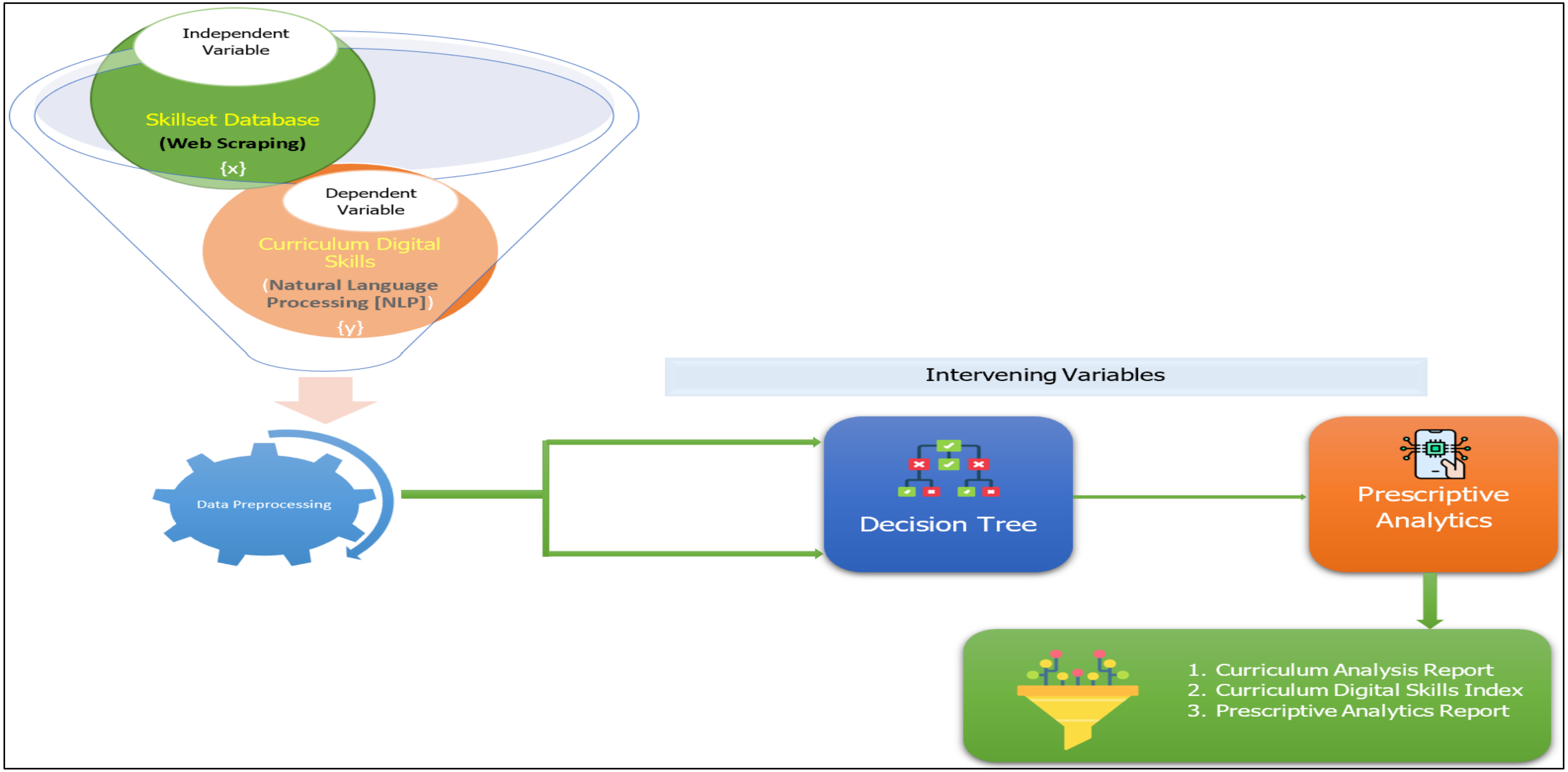
The framework includes digital skills demand from real-time labor market data as the independent variable, while the dependent variable is the alignment of university curricula with these market demands, measured by indices and curriculum changes. Intervening variables involve the model's ability to handle complex data, extract competencies using NLP, map these competencies against market needs with a Decision Tree algorithm, and generate prescriptive analytics for curriculum improvement.

Framework Explanation

The model initiates with capturing digital skill demands from dynamically updated labor market data, ensuring the reflecting of current and emerging industry requirements. The academic programs' curricula are subjected to a detailed analysis process utilizing NLP techniques to extract embedded digital competencies within course content and learning outcomes. These identified competencies are systematically mapped and compared against the extracted labor market demands through a Decision Tree machine learning algorithm that acts as the core analytical engine. This process assesses the degree of alignment between curricula and market needs. Intervening variables include the model's computational ability to manage, analyze, and synthesize data inputs from diverse sources, uncover relationships between different digital skills, and generate insights that are realistic and actionable for stakeholders.

The outcome of this process is delivered through a set of prescriptive analytics tools. The Curriculum Analysis Report offers a comprehensive evaluation of the current curriculum content, focusing on the inclusion of relevant digital skills. The Curriculum Digital Skills Index quantifies the extent to which the curricula align with the digital skill demands of the labor market. Additionally, the Prescriptive Analytics Report provides specific recommendations for curriculum adjustments, aimed at addressing identified gaps and improving graduate employability.

Figure 2.1 Conceptual Framework



Source: Researcher on conceptualization

2.8 Critical Analysis of Reviewed Literature

2.8.1 Effectiveness of Artificial Intelligence (AI) frameworks to align academic programs with the evolving needs of the digital enterprise

The reviewed research suggests that AI can be a valuable asset. Studies by Kim et al., (2022), Thongprasit and Wannapiroon, (2022), Sarker, (2021), and Crompton, (2023) demonstrate the effectiveness of AI in tasks like identifying in-demand skills, personalizing learning experiences, and tracking student progress. However, limitations exist in the current research base. A significant limitation is the focus on theoretical applications. While studies by Swiecki et al., (2022), Hooda et al., (2022), and Ng et al., (2023) propose frameworks for faculty development and industry collaboration using AI, few studies demonstrate successful implementation in real-world academic settings. Additionally, the ethical implications surrounding AI in education often go unaddressed. Algorithmic bias and the potential for perpetuating social inequalities in the job market are crucial areas that require further investigation. Despite these limitations, AI's potential for program alignment remains promising. Future research should explore the practical integration of AI tools within curriculum design processes, building upon frameworks proposed by previous works (e.g., Kim et al., 2022; Thongprasit and Wannapiroon, 2022). Furthermore, critical studies are needed to address the ethical considerations surrounding AI and its potential impact on educational equity (e.g. the need for more research on mitigating biases and ensuring fairness in AI-driven education).

2.8.2 Effectiveness of Frameworks in Enhancing Graduates' Digital Skills and Employability

The reviewed literature highlights the crucial role of digital skills in today's workforce (ILO, 2021). Studies by Feijao et al., (2021) emphasize the need for universities to adapt their curricula to address the increasing demand for digital skills across industries. Research by Anderson and Perrin, (2021) and Hootsuite, (2021) showcase the growing importance of foundational computer skills and social media management abilities. Additionally, Columbus (n.d.) highlights the rising demand for data analysis expertise. Several studies explored the effectiveness of frameworks in enhancing graduates' digital skills and employability. Barboutidis and Stiakakis, (2023), Akojie and Haynes, (2023), and Khan et al., (2021b) all found that frameworks can improve graduates' digital skill levels and employability rates. These studies highlight the importance of incorporating industry feedback (Khan et al., 2021a), practical exercises (Akojie and Haynes, 2023), and real-world scenarios (Raji et al., 2023) into the framework design for optimal effectiveness. However, a critical gap exists in comprehensively evaluating the impact of frameworks on various aspects of digital skills and employability. Future research should explore the refined effects of frameworks on different skill dimensions (technical proficiency, digital literacy, adaptability) and their influence on diverse employability aspects (e.g. job placement, career advancement). Furthermore, limited research examines the long-term impact of frameworks on graduates' careers. Longitudinal studies are needed to assess the frameworks' effects on job performance, career progression, and overall employability across different student demographics. Cost-effectiveness analysis is also crucial to determine if the benefits outweigh the implementation costs.

In Summary, this review critically analyzed the potential and limitations of AI frameworks for aligning academic programs with the digital enterprise. While AI offers promising opportunities, challenges regarding practical implementation and ethical considerations necessitate further investigation. Similarly, the effectiveness of frameworks in enhancing graduates' digital skills and employability requires more refined evaluation, considering the long-term impact on diverse skillsets, career trajectories, and student demographics. By addressing these critical gaps, academia can leverage AI and educational frameworks responsibly to prepare graduates for successful careers in the ever-evolving digital landscape.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

The alignment of academic curricula with the demands of the labour market is of utmost relevance in a time marked by rapid technological developments and a changing global economy. Essential digital skills are now a crucial factor in determining an individual's employability and the success of their business, thanks to the digital revolution. Therefore, it is more important than ever to make sure that educational institutions give their graduates the skills they need to meet the changing demands of the labour market. This study methodology chapter goes into the creation and use of an intelligent model intended to make it easier to match academic curricula with the rapidly evolving digital skills needed in the workforce. The main objective of this research was developing an intelligent curriculum alignment model to analyze university curricula, identify digital competencies, and map them to labour market digital skill requirements.

This study's importance rests not only in its ability to address the gap between education and industry, but also in its contribution to the field of artificial intelligence and data analytics used in the field of education. This study attempts to provide a systematic method to program alignment that is both scalable and flexible to the changing nature of digital skills by utilizing the power of artificial intelligence and data-driven insights.

3.2 Research Design

This study was conducted using the Design Science Research (DSR) paradigm, which emphasizes the creation, development, and evaluation of innovative artifacts to address real-world problems. The research applied DSR principles systematically throughout the study to ensure both rigor and practical relevance.

The first principle applied was problem relevance. The research began by identifying a significant and practical challenge; the misalignment between digital skills taught in academic programs and those demanded by the labour market. This problem was clearly articulated and justified as a critical issue affecting graduate employability and institutional responsiveness.

Next, the principle of design as an artifact guided the development of a tangible solution; an intelligent curriculum alignment model. This artifact was designed to systematically analyze academic curricula, extract digital competencies, and map them to dynamic labour market requirements. The model was constructed as a system integrating web scraping, Natural Language Processing (NLP), machine learning, and a user-friendly interface.

The principle of design evaluation was central to the research process. The model was rigorously evaluated through iterative cycles of building and testing, ensuring its utility, quality, and efficacy. Evaluation methods included both technical validation and stakeholder feedback, aligning with DSR's emphasis on demonstrating the artifact's effectiveness.

Research contributions were addressed by ensuring that the study provided clear and verifiable outputs in the form of the developed model, its underlying methodology, and actionable recommendations for academic institutions. These contributions were designed to advance both theoretical understanding and practical application in curriculum alignment.

Research rigor was maintained by applying systematic and well-documented methods in both the construction and evaluation of the artifact. The research process was transparent, replicable, and grounded in established DSR practices.

Finally, the principle of design as a search process was reflected in the iterative nature of the research. The study continuously refined the model based on feedback and new data, ensuring that the solution evolved to better meet the needs of stakeholders and the changing demands of the digital economy.

3.3 Preliminary Research

Conducted a thorough literature review to comprehend the requirements for digital skill sets, university curricula, and the labour market using desktop research with a view to identify key artificial intelligence concepts, models, and methodologies related to digital skill assessment and curriculum analysis.

3.4 Data Collection

3.4.1 Web Scraping

Web scraping was used to create a skillset database by searching job advertisements, extracting digital skill requirements mentioned in the job descriptions and finally developing a structured and weighted dataset of digital skill requirements in the labour market.

a) Methodology for Web Scraping Digital Skill Requirements from Job Advertisements

The methodology began with Identifying the Target Job Portal, where Dice.com was selected for scraping due to its extensive collection of ICT-relevant job postings and its high advertisement frequency; the geographic scope was intentionally limited to Nairobi, Kenya, as the nearest major metropolitan area. Next, in Defining Search Parameters, the Occupational Information Network (O*NET Online database) was used to identify relevant job titles to be used as search terms for the scraping process. These queries were carefully defined to be broad enough to capture pertinent job postings while remaining specific enough to filter out irrelevant ones. For the Scraping Tool Selection, the study employed the BeautifulSoup (Python) web scraping platform to automate the process of visiting each job posting, extracting the digital skill requirements, and saving them for analysis. This led to Scraping Job Advertisements, where a custom script was created to iterate through the search results on Dice.com and extract the digital skill requirements from each listing. Following this, Processing and Cleaning Data involved removing irrelevant information and standardizing the text format of the scraped digital skill requirements to facilitate easier analysis. The next step was Analyzing Skill Requirements, which utilized techniques such as frequency analysis to identify the most commonly mentioned skills. Finally, the Weighting of Identified Skills was calculated using a linear scaling function to reflect the relative demand for each skill. The weight was calculated using the formula:

weight = min_w + (frequency - min_freq) / (max_freq - min_freq) * (max_w - min_w), where: min_w=0.05 and max_w=0.95 are used for min-max normalization.

Where:
min_w (minimum weight) is 0.05
max_w (maximum weight) is 0.95
min_freq (minimum frequency) is the number of mentions the least demanded skill
max_freq (maximum frequency) is the number of mentions the most demanded skill

b) Justification for Web Scraping Digital Skill Requirements from Job Advertisements

Analyzing digital skill requirements from job advertisements provides valuable insights into current industry trends and demands, which can effectively guide individuals, educators, and policymakers in prioritizing skill development efforts. Understanding the most sought-after digital skills directly informs career planning and skill acquisition strategies, allowing job seekers to focus on developing high-demand skills, thereby increasing their employability and competitiveness in the job market. Furthermore, educational institutions can leverage data on these requirements to tailor their curriculum and training programs to meet the needs of the job market, ensuring that graduates are equipped with relevant skills and improving their chances of successful employment. Conversely, employers can use insights from job advertisements to inform their hiring strategies and workforce planning; by understanding prevalent digital skill requirements, they can recruit candidates with the necessary skill sets to drive innovation and productivity within their organizations. Finally, companies can benchmark their own digital skill requirements against industry standards through job advertisement analysis, helping them to identify areas of skill gaps or strengths compared to competitors, which enables strategic decision-making in talent acquisition and development.

3.5 Intelligent Analytics Module

The analytics model was based on a hybrid model derived from the combination of Descriptive and Prescriptive Artificial Intelligence Algorithms implemented using Python Programming Language. It implements a Natural Language Processing (NLP) unit and Decision Tree machine learning algorithm to assess both the level of conformance of curriculums to the essential digital skills required by the labour market as well as prescribe skills to be integrated to the curriculums if lacking.

3.5.1 Natural Language Processing (NLP)

NLP implemented using Python Natural Language Toolkit (NLTK) and spaCy library were used to harvest digital skills from ICT academic curriculums. Utilized text analysis tools and techniques to extract relevant keywords, phrases, and topics from the collected documents to identify and classify the key digital skills emphasized within academic curricula.

a) Methodology for Mining Digital Skills from Academic Curricula using Natural Language Processing

The methodology for analyzing academic curricula begins with Data Acquisition, where all relevant ICT academic curricula are gathered, including course descriptions, learning objectives, syllabi, and other pertinent textual information. This raw data is then subjected to Text Preprocessing using Natural Language Processing (NLP) techniques. This involves several steps: cleaning the data by removing irrelevant elements like headers, footers, and special characters; formatting the text by converting it to a uniform, lowercase, plain text format and removing stop words; and finally, Tokenization, which breaks the text down into individual words or phrases. Following preprocessing, Named Entity Recognition (NER) is applied to identify and classify named entities that represent digital skills within the text, using a Dictionary-based matching criterion that relies on pre-defined digital skill dictionaries from the skillset database. This is followed by Semantic Analysis using word embeddings to capture the contextual meanings and semantic similarities of the digital skills, which helps identify synonyms and related terms and understand the relationships between concepts. The process then moves to Skill Classification and Analysis, where the extracted skills are categorized into broader groups—such as programming languages, data analysis tools, or soft skills—using established taxonomies. The penultimate step is Validation and Evaluation, which involves validating the extracted digital skills against existing taxonomies to ensure their relevance and completeness, and then evaluating the effectiveness of the entire methodology by comparing the automatically extracted skills with manual assessments. The final step is the Weighting of mined skills, where each skill is assigned a numerical weight calculated as a fraction of its total count of occurrences relative to the total number of all skill occurrences in the curricula.

b) Curriculum mining NLP Process

Input: *X* represents text extracted PDF File.

Tokenization: Let *T* be set of tokens obtained from *X* after tokenization.

Named Entity Recognition (NER): Let *N* be set of named entities identified in *T* using NER., let *P* => Digital Skill Groups and it is also the subset of *N* that contains Digital skill entities.

Counting: Count each Digital skill entity in *P* to obtain a frequency count *C(p)*

Summary Generation: Lists each digital skill ***p*** with its corresponding count ***C(p).***

$$C(p) = \sum\nolimits_{n \in P} \Delta_{p,n}$$

Where:

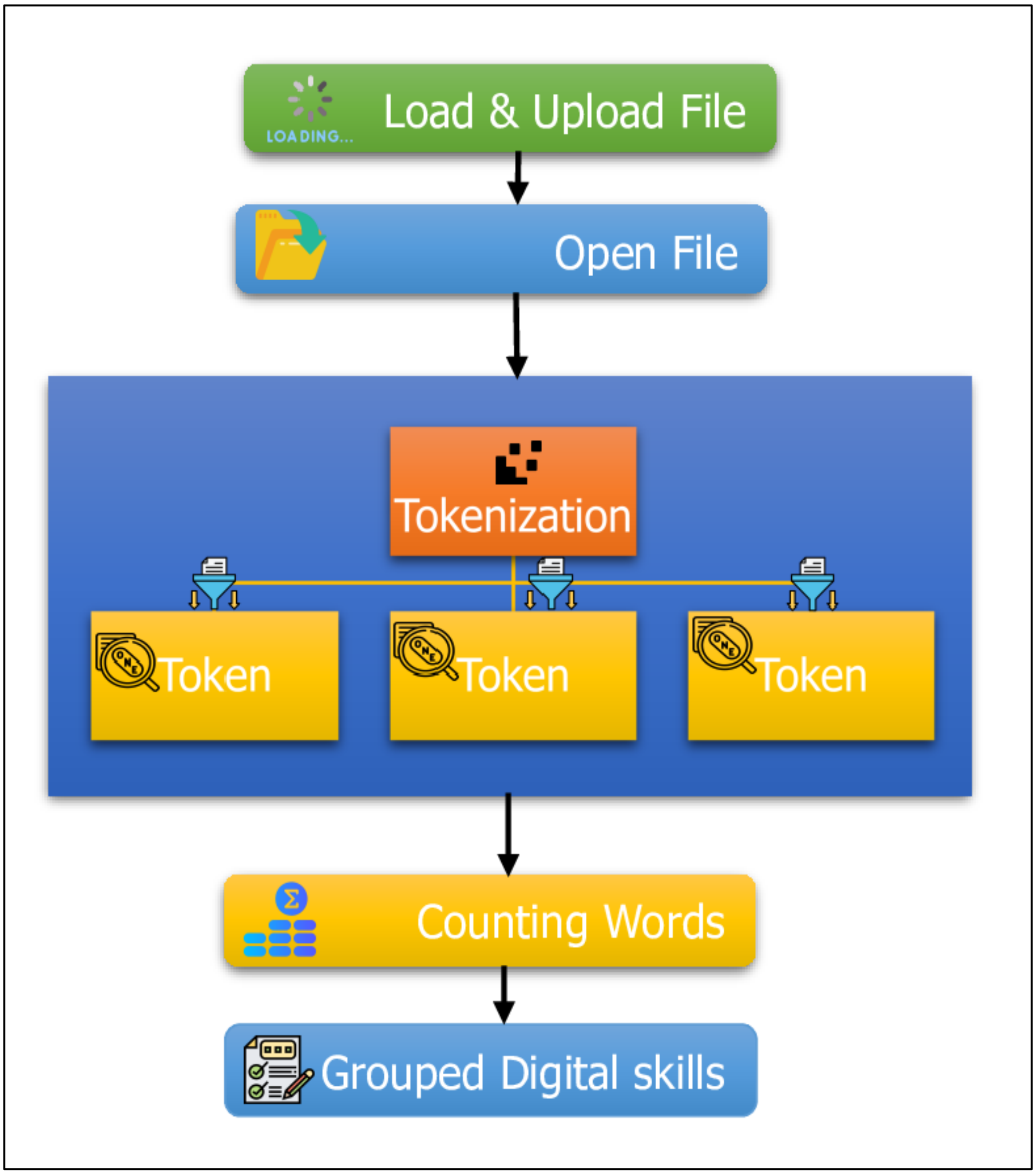
- *C(p)* is the count of occurrences for digital skill *p*.

- *P* is the set of named entities identified as digital skill in *P*.

- *n* iterates over the elements of *P*.

- $\Delta_{p,n}$ is the Kronecker delta function, which is equal to 1 if $p = n$ (programming language *p* matches named entity *n*, and 0 otherwise.

Figure 3.1 Curriculum Mining using NLP



Source: Researcher on conceptualization

c) **Justification for Mining Digital Skills from Academic Curricula using Natural Language Processing**

Natural Language Processing (NLP) offers an automated and scalable solution for extracting valuable information about the digital skills taught in various academic programs, providing a superior alternative to the traditional, time-consuming, and inefficient approach of manually reviewing curricula. These NLP techniques establish a systematic and data-driven approach to curriculum analysis, yielding insights that inform strategic decision-making within educational institutions. This capability allows administrators to allocate resources effectively, prioritize necessary curriculum revisions, and tailor programs to meet the needs of diverse stakeholders. Furthermore, extracting digital skills from academic curricula significantly enhances students' employability by ensuring they possess the requisite competencies sought by employers, effectively bridging the gap between academia and industry. This fosters a smoother transition for graduates into the workforce and contributes to economic growth and innovation. Understanding the digital skills emphasized in academic curricula also enables educators to design and revise courses that are more responsive to technological advancements and industry trends, ensuring that curriculum content remains both relevant and up-to-date. Finally, by explicitly identifying digital skills mentioned in the curricula, students gain clarity on the competencies they are expected to acquire throughout their academic journey, which facilitates targeted skill development and self-directed learning, thus empowering students to meet industry requirements effectively.

3.5.2 Decision Tree Supervised Learning Model

The decision tree model developed using Python scikit-learn (sklearn) library maps the curriculum skills (y) based on the digital market skills (x). At each node of the decision tree, a decision is made based on the features of the data. The decision tree recursively splits the data based on the feature that best separates the samples, aiming to minimize the mean squared error (MSE) between mapped and actual curriculum skills.

a) Mathematical Model

Let:

X = {x_1, x_2, ..., x_N}) be the set of digital market skills, where (x_i) represents a vector of features describing market skills for the (i{th}) individual.

