

Career Guidance Tool Using Deep Learning.

Alice Kabui

8755.

**Submitted in partial fulfillment of the requirements for the Degree of
Master of Science in Information Technology at Strathmore University**

**School of Computing and Engineering Sciences
Strathmore University
Nairobi, Kenya**

Declaration

I hereby affirm this thesis proposal is my own work, not submitted elsewhere. It's for the partial fulfilment of an MSc.IT degree at Strathmore University. Except for cited references, the examination submission will be solely based on my research, and the expressed views are mine.


Student name:Alice Kabui 

Signature:..... Date:.....08/04/2024.....

Approval

This thesis proposal of Alice Kabui has been submitted for examination with my approval as University Supervisor.

Lecturer name: Prof. Ismail Ateya.....

Signature:..... 

Date:.....08/04/2024.....

Abstract

The selection of a suitable career is a significant societal concern, with career counselling playing a vital role in addressing this issue. Students pursuing their undergraduate studies encounter the challenge of selecting a career path. Embarking on a career path not aligned with a students' preferences can result in diminished job satisfaction, a deficit in skills, and mental strain. This study introduces innovative approaches involving deep learning techniques that has the potential to benefit students. Within the field of education, a substantial volume of data exists, although its utility for students is limited. This research aims to assess student data and provide guidance on the most suitable career for their sustained success. The study employed machine learning algorithms to provide career options for students with emphasis being placed on students pursuing or have graduated in Information Technology (IT) in institutions of higher learning. The research used agile software development methodology as the preferred software development methodology. The model was developed using MLP classifier a class of deep learning algorithms. The developed model achieved an accuracy of 99.93% surpassing the accuracy of all prediction models in career guidance domain. The developed tool will exert a significant influence on both the institutions and society at large.

Keywords: Career Guidance, Data Mining, Deep Learning, Information Technology, Machine Learning, Undergraduates.

Table of Contents

Declaration	ii
Approval.....	iii
Abstract	iv
Table of Contents	v
List of Figures	ix
Chapter 1: Introduction	1
1.1 Background	1
1.2 Problem Statement	3
1.3 Research Objective.....	4
1.3.1 Aim.....	4
1.3.2 Specific Objectives.....	4
1.4 Research Questions	4
1.5 Scope of the study	4
1.6 Significance of the study	5
Chapter 2: Literature Review	6
2.1 Introduction	6
2.2 Empirical Literature	7
2.3 Theoretical Frameworks.....	11
2.3.1 Trait and Factor Theory	11
2.3.2 Ginzberg's General Theory of Occupational Choice	12
2.3.3 Super's Developmental Self-Concept Theory	14
2.4 Factors Influencing Career Selection Among Students	17
2.4.1 Interest.....	17
2.4.2 Self-Capacity.....	17
2.4.3 Career opportunities	18

2.5 Techniques Used For Career Guidance.....	18
2.5.1 Expert Systems.....	18
2.5.2 Machine Learning	20
2.6 Models and Frameworks	24
2.6.1 Random Forests.....	24
2.6.2 Artificial Neural Networks.....	24
2.6.3 Support Vector Machines.....	26
2.6.4 K-Means	26
2.6 Gaps in the Existing Systems.....	29
2.7 Conceptual Model	29
Chapter 3: Research Methodology.....	31
3.1 Introduction.....	31
3.2 Research Design.....	31
3.3 Population and Sampling	32
3.4 Data Collection.....	33
3.5 Research Quality and Reliability	33
3.6 System Development Methodology	33
3.7 Utilization and Dissemination of Research Results	35
3.8 Ethical Considerations / Issues.....	35
Chapter 4: System Analysis and Design	35
4.1 Introduction	35
4.2 Requirement Specifications	36
4.2.1 Functional Requirements	36
4.2.2 Non-Functional Requirements	36
4.3 System Architecture	36
4.4 System Design.....	37
4.4.1 Use Case Diagram.....	37