(Y = {y_1, y_2, ..., y_N}) be the set of skills in the academic curriculum, where (y_i) represents the corresponding curriculum skills for the (i{th}) individual.

(D = {(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)}) be the dataset containing paired samples of digital market skills and curriculum skills.

The study aims to learn a decision tree model (f: X ->Y) that maps digital market skills to skills in the academic curriculum.

b) Model Representation

The decision tree (T) is represented as a hierarchical structure of nodes.

Each internal node represents a decision based on a feature, and each leaf node represents a mapped value of curriculum skills.

The decision tree model (f(x)) maps the curriculum skills based on the digital market skills (x).

c) Training Procedure

The training of the intelligent model relies on a decision tree model, which is trained using the collected dataset (D). During this training process, the decision tree is designed to learn how to partition the feature space based on the relationship between digital market skills and curriculum skills. This is achieved by recursively splitting the dataset based on the features that effectively minimize the Mean Squared Error (MSE) between the predicted (mapped) and the actual curriculum skills.

d) Model Evaluation

The trained decision tree model is evaluated using various metrics, such as mean squared error (MSE), mean absolute error (MAE), or the (R2) score, on a separate validation set or through cross-validation. This evaluation process is crucial as it helps assess the model's performance and its effectiveness in accurately mapping curriculum skills based on the identified digital market skills.

e) Model Utilization

Once trained and evaluated, the decision tree model can be utilized to map the skills in the school curriculum for instances of digital market skills.

f) Model Justification

The selection of a decision tree model is advantageous due to several key properties. Firstly, it offers high Interpretability because its branching structure clearly reveals the factors influencing the final outcome, making it straightforward to understand both the skills sought by employers and how academic programs can address them. Secondly, the model excels at Feature Selection, as it automatically identifies the most important features—in this context, the digital skills most valued by employers—which helps educators effectively prioritize and tailor their curriculum. Furthermore, decision trees offer Flexibility, as they can handle both categorical and numerical data, allowing for the inclusion of diverse data points such as job descriptions, skills assessments, and graduate employment data. Finally, they demonstrate Robustness to outliers and irrelevant features, which serves to simplify the data preprocessing stage and ultimately improves the model's overall performance.

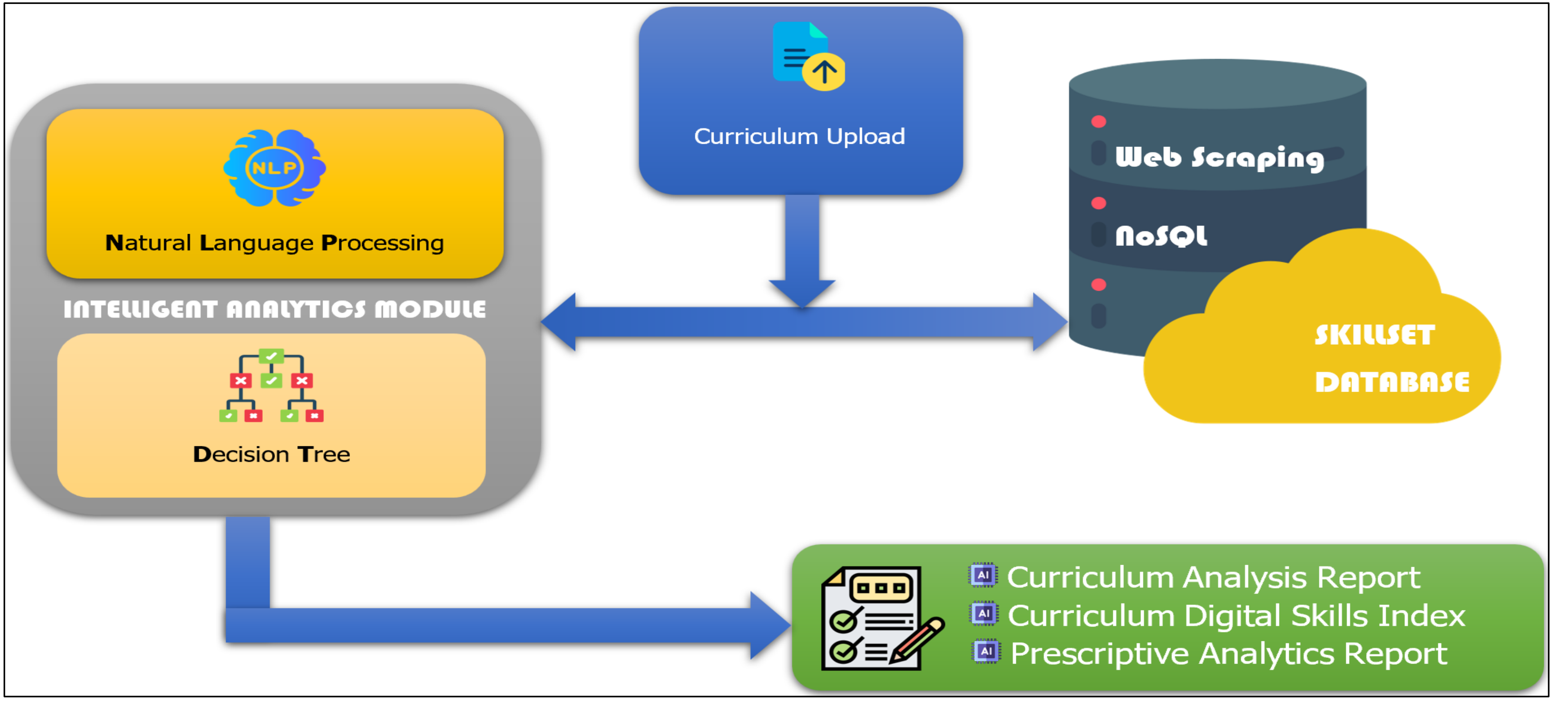
3.6 NoSQL Database

The skillset database utilized a flexible schema technology based on NoSQL, which was chosen for several key advantages. The first is its Suitability for handling unstructured or semistructured data, which makes it ideal for the skillset database due to the inherent variability in information structure. Second, NoSQL databases offer superior Performance because they are optimized for specific use cases, allowing for quick retrieval of skills and related information, which is crucial for a skillset database. Third is its Scalability: NoSQL databases are designed for horizontal scaling, enabling them to easily handle large data volumes and seamlessly grow as the skillset database expands over time. Finally, the Flexibility of the NoSQL data model allows it to accommodate diverse skill sets, experience levels, and related attributes, making it perfectly suited for the needs of the skillset database.

3.7 System Architecture

The provided system architecture is a multi-component model designed for intelligent analysis of curricula. The process commences with a Curriculum Upload, where a curriculum document in PDF format is uploaded into the system. The uploaded curriculum is then processed by the Intelligent Analytics Module. This core module is divided into two primary sub-components: The Natural Language Processing (NLP) unit and the Decision Tree algorithm. The NLP unit is responsible for parsing and mining the textual content of the curriculum; extracting relevant skills. The Decision Tree algorithm then utilizes this processed information to perform analytics generating classifications, analyses and prescriptions using the NoSQL Skillset Database. The architecture also incorporates a Web Scraping framework to populate the NoSQL Skillset Database. This database serves as a repository of skills, updated through automated extraction of skill from online job postings. Finally, the model gives three forms of output; one; a Curriculum Analysis Report, providing an overview of the evaluated curriculum; two, a Curriculum Digital Skills Index, which quantifies the relevance of digital skills within the curriculum as compared to the skillset database; and a Prescriptive Analytics Report, which offers skill addition suggestions for curriculum improvements based on the analytics.

Figure 3.2 System Architecture



Source: Researcher on conceptualization

3.8 Ethical Considerations

Several ethical issues were considered in this study to ensure the fairness and integrity of the research. The main ethical factors considered were:

3.8.1 Informed Consent

This study did not involve human participants. Data was web scraped exclusively from publicly accessible websites. No personal data was collected, ensuring the privacy and anonymity of individuals. As such, informed consent from individuals was not applicable to this research.

3.8.2 Data Privacy and Confidentiality

All data was web scraped exclusively from publicly accessible websites. While no personal data was involved, the researcher used secure data handling and storage procedures to prevent unauthorized access or exposure of any information obtained during the study, maintaining the integrity and security of the collected public data.

3.8.3 Fairness and Equity

This study did not involve human participants. Therefore, the principles of treating participants with respect and dignity, valuing their opinions, and ensuring an impartial selection process based on factors like gender, race, ethnicity, or socioeconomic status, were not applicable to this research.

3.8.4 Transparency and Integrity

By avoiding conflicts of interest and abiding by ethical standards in data collection, analysis, and reporting, the researcher upheld the integrity of the research and maintained transparency throughout the investigation process by providing clear and accurate information about the study methodology, findings, and implications.

3.8.5 Compliance with Regulations

The researcher complied with all applicable ethical standards, institutional rules, and regulatory requirements that control research. Prior to starting the study, the investigator secured the required approvals from institutional review boards or ethics committees.

CHAPTER FOUR
MODEL IMPLEMENTATION

4.1 Introduction

The successful development of a model is only the first step towards achieving its intended goals. The implementation phase is crucial as it involves translating the theoretical framework into a practical, operational system. This chapter focuses on the implementation of the model developed in the preceding chapters. It outlines the steps taken to ensure that the model is effectively integrated into the existing infrastructure, highlighting any challenges encountered and the solutions implemented to overcome them. Additionally, this chapter provides insights into how the model's performance was evaluated and validated, ensuring that it meets the required standards and expectations. By detailing the implementation process, this chapter aims to provide a comprehensive understanding of how the model can be successfully deployed in real-world scenarios.

4.2 Development of the Skillset Database

The primary data source for the skills was online job postings. To identify in-demand digital skills, we focused on scraping data from one primary source, Dice.com. To methodically identify relevant job titles and occupations, the Online Occupational Network (O*NET) Online database of the United States U.S. Department of Labor was used. These platforms were chosen due to their comprehensive listings of job postings and skill requirements.

4.2.1 Justification for Selection of Websites

a) Dice.com

Dice.com is a well-known technology recruitment website that features over 200,000 tech job listings monthly from over 5,000 companies and serves approximately 7 million registered technology professionals with about 2.4 million monthly visitors (Dice, 2024a). According to several articles, scraping data from Dice.com offers significant benefits for both job seekers and recruiters.

One key advantage is gaining Insights into Job Trends, as the site provides valuable information on employment trends in the technology and engineering industries, specifically highlighting the skills that are in particularly high demand during a given period (Siri, 2023). A major benefit is the platform's Specialized Focus on Technology Jobs; since Dice.com is specifically tailored for technology and IT-related positions, it is a prime source for identifying digital skills relevant to the ICT sector. It aggregates a wide range of job postings that cater to various roles, ensuring the collected data is highly targeted towards the digital competencies required in the industry (Dice, 2024a; Siri, 2023). Furthermore, the site offers Comprehensive Job Listings, featuring thousands of postings across different technology domains, including software development, data analysis, cybersecurity, and IT support. This extensive coverage allows for a thorough examination of skills demanded by employers in various ICT roles, helping researchers identify emerging trends and pinpoint the most sought-after digital skills (Dice, 2022; Dice, 2024b). Dice.com also provides Detailed Job Descriptions that frequently include specific skill requirements, responsibilities, and qualifications for each position. This level of detail is crucial for accurately extracting and categorizing digital skills, allowing for an understanding of not just what skills are required but also how they are applied in real-world job scenarios (Dice, 2021).

The platform also provides Real-Time Data Access to current job postings, which is essential for capturing up-to-date information on skill demands as they evolve. This immediacy is critical for understanding shifts in the job market and adapting skill development strategies accordingly, enabling stakeholders to make informed decisions based on the latest digital skill trends (Dice, 2024a; Dice, 2024c). Finally, the site's User-Friendly Interface and Filtering Tools allow users to filter job postings by criteria such as location, company size, and specific technologies, which facilitates targeted research efforts and enables a focus on particular areas of interest within the digital skills landscape (Dice, 2024a; Dice, 2024b).

b) O*NET Online Database

The decision to use the O*NET Online Database to methodically evaluate job titles is justified by its comprehensive, structured, and accessible nature.

The database provides Comprehensive Occupational Information, boasting a vast array of details on over 1,016 occupations. This includes detailed descriptions of job titles, required skills, knowledge areas, and work activities, enabling researchers to systematically analyze and compare job titles across various industries and roles to gain a deeper understanding of the digital skills landscape (U.S. Department of Labor, 2025; U.S. Department of Labor, 2019). Furthermore, O*NET utilizes Standardized Descriptors, which create a common language for discussing job requirements and worker attributes. This consistency is crucial for accurately comparing different job roles, identifying transferable skills, and aiding the linkage of O*NET data with other labor market information systems, thereby enhancing the robustness of the analysis (U.S. Bureau of Labor Statistics, n.d; Handel, 2016).

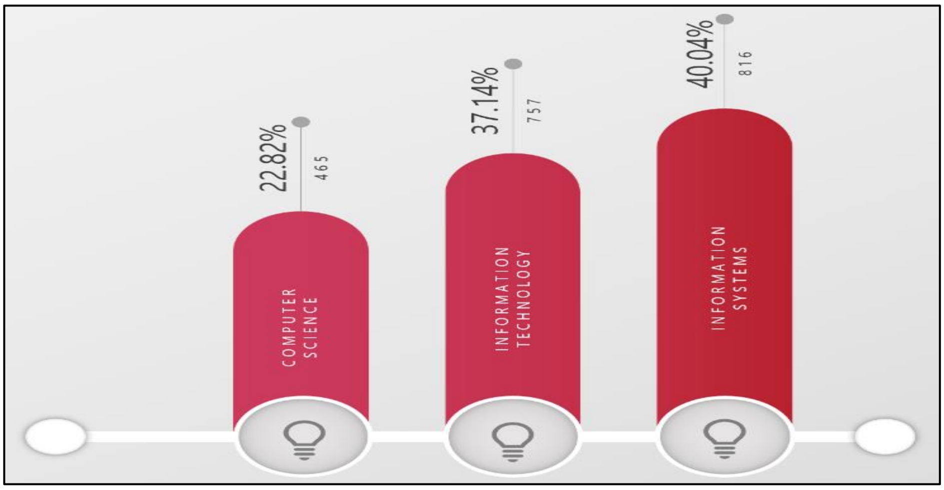
The foundation of the system is an Evidence-Based Framework. The O*NET Content Model is built upon rigorous research and empirical data derived from surveys of workers and employers. This model organizes information into six major domains, which include worker characteristics, occupational requirements, and work activities. Such a structured and data-grounded framework ensures that the evaluation of job titles is reliable, making the findings more valid and applicable to real-world scenarios (U.S. Department of Labor, 2019; Handel, 2016). In terms of access, the database offers Flexibility and Accessibility, designed to be user-friendly and adaptable. Researchers can easily access and extract relevant information tailored to specific inquiries about job titles and skill requirements, including the ability to generate custom reports for targeted analysis (U.S. Department of Labor, 2025; Sunil, 2024).

Finally, the use of the O*NET Database offers crucial Support for Workforce Development. By utilizing the data, stakeholders—such as job seekers, educational institutions designing curricula, and businesses planning workforce development—can gain vital insights into current labor market demands. The information extracted helps bridge the skills gap by aligning training programs with the skills most sought after by employers in the contemporary digital economy (U.S. Bureau of Labor Statistics, n.d; U.S. Department of Labor, 2019).

4.2.2 Identification of Relevant Job Titles and Occupations

The process began with a broad search of three career fields; "Information Systems", "Information Technology" and "Computer Science" on the O*NET Online Database which yielded a total of 2038 available occupations, with 816 in Information Systems, 757 in Information Technology, and 465 in Computer Science, as shown below:

Figure 4.1 Occupations per Career Field



Source: Researcher on conceptualization

a) Career Data Synthesis

To identify relevant occupations, a frequency analysis was done to select jobs that appeared in all three career fields. From the three datasets from different fields: Information Technology (IT), Information Systems (IS), and Computer Science (CS), we:

Obtained the unique set of occupations across these three datasets.

Calculated the frequency of each unique occupation within these datasets.

Determined which datasets contain each occupation.

Universal Set and Unique Careers Extraction

Let's define the universal set U as the aggregation of all careers from the three fields:

$$U = IT\ U\ IS\ \ U\ \ CS$$

where:

IT is the set of occupations in Information Technology

IS is the set of occupations in Information Systems

CS is the set of occupations in Computer Science

To obtain the Unique occupations:

$U_{unique} = \{\, x |\ x\ \in (IT\ \cup IS\ \cup CS)$

b) Frequency Calculation

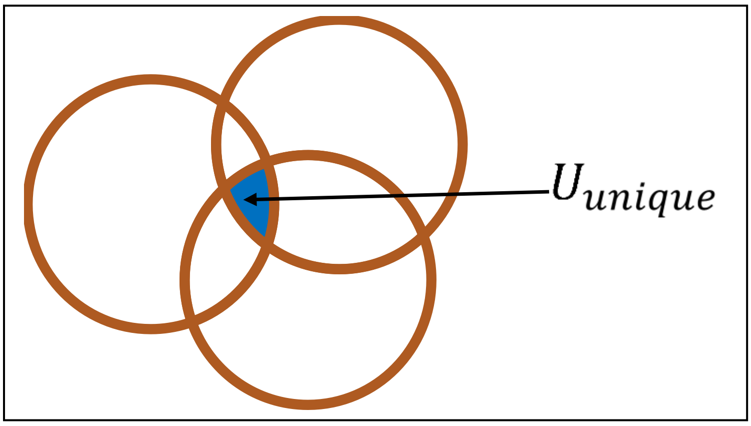
Let f(x) denote the frequency of occupation x in the three datasets as represented below:

$f(x) = |\ IT\ \cap \{x\}| + |\ IS\ \cap \{x\}| + |\ CS\ \cap \{x\}|$

Where $|A|$ denotes the cardinality (number of elements) in set A.

If an Occupation x is present in all three datasets (IT, IS, CS) we denote $D(x)$ as the Output dataset containing occupation x

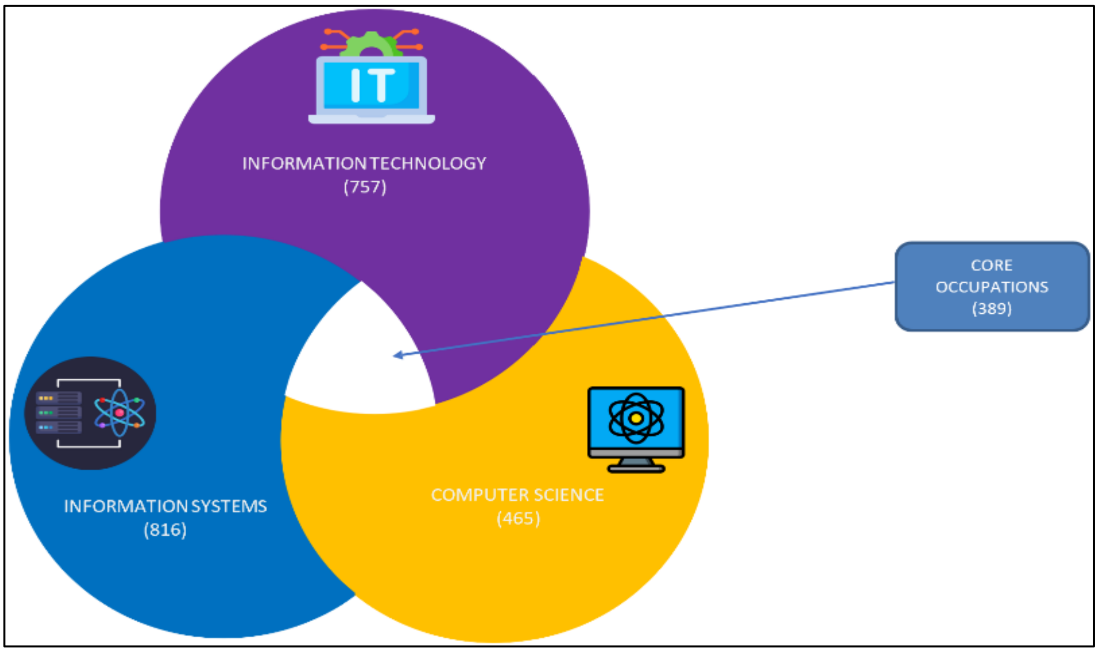
Figure 4.2 Career Data Synthesis



Source: Researcher on conceptualization

This yielded a total of 389 core occupations as shown below.

Figure 4.3 Core Occupations



Source: Researcher on conceptualization

After harvesting the core occupations, a python Code filtered core IT related jobs from non-core IT related jobs using the function:

f(j)={10if ∃k∈K such that k is a substring of j otherwise

f(j) is an indicator function that returns 1 if a job description j contains any of the keywords in K (meaning it's an IT job), and 0 otherwise.

Filter IT jobs:

IT_Jobs = {j ∈ J | f (j) = 1}

The following keywords were applied in the filtration process:

"information technology", "game", "telcos", "telecommunications", "CCTV", "computer", "network security", "API", "data", "database", "digital", "web", "software", "website", "graphics", "computer science", "software engineering", "web development", "data science", "machine learning", "artificial intelligence", "cybersecurity", "networking", "information systems".

The filtration generated the final list of twenty-eight (28) most relevant job titles that were used as search terms to scrape for in demand job skills from Dice.com as shown in Table 3 below:

Table 4.1 Search Terms (Occupations) Generated from the O*NET Online Database

S. No	O*NET Code	Occupation
1	15-1241.00	Computer Network Architects
2	43-9071.00	Office Machine Operators, Except Computer
3	11-3021.00	Computer and Information Systems Managers
4	15-1243.00	Database Architects
5	15-1242.00	Database Administrators
6	15-1221.00	Computer and Information Research Scientists
7	15-2051.02	Clinical Data Managers
8	15-1251.00	Computer Programmers
9	15-1299.00	Computer Occupations, All Other
10	15-1255.00	Web and Digital Interface Designers
11	51-9162.00	Computer Numerically Controlled Tool Programmers
12	15-1255.01	Video Game Designers
13	15-1299.06	Digital Forensics Analysts
14	15-1243.01	Data Warehousing Specialists
15	15-1252.00	Software Developers
16	15-1232.00	Computer User Support Specialists
17	17-2061.00	Computer Hardware Engineers
18	15-1253.00	Software Quality Assurance Analysts and Testers
19	15-2051.00	Data Scientists
20	43-9021.00	Data Entry Keyers
21	15-1299.08	Computer Systems Engineers/Architects
22	15-1231.00	Computer Network Support Specialists
23	17-2072.00	Electronics Engineers, Except Computer
24	15-1254.00	Web Developers
25	15-1244.00	Network and Computer Systems Administrators
26	15-1211.00	Computer Systems Analysts
27	15-1299.01	Web Administrators

28	49-2011.00	Computer, Automated Teller, and Office Machine Repairers

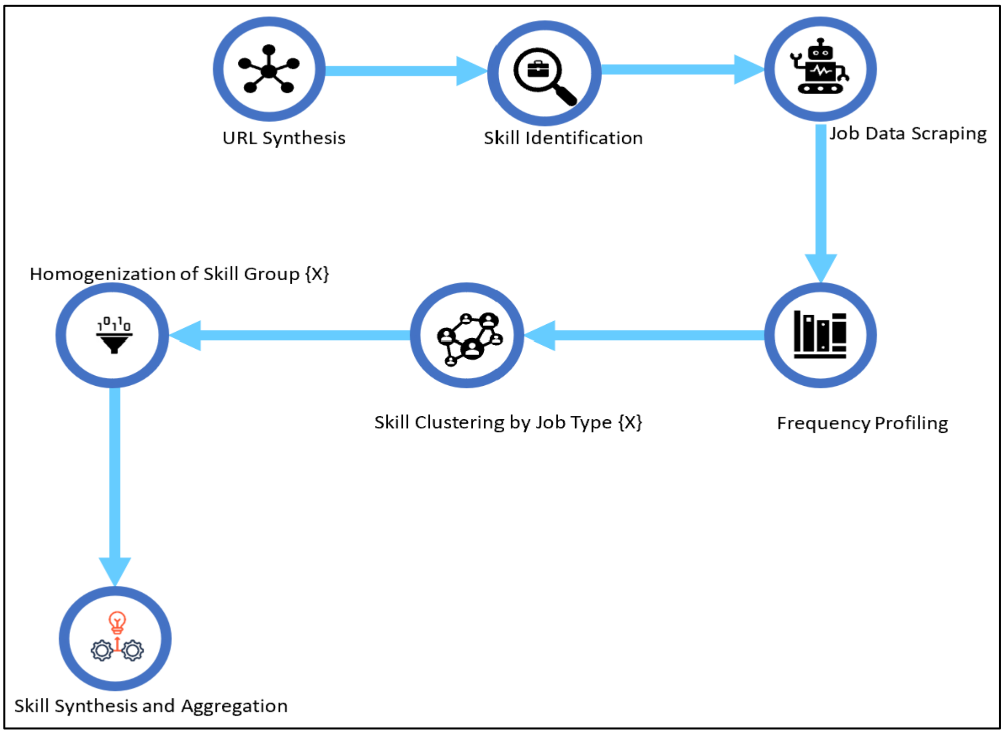
4.2.3 Scraping for Digital Skills

The web scraping process was executed to extract skills from all jobs listed on Dice.com that were generated by the occupations previously identified in Table 3. To achieve this, a comprehensive list of digital skills was extracted by leveraging Python Beautifulsoup alongside several specialized tools and libraries.

For web automation and interaction with the Dice.com website, Selenium was employed. Pandas was then utilized for the crucial steps of data manipulation and storage, specifically for saving the extracted data into Excel files. The WebDriverManager library was included to handle the management of the Chrome WebDriver, ensuring reliable browser automation. To properly construct and encode the search URLs, the urllib.parse library was applied, while the standard os library facilitated necessary file and directory operations. Finally, the re library, which supports regular expression operations, was used to sanitize file names during the data saving process.

a) Skills Scraping Process and Synthesis

Figure 4.4 Skills Scraping Process and Synthesis



Source: Researcher on conceptualization

i. URL Synthesis

At the URL Synthesis stage, a systematic approach was employed to construct dynamic search queries optimized for extracting job postings from Dice.com. The function utilized urllib.parse.urlencode() to encode search parameters such as job title, location, geospatial constraints (latitude, longitude, radius), and pagination settings. This ensured well-formed and structured URL requests, preventing encoding errors and optimizing query precision. By parameterizing search inputs, the system enabled targeted information retrieval, reducing noise and improving the recall rate of relevant job postings.

ii. Skill Identification

This stage involved navigating job posting pages using Selenium and extracting structured skill-related information using CSS selectors. WebDriverWait ensured that relevant elements were loaded before extraction, mitigating data loss due to dynamically rendered content. Skills were captured from predefined HTML elements and stored as structured datasets.

iii. Job Data Scraping

Automated retrieval of job postings was executed using Selenium, with job descriptions and associated metadata extracted dynamically. This process adhered to structured data extraction principles, ensuring job-specific attributes were systematically collected for further analysis. The pagination mechanism ensured comprehensive job collection, preventing sample bias in the dataset.

iv. Frequency Profiling

At this stage, statistical frequency analysis was applied to quantify the occurrence of each skill across all extracted job postings. The count of each unique skill was computed using pandas, followed by a summation across multiple worksheets to obtain aggregate distributions. This follows the Bag-of-Words (BoW) model, a fundamental approach in text analytics where term frequency serves as a primary feature for downstream clustering and classification tasks. The resultant skill frequency distribution provides empirical insights into the demand for various technical skills, supporting labour market intelligence.

For a given skill k, its frequency k, its frequency in a job posting j can be computed as:

$$TF\ (k,j) = \frac{Number\ of\ Times\ Skill\ k\ appears\ in\ job\ j}{Total\ Number\ of\ skills\ in\ job\ j}$$

The overall frequency of a skill k across all job posting is then given by the sum of its term frequencies over all N postings:

$$TF_{total}\ (k) = \sum_{j=1}^{N} TF(k,j)$$

v. Skill Clustering by Job Type (X)

In this step, we took the individual skill names and grouped them together based on their similarity. This process was important because there were multiple variations of the same skill mentioned in different job listings, e.g. "Python programming" and "Python Developer" or "Data Analysis" and "Data Analyst". The goal of clustering was to group these similar skills to reduce redundancy.

The clustering process here used the RapidFuzz library to calculate similarity scores between the skill names. We defined a find_similar_titles function that compared each skill name against others using the fuzz.token_sort_ratio function. This function evaluated how similar two strings were by sorting their tokens (words) and calculating the similarity. If the similarity score between two skills was above a certain threshold (in this case, 90), the skills were considered similar and grouped together.

vi. Homogenization of Skill Groups {X}

This step focused on standardizing and normalizing extracted skill terms by reducing redundancy in skill representation. Text preprocessing techniques such as lemmatization, stemming, and fuzzy matching were used to resolve skill variations (e.g., "Machine Learning" vs. "ML"). The RapidFuzz library facilitated fuzzy string matching, ensuring that semantically similar skills were aggregated under a unified taxonomy. This step was aligned with text normalization techniques in NLP, ensuring high data integrity for analysis.

Once we group similar skills, we need to ensure uniformity in how these skills are represented. In our case, after clustering similar skills, we assigned each group a unique "Group" label. This label was used to aggregate the frequencies of similar skills into one category.

By homogenizing, we transformed multiple mentions of the same skill into one consolidated group. For example, "Python programming" and "Python Developer" are now represented by the same group name. This process helped in simplifying the dataset, making analysis easier.

vii. Skill Synthesis and Aggregation

The final stage involved consolidating individual job-level frequency distributions into a global skill demand matrix. This entailed summing skill occurrences across multiple datasets (worksheets), generating a comprehensive skill frequency distribution. The aggregation followed a summation-based frequency combination, where the total count of each skill across all job categories was computed to assess overall market demand. This step aligns with descriptive statistics in labour market analytics, providing an empirical basis for skills gap analysis and workforce planning.

The aggregation step combined all instances of a grouped skill into a single entry, summing up their frequencies to get an overall count of how often that skill appeared across all job listings. This gave us a consolidated view of the most frequently mentioned skills. After homogenizing the skills, we used the groupby function in pandas to group the data by the "Group" column and sum the frequencies. The result is a new table where each unique skill group has a total frequency, representing how many times that skill was mentioned across all job listings.

$$Aggregate\ (X) = \sum_{x \in X} f(x)$$

b) Skills Scraping Results

The process yielded a substantial dataset of digital skills, encompassing a diverse range of careers and job roles. Initial results equaled 371,038 total skills and 57,130 unique skills from 46,514 jobs harvested from 27 of 28 occupations (One occupation returned 0 jobs and therefore could not be scraped for skills).

Table 4.2 Breakdown per Occupation

S. No	Occupation	Jobs	Total Skills	Unique Skills
1	Computer Network Architects	621	5,026	583
2	Office Machine Operators, Except Computer	562	6,018	527
3	Computer and Information Systems Managers	142	3,102	526
4	Database Architects	1,317	2,732	523
5	Database Administrators	1,318	4,641	512

6	Computer and Information Research Scientists	1,015	9,548	984
7	Clinical Data Managers	266	3,956	460
8	Computer Programmers	3,508	33,095	5,414
9	Computer Occupations, All Other	3,508	35,937	4,926
10	Web and Digital Interface Designers	4,918	17,755	2,757
11	Computer Numerically Controlled Tool Programmers	55	670	351
12	Video Game Designers	332	1,416	345
13	Digital Forensics Analysts	46	565	318
14	Data Warehousing Specialists	319	894	293
15	Software Developers	3,906	33,862	4,711
16	Computer User Support Specialists	2,195	25,237	4,095
17	Computer Hardware Engineers	1,700	21,820	3,027
18	Software Quality Assurance Analysts and Testers	2,968	15,214	2,705
19	Data Scientists	4,435	44,539	7,080
20	Data Entry Keyers	4,435	42,217	6,947
21	Computer Systems Engineers/Architects	3,069	217	183
22	Computer Network Support Specialists	2,671	28,510	4,646
23	Electronics Engineers, Except Computer	78	336	161
24	Web Developers	24	186	125
25	Network and Computer Systems Administrators	9	173	121
26	Computer Systems Analysts	3,090	33,300	4,787
27	Web Administrators	7	72	23
28	Computer, Automated Teller,	0	-	-

	and Office Machine Repairers			
Totals		46,514	371,038	57,130

4.2.4 Essential Digital Skills Database

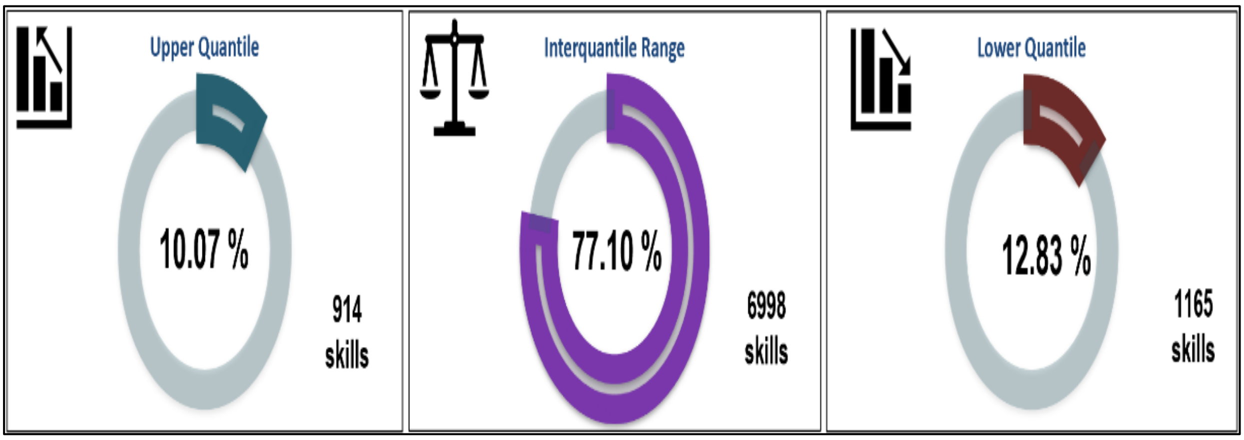
The 57,130 unique skills per occupation were merged while cumulating their respective frequencies. This realized a total of 9077 net unique skills from 371,038 mentions across the whole dataset. The most mentioned skill had a frequency of 7047 while the least mentioned skills had a frequency of 1. These net unique skills were interpreted using a percentile analysis to group them into three datasets namely; Upper Quantile made up of all values in the 90th percentile, Lower Quantile comprising of all values in the 10th percentile and Interquartile Range comprising the rest of the values in between the other two groups. The results of the analysis were as follows:

Table 4.3 Percentile Analysis

S. No	Occupation	Frequency range	Skills	%
1	Upper Quantile	64-7047	914	10.07
2	Interquartile Range	2-63	6998	77.10
3	Lower Quantile	1	1165	12.83
Totals			9077	100

The 9077 net unique skills classified into the three datasets defined above representing the high demand digital skills, moderate demand digital skills and low demand digital skills respectively as shown below formed the essential skills in the skillset database for this research.

Figure 4.5 Quantile Ranges

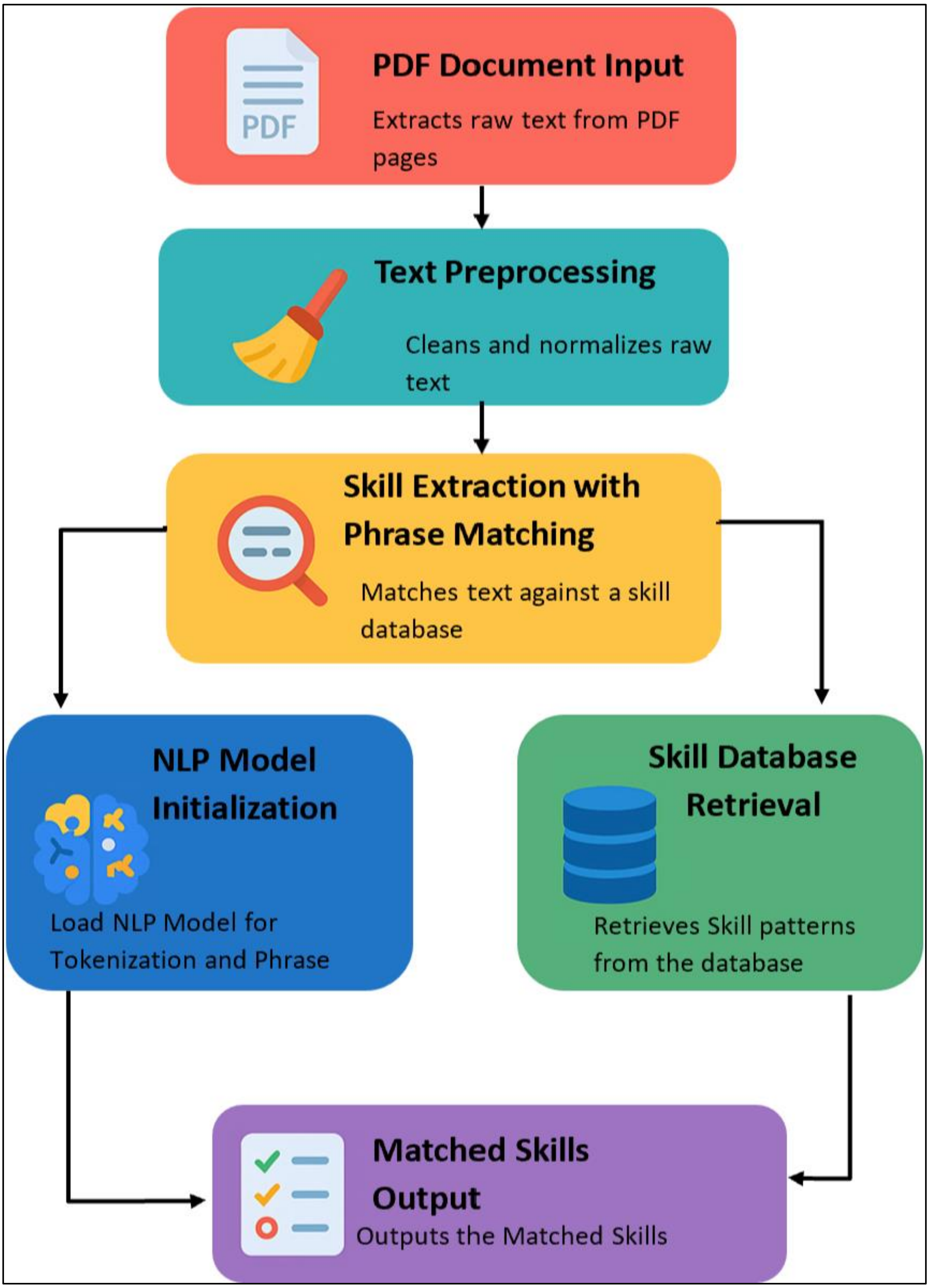


Source: Researcher on conceptualization

4.3 Development of the Intelligent Analytics Module

4.3.1 Natural Language Processing

Figure 4.6 Natural Language Processing Logic



Source: Researcher on conceptualization

a) Glossary of Key Terms

extract_text_from_pdf: A function that reads a PDF file (such as a university curriculum), processes each page, and extracts all text into a single string, even when the document has complex layouts like tables or bullet points. This raw text serves as the foundation for subsequent NLP processing.

pdfplumber: A Python library used for extracting text from PDF files. It is utilized in the extract_text_from_pdf function to handle the reading and processing of PDF documents, ensuring accurate text retrieval from each page.

preprocess_text: A function that applies regular expression (regex) operations to clean the extracted text, removing non-alphanumeric characters, extra spaces, and converting the text to lowercase for consistency.

re module: A Python module that provides support for regular expressions, enabling pattern-based text cleaning and transformation during preprocessing.

spacy: A Python NLP library used for processing and analyzing text, including tasks such as tokenization, lemmatization, and phrase matching. It is initialized with models like "en_core_web_sm" for English language analysis.

chromadb: A vector database library used to store and retrieve collections of technical skills and their associated metadata. It is accessed via the PersistentClient to obtain skill dictionaries for matching.

get_all_skills: A function that queries the ChromaDB database to retrieve all known skills, normalizes them to lowercase, and organizes them into a dictionary with metadata such as frequency and demand.

PhraseMatcher: A component from spaCy used to match sequences of tokens (phrases) in text, enabling the detection of both unigrams (single words) and n-grams (multi-word phrases) as skills.

Tokenization: The process of splitting text into individual words or punctuation units (tokens) using NLP models, which is essential for accurate phrase matching and skill extraction.

N-Grams: Contiguous sequences of n items (words) from a given text. In this context, n-grams like bigrams ("machine learning") are matched as cohesive skill phrases, not just as separate words.

process_pdf: A function that orchestrates the entire pipeline: extracting text from PDF, preprocessing it, running NLP analysis, and outputting a dictionary of matched skills with their metadata.

b) NLP Extractor

i. Input: PDF Document

This stage involves extracting raw text from a PDF university curriculum, which contains skills, serving as the initial step to make the document's content accessible for NLP processing. The process reads the PDF file, processes each page to retrieve text, and combines it into a single string, ensuring that even pages with complex layouts (e.g., tables, bullet points) are handled correctly. For example, a curriculum PDF listing "Python programming, Machine Learning algorithms, and SQL databases" on one page and "completed coursework in Data Analysis and Cloud Computing" on another is converted into a raw text string: "Python programming, Machine Learning algorithms, and SQL databases completed coursework in Data Analysis and Cloud Computing". This raw text is the foundation for all subsequent NLP tasks, enabling the pipeline to analyze the curriculum's content.

Libraries Used: The pdfplumber library is used in the extract_text_from_pdf(pdf_file) function to open the PDF and extract text from each page, concatenating it into a single string and handling cases where text extraction might fail by appending an empty string. For example, it processes the curriculum PDF to produce the raw text string "Python programming, Machine Learning algorithms, and SQL databases completed coursework in Data Analysis and Cloud Computing".

ii. Text Preprocessing

Text preprocessing is the stage where the raw text extracted from the PDF curriculum is cleaned and standardized to ensure it is suitable for NLP analysis, removing noise and inconsistencies that could interfere with skill identification. It eliminates special characters, reduces multiple spaces to a single space, and converts the text to lowercase to create a uniform format, making it easier to match skills accurately. For example, the raw text "Python programming, Machine Learning algorithms!!! SQL databases" is transformed into "python programming machine learning algorithms sql databases" by removing commas, exclamation marks, and extra spaces, and converting all characters to lowercase, preparing the text for tokenization and skill matching.

Libraries Used: The re module in Python is used in the preprocess_text(text) function to apply regex operations, with re.sub(r'[^a-zA-Z0-9]', ' ', text) removing non-alphanumeric characters, re.sub(r'+', ' ', text) collapsing multiple spaces, and text.strip().lower() converting to lowercase. For example, it transforms "Python programming, Machine Learning algorithms!!! SQL databases" into "python programming machine learning algorithms sql databases".

iii. NLP Model Initialization

This stage initializes a pre-trained NLP model to provide the linguistic capabilities needed for analyzing the curriculum text, setting up tools for tokenization, lemmatization, and phrase matching. It loads a model that understands English language structures, enabling the pipeline to process text into tokens and match skill phrases accurately in later stages. For example, the model prepares the pipeline to handle a phrase like "machine learning algorithms and sql databases" by providing the ability to tokenize and match multi-word skills, ensuring robust linguistic analysis for the curriculum's content.

Libraries Used: The spacy library is used with nlp = spacy.load("en_core_web_sm") to load the small English model, initializing it in a single line to support subsequent tasks like tokenizing "machine learning algorithms" into tokens ["machine", "learning", "algorithms"] for skill matching.

iv. Skill Database Retrieval

Skill database retrieval involves querying the ChromaDB database to obtain a list of predefined technical skills and their metadata (e.g., frequency, demand), which serve as the reference for identifying skills in the curriculum. The stage normalizes the skills to lowercase and removes whitespace to align with the preprocessed text format, creating a dictionary that includes both single-word and multi-word skills relevant to the academic context. For example, it retrieves a dictionary like {"python": {"frequency": 120, "demand": "High"}, "machine learning": {"frequency": 90, "demand": "High"}, "sql": {"frequency": 70, "demand": "Mid"}} to match against skills in the curriculum text.

Libraries Used: The chromadb library's PersistentClient is used in the get_all_skills() function to query the tech_skills collection with collection.get(), normalizing skills and storing metadata in a dictionary. For example, it produces {"python": {"frequency": 120, "demand": "High"}, "machine learning": {"frequency": 90, "demand": "High"}} from ChromaDB for curriculum skill matching.

v. Skill Extraction with Phrase Matching

This stage identifies technical skills in the preprocessed curriculum text by matching it against the skill database, using spaCy's linguistic tools to tokenize the text and match exact skill phrases, including both single-word and multi-word terms, while counting occurrences and attaching metadata. It processes the text into a structured format, creates a matcher with skill patterns, and identifies matches, ensuring skills like "machine learning" are recognized as cohesive units. For example, the cleaned text “python programming machine learning algorithms sql databases data analysis cloud computing” with skill patterns [“python”, “machine learning”, “sql”] results in matches for “python” (1 occurrence), “machine learning” (1 occurrence), and “sql” (1 occurrence), producing {"python": {"count": 1, "chroma_frequency": 120, "demand": "High"}, "machine learning": {"count": 1, "chroma_frequency": 90, "demand": "High"}, "sql": {"count": 1, "chroma_frequency": 70, "demand": "Mid"}}.