4.4.2 Class Diagram	40
4.4.3 Sequence Diagram	40
4.4.4 Database Schema	41
4.5 Wireframes	42
4.5.1 Home Page Wireframe	42
4.5.2 Login Wireframe	42
4.5.3 Register Wireframe	43
4.5.4 Career Evaluation Page Wireframe.....	43
4.5.5 Results Page Wireframe	44
4.5.6 History Wireframe	45
Chapter 5: System Implementation and Testing	46
5.1 Introduction	46
5.2 Model Components	46
5.2.1 Multi-Layer Perceptron (MLP)	46
5.3 Web Application Components	47
5.3.1 Home Page	47
5.3.2 Career Evaluation Interface.....	47
5.3.3 Prediction Results	48
5.4 System Implementation.....	48
5.4.1 Development Environment	49
5.4.2 Career Data Collection.....	49
5.4.3 Data Pre-processing	50
5.4.4 Training Model.....	51
5.4.5 Creating Model API	51
5.4.6 Career Guidance Tool	52
5.5 System Testing	52
5.5.1 Test on model accuracy.....	53

Chapter 6: Discussions	54
6.1 Review of Research Objectives	54
6.2 To Investigate Factors Influencing Career Selection Among Students.	54
6.3 To Analyze The Methods Currently Used In Career Guidance.....	55
6.4 To Review The Existing Algorithms and Models Used In Career Guidance..	55
6.4 To Develop a Career Guidance Tool Using Deep Learning.....	56
6.5 To Test and Validate The Developed Tool.....	56
Chapter 7: Conclusion and Recommendation.....	57
7.1 Conclusions	57
7.2 Recommendations	58
7.3 Limitations of the Study.....	59
7.3 Future work	59
References	61
Appendices.....	67
Appendix A: Similarity Report	67
Appendix B: Ethical Clearance Letter	67

List of Figures

Figure 1.1 Standard Approach for Career Guidance (Guleria & Sood, 2022).....	2
Figure 2.1 Rule Based Expert System Architecture (Gunwant et al., 2022).	19
Figure 2.2 Types of Machine Learning (Pugliese et al., 2021).....	22
Figure 2.3 Biological Neuron (López et al., 2022).	25
Figure 2.4 Conceptual Model.....	30
Figure 3.1 Agile Methodology	34
Figure 4.1: System Architecture	37
Figure 4.2: Use Case Diagram	38
Figure 4.3 Class Diagram.....	40
Figure 4.4: Sequence Diagram	41
Figure 4.5: Database Schema	41
Figure 4.6: Landing Page Wireframe	42
Figure 4.7: Login Wireframe	43
Figure 4.8: Register Wireframe.....	43
Figure 4.9: Career Evaluation Form Wireframe	44
Figure 4.10: Evaluation Results Wireframe.....	44
Figure 4.11: Career Evaluation History Wireframe.....	45
Figure 5.1: Schematic Representation of MLP	46
Figure 5.2: Home Page.....	47
Figure 5.3: Career Evaluation Form	48
Figure 5.4: Predicted Career	48
Figure 5.5: Load CSV file.....	50
Figure 5.6 One hot encoding	51
Figure 5.7: Model Training	51
Figure 5.8: API.....	52
Figure 5.9: Model Evaluation	53

List of Acronyms and Abbreviations

ANN	Artificial Neural Networks
DL	Deep Learning
KNN	K-Nearest Neighbor
KCSE	Kenya Certificate of Secondary Education
ML	Machine Learning
SHAP	Shapley Additive Explanation
SVM	Support Vector Machine
RF	Random Forest
VUCA	volatility, uncertainty, complexity, and ambiguity

Acknowledgments

I express my gratitude to the Almighty for providing me with guidance and strength throughout the completion of this thesis. My hope is that the insights and impact of this research will contribute to the betterment of humanity, particularly in the realms of education and industry. I extend heartfelt appreciation to my dedicated supervisor, Professor Ismail Ateya. His unwavering patience and invaluable guidance have been the bedrock of this research, from its inception to its realisation. I am profoundly thankful for the collaborative journey we have undertaken. May the blessings of the Almighty be upon each and every one of you.