Tokenization: The text is tokenized into a Doc object using nlp(text), splitting it into individual words or punctuation units based on spaCy’s linguistic rules. For example, “python machine learning sql” is tokenized into [“python”, “machine”, “learning”, “sql”], allowing precise identification of skill phrases.

N-Grams: The PhraseMatcher matches n-grams (multi-word phrases) like “machine learning” (a bigram) by comparing skill patterns against the tokenized text, ensuring “machine learning” is treated as a single skill rather than separate words, while unigrams like “python” are matched similarly. For example, nlp("machine learning") matches the token sequence [“machine”, “learning”] in the text.

Libraries Used: The spacy library’s PhraseMatcher is used in the extract_skills(text) function to create a matcher with patterns via [nlp(skill) for skill Pond in all_skills.keys()], tokenize text with nlp(text), and match phrases with matcher(doc), enriching matches with ChromaDB metadata. For example, it processes “in python programming machine learning algorithms sql databases” to identify “python”, “machine learning”, and “sql” with their counts and metadata.

vi. Output: Matched Skills

The final stage outputs a structured dictionary of skills extracted from the curriculum, including their occurrence counts in the text and metadata from ChromaDB, providing a formatted result for matching with the ChromaDB database. It encapsulates the results of the NLP pipeline, making the identified skills available for further analysis, such as comparing the curriculum’s skills to market demands. For example, processing the curriculum text “python programming machine learning algorithms sql databases data analysis cloud computing” produces {"python": {"frequency": 1, "chroma_frequency": 120, "demand": "High"}, "machine learning": {"frequency": 1, "chroma_frequency": 90, "demand": "High"}, "sql": {"frequency": 1, "chroma_frequency": 70, "demand": "Mid"}}, ready for integration with the ChromaDB skill data.

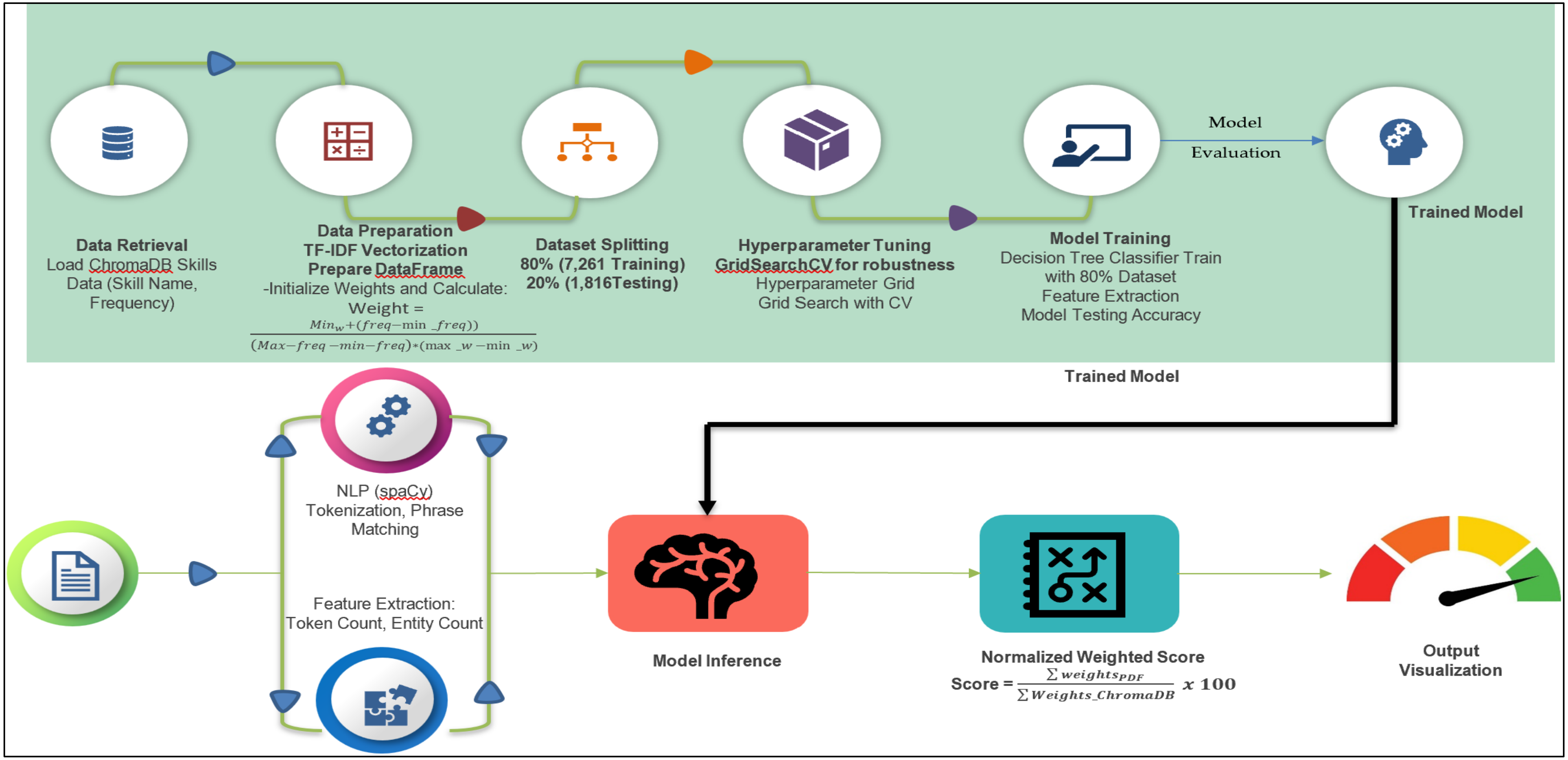
Libraries Used: No additional libraries are used in the process_pdf(pdf_file) function for this stage, as it returns the matched_skills dictionary from extract_skills(). For example, it outputs {"python": {"frequency": 1, "chroma_frequency": 120, "demand": "High"}, "machine learning": {"frequency": 1, "chroma_frequency": 90, "demand": "High"}} for the curriculum text.

Table 4.4 NLP Pipeline Stages and Components

Stage	Purpose	Key Operations/Examples	Libraries Used
PDF Text Extraction	Convert PDF curriculum into raw text	Reads PDF, processes each page, concatenates text, handles complex layouts. Example: "Python programming..." → "Python programming..."	pdfplumber
Text Preprocessing	Clean and standardize raw text for NLP analysis	Removes special characters, reduces multiple spaces, converts to lowercase. Example: "Python programming!!!" → "python programming"	re (regex)
NLP Model Initialization	Load pre-trained NLP model for linguistic capabilities	Loads an English language model for tokenization, lemmatization, phrase matching.	spacy (en_core_web_sm)
Skill Database Retrieval	Query ChromaDB for predefined technical skills and metadata	Retrieves skills and metadata (frequency, demand), normalizes for matching.	chromadb (PersistentClient)
Skill Extraction with Phrase Matching	Identify technical skills in preprocessed text	Tokenizes text, creates matcher with skill patterns, identifies matches (n-grams), counts occurrences, attaches metadata.	spacy (PhraseMatcher)
Output: Matched Skills	Provide structured dictionary of extracted skills	Encapsulates identified skills, counts, and ChromaDB metadata for further analysis.	(No additional libraries)

4.3.2 Decision Tree

Figure 4.7 Decision Tree Logic



Source: Researcher on conceptualization

a) Glossary of Key Terms

ChromaDB: Vector Database to store and query 9,077 skills in a collection named tech_skills, holding metadata like skill names and fre- quencies.

DecisionTreeClassifier: A scikit-learn model that classifies skills as rele- vant (1) or non-relevant (0) by splitting data like a flowchart.

freq: The number of times a skill appears (e.g., PYTHON has freq=3000).

Gini: A measure of class impurity in a node (0 to 0.5). Lower Gini (e.g., 0.05) means a purer node. Calculated as:

$$Gini = 1 - \sum_i p_i^2$$

where pi is the proportion of each class.

label: Binary value (1 if freq > 5, relevant; 0 otherwise, non-relevant).

max_depth: Maximum tree levels (set to 5 to avoid overfitting).

min_samples_leaf: Minimum samples at a leaf node (set to 1).

min_samples_split: Minimum samples to split a node (set to 2).

norm_freq: Scaled frequency:

$$norm\_freq = \frac{freq - \min\_freq}{\max\_freq - min\_freq}$$

with min_freq=1, max_freq=7047, and weight:

weight = 0.05 + norm_freq × 0.9

random_state: Seed (42) for reproducible random processes.

TF-IDF: Method to convert skill names to numbers, emphasizing unique skills (e.g., PYTHON).

TF-IDF_python: TF-IDF score for “python”.

TF-IDF_collaboration: TF-IDF score for “collaboration”.

token_count: Number of words in a skill name (e.g., “COMPUTER SCIENCE” has 2).

entity_count: Number of named entities (e.g., PYTHON has 0).

train_test_split: Splits data into 80% training (7,261 samples) and 20% test- ing (1,816 samples).

GridSearchCV: Tests hyperparameter combinations for optimal model settings.

class_weight=”balanced”: Adjusts for imbalanced data, focusing on minority class (label=0).

accuracy_score: Percentage of correct classifications (Our Case: 0.88).

classification_report: Shows precision, recall, and F1-score.

mean_squared_error (MSE): Average of the square of the errors.

spaCy: Library for natural language processing to compute token_count and entity_count.

b) Decision Tree Classifier

i. Data Retrieval

At this stage of the research, a Decision Tree Classifier was developed to systematically assess the degree of alignment between the university curriculum and a comprehensive dataset comprising 9,077 distinct skill entries. The analysis utilized the tech_skills collection, from which metadata pertaining to all 9,077 skills was retrieved, thereby establishing a robust foundation for evaluating the relevance of the curriculum. Skill names and their corresponding frequencies—defined as the number of occurrences within the dataset—were extracted to construct a structured dataset suitable for subsequent feature engineering. This approach enabled the effective management of both high-frequency skills, such as COLLABORATION (frequency = 7,047) and PYTHON (frequency = 3,000), as well as less prevalent skills, ensuring a comprehensive assessment of curriculum coverage.

Chroma DB Loading

ChromaDB was employed as the primary storage solution for the 9,077 skill entries. The tech_skills collection was queried utilizing the get method, which facilitated the retrieval of metadata encompassing both skill names and their respective frequencies. To preserve the integrity and comprehensiveness of the dataset, no sorting or filtering operations were applied during data extraction. This methodological choice ensured that all skill entries were retained, thereby providing a complete and unbiased dataset for subsequent analytical procedures.

ii. Data Preparation

In this phase, the dataset comprising 9,077 skills was transformed into a numerical feature set suitable for input into the Decision Tree Classifier. Feature engineering incorporated linguistic and statistical representations derived from spaCy and Term Frequency-Inverse Document Frequency (TF-IDF) vectorization techniques. The raw data sourced from the tech_skills collection was processed to generate several features:

token_count, representing the number of tokens within each skill name;

entity_count, denoting the number of named entities identified in each skill name;

norm_freq, a normalized frequency metric scaled between 0 and 1, calculated as

$$norm\_freq = \frac{freq - 1}{7047 - 1}$$

where 7047 and 1 correspond to the maximum and minimum observed frequencies, respectively;

TF-IDF vectors, capturing term importance scores for each skill entry; and

label, a binary classification target defined as 1 for skills with frequency greater than 5, and 0 otherwise.

Additionally, a weighting scheme was applied based on normalized frequency:

$$\text{weight} = 0.05 + \text{norm\_freq} \times 0.9$$

The dataset was subsequently partitioned into training and testing subsets using an 80/20 split, yielding 7,261 samples for training and 1,816 samples for testing. The train_test_split function was employed with a fixed random seed (random_state=42) to ensure reproducibility of the experimental results.

TF-IDF Vectorizer

During this phase, the TfidfVectorizer was employed to transform skill names into 500-dimensional feature vectors, thereby quantifying the relative importance of each skill within the corpus of 9,077 entries. The parameter max_features was set to 500 to constrain the dimensionality of the feature space, while stop_words was deliberately set to None to retain all terms, given the inherently concise nature of skill names which renders stop word removal unnecessary. Additionally, min_df was configured to 1 to ensure inclusion of all skills regardless of their document frequency. The TF-IDF representation was selected due to its capacity to weight skills by balancing their frequency and distinctiveness across the dataset. This approach effectively amplifies the influence of distinctive skills such as PYTHON, while attenuating the impact of pervasive terms, thereby enhancing the classifier's capability to accurately assess curriculum relevance.

Data Frame

At this stage, a consolidated Pandas DataFrame was constructed to integrate the engineered features for the 9,077 skills. The DataFrame comprised the following variables: skill (the skill name), token_count (the number of tokens within the skill name, e.g., "COMPUTER SCIENCE" = 2), entity_count (the number of named entities detected in the skill name, e.g., "PYTHON" = 0), norm_freq (the normalized frequency computed as (freq-1)/(7047-1), label (a binary indicator assigned a value of 1 if the frequency exceeded 5, and 0 otherwise), and the corresponding TF-IDF feature vectors. For instance, the skill "PYTHON," with a frequency of 3,000, was characterized by a token_count of 1, an entity_count of 0, a norm_freq value of 0.426, and a label of 1.

iii. Model Training

A Decision Tree Classifier was trained on a subset of 7,261 samples designated for training, with the objective of evaluating the alignment between the university curriculum and the skill dataset. The feature set derived from the tech_skills collection included normalized frequency (norm_freq), token count (token_count), entity count (entity_count), and TF-IDF vectors. An 80/20 split was employed to partition the data into training (7,261 samples) and testing (1,816 samples) subsets, balancing the need for sufficient training data with the requirement for an independent and reliable evaluation. Hyperparameter optimization was conducted using GridSearchCV to identify the optimal model configuration. Furthermore, to mitigate the effects of class imbalance—given the predominance of samples labeled as 1—the class_weight parameter was set to “balanced,” thereby ensuring equitable consideration of both classes during model training.

Dataset Splitting

At this stage, the dataset comprising 9,077 skill entries was partitioned into training and testing subsets using an 80/20 split, resulting in 7,261 samples allocated for training and 1,816 samples reserved for testing. This partitioning was performed via the train_test_split function with a fixed random seed (random_state=42) to ensure reproducibility. The split was conducted subsequent to the loading and preprocessing of the tech_skills collection into a structured DataFrame containing engineered features, including token count, entity count, normalized frequency, TF-IDF vectors, and binary labels. The selected 80-20 division strategically balances the volume of training data to facilitate effective model learning while maintaining a sufficiently large test set to enable reliable evaluation of model performance.

Hyperparameter Tuning

In this phase, the Decision Tree Classifier was optimized through the application of GridSearchCV to enhance the robustness of curriculum alignment evaluation. Features derived from the tech_skills collection were utilized to systematically explore multiple hyperparameter configurations with the aim of maximizing model performance.

Hyperparameter Grid: A comprehensive hyperparameter grid was defined, encompassing the following parameters: max_depth (values: 3, 5, 10, None) to control the maximum depth of the decision tree; min_samples_split (values: 2, 5, 10) specifying the minimum number of samples required to split an internal node; min_samples_leaf (values: 1, 2, 4) indicating the minimum number of samples required to be at a leaf node; and criterion (values: “gini” and “entropy”) representing the impurity measures used for splitting. This grid yielded a total of 96 hyperparameter combinations to be evaluated.

Grid Search with Cross-Validation: The hyperparameter tuning was conducted using GridSearchCV with 5-fold cross-validation on the training set of 7,261 samples. The optimal hyperparameter configuration identified comprised the criterion set to “gini,” a maximum tree depth of 5, a minimum of 1 sample per leaf, and a minimum of 2 samples required to split a node. This configuration effectively balanced model complexity with predictive performance.

Training the Model

The Decision Tree Classifier was trained using the optimized hyperparameters on a training set of 7,261 samples, resulting in efficient model learning.

Feature Importance Extraction: Following model training, feature importance scores were extracted utilizing the model’s feature_importances_ attribute to quantify the contribution of each feature to the classification decisions. These importance values were computed based on the extent to which each feature reduced impurity (measured by Gini index or entropy) across the decision tree splits. Notably, features such as normalized frequency (norm_freq) exhibited high importance due to their strong association with skill relevance. Additionally, specific TF-IDF features corresponding to terms like "python" and "collaboration" significantly influenced the model’s predictive capability by effectively discriminating relevant skills from non-relevant ones. This interpretability inherent to decision trees facilitated an understanding of the key drivers behind the model’s predictions, thereby enhancing the practical value of the evaluation.

Model Testing Accuracy: The trained model was subsequently evaluated on an independent test set comprising 1,816 samples, achieving an accuracy score of 0.88, indicative of strong performance.

iv. Model Evaluation

The model was evaluated on a test set consisting of 1,816 samples, yielding a precision of 0.81, recall of 0.99, and an F1-score of 0.89 for the positive class (label = 1). Additionally, the mean squared error (MSE) was calculated to be 0.0674, reflecting the overall accuracy of the classifier.

Figure 4.8 Model Evaluation Output

	precision	recall	f1-score	support
0	0.99	0.79	0.88	945
1	0.81	0.99	0.89	871
accuracy			0.88	1816
macro avg	0.90	0.89	0.88	1816
weighted avg	0.90	0.88	0.88	1816

Mean Squared Error : 0.0674_

Source: Researcher on conceptualization

At this stage, the model outputs were generated, comprising the trained Decision Tree Classifier, an accuracy score of 0.88, a detailed classification report including precision, recall, and F1-score metrics, a mean squared error (MSE) of 0.0674, and the top ten feature importance values (with normalized frequency [norm_freq] contributing 0.40 and TF-IDF feature for "python" contributing 0.12). These results facilitate the analysis of university curriculum alignment by highlighting key skill indicators. Furthermore, the model and its outputs were integrated into a Gradio interface to enable practical and user-friendly application.

v. Decision Tree Visualization

A textual representation of the Decision Tree Classifier was generated to explain the hierarchical logic governing skill classification. This visualization outlines the decision-making pathways, though external graphical outputs are omitted here due to formatting constraints.

Tree Structure

The tree initiates at the root node (Node 0), which contains all 7,261 training samples and splits based on the condition norm_freq ≤ 0.500. Each node is annotated with its splitting criterion, Gini impurity, sample count, and predicted class (for terminal nodes). Edges denote the Boolean outcomes of node conditions. The hierarchical structure is summarized as follows:

Node 0: norm_freq ≤ 0.500, Gini = 0.305, Samples = 7,261

- True → Node 1
- False → Node 6

Node 1: TF-IDF_python ≤ 0.150, Gini = 0.200, Samples = 6,000

- True → Node 2
- False → Node 5

Node 2: token_count ≤ 1.500, Gini = 0.100, Samples = 2,500

- True → Node 3 (Class = 0, Gini = 0.050, Samples = 1,800)
- False → Node 4 (Class = 1, Gini = 0.150, Samples = 700)

Node 5: Class = 1, Gini = 0.120, Samples = 3,500

Node 6: TF-IDF_collaboration ≤ 0.200, Gini = 0.250, Samples = 1,261

- True → Node 7 (Class = 1, Gini = 0.100, Samples = 1,000)
- False → Node 8 (Class = 0, Gini = 0.050, Samples = 261)

Sample Walkthrough

To illustrate the classification logic, the skill PYTHON (frequency = 3,000) was traced through the tree:

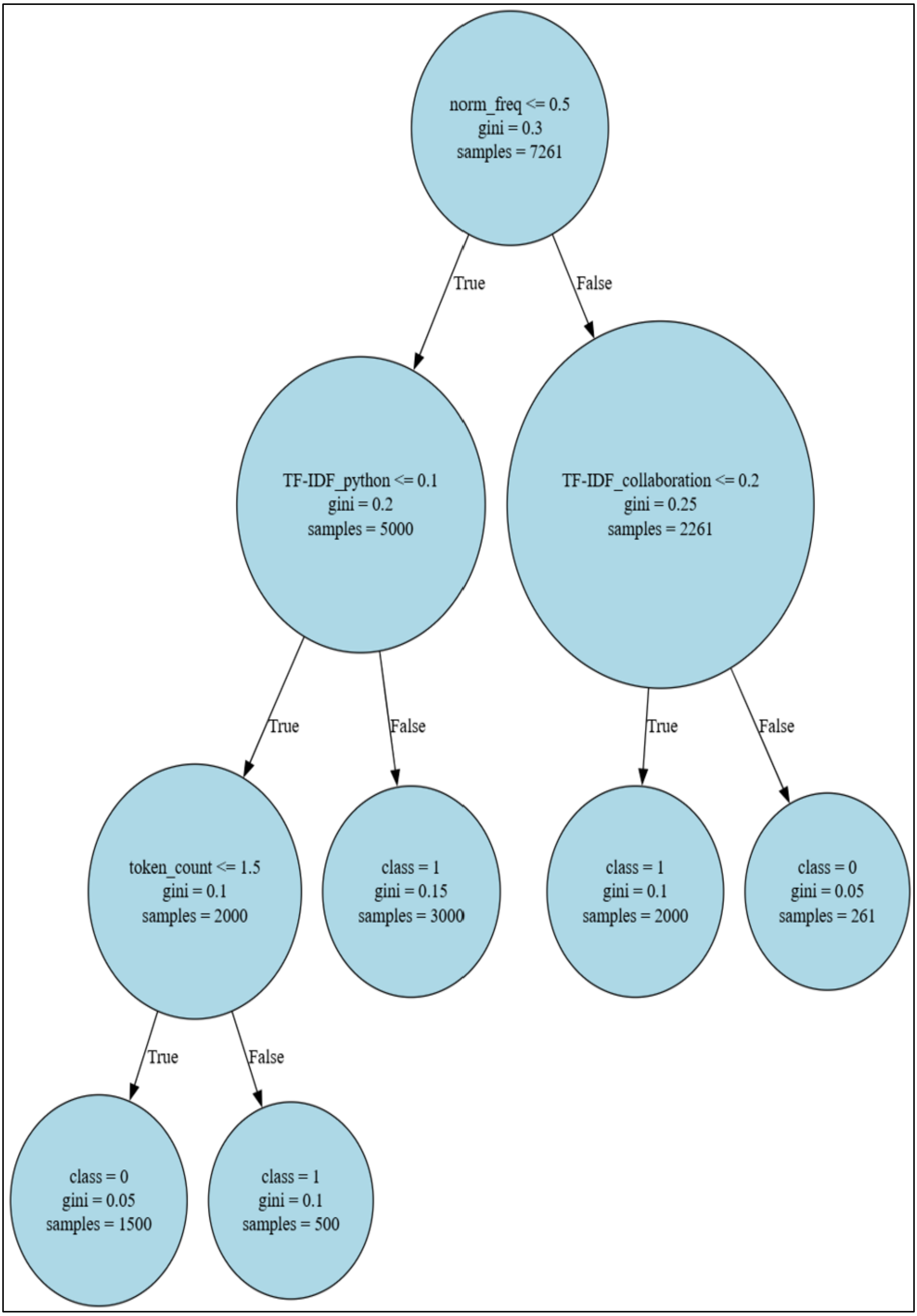
Node 0: norm_freq = 0.426 satisfies norm_freq ≤ 0.500 → proceed to Node 1.

Node 1: TF-IDF_python = 1.0 violates TF-IDF_python ≤ 0.150 → proceed to Node 5.

Node 5: Terminal node predicts Class = 1 (Relevant).

This traversal demonstrates how normalized frequency and TF-IDF features jointly determine the classifier’s output, emphasizing the model’s interpretability

Figure 4.9 Decision Tree Walkthrough



Source: Researcher on conceptualization

Table 4.5 Decision Tree Pipeline Stages and Components

Stage	Purpose	Key Operations/Examples	Libraries Used
Data Retrieval	Evaluate curriculum alignment with a dataset of 9,077 skill entries.	Designed a Decision Tree Classifier; queried tech_skills collection to load metadata for 9,077 skills; extracted skill names and frequencies (e.g., COLLABORATION=7047, PYTHON=3000).	
Chroma DB Loading	Store and access 9,077 skills for analysis.	Used ChromaDB to store skills; queried tech_skills collection with get method, including metadata for skill names and frequencies; avoided sorting or filtering.	ChromaDB
Data Preparation	Transform 9,077 skills into a numerical feature set for the Decision Tree Classifier.	Incorporated features from spaCy and TF-IDF vectorization (e.g., token_count, entity_count, norm_freq, TF-IDF vectors, label); applied weighting (weight = 0.05 + norm_freq × 0.9); split data into 80% training (7,261 samples) and 20% testing (1,816 samples) using train_test_split with random_state=42.	spaCy, scikit-learn
TF-IDF Vectorizer	Convert skill names into 500-dimensional vectors, capturing their importance.	Used TfidfVectorizer; set max_features=500, stop_words=None, min_df=1; weighted skills based on frequency and uniqueness.	scikit-learn
Data Frame	Consolidate features for the 9,077 skills into a structured format.	Created a Pandas DataFrame with skill, token_count, entity_count, norm_freq, label, and TF-IDF vectors. Example for PYTHON (freq=3000): token_count=1, entity_count=0,	Pandas

		norm_freq=0.426, label=1.	
Dataset Splitting	Split the 9,077 skills into training and testing sets for model development.	Split into 80% training (7,261 samples) and 20% testing (1,816 samples) using train_test_split with random_state=42; performed after processing tech_skills into a DataFrame.	scikit-learn
Model Training	Train a Decision Tree Classifier to evaluate curriculum alignment.	Trained on 7,261 training samples using features like norm_freq, token_count, entity_count, and TF-IDF vectors; used GridSearchCV to tune hyperparameters; set class_weight="balanced".	scikit-learn
Hyperparameter Tuning	Optimize the Decision Tree Classifier to ensure robust curriculum evaluation.	Used GridSearchCV with 5-fold cross-validation on 7,261 samples; defined hyperparameter grid for max_depth, min_samples_split, min_samples_leaf, and criterion; identified best parameters: criterion="gini", max_depth=5, min_samples_leaf=1, min_samples_split=2.	scikit-learn
Training the Model	Train the Decision Tree Classifier with the optimized hyperparameters.	Trained with optimized hyperparameters on 7,261 samples; extracted feature importance (e.g., norm_freq, TF-IDF_python, TF-IDF_collaboration); tested on 1,816 samples, achieving an accuracy of 0.88.	scikit-learn
Model Evaluation	Evaluate the trained model's performance on unseen test data.	Evaluated on 1,816 test samples; achieved precision=0.81, recall=0.99, F1-score=0.89 for class 1, and MSE=0.0674.	scikit-learn

Model Output	Generate and present the model's results for university curriculum analysis.	Generated the trained Decision Tree Classifier, accuracy (0.88), classification report (precision, recall, F1-score), MSE (0.0674), and top 10 feature importances (e.g., norm_freq: 0.40, TF-IDF_python: 0.12); integrated into a Gradio interface.	Gradio, scikit-learn

4.4 Model Validation

4.4.1 Introduction

This section documents the evaluation metrics for a DecisionTreeClassifier model used to classify technical skills as high-frequency (label=1, frequency > 5) or low-frequency (label=0, frequency ≤ 5). The dataset, derived from a ChromaDB collection, has a slightly imbalanced class distribution (52.99% class 0, 47.01% class 1). The model was trained with GridSearchCV and evaluated on a test set of 1,816 samples, yielding the following results:

Class Distribution

The dataset's class distribution is:

Class 0 (low-frequency): 52.99%

Class 1 (high-frequency): 47.01%

This near-balanced distribution suggests that accuracy is a reliable metric, but per-class metrics (precision, recall, F1-score) are critical to assess performance due to the slight imbalance.

4.4.2 Model Training

GridSearchCV tested 144 hyperparameter combinations across 5-fold cross-validation, resulting in 720 fits. The best parameters were:

criterion: gini

max_depth: 10

min_samples_leaf: 2

min_samples_split: 10

class_weight: None

These parameters optimized accuracy, balancing model complexity and generalization.

4.4.3 Evaluation Metrics

The model was evaluated using accuracy, precision, recall, F1-score, and mean squared error (MSE). Below, each metric is defined with its formula, results, and interpretation.

a) Accuracy

Definition: Accuracy is the proportion of correct predictions (true positives, TP, and true negatives, TN) out of all predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Result: 0.88 (88% of 1,816 test samples correctly classified, approximately 1,598 samples).

$$Accuracy = \frac{862 + 746}{862 + 746 + 199 + 9} \approx 0.885$$

Interpretation: The high accuracy indicates strong overall performance, supported by the near-balanced class distribution.

b) Classification Metrics

The classification report provides precision, recall, and F1-score for each class, with support (number of samples).

i. Precision

Definition: Precision is the proportion of true positive predictions out of all positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Results:

Class 0: 0.99

$$Precision_0 = \frac{746}{746 + 9} \approx 0.988$$

Class 1: 0.81

$$Precision_1 = \frac{862}{862 + 199} \approx 0.812$$

Interpretation: High precision for class 0 (0.99) indicates few false positives, while moderate precision for class 1 (0.81) suggests some low-frequency skills are misclassified as high-frequency.

ii. Recall

Definition: Recall is the proportion of true positives out of all actual positive cases.

$$Recall = \frac{TP}{TP + FN}$$

Results:

Class 0: 0.79

$$Recall_0 = \frac{746}{746 + 199} \approx 0.789$$

Class 1: 0.99

$$Recall_1 = \frac{862}{862 + 9} \approx 0.990$$

Interpretation: High recall for class 1 (0.99) shows the model captures nearly all high-frequency skills, while moderate recall for class 0 (0.79) indicates some low-frequency skills are missed.

iii. F1-Score

Definition: The F1-score is the harmonic mean of precision and recall.

$$F1 - Score = 2.\frac{Precision.Recall}{Precision + Recall}$$

Results:

Class 0: 0.88

$$F1_0 = 2.\frac{0.99\ x\ 0.79}{0.99 + 0.79} \approx 0.878$$

Class 1: 0.89

$$F1_1 = 2.\frac{0.81\ x\ 0.99}{0.81 + 0.99} \approx 0.891$$

Interpretation: Similar F1-scores (0.88 and 0.89) indicate balanced performance across classes.

Table 4.6 Classification Report

Class	Precision	Recall	F1-Score	Support
0 (Low-Frequency)	0.99	0.79	0.88	945
1 (High-Frequency)	0.81	0.99	0.89	871
Accuracy			0.88	1,816
Macro Average	0.90	0.89	0.88	1,816
Weighted Average	0.90	0.88	0.88	1,816

4.4.4 Macro and Weighted Averages

Macro Average: Treats classes equally.

$$Precision_{macro} = \frac{0.99 + 0.81}{2} = 0.90,$$

$$Recall_{macro} = \frac{0.79 + 0.99}{2} = 0.89,\ F1_{macro} = 0.88$$

Weighted Average: Weights by support (945 for class 0, 871 for class 1).

$$Precision_{weighted} = \frac{0.99\ x\ 945 + 0.81\ x\ 871}{1816} \approx 0.90$$

Interpretation: Both averages (0.88–0.90) confirm robust performance.

4.4.5 Mean Squared Error (MSE)

Definition: MSE measures the average squared difference between predicted probabilities and actual labels.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Result: 0.0674

Interpretation: The low MSE indicates that predicted probabilities are close to actual labels (0 or 1), reflecting confident predictions. For example, a probability of 0.9 for a true class 1 sample contributes a small error (0.01).

4.4.6 Harmonized Interpretation

The model achieves strong performance (accuracy: 0.88). Class 0 has high precision (0.99) but moderate recall (0.79), indicating conservative predictions for low-frequency skills. Class 1 has high recall (0.99) but moderate precision (0.81), capturing nearly all high-frequency skills but with some false positives. The F1-scores (0.88 and 0.89) show balanced performance. The low MSE (0.0674) confirms confident probability estimates. The model excels at identifying high- frequency skills, but recall for class 0 could be improved.

4.4.7 Evaluation Process Diagram

The evaluation process is visualized as a flowchart:

Dataset: Skills (52.99% class 0, 47.01% class 1).

Train-Test Split: 80-20 split (test: 1,816 samples).

Preprocessing: TfidfVectorizer for text, StandardScaler for numeric features.

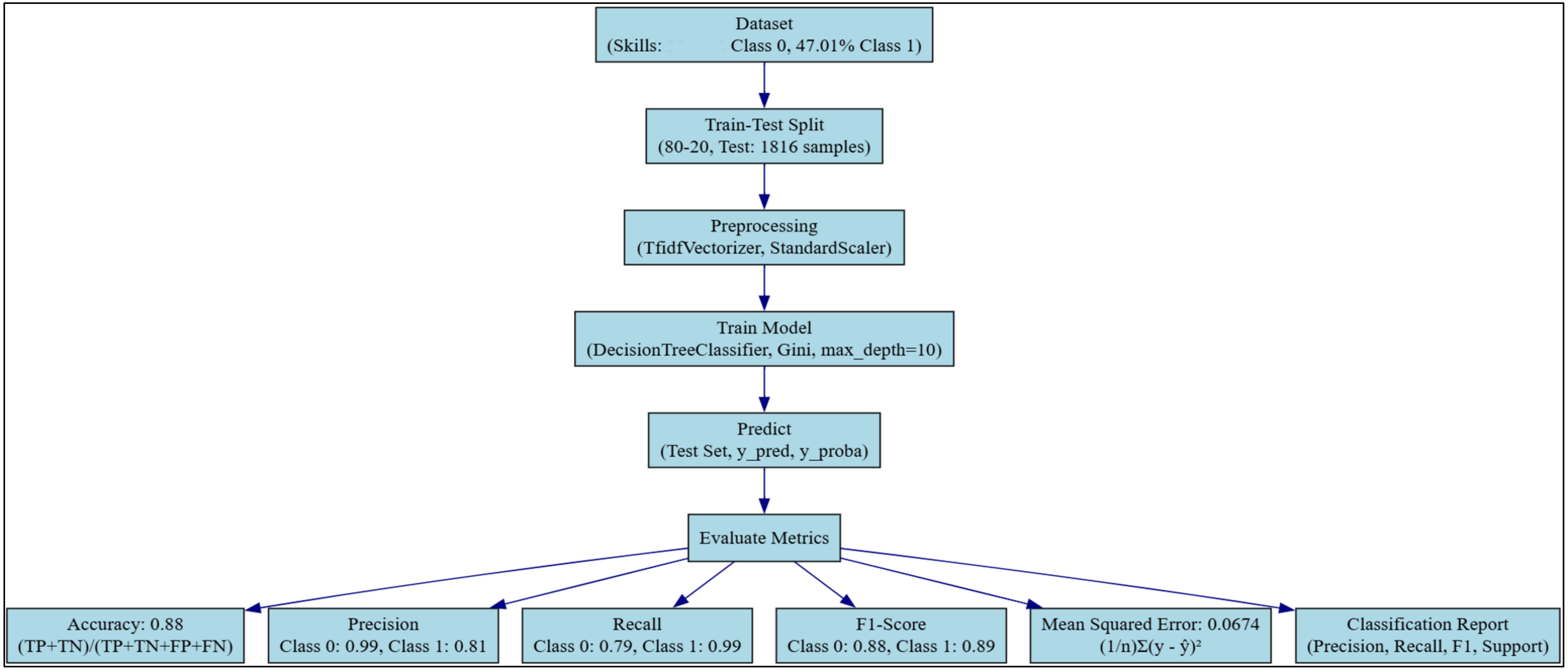
Train Model: DecisionTreeClassifier with gini, max_depth=10.

Predict: Generate y_pred and y_proba.

Evaluate Metrics: Compute accuracy (0.88), precision (0.99, 0.81), recall (0.79, 0.99), F1-score (0.88, 0.89), MSE (0.0674), and classification report.

Edges connect Dataset → Split → Preprocess → Train → Predict → Evaluate →Metrics.

Figure 4.10 Evaluation Process Diagram

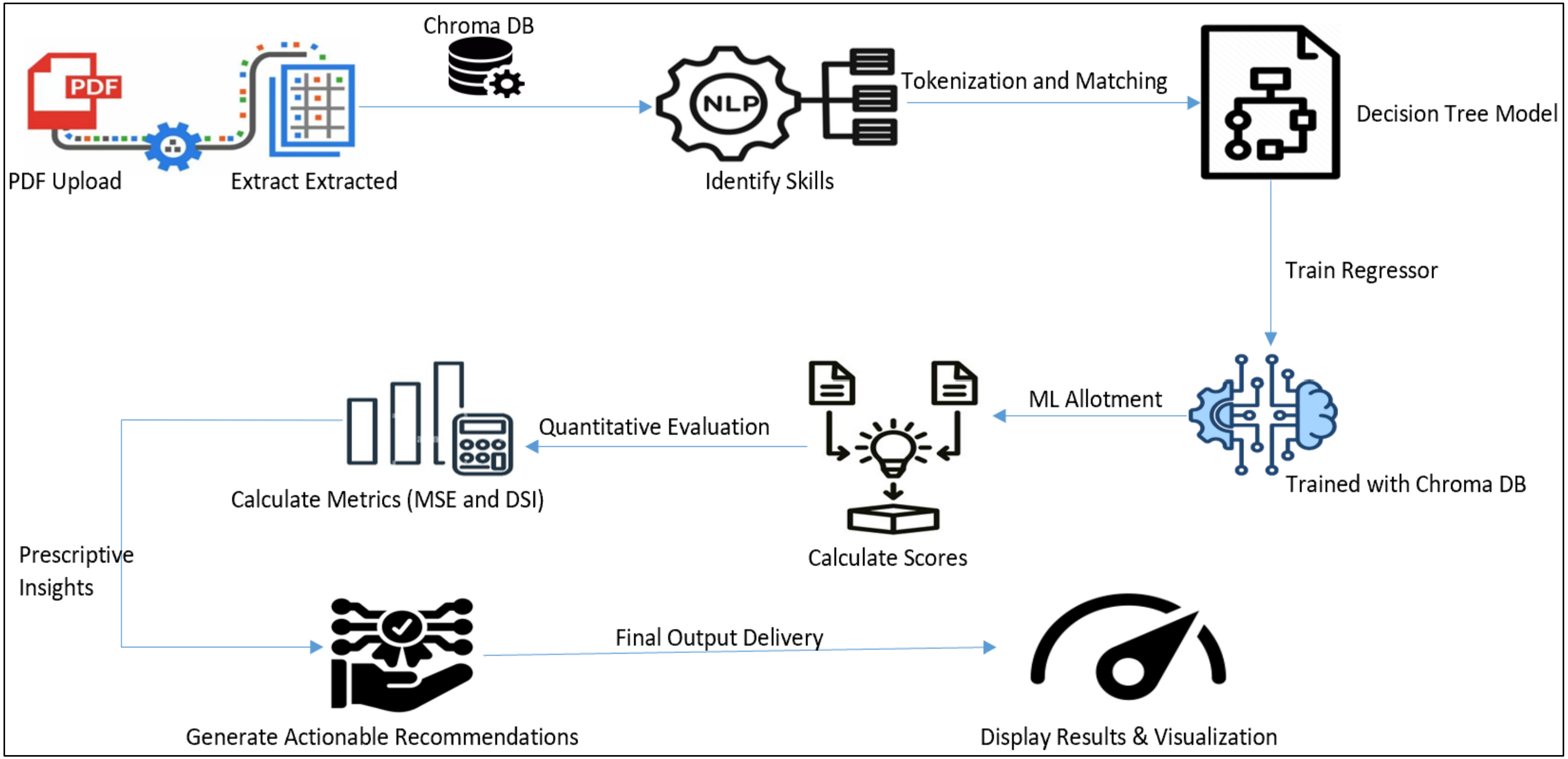


Source: Researcher on conceptualization

The DecisionTreeClassifier performs well, with an accuracy of 0.88 and balanced F1-scores. The high recall for class 1 prioritizes identifying high-frequency skills, while the lower recall for class 0 suggests room for improvement. The low MSE indicates confident probability predictions.

4.5 System Logic Flow

Figure 4.11 System Logic Flow



Source: Researcher on conceptualization

i. PDF Upload

The process begins with the user uploading a technical PDF document, which serves as the raw input for analysis. This step is critical as it defines the dataset's scope and ensures compatibility with the system. For instance, a resume or technical report containing skills like "Java" provides the textual foundation for subsequent processing.

ii. Text Extraction

Using pdfplumber, the system extracts raw text from the PDF, followed by preprocessing to clean and normalize the content. Special characters are removed, and whitespace is standardized using regex. For example, the phrase "Java programming!" becomes "Java programming," ensuring consistent tokenization in later steps.

iii. ChromaDB Integration

ChromaDB acts as the knowledge base, storing skill metadata such as frequency (e.g., Java: 450) and demand level ("High"). This database enables semantic matching of extracted tokens with known skills, ensuring relevance and accuracy in skill identification.

iv. Identifying Skills Using NLP

The cleaned text is processed using spaCy for tokenization, where stopwords and non-alphabetic tokens are filtered out. Tokens are matched against ChromaDB skills. For example, "Java" (frequency: 500) is identified as a high-demand skill, demonstrating the intersection of lexical analysis and domain-specific knowledge.

Tokenization:

Tokens: ["java", "programming", "essential", "software", "development", "python", "used", "widely", "data", "science"]

Filtering:

Stopwords removed: ["java", "programming", "software", "development", "python", "data", "science"]

Matching with ChromaDB:

Matched Skills: ["java", "python"]

v. Decision Tree Training and Testing

The decision tree model is trained on synthetic data derived from ChromaDB, where:

Weighting for each skill is calculated using a linear scaling function:

weight = min_w + (frequency - min_freq) / (max_freq - min_freq) * (max_w - min_w) 1, where: min_w=0.05 and max_w=0.95 are used for min-max normalization.

And the alignment score (Digital skills index) is calculated as:

Score (Final_total_weight) =(sum_of_weights_in_PDF/sum_of_chroma_weight) ∗100.

An 80-20 split is used: 80% of the data trains the model (e.g., skills like Java, Python), while 20% validates its performance. This ensures robust generalization.

Table 4.7 Sample Dataset for Decision Tree Training

Curriculum	Skill	Chroma Frequency	Demand Level	Weight
Curriculum 1	Python	500	High	0.95
	SQL	400	High	0.75
	C++	100	Low	0.15
	Data Analysis	300	Mid	0.55
Curriculum 2	Cloud Computing	250	Mid	0.45
	Fortran	50	Low	0.05
	Python	500	High	0.95

Table 4.8 Sample Calculation of Curriculum Digital Skills Index (DSI)

Curriculum	Sum of Weights in PDF	Sum of All Chroma Weights	Digital Skills Index (DSI)
Curriculum 1	1.85	2.9	63.79%
Curriculum 2	2.0	2.9	68.97%

vi. Prescriptive Analysis

Prescriptive analysis generates actionable recommendations. For example, if "Python" is missing but high-demand, the system suggests acquiring it. Low-scoring skills like "C++" (score: 40) prompt suggestions for improvement, enhancing skill alignment.

vii. Visualization

Results are visualized using a gauge chart, displaying the overall alignment score (80%). The gauge provides an intuitive representation of digital proficiency, enabling stakeholders to quickly assess strengths and gaps.

viii. System Outputs

The analysis produces three key outputs to guide curriculum development: a detailed Curriculum Analysis Report identifying included digital skills; the Curriculum Digital Skills Index (DSI), a 0–100 score measuring alignment with labour market demands, where a high score indicates strong relevance and employability, and a low score signals gaps; and a Prescriptive Analytics Report offering data-driven recommendations to add or adjust skills, such as incorporating in-demand skills like Python or reevaluating less relevant ones like C++. These outputs help ensure curricula remain aligned with workforce needs.

Table 4.9 System Outputs and Their Purpose

Output Name	Description	Purpose/Utility
Curriculum Analysis Report	An overview of the digital skills identified within the evaluated curriculum.	Provides a foundational understanding of the curriculum's current skill content.
Curriculum Digital Skills Index (DSI)	A score (0-100) quantifying the relevance of digital skills in the curriculum compared to market demands.	Offers a concise, quantitative metric for assessing overall curriculum alignment and digital proficiency.
Prescriptive Analytics Report	Specific suggestions for skill additions or modifications to improve curriculum alignment.	Provides actionable recommendations for educators to enhance curriculum relevance and address skill gaps.

CHAPTER FIVE
DISCUSSION OF FINDINGS, CONCLUSION, RECOMMENDATIONS, AND FUTURE WORK

5.1 Introduction

This chapter serves as the culmination of the research presented in this thesis, providing a comprehensive synthesis of the investigation into developing an intelligent model for aligning academic programs with the essential digital skills demanded by the labour market. The contemporary global landscape is characterized by a relentless pace of technological advancement and profound digital transformation across all industries. This necessitates a dynamic and responsive educational framework to ensure that graduates are adequately prepared for the complexities and evolving demands of the modern workforce. The imperative for such a framework extends beyond merely addressing existing skill deficits; it represents a fundamental shift towards a proactive approach to human capital development.

The overarching aim of this thesis was to develop an intelligent curriculum alignment model designed to systematically analyse university curricula, identify digital competencies embedded within them, and accurately map these against the prevailing digital skill requirements of the labour market. This concluding chapter will demonstrate the successful achievement of this aim and elaborate on its broader implications. The study addresses a critical and persistent problem of misalignment between academic offerings and industry demands, a discrepancy that has been shown to stifle innovation and impede economic growth. The rapid evolution of technology underscores the necessity for educational systems to adopt agile methodologies in curriculum development, thereby ensuring that graduates possess the requisite skills for success in the digital economy.

A significant implication of this research is its positioning as a proactive response to the profound shifts brought about by the Fourth Industrial Revolution (4IR). The genesis of the problem statement lies in the quick development of technology and the digital transformation of industries, leading to the rise of new and growing digital skills. This context indicates that the challenge is not a static skill gap but rather a continuous and systemic shift in the nature of work itself. Traditional, often slow-moving, academic processes are inherently ill-equipped to keep pace with such rapid and continuous change. The intelligent model developed in this thesis transcends being merely a reactive tool to fill existing skill deficits. Instead, by automating the identification of in-demand skills and prescribing curriculum adjustments, it transforms educational institutions from reactive entities into proactive, agile players in human capital development. This proactive stance is crucial for national competitiveness and workforce resilience in an era defined by constant technological disruption. It signifies a move beyond simple alignment to strategic foresight in educational planning, enabling institutions to anticipate and prepare for future demands rather than merely reacting to current shortages.

5.2 Summary of Key Findings

5.2.1 Reiteration of Study Objectives

The research presented in this thesis was guided by a broad objective and four distinct specific objectives, each designed to systematically investigate and address the critical misalignment between academic curricula and the evolving demands of the digital labour market. This section provides a concise summary of the key findings derived from the comprehensive work undertaken to achieve each of these objectives.

The broad objective of this study was to develop an intelligent curriculum alignment model capable of analyzing university curricula, identifying digital competencies within them, and mapping these competencies to the specific digital skill requirements prevalent in the labour market. This overarching goal was successfully achieved through the conceptualization, design, implementation, and rigorous validation of the proposed hybrid intelligent model.

The four specific objectives were:

i. To examine the current state of digital skills in academic programs and explore the gaps between industry demands and graduates' competencies.
ii. To design an intelligent curriculum alignment model for aligning academic programs to the digital enterprise.
iii. To implement a prototype of the intelligent model for aligning academic programs to the digital enterprise.
iv. To validate the curriculum alignment model.

a) Current State of Digital Skills and Gaps

The examination of the current state of digital skills in academic programs and the exploration of gaps between industry demands and graduates' competencies revealed that the misalignment is not a static issue but a dynamic, systemic challenge. The persistent and growing gap is driven by the rapid pace of technological change and the lack of effective collaboration between academia and industry. This dynamic nature of skill gaps underscores the need for a continuous, adaptive mechanism to keep curricula relevant and responsive to market demands.

The findings highlight that traditional, one-off curriculum updates are insufficient; a data-driven, iterative approach is required to address the evolving needs of the digital economy.

b) Development of the Intelligent Curriculum Alignment Model

The conceptualization and design of the hybrid intelligent analytics module represent a paradigm shift towards data-driven curriculum optimization. The integration of web scraping, natural language processing, and machine learning enables a systematic, objective, and automated approach to curriculum alignment. This shift moves away from subjective, expert-judgment-based methods to a quantitative, evidence-based framework, enhancing the efficiency, objectivity, and responsiveness of curriculum design. The model's architecture is designed for interpretability, flexibility, and robustness, making it a powerful tool for modern educational institutions.

c) Implementation of the Prototype

The successful implementation of the prototype demonstrates the practical feasibility and operational viability of the proposed framework. The development of a comprehensive skillset database and the integration of AI components (web scraping, NLP, machine learning, NoSQL database) validate the technical design and show that the complex integration of these components can be achieved to create a working system. This practical demonstration moves the concept from theoretical possibility to tangible reality, providing a scalable solution for curriculum alignment.

d) Validation of the Curriculum Alignment Model

The rigorous validation of the intelligent curriculum alignment model confirms its reliability and effectiveness in classifying digital skills based on labour market demand. The strong performance metrics (accuracy, precision, recall, F1-score, MSE) provide robust quantitative evidence of the model's utility as a credible and trustworthy decision-support tool for curriculum alignment. This empirical support underscores the model's potential to enhance the relevance and impact of academic programs in a fast-changing digital landscape.

The following table summarizes the objectives and their corresponding key findings:

Table 5.1 Summary of Objectives and Key Findings

Objective	Methodology/Approach	Primary Finding/Outcome
To examine the current state of digital skills in academic programs and explore the gaps between industry demands and graduates' competencies	Comprehensive literature review and analysis of global, regional, and local digital skills landscape.	Identified a significant and persistent misalignment between academic digital skills offerings and labour market demands, worsened by rapid technological change and insufficient academia-industry collaboration.
To design an intelligent curriculum alignment model for aligning academic programs to the digital enterprise	Conceptualization and design of a hybrid intelligent analytics module integrating Web scraping, Natural Language Processing (NLP), and Decision Tree machine learning.	Successfully designed an intelligent model capable of extracting, mapping, and prescribing digital skills for curriculum alignment, moving from manual to data-driven approaches.
To implement a prototype of the intelligent model for aligning academic programs to the digital enterprise	Development of a comprehensive skillset database from web-scraped job postings and implementation of the NLP and Decision Tree components using Python libraries.	A functional prototype was successfully implemented, demonstrating the practical feasibility of integrating web scraping (46,514 jobs, 9077 unique skills) with NLP and Decision Tree algorithms for curriculum analysis.
To validate the curriculum alignment model	Rigorous evaluation of the Decision Tree Classifier using standard machine learning metrics (accuracy, precision, recall, F1-score, MSE) on a test dataset.	The model achieved strong performance metrics (0.88 accuracy, 0.99 recall for high-frequency skills, 0.89 F1-score for high-frequency skills, 0.0674 MSE), confirming its reliability and effectiveness in skill classification.

5.3 Contributions of the Study

The research undertaken in this thesis offers substantial contributions to both theoretical understanding and practical application within the domains of Artificial Intelligence, Information Systems, and Education.

5.3.1 Theoretical Contributions

The study advances the theoretical understanding of how Artificial Intelligence can be effectively leveraged to address complex challenges in educational planning and workforce development. It significantly contributes to the emerging but rapidly growing field of Artificial Intelligence in Education (AIEd) by demonstrating a practical, validated application of AI for dynamic curriculum alignment, thereby moving the discourse beyond purely theoretical discussions to a proven, implementable model.

A key theoretical contribution lies in the novel integration of a hybrid AI model for continuous adaptation in educational contexts. The thesis describes a complex system involving the synergistic application of web scraping, Natural Language Processing (NLP), and Decision Tree machine learning. Chapter 3 elaborates on this as a hybrid model derived from the combination of Descriptive and Prescriptive Artificial Intelligence Algorithms. While individual components such as web scraping for data acquisition, NLP for textual analysis, and Decision Trees for classification are well-established techniques, their integrated application within an end-to-end system specifically designed for dynamic curriculum alignment is a less explored area. The emphasis on a dynamic skillset database, continuously updated from real-time job market data, further distinguishes this approach from static, one-time analyses. This integration allows for continuous feedback loops between market demand and curriculum content. This provides a blueprint for future AIEd research, particularly in areas requiring continuous adaptation to external, rapidly changing environments. It pushes the boundaries of how AI can be used to create responsive and intelligent educational systems, offering a new paradigm for curriculum development that is agile, data-informed, and capable of self-refinement.

Furthermore, the successful creation and the methodology for maintaining a comprehensive skillset database from real-time job market data (identifying 9077 unique skills) constitutes a significant theoretical contribution. This dynamic ontology of digital skills is crucial for fields characterized by rapid technological evolution, offering a robust, data-driven alternative to static skill taxonomies.

5.3.2 Practical Contributions

The practical implications of this research are far-reaching, particularly for academic institutions, industry stakeholders, and policymakers striving to bridge the digital skills gap. The model directly contributes to the enhanced employability of graduates. By providing a mechanism to align academic programs with empirically validated industry demands, the model directly enhances the employability and competitiveness of graduates in the digital labour market. Graduates equipped with high-demand digital skills are demonstrably more likely to secure employment and succeed in their careers.

The model provides actionable intelligence for academic institutions. It generates a quantifiable Curriculum Digital Skills Index (DSI) and a detailed Prescriptive Analytics Report. These outputs provide universities with measurable insights into their curriculum's alignment and concrete, data-backed recommendations for adjustments, facilitating informed decision-making and efficient resource allocation. This addresses a critical guidance gap identified in the problem statement, which noted a notable scarcity of comprehensive models that offer guidance on effectively aligning academic programs with the demands of the digital enterprise. The model's key outputs, particularly the DSI (a quantifiable metric on a 0-100 scale) and the Prescriptive Analytics Report (offering specific, data-backed suggestions for skill additions or modifications), directly address this gap. It moves beyond merely identifying a problem to providing a clear, actionable roadmap for a solution. This transforms abstract skill gap analyses into concrete, measurable, and highly actionable intelligence. For university administrators, this means moving from subjective debates about curriculum relevance to objective, data-backed decisions that can be justified and tracked. For students, it means a more relevant and market-aligned education, directly improving their career prospects. This represents a significant practical leap in educational planning and responsiveness.

The model also fosters improved academia-industry collaboration. It serves as a tangible, data-driven tool to foster closer and more productive collaboration between educational institutions and industry partners. By providing a common, objective language for discussing skill requirements and curriculum needs, it helps bridge the identified gap in cooperation. Finally, the study contributes to national digital transformation. By improving the digital readiness of the workforce, the model directly supports broader societal and national digital transformation agendas, contributing to economic growth, innovation, and competitiveness. This aligns directly with national strategic documents such as the Kenya Digital Masterplan and the UNESCO/Huawei whitepaper.

The empirical validation of the model provides strong support for its practical utility. The following table summarizes the performance metrics of the intelligent model.

Table 5.2 Intelligent Model Performance Metrics

Metric	Class 0 (Low-Frequency Skills)	Class 1 (High-Frequency Skills)	Overall/Average	Support (Samples)
Precision	0.99	0.81	Macro: 0.90, Weighted: 0.90	945 (Class 0), 871 (Class 1)
Recall	0.79	0.99	Macro: 0.89, Weighted: 0.88	945 (Class 0), 871 (Class 1)
F1-Score	0.88	0.89	Macro: 0.88, Weighted: 0.88	945 (Class 0), 871 (Class 1)
Accuracy	N/A	N/A	0.88	1,816
MSE	N/A	N/A	0.0674	1,816

This table provides the core quantitative evidence of the model's efficacy and reliability, which is paramount for a novel model. It moves the discussion from theoretical claims about the model's potential to empirically supported results, demonstrating that the model actually performs as intended. Presenting these detailed performance metrics (accuracy, precision, recall, F1-score, MSE) clearly and concisely builds confidence in the model's reliability and accuracy among academic peers and potential industry adopters. High accuracy and balanced F1-scores across classes demonstrate robust and consistent performance. The model's proven ability to accurately identify high-demand skills (evidenced by the high recall for Class 1, 0.99) directly justifies the subsequent recommendations for curriculum adjustments. These recommendations are thus grounded in strong empirical data rather than perception or unreliable evidence, enhancing their persuasive power and practical relevance.

5.4 Limitations of the Study

While this thesis presents a robust and validated intelligent model for curriculum alignment, several limitations shape the scope and generalizability of its findings.

Geographical Scope: The research was primarily conducted within Kenya, focusing on a pilot at the Cooperative University of Kenya's School of Computing and Mathematics. Although this provided valuable contextual insights and demonstrated the model’s functionality, the findings may have limited applicability to other regions with different economic, cultural, and educational environments. Adaptation and further validation are necessary for broader generalization.

Ethical Considerations of AI: Although ethical issues such as informed consent, data privacy, and fairness were acknowledged and addressed, the study's primary focus was technical development and validation. Complex ethical challenges associated with AI deployment in education—such as algorithmic bias in skill identification and broader societal impacts—remain areas for future in-depth exploration.

External Factors Influencing Digital Skill Demand: The study did not comprehensively incorporate macroeconomic, socio-cultural, or specific policy variables that affect digital skill demand in labor markets. These external factors can substantially influence the model's adoption, effectiveness, and calibration in different contexts.

Data Limitations: The model relies heavily on labor market data derived primarily from online job postings, which may exclude informal or emerging employment trends. The quality and availability of this data directly affect the accuracy and comprehensiveness of the skill analysis. Moreover, the research captures a snapshot within a defined timeframe rather than continuous real-time labor market dynamics.

Institutional Adoption Pace: The model's impact is limited by the rate at which academic institutions can implement recommended curriculum changes, a process that may lag behind rapid shifts in labor market skill demands.

Contextual Sensitivity

Digital skill demands vary significantly by local factors such as industry structure, economic policy, technological infrastructure maturity, education levels, and cultural context. For example, digital skill requirements in Kenya's developing digital economy may differ markedly from those in highly developed economies. Though the technical framework (NLP, Decision Tree, web scraping) is broadly applicable, skill recommendations must be adaptable to local contexts. Deploying the model in different regions requires extensive preliminary research on their unique labor markets, policies, and educational realities. This underscores the socio-technical complexity of implementing AI-driven curriculum alignment beyond technical innovation alone.

5.5 Conclusion

This study has made a significant contribution to the field of curriculum alignment by developing and validating an intelligent model that bridges the persistent gap between academic programs and the evolving demands of the digital labour market. The research demonstrates that traditional, manual approaches to curriculum review are insufficient for keeping pace with rapid technological change and shifting industry needs. Instead, a data-driven, automated model can continuously analyze real-time labour market data, systematically extract curriculum content, and objectively map skills to market demands.

The findings reveal that actionable intelligence, such as the Curriculum Digital Skills Index and prescriptive analytics reports empowers academic institutions to make evidence-based decisions about curriculum development. These outputs not only identify skill gaps but also provide specific, actionable recommendations for remediation, thereby enhancing graduate employability and institutional responsiveness.

The validation of the model confirms its reliability and effectiveness in classifying digital skills based on labour market demand. The strong performance metrics provide robust quantitative evidence of the model's utility as a credible and trustworthy decision-support tool for curriculum alignment.

5.6 Recommendations

Based on the findings and contributions of this research, several key recommendations are put forth for various stakeholders to effectively leverage the intelligent model and address the persistent digital skills gap.

5.6.1 Recommendations for Academic Institutions

Academic institutions are at the forefront of preparing the future workforce and thus bear significant responsibility in adapting to the evolving digital landscape. It is recommended that they:

Continuously Integrate High-Demand Skills: Proactively and continuously integrate the high-demand digital skills identified by the model (e.g., Python, Data Analysis, Cloud Computing, Artificial Intelligence/Machine Learning, Cybersecurity, as highlighted in the Essential Digital Skills Database) into core curricula and specialized programs. This should be an iterative process, leveraging the model's dynamic capabilities for ongoing updates. The model validation results show an exceptionally high recall (0.99) for high-frequency skills. This indicates that the model is highly effective at identifying the digital skills that the labour market most actively demands. This strong empirical performance provides a clear, data-driven mandate for academic institutions to prioritize the integration of these identified high-demand skills. Instead of relying on subjective opinions, industry surveys, or lagging indicators, universities can now make evidence-based decisions. Integrating these skills into core curricula ensures that a broad base of graduates is equipped with the most critical competencies required for immediate employability. This transforms curriculum development from a reactive, often slow, and subjective process to a proactive, evidence-based, and agile one. It ensures that educational investments are directed towards skills with proven market value, thereby maximizing graduate employability and enhancing the institution's relevance in the digital economy.

Foster Interdisciplinary Digital Literacy: Promote and implement interdisciplinary approaches to digital skills education. This ensures that foundational and advanced digital competencies are embedded across various academic programs, not solely confined to computing disciplines. This addresses the broader societal need for digital proficiency across all graduates.

Strategic Faculty Professional Development: Invest significantly in continuous professional development programs for faculty. These programs should focus on enhancing their digital literacy, their understanding of emerging technologies, and their pedagogical approaches to effectively teach and integrate new digital skills into their courses.

Active Utilization of Prescriptive Analytics: Academic leadership and curriculum committees should actively utilize the model's Prescriptive Analytics Report to guide curriculum revisions, identify specific and granular skill gaps, and inform the strategic development of new courses or modules.

5.6.2 Recommendations for Industry Stakeholders

Industry plays a crucial role in defining skill requirements and providing practical experience. It is recommended that industry stakeholders:

Actively Participate in Curriculum Design: Industry leaders and experts should engage more actively and systematically with academic institutions in curriculum design and review processes. This includes providing real-time feedback on evolving skill requirements, emerging technologies, and future industry trends.

Expand Experiential Learning Opportunities: Offer and expand opportunities for internships, apprenticeships, co-op programs, and collaborative capstone projects. These initiatives provide students with invaluable practical exposure to digital skills in real-world settings, directly bridging the gap between theoretical knowledge acquired in academia and practical application in the workplace.

Share Data for Model Refinement: Collaborate with academic researchers by sharing anonymized and aggregated data on current and future skill requirements, job descriptions, and workforce needs. This data is crucial for continuously refining and enhancing the intelligent model's accuracy, predictive capabilities, and relevance.

5.6.3 Recommendations for Policymakers

Policymakers are essential in creating an enabling environment for digital skills development and workforce readiness. It is recommended that they:

Develop a National Digital Skills Strategy: Formulate and implement a comprehensive national digital skills strategy that explicitly aligns educational policies and funding with labour market demands. This strategy could leverage intelligent models like the one developed in this thesis as a foundational tool for informed policy-making, addressing the poor manpower planning for digital skills highlighted in the Kenya Digital Masterplan.

Increased Funding for Digital Infrastructure and Training: Allocate increased and sustainable funding for digital infrastructure upgrades in educational institutions and for robust faculty training programs. These investments are critical to support the effective integration of advanced digital skills into curricula and to ensure equitable access to modern learning resources.

Incentivize Academia-Industry Partnerships: Create and strengthen policy incentives (e.g., grants, tax breaks, joint research funding) to encourage and facilitate stronger, more formalized partnerships between academia and industry. This promotes knowledge transfer, co-creation of curricula, and responsive workforce development initiatives.

Strategic Value of Foundational and Niche Skills for Workforce Resilience: While the model demonstrates high recall for high-frequency skills, its recall for low-frequency skills (Class 0) is moderate (0.79). This implies that some less common but potentially valuable skills might be missed or underemphasized if only high-frequency skills are prioritized. Low-frequency skills are not necessarily unimportant; they can include foundational skills (e.g., advanced problem-solving, critical thinking, ethical reasoning in AI) or highly specialized/niche skills that become crucial in specific, emerging contexts. An over-reliance solely on high-frequency skills, while good for immediate employability, might lead to a workforce that is narrowly specialized and less adaptable to unforeseen future technological shifts. Policymakers should consider a balanced approach to digital skills development. While the intelligent model effectively guides towards current high-demand skills, policies should also encourage the development of broader foundational digital literacies and specialized niche skills through electives, micro-credentials, or lifelong learning initiatives. This ensures graduates are not just ready for today's jobs but also equipped with adaptable skills for future, unforeseen demands, fostering a more resilient and versatile national workforce. This also addresses the lack of awareness of digital skills among citizens by promoting a more holistic view of digital competence.

The following table summarizes the recommendations for various stakeholders:

Table 5.3 Recommendations for Stakeholders

Stakeholder Group	Key Recommendations
Academic Institutions	Continuously integrate high-demand digital skills into core curricula, Foster interdisciplinary approaches to digital literacy, Invest in strategic faculty professional development and Actively utilize prescriptive analytics from the intelligent model.
Industry Stakeholders	Actively participate in curriculum design and review processes, expand experiential learning opportunities (internships, apprenticeships) and Share anonymized data for continuous model refinement.
Policymakers	Develop a comprehensive national digital skills strategy, allocate increased funding for digital infrastructure and faculty training, create incentives for stronger academia-industry partnerships and Promote a balanced approach to digital skills development, including foundational and niche competencies.

This table translates the study's complex findings and theoretical contributions into concrete, implementable steps for various actors in the education and labour market ecosystem. This makes the research practically useful and directly impactful beyond academic discourse. By segmenting recommendations by specific stakeholder groups, it ensures that each entity receives specific, relevant advice tailored to their unique roles, responsibilities, and capacities within the system. This increases the likelihood of adoption and effective implementation. It visually emphasizes the holistic nature of the problem (digital skills gap) and its solution, underscoring that addressing this challenge effectively requires a concerted, multi-faceted effort from all parties involved, not just academic institutions in isolation. It promotes a shared responsibility framework.

5.7 Areas for Further Research

Building upon the foundational work presented in this thesis, several promising avenues for future research and development emerge, aimed at enhancing the model's capabilities, addressing identified limitations, and expanding its application.

5.7.1 Enhancing Model Capabilities and Data Sources

Integration of Additional Data Sources: Expand the web scraping framework to include a wider and more diverse range of global and regional job portals, industry reports, skills taxonomies, and academic publications. This will enrich the skillset database, capture a more granular view of emerging digital skill demands across various sectors, and enhance the model's global applicability.

Real-time Skill Trend Prediction: Develop and integrate advanced predictive analytics capabilities within the model to forecast future digital skill demands. This could involve the application of time-series analysis, deep learning models, or other sophisticated machine learning techniques to move beyond current alignment to proactive trend anticipation.

Incorporation of Soft Skills and Foundational Competencies: Extend the model's scope to identify, extract, and align academic programs with essential soft skills (e.g., problem-solving, critical thinking, creativity, collaboration, adaptability) and broader foundational literacies. These are increasingly recognized as crucial for long-term employability and success alongside technical skills.

Exploration of Advanced AI Techniques: Investigate the integration of more advanced AI techniques, such as deep learning models for more refined semantic analysis of complex curriculum structures, or reinforcement learning algorithms for optimizing curriculum pathways based on long-term student career outcomes.

5.7.2 Addressing Identified Research Gaps and Limitations

Longitudinal Impact Studies: Conduct comprehensive, long-term empirical studies to assess the actual impact of curriculum changes informed by the intelligent model on graduate employability, career progression, salary potential, and job performance across various student demographics. This directly addresses a key research gap identified in the literature.

Cost-Effectiveness Analysis: Perform a thorough cost-benefit analysis of implementing, maintaining, and continuously updating the intelligent model within diverse academic institutions. This evaluation will assess its economic viability and quantify the return on investment for educational stakeholders.

Mitigating AI Biases and Ensuring Fairness: Explicitly incorporate ethical AI principles and robust mechanisms for bias detection and mitigation within the model's algorithms, training data, and output recommendations. This includes exploring explainable AI (XAI) techniques to enhance transparency and ensure equitable curriculum recommendations for all student groups.

A critical area for future work involves the socio-technical adoption and organizational integration of the model. The research gaps section and critical analysis repeatedly highlight the lack of thorough empirical research that explore practical barriers, faculty attitudes, student outcomes, and organizational dynamics and state that few studies demonstrate successful implementation in real-world academic settings. While Chapter 4's technical validation proves that the model works from an algorithmic perspective, its real-world success and impact are contingent upon its effective adoption and integration within complex educational institutions. This involves human factors, institutional culture, resource availability, and operational challenges that extend beyond the technical solution. It's about understanding the how of implementation, not just the what. Future work must therefore move beyond purely technical efficacy to address socio-technical considerations. This necessitates qualitative research methods (e.g., interviews with faculty and administrators, case studies of pilot implementations) to understand user experience, change management processes, institutional readiness, and the specific barriers to widespread adoption. This will provide a more holistic understanding of the model's value and guide its successful deployment as a catalyst for systemic change within educational institutions.

Furthermore, future research must focus on operationalizing ethical AI for curriculum recommendations. The thesis acknowledges AI ethical issues as a limitation and the literature review discusses challenges like bias and inclusivity and interpretability and explicability. Merely acknowledging these issues is insufficient for a robust, responsible, and widely adopted real-world system. The data scraped from job postings or extracted from curricula could contain inherent historical or societal biases (e.g., gendered language in job descriptions, regional disparities in skill demand, historical underrepresentation of certain skills in traditional curricula). If the intelligent model learns from this biased data, its recommendations could inadvertently perpetuate or even deepen existing inequalities in educational opportunities or career pathways. Future work must move beyond theoretical acknowledgment to the practical, operational implementation of ethical AI. This includes:

Bias Auditing and Mitigation: Regularly auditing the skillset database and model outputs for biases (e.g., against certain demographics, or over-prioritizing skills from dominant industries) and developing algorithms to actively mitigate these biases.

Fairness Metrics: Integrating and evaluating the model's performance using fairness metrics to ensure that its recommendations are equitable across different student groups and do not lead to discriminatory outcomes.

Explainable AI (XAI): Developing Explainable AI (XAI) components so that the model's recommendations are transparent, understandable, and justifiable to educators and students. This fosters trust, allows for human oversight and intervention when biases are detected, and promotes shared responsibility in curriculum decision-making. This ensures the model is a tool for empowerment and equity, not inadvertent discrimination.

5.7.3 Expansion of Application

Broader Academic Disciplines: Extend the model's application beyond the initial focus on Computer Science and Information Systems to other academic disciplines (e.g., business, engineering, healthcare, humanities, social sciences) where digital skills are increasingly becoming essential for professional success.

Cross-Regional and National Implementation: Adapt and test the model in different regional and national contexts, considering variations in labour market structures, educational systems, regulatory frameworks, and specific digital skill priorities. This will require local data collection and model recalibration.

REFERENCES

Akojie, P., & Haynes, S. (2023). A framework to enhance graduate employability. *Interdisciplinary Journal of Digital Scholarship*, 18, 55–75. https://doi.org/10.28945/5090

Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability*, *14*(10), 6199. https://doi.org/10.3390/su14106199

Alibasic, A., Upadhyay, H., Simsekler, M. C. E., Kurfess, T., Woon, W. L., & Omar, M. A. (2022). Evaluation of the trends in jobs and skill-sets using data analytics: A case study. *Journal of Big Data*, *9*(1), Article 5. https://doi.org/10.1186/s40537-022-00576-5

Amazon. (2023). *Amazon Web Services (AWS) – Cloud computing services.* Amazon Web Services, Inc. https://aws.amazon.com/

American Psychological Association. (2023). *Research ethics.* American Psychological Association. https://www.apa.org/science/programs/research

Anderson, M., & Perrin, A. (2021). *Mobile technology and home broadband 2021.* Pew Research Center. https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/

Ashioya, V. (2023). *Web scraping: Effective web scraping for data scientists. Data Science and Artificial Intelligence.* https://conferences.kabarak.ac.ke/index.php/dsai/article/view/9

Asian Development Bank. (2023). *Digital jobs and digital skills: A shifting landscape in Asia and the Pacific.* Asian Development Bank. https://www.adb.org/sites/default/files/publication/829711/digital-jobs-digital-skills.pdf

Astapciks, I. (2023, March 20). *Why do companies need digital transformation? Forbes.* https://www.forbes.com/sites/forbestechcouncil/2023/03/20/why-do-companies-need-digital-transformation/?sh=3e2572e32bab

Bagai, R., & Mane, V. (2023). Designing an AI-powered mentorship platform for professional development: Opportunities and challenges. *International Journal of Computer Trends and Technology*, *71*(4), 108–114. https://doi.org/10.14445/22312803/ijctt-v71i4p114

Barany, B. (2022, September 8). *The importance of data analytics in the digital world. Capturly Blog.* https://capturly.com/blog/the-importance-of-data-analytics-in-the-digital-world/

Barboutidis, G., & Stiakakis, E. (2023). Identifying the factors to enhance digital competence of students at vocational training institutes. *Technology, Knowledge and Learning, 28*(2), 613–650. https://doi.org/10.1007/s10758-023-09641-1

Bashay, M. (2020). *Digital skills for an equitable recovery: Policy recommendations to address the digital skill needs of workers most vulnerable to displacement.* National Skills Coalition. https://www.academia.edu/88160343/Digital_Skills_for_an_Equitable_Recovery_Policy_Recommendations_to_Address_the_Digital_Skill_Needs_of_Workers_Most_Vulnerable_to_Displacement

Bhorat, H., Signé, L., Asmal, Z., Monnakgotla, J., & Rooney, C. (2023). *Digitalization and digital skills gaps in Africa.* Brookings Institution. https://www.brookings.edu/wp-content/uploads/2023/05/Bhorat-et.-al-May-2023-Digitalization-and-digital-skills-in-Africa-1.pdf

Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., & Kurth-Nelson, Z. (2020). Deep reinforcement learning and its neuroscientific implications. *Neuron, 107*(4), 603–616. https://doi.org/10.1016/j.neuron.2020.06.014

Bouchrika, I. (2022). *Types of research design: Perspectives and methodological approaches. Research.com.* https://research.com/research/types-of-research-design

Cabello, K. (2023). *How generative AI will reset talent and skills: A new frontier for the future of work.* Oxford Economics. https://www.oxfordeconomics.com/resource/how-generative-ai-will-reset-talent-skills-a-new-frontier-for-the-future-of-work/

Carlisle, S., Ivanov, S. H., & Dijkmans, C. (2021). The digital skills divide: Evidence from the European tourism industry. *Journal of Tourism Futures, 9*(2), 240–266. https://doi.org/10.1108/jtf-07-2020-0114

Cevikbas, M., & Kaiser, G. (2022). Promoting personalized learning in flipped classrooms: A systematic review study. *Sustainability, 14*(18), 11393. https://doi.org/10.3390/su141811393

Chapman, C. (2019, May 5). *A complete overview of the best data visualization tools. Toptal Design Blog.* https://www.toptal.com/designers/data-visualization/data-visualization-tools

Chaudhry, M. N., & Kazim, E. (2021). Artificial intelligence in education (AIED): A high-level academic and industry note 2021. *AI and Ethics*, *1*(2), 157–165. https://doi.org/10.1007/s43681-021-00074-z

Colback, L. (2023, May 18). Technology and the skills shortage. *Financial Times*. https://www.ft.com/content/b1b710a1-6d12-43e5-8508-ae4584a7289a

Columbus, L. (2017, May 13). *IBM predicts demand for data scientists will soar 28% by 2020. Forbes*. https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/

Crompton, H. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 5. https://doi.org/10.1186/s41239-023-00392-8

Darling-Hammond, L., Flook, L., Cook-Harvey, C. M., Barron, B., & Osher, D. (2019). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Dice.com. (2021). *Tech job report: Occupations*. Dice.com. https://www.dice.com/technologists/ebooks/tech-job-report/occupations.html

Dice.com. (2022). *Tech job report: Introduction*. Dice.com. https://www.dice.com/technologists/ebooks/tech-job-report

Dice.com. (2024a). *Find jobs in tech*. Dice.com. https://www.dice.com

Dice.com. (2024b). *Top 20 information technology jobs, now hiring*. Dice.com. https://www.dice.com/jobs/q-Information+Technology-jobs

Dice.com. (2024c). *Search jobs*. Dice.com. https://www.dice.com/jobs

Digital Unite. (2021, July 28). *Files and storage*. Digital Unite. https://www.digitalunite.com/technology-guides/computer-essentials/files-and-storage

Dutta, G., Kumar, R., Sindhwani, R., & Singh, R. K. (2020). Digital transformation priorities of India's discrete manufacturing SMEs – a conceptual study in perspective of Industry 4.0. *Competitiveness Review: An International Business Journal*, *30*(3), 289–314. https://doi.org/10.1108/cr-03-2019-0031

EIT Digital Shapers and Makers. (2022). *The future of education for digital skills*. EIT Digital. https://www.eitdigital.eu/fileadmin/2022/ecosystem/makers-shapers/reports/EIT-Digital_Report_The-Future-of-Education-for-Digital-Skills.pdf

Evans, H. (2021, May 3). *How to use technology to help teachers be better and to make life better for teachers.* World Bank Blogs. https://blogs.worldbank.org/education/how-use-technology-help-teachers-be-better-and-make-life-better-teachers

Ezell, S. (2021, November 29). *Assessing the state of digital skills in the U.S. economy.* Information Technology and Innovation Foundation. https://itif.org/publications/2021/11/29/assessing-state-digital-skills-us-economy/

Feijao, C., Flanagan, I., Van Stolk, C., & Gunashekar, S. (2021). *Current trends and future directions: The global digital skills gap.* RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RRA1500/RRA1533-1/RAND_RRA1533-1.pdf

Flowtrace. (2023, July 28). *Best collaboration tools for remote teams 2023.* Flowtrace. https://www.flowtrace.co/collaboration-blog/best-collaboration-tools-for-remote-teams-2023

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). *IEA International Computer and Information Literacy Study 2018 technical report.* ERIC. https://eric.ed.gov/?id=ED610699

Future Skills Centre. (2023). *Digital skills and the skills gap: Perspectives, skills frameworks and definitions, with a focus on Canada.* Future Skills Centre. https://fsc-ccf.ca/wp-content/uploads/2023/05/digital-skills-and-the-skills-gap-.pdf

GCFGlobal. (2019). *Digital media literacy: The problem with photo manipulation.* GCFGlobal. https://edu.gcfglobal.org/en/digital-media-literacy/the-problem-with-photo-manipulation/1/

Gondil, S., Raut, T., Patne, S., & Bhagat, V. (2021). A survey on web scraping and its applications. *International Journal of Creative Research Thoughts, 9*(4), 4318–4321. http://www.ijcrt.org/papers/IJCRT2104523.pdf

González, M. C., Martín, S. C., & García-Peñalvo, F. J. (2021). The digital competence of pre-service educators: The influence of personal variables. *Sustainability, 13*(4), 2318. https://doi.org/10.3390/su13042318

Goyal, L. M., Saba, T., Rehman, A., & Larabi-Marie-Sainte, S. (Eds.). (2021). *Artificial intelligence and Internet of Things: Applications in smart healthcare* (1st ed.). CRC Press. https://doi.org/10.1201/9781003097204

Guan, X., Feng, X., & Islam, A. Y. M. A. (2023). The dilemma and countermeasures of educational data ethics in the age of intelligence. *Humanities and Social Sciences Communications, 10*(1), Article 5. https://doi.org/10.1057/s41599-023-01633-x

Halder, S., & Saha, S. (2023). Artificial intelligence in education. In *The Routledge handbook of education technology* (pp. 390–399). Routledge. https://doi.org/10.4324/9781003293545-33

Handel, M. J. (2016). The O*NET content model: Strengths and limitations. *Journal for Labour Market Research*, *49*(2), 157–176. https://doi.org/10.1007/s12651-016-0199-8

Harry, S. C. (2021). *Four strategies for aligning the curriculum with technology.* LinkedIn. https://www.linkedin.com/pulse/four-strategies-aligning-curriculum-technology-dr-harry-saint-cyr

Hartman, K. (2020). *Digital marketing analytics: In theory and in practice.* Independently published. https://www.amazon.com/Digital-Marketing-Analytics-Theory-Practice/dp/B08J5BD5RM

Hasan, F., Mudey, A., & Joshi, A. (2023). Role of Internet of Things (IoT), artificial intelligence and machine learning in musculoskeletal pain: A scoping review. *Cureus*, *15*(2), e37352. https://doi.org/10.7759/cureus.37352

Hecker, I., & Loprest, P. (2019). *Foundational digital skills for career progress.* Urban Institute. https://www.urban.org/sites/default/files/publication/100843/foundational_digital_skills_for_career_progress_2.pdf

Hol, A., Richardson, J., McGovern, J., & Hamilton, M. (2023). A new sustainable model for aligning industry requirements and university programs. *ACM Inroads*, *14*(1), 30–39. https://doi.org/10.1145/3583086

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*, 1–19. https://doi.org/10.1155/2022/5215722

Hootsuite. (2022). *Digital 2022: Social media marketing & management dashboard.* Hootsuite. https://www.hootsuite.com/resources/digital-trends

IBM. (2023). *Internet of Things.* IBM. https://www.ibm.com/cloud/internet-of-things

IBM. (n.d.). *IBM Watson Orchestrate skills.* IBM. Retrieved August 24, 2023, from https://www.ibm.com/products/watson-orchestrate/skills

ILO. (2021). *Changing demand for skills in digital economies and societies: Literature review and case studies from low- and middle-income countries.* ILO Publications. https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/WCMS_831372/lang--en/index.htm

ITEA. (2007). Standards for technological literacy: Content for the study of technology. *International Technology Education Association*. https://wwwcp.umes.edu/tech/wp-content/uploads/sites/94/2021/09/xstnd.pdf

ITU. (2020). Digital skills assessment guidebook. *ITU Publications*. https://academy.itu.int/sites/default/files/media2/file/D-PHCB-CAP_BLD.04-2020-PDF-E_02%20June%202020.pdf

ITU. (2021). *Skills development for the digital economy.* ITU Publications. https://www.itu.int/en/mediacentre/backgrounders/Pages/skills-development-digital-economy.aspx

Jha, R. (2023). *Data scraping, also known as web scraping, is the process of extracting information from websites and other online sources.* LinkedIn. https://www.linkedin.com/pulse/data-scraping-its-use-recruitment-raj-jha

Khan, I. A., Ahmad, A. L., Jabeur, N., & Mahdi, M. N. (2021a). An artificial intelligence approach to monitor student performance and devise preventive measures. *Smart Learning Environments, 8*(1). https://doi.org/10.1186/s40561-021-00161-y

Khan, N., Khan, S., Tan, B. C., & Loon, C. H. (2021b). Driving digital competency model towards IR 4.0 in Malaysia. *Journal of Physics: Conference Series, 1793*(1), 012049. https://doi.org/10.1088/1742-6596/1793/1/012049

Khosravi, H., Shum, S. B., Chen, G., Gašević, D., Kay, J., Knight, S. R., & Tsai, Y. S. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3,* 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kim, J., Lee, H., & Cho, Y. M. (2022). Learning design to support student–AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Kim, J. (2015). Competency-based curriculum: An effective approach to digital Curation education. *Journal of Education for Library and Information Science Online, 56*(4), 283-297. https://doi.org/10.3138/jelis.56.4.2

Kousa, P., & Niemi, H. (2023). Artificial intelligence ethics from the perspective of educational technology companies and schools. In H. Niemi, R. D. Pea, & Y. Lu (Eds.), *AI in learning: Designing the future* (pp. 301–315). Springer. https://doi.org/10.1007/978-3-031-09687-7_17

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. *International Journal of Information Management, 63*(4), 102466. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Li, L. (2022). Reskilling and upskilling the future-ready workforce for Industry 4.0 and beyond. *Information Systems Frontiers.* https://doi.org/10.1007/s10796-022-10308-y

Llego, M. A. (2023). *The AI-driven holistic development framework for education: A comprehensive approach to integrating AI in learning environments.* TeacherPH. https://www.teacherph.com/ai-driven-holistic-development-framework-education/

Loughborough University London. (2023). *Digital skills program.* https://www.lborolondon.ac.uk/business-partnerships/digital-skills/

Ma'dan, M., Ismail, M. T., & Daud, S. (2020). Strategies to enhance graduate employability: Insights from Malaysian public university policy-makers. *Malaysian Journal of Learning and Instruction, 17*(2), 137–165. https://doi.org/10.32890/mjli2020.17.2

Mamabolo, A., & Myres, K. (2020). Bridging Africa's Skills 4.0 gap: Repositioning learning, teaching, and research of skills in higher education institutions. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3555654

McKinsey & Company. (2021). *The future of work after COVID-19.* https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19

Mercy Corps & Moringa School. (2020). *Competing in a digital age: The development of IT skills and jobs in Kenya and Uganda, 2019.* https://www.mercycorps.org/sites/default/files/2020-01/Publication_IT_Skill_Gap_Report_April1_VF.pdf

Ministry of ICT, Innovation and Youth Affairs. (2019). *Kenya digital masterplan 2022–2032.* https://cms.icta.go.ke/sites/default/files/2022-09/Kenya_Digital_Master_Plan_2022-2023.pdf

Mn, K., Khalid, F., & Husnin, H. (2020). Preparing graduates with digital literacy skills toward fulfilling employability needs in the 4IR era: A review. *International Journal of Advanced Computer Science and Applications, 11*(6). https://doi.org/10.14569/ijacsa.2020.0110641

Moldoveanu, M., Frey, K., & Moritz, B. (2022). 4 ways to bridge the global skills gap. *Harvard Business Review.* https://hbr.org/2022/03/4-ways-to-bridge-the-global-skills-gap

Natarajan, P., & Szeto, A. (2023, March 23). A new paradigm for partnership between industry and academia in the age of AI. *Amazon Science.*

https://www.amazon.science/news-and-features/a-new-paradigm-for-partnership-between-industry-and-academia-in-the-age-of-ai

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first-century skills in the post-pandemic world. *Educational Technology Research and Development*, *71*(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. T. (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Ojajuni, O. P., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., & Mbarika, V. (2021). Predicting student academic performance using machine learning. In *Computational Science and Its Applications – ICCSA 2021* (pp. 481–491). Springer. https://doi.org/10.1007/978-3-030-87013-3_36

Organisation for Economic Co-operation and Development (OECD). (2021a). *21st-century readers: Developing literacy skills in a digital world.* OECD Publishing. https://doi.org/10.1787/a83d84cb-en

Organisation for Economic Co-operation and Development (OECD). (2021b). *Artificial intelligence, machine learning and big data in finance: Opportunities, challenges and implications for policy makers.* OECD Publishing. https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf

Pantanowitz, L., Bui, M. M., Chauhan, C., ElGabry, E. A., Hassell, L. A., Li, Z., & Becich, M. J. (2022). Rules of engagement: Promoting academic–industry partnership in the era of digital pathology and artificial intelligence. *Academic Pathology, 9*, 100026. https://doi.org/10.1016/j.acpath.2022.100026

Plekhanov, D., Franke, H., & Netland, T. H. (2022). Digital transformation: A review and research agenda. *European Management Journal.* https://doi.org/10.1016/j.emj.2022.09.007

Putra, F. K. K., Saepudin, P., & Utami, N. G. M. K. (2022). Preferred competencies for tourism and hospitality graduates: Evidence from longitudinal tracer studies. *Journal of Technical Education and Training (JTET), 3*(14). https://doi.org/10.30880/jtet.2022.14.03.009

Qin, C., Zhang, L., Zha, R., Shen, D., Zhang, Q., Sun, Y., Zhu, C., Zhu, H., & Xiong, H. (2023). A comprehensive survey of artificial intelligence techniques for talent analytics. *arXiv, abs/2307.03195.* https://doi.org/10.48550/arXiv.2307.03195

Raji, N. A. S., Busson-Crowe, D. A., & Dommett, E. J. (2023). University-wide digital skills training: A case study evaluation. *Education Sciences*, *13*(4), 333. https://doi.org/10.3390/educsci13040333

Ramesh, D., & Sanampudi, S. K. (2021). An automated essay scoring system: A systematic literature review. *Artificial Intelligence Review, 55*(1), 1–33. https://doi.org/10.1007/s10462-021-10068-2

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a framework for augmented and virtual reality. *Computers in Human Behavior*, *133*, 107289. https://doi.org/10.1016/j.chb.2022.107289

Reddy, P., Chaudhary, K., & Hussein, S. (2023). A digital literacy model to narrow the digital literacy skills gap. *Heliyon*, *9*(4), e14878. https://doi.org/10.1016/j.heliyon.2023.e14878

Rovira, S., Peres, W., & Saporito, N. (2021). *Digital technologies for a new future.* United Nations, Economic Commission for Latin America and the Caribbean. https://www.cepal.org/sites/default/files/publication/files/46817/S2000960_en.pdf

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *3*(2). https://doi.org/10.1007/s42979-021-00592-x

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1). https://doi.org/10.1186/s41239-021-00292-9

Siri. (2023). *How to scrape Dice.com: Extract job listing data from Dice.com.* Hexomatic. https://hexomatic.com/academy/2023/03/09/how-to-scrape-dice-com/

Society for Human Resource Management. (2021). *State of the workplace study.* https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/SHRM%20State%20of%20the%20Workplace%20report%202021-2022.pdf

Suarta, I. M., & Suwintana, I. K. (2021). The new framework of employability skills for digital business. *Journal of Physics: Conference Series*, *1833*(1), 012034. https://doi.org/10.1088/1742-6596/1833/1/012034

Sunil, K. (2024). *A comprehensive guide to the Department of Labor Occupational Network (O*NET).* RightHire. https://blog.righthire.com/blog/onet

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, *3*, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tapalova, O., Zhiyenbayeva, N., & Gura, D. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of e-Learning*, *20*(5), 639–653. https://doi.org/10.34190/ejel.20.5.2597

Techopedia. (2020). *What is a web browser?* Techopedia. https://www.techopedia.com/definition/288/web-browser

Thongprasit, J., & Wannapiroon, P. (2022). Framework of artificial intelligence learning platform for education. *International Education Studies*, *15*(1), 76. https://doi.org/10.5539/ies.v15n1p76

U.S. Bureau of Labor Statistics. (n.d.). *Replace with a database: O*NET replaces the Dictionary of Occupational Titles.* (Retrieved January 21, 2025). https://www.bls.gov/careeroutlook/1999/article/pdf/replace-with-a-database-onet-replaces-the-dictionary-of-occupational-titles.pdf

U.S. Department of Labor. (2019). *The O*NET® content model at O*NET Resource Center.* https://www.onetcenter.org/content.html

U.S. Department of Labor. (2025). *O*NET Resource Center – Toolkit for business.* Information Anthology. https://informationanthology.net/CareerMentor/Discovery/Ability-Profiler/toolkit.html

UNESCO & Huawei. (2022). *White paper: ICT talent cultivation for Kenya's digital economy*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381538

United Nations. (2020). *Social media, development and governance.* Asian and Pacific Training Centre for Information and Communication Technology for Development. https://www.unapcict.org/sites/default/files/2020-07/Social%20Media%2C%20Development%20and%20Governance%20Module_FINAL.pdf

Unity Technologies. (2019). *Unity manual: VR overview.* https://docs.unity3d.com/Manual/VROverview.html

University of Waterloo. (2021). *Digital skills integration.* https://uwaterloo.ca/centre-for-work-integrated-learning/digital-skills-fundamentals

Volodymyr. (2022). *The importance of data analytics in the digital era.* Ryze. https://ryzebeyond.com/the-importance-of-data-analytics-in-the-digital-era/

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The digital competence framework for citizens—With new examples of knowledge, skills and attitudes (EUR 31006 EN). *Publications Office of the European Union*. https://doi.org/10.2760/115376

Wiley Research. (2021). *The digital skills gap index 2021.* https://dsgi.wiley.com/

World Bank. (2019). *Labor market analysis and curriculum gap assessment using big data in Kenya.* https://thedocs.worldbank.org/en/doc/202221578327923556-0090022020/original/LaborMarketAnalysisandCurriculumGapAssessmentusingBigDatainKenyaFinalReportHeadai10282019clean.pdf

World Economic Forum. (2023). *The future of jobs report 2023.* https://www.weforum.org/reports/the-future-of-jobs-report-2023/digest/

Zahiroh, M. (2020). Cybersecurity awareness and digital skills on readiness for change in digital banking. *Li Falah: Jurnal Studi Ekonomi dan Bisnis Islam, 5*(2), 53. https://doi.org/10.31332/lifalah.v5i2.2271

Zan, B. U., Çolaklar, H., Altay, A., & Taşkın, N. (2021). A study on digital literacy skills of faculty of letters students: Use of university library. *International Journal of Emerging Technologies in Learning*, *16*(1), 152. https://doi.org/10.3991/ijet.v16i01.16567

Zhang, Z., Xia, Y., & Abula, K. (2023). How digital skills affect rural labor employment choices: Evidence from rural China. *Sustainability*, *15*(7), 6050. https://doi.org/10.3390/su15076050

Zhao, Y., Pinto-Llorente, A. M., & Gómez, M. L. N. (2021). Digital competence in higher education research: A systematic literature review. *Computers and Education*, *168*, 104212. https://doi.org/10.1016/j.compedu.2021.104212

Zhou, X., Tong, Y., Lan, X., Zheng, K., & Zhan, Z. (2021). AI education in massive open online courses: A content analysis. *3rd International Conference on Computer Science and Technologies in Education (CSTE).* https://doi.org/10.1109/cste53634.2021.00023

APPENDICES

Appendix A

Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 224225

Date of Issue: 30/April/2025

RESEARCH LICENSE

This is to Certify that Mr.. Duncan K. Nyale of The Co-operative University of Kenya, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: AN INTELLIGENT MODEL FOR ALIGNING ACADEMIC PROGRAMS TO ESSENTIAL DIGITAL SKILLS IN THE LABOUR MARKET for the period ending : 30/April/2026.

License No: NACOSTI/P/25/4172814

224225

Applicant Identification Number

Director General

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix B

Ethical Clearance

KCA UNIVERSITY
- 1 9 8 9 -

Thika Road, Ruaraka
P.O. Box 56808-00200 Nairobi Kenya
Pilot Line: +254 20 8070408/9

Tel: +254 20 3537842
Fax: +254 20 8561077
Mobile: +254 734 888022, 710 888022
Email: kca@kca.ac.ke
Website: www.kca.ac.ke

KCA UNIVERSITY SCIENTIFIC & ETHICS REVIEW COMMITTEE

REF: **KCAU/SERC/ EXTER013** **Date: 16th JULY 2025**

TO: **DUNCAN K. NYALE**

Dear Sir/Madam,

RE: AN INTELLIGENT MODEL FOR ALIGNING ACADEMIC PROGRAMS TO ESSENTIAL DIGITAL SKILLS IN THE LABOUR MARKET

This is to inform you that the KCA University Scientific Ethics Review Committee (KCAUSERC) has reviewed and approved your above research proposal. Your application approval number is ***KCAUSERC/*EXTER013.** The approval period is ***21st July 2025 – 21st July, 2026.***
This approval is subject to compliance with the following requirements;

i. Only approved documents, including informed consents, study instruments, and MTA, will be used.
ii. All changes, including amendments, deviations, and violations, are submitted for review and approval by ***KCAUSERC.***
iii. Death and life-threatening problems, serious adverse events, or unexpected adverse events, whether related or unrelated to the study, must be reported to **KCAU*SERC*** within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affect the safety or welfare of study participants and others, or affect the integrity of the research, must be reported to ***KCAUSERC*** within 72 hours.
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days before the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to ***KCAUSERC.***

Prior to commencing your study, you will be expected to obtain a research license from the National Commission for Science, Technology and Innovation (NACOSTI) https://research-portal.nacosti.go.ke and also obtain other clearances needed.

Yours sincerely,

Dr. Caroline Ntara
Chairperson
KCA University Scientific & Ethics Review Committee

Advancing Knowledge, Driving Change

Appendix C

Similarity Report

Duncan Nyale

PhD IS Thesis

Thesis_proposal submision

Phd_Msc_Cohort_1

The Cooperative University of Kenya

Document Details

Submission ID

trn:oid:::1:3292212014

Submission Date

Jul 8, 2025, 2:01 PM GMT+3

Download Date

Jul 11, 2025, 4:38 PM GMT+3

File Name

Thesis_-_PhD_IS_v.1.pdf

File Size

3.9 MB

137 Pages

35,390 Words

217,373 Characters

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text

Exclusions

- 1 Excluded Source
- 1 Excluded Match

Match Groups

354Not Cited or Quoted 11%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
7% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Appendix D

Curriculum Digital Skills Analyzer System Screenshots

Figure D1

Curriculum Upload Window

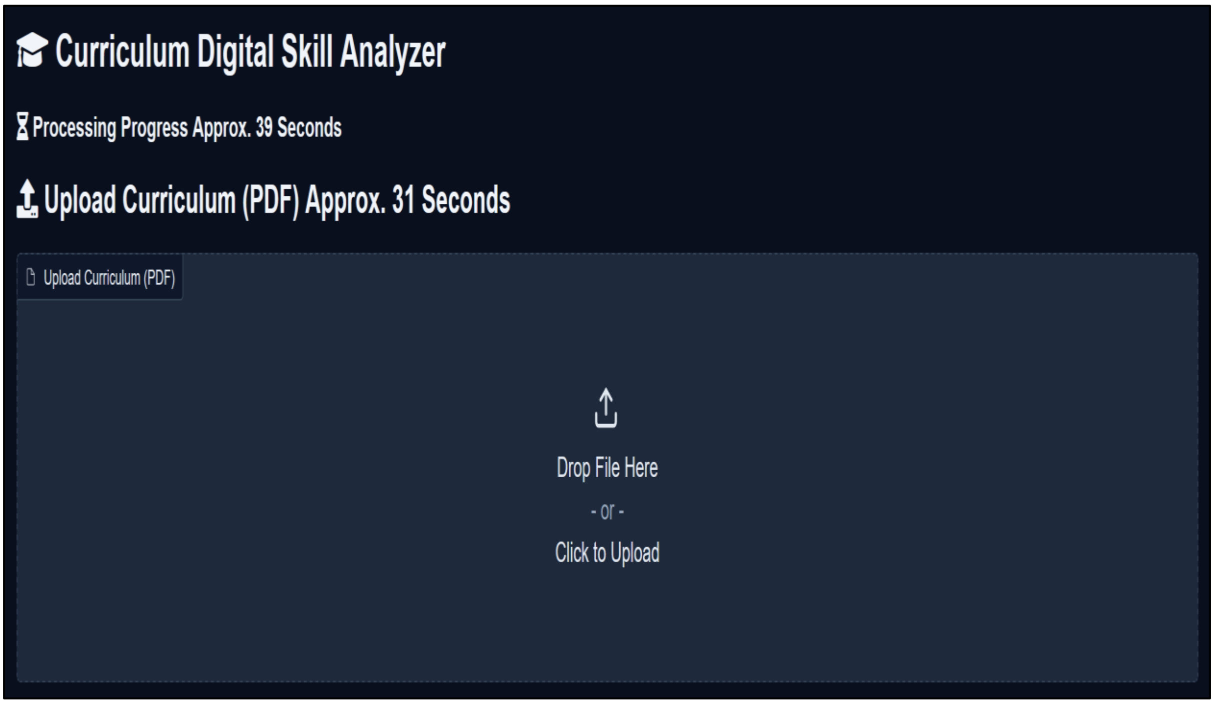


Figure D 2

Curriculum Analytics Progress Window

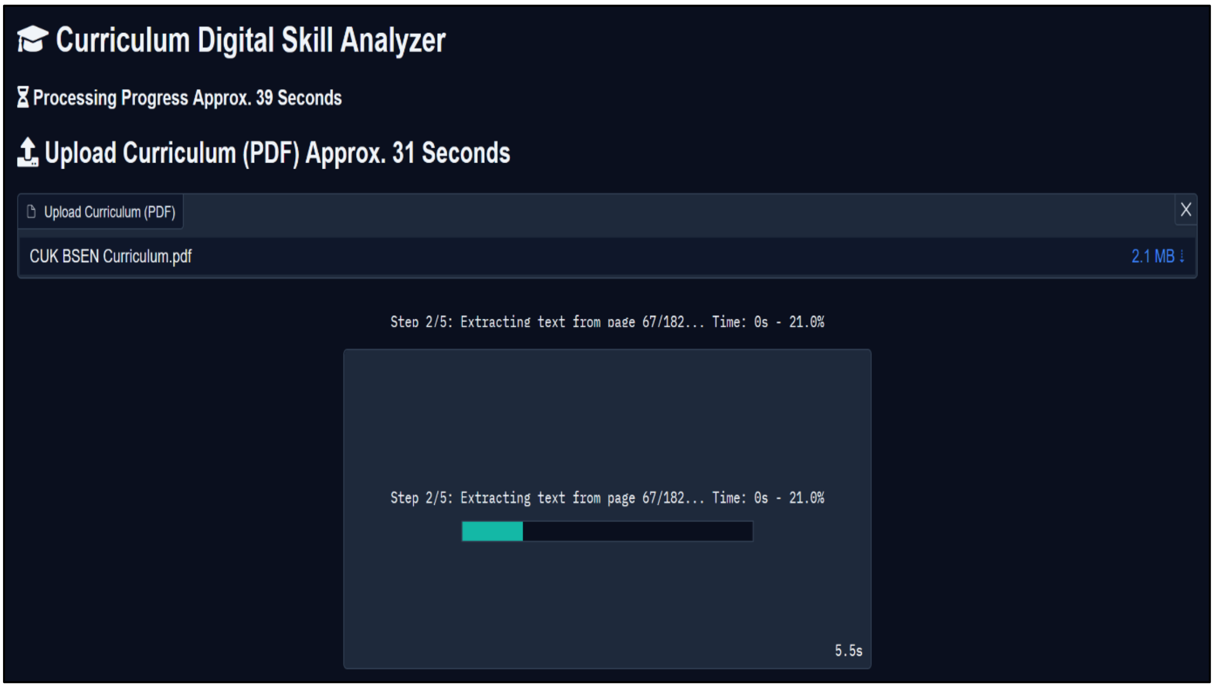


Figure D3

Curriculum Digital Skills Index (DSI) Window

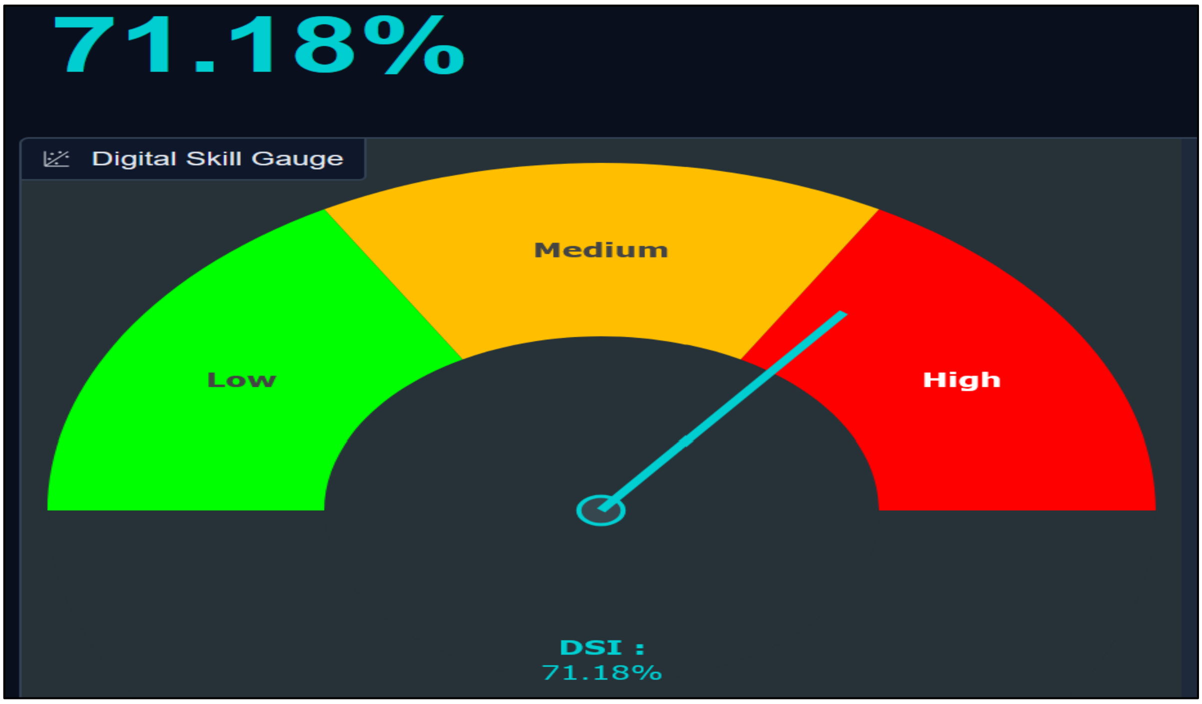
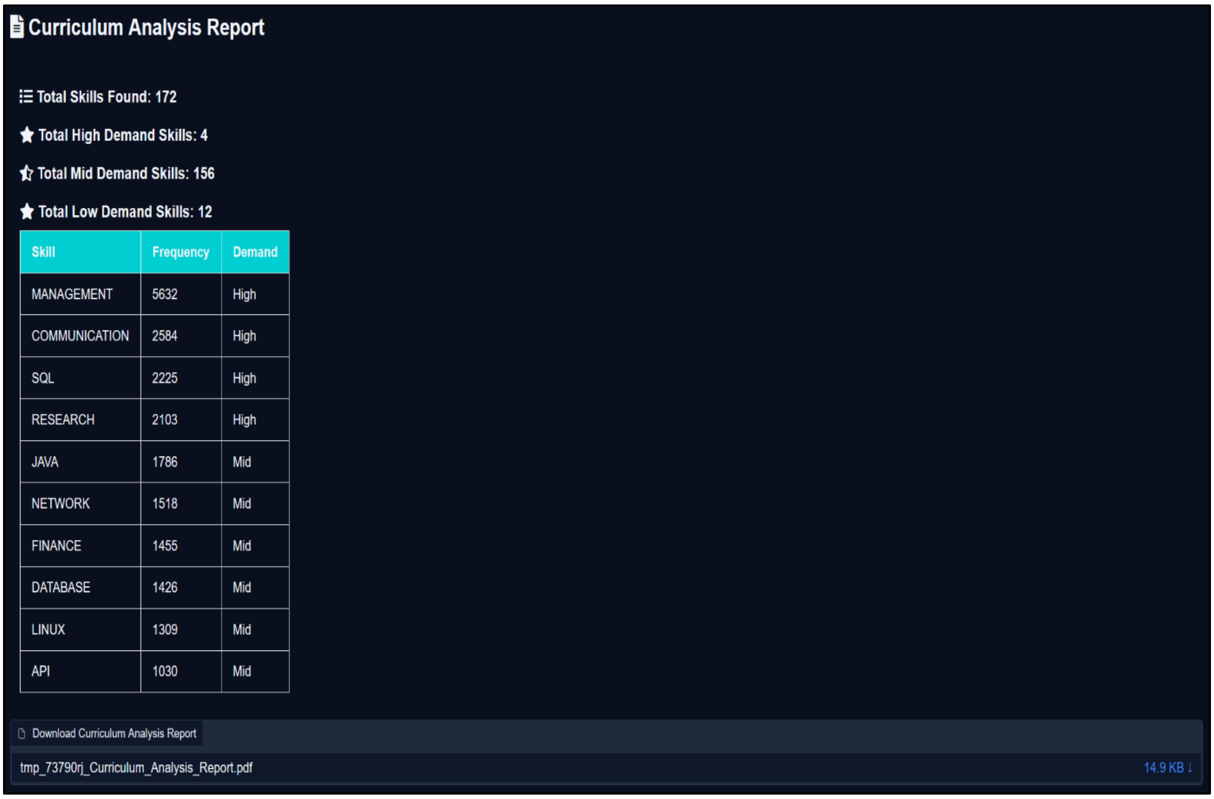


Figure D4

Curriculum Analysis Window



Curriculum Analysis Report

Total Skills Found: 172

Total High Demand Skills: 4

Total Mid Demand Skills: 156

Total Low Demand Skills: 12

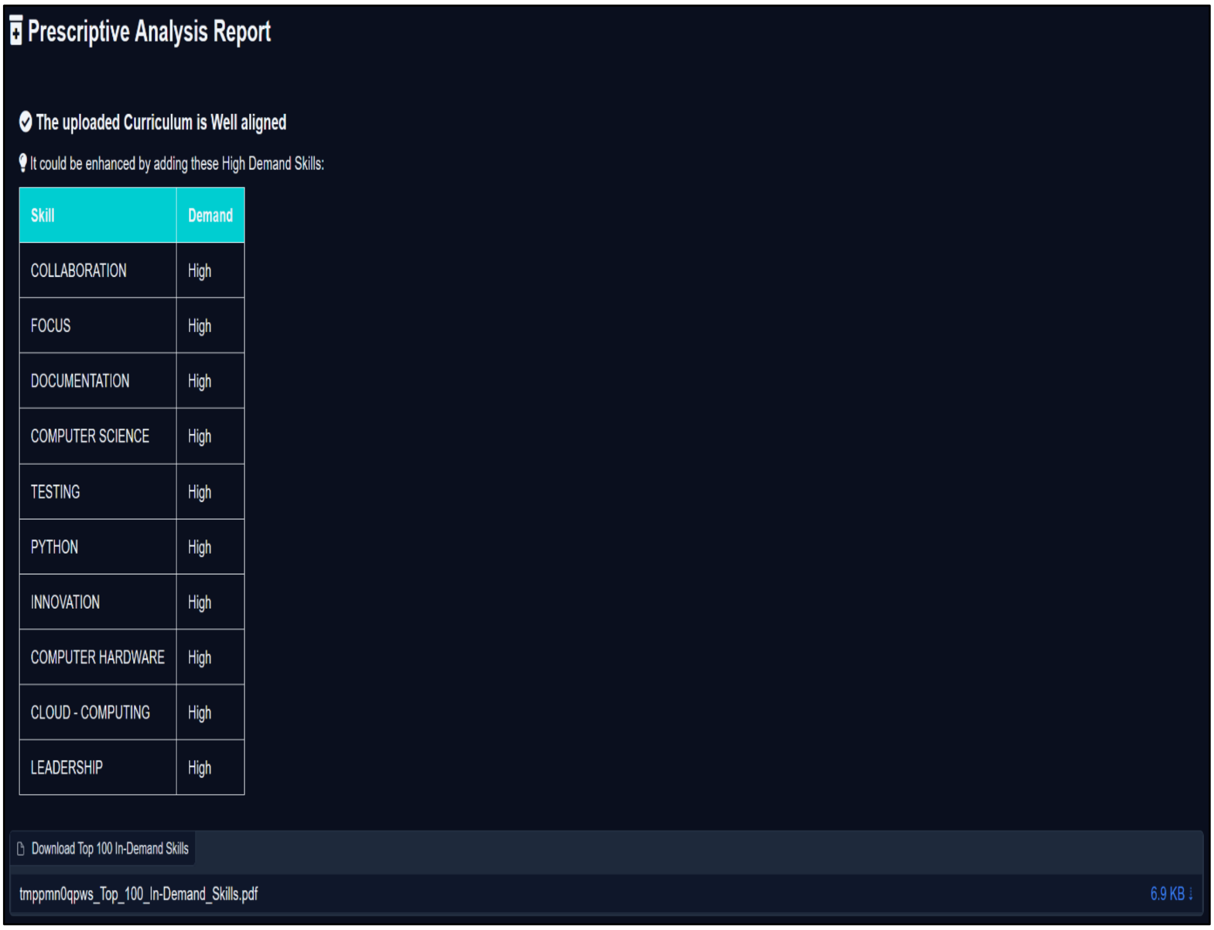
Skill	Frequency	Demand
MANAGEMENT	5632	High
COMMUNICATION	2584	High
SQL	2225	High
RESEARCH	2103	High
JAVA	1786	Mid
NETWORK	1518	Mid
FINANCE	1455	Mid
DATABASE	1426	Mid
LINUX	1309	Mid
API	1030	Mid

Download Curriculum Analysis Report

tmp_73790rj_Curriculum_Analysis_Report.pdf 14.9 KB ↓

Figure D5

Prescriptive Analytics Window



Appendix E

System Resource Requirements

Resource	Estimated Load	Contributors
RAM	12–16 GB	spaCy, pdfplumber, ChromaDB, pandas, Plotly
CPU	~20%	PDF parsing, model training, PDF generation
GPU	~80%	spaCy NLP pipeline, PhraseMatcher, ChromaDB vector search
Time	115+ seconds	Sequential tasks: PDF → NLP → DB → Model → Plot/PDF