Chapter 1: Introduction

1.1 Background

Career guidance in the era of life-long learning faces challenges related to building accessible services that bridge education and employment services. So far, only limited research has been conducted on using artificial intelligence to support guidance across working life. In light of technological advancements and societal progress, our world is currently navigating a landscape that is increasingly challenging and uncertain. We find ourselves in what can be termed as the VUCA era, marked by volatility, uncertainty, complexity, and ambiguity. Recent years have witnessed frequent "black swan events," and 2020 global upheaval due to the COVID-19 pandemic stands out as a prominent example (Wang et al., 2022). In this era of heightened unpredictability, characterized by the VUCA paradigm, it becomes imperative for individuals to discern their own positioning and chart a course for future career development.

students often find themselves grappling with the daunting task of selecting the most suitable career path. This dilemma frequently leads them to pursue careers for which they harbor minimal interest. Amidst the ever-intensifying competition prevalent in today's technical society, the need for students to cultivate a sense of organization and strategic planning from the outset of their educational journey becomes paramount (Gavhane et al., 2020). Guiding them towards their desired career trajectories necessitates continuous self-assessment, proactive motivation, and diligent monitoring of their progress. It is imperative to consistently evaluate their performance, discern their passions, and ascertain their proximity to their envisioned goals. Such vigilance ensures that they remain on the appropriate trajectory towards their desired destinations. The conventional method of career counseling, which does not utilize advanced tools and strategies, is depicted in Figure 1.1.

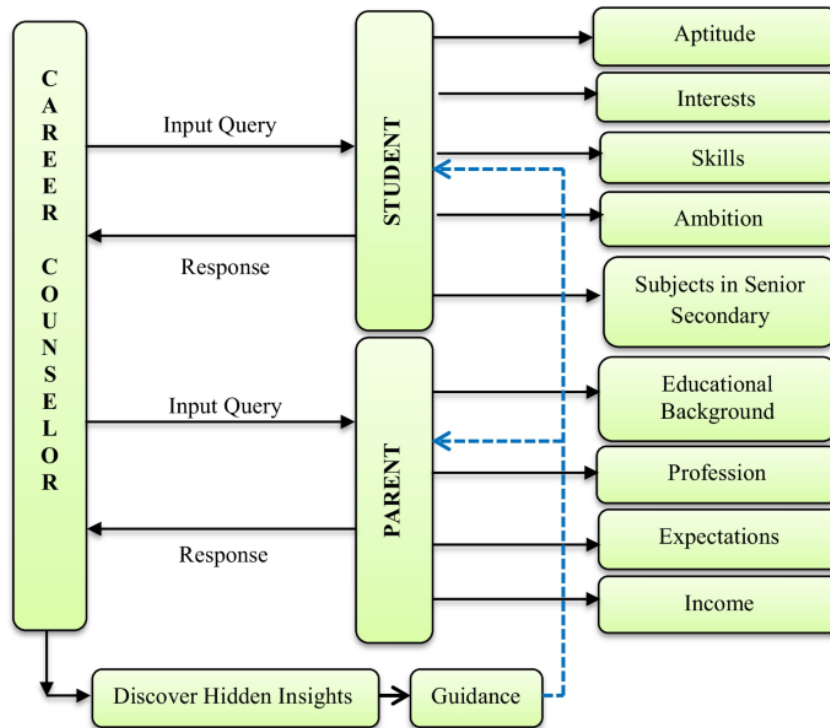


Figure 1.1 Standard Approach for Career Guidance (Guleria & Sood, 2022)

This technique involves career counselors gathering information from students regarding their interests, aptitude abilities, goals, and other relevant factors. Likewise, information regarding the educational history, occupation, income, and other relevant details of the parents is gathered (Guleria & Sood, 2022). Afterwards, the counselor endeavors to uncover the concealed perceptions from the feedback provided by parents and students and offers suitable career recommendations or recommends certain fields of study such as arts, finance, engineering, medical, etc., for the children. The conventional counselling strategy has a restricted scope and is confined to a particular subject domain region. There is a lack of advanced methodologies, repository of information, logical reasoning system, and storage system for the collected data. Without available data, machine learning techniques cannot be utilized to deduce outcomes and do predictive analytics.

Machine learning approaches can assist in resolving inquiries from students and parents regarding the optimal career selection through the use of data mining methodologies and algorithms to get knowledge from past occurrences. The ML-based system employs advanced techniques such as Artificial Intelligence, Deep

Learning, Neural Networks, and Natural Language Processing to make informed decisions regarding students' career guidance. Future counseling assists students in making informed decisions regarding their future path, enabling them to effectively utilize their interests and abilities. Furthermore, with the proliferation of educational institutions and the intensification of competition, selecting the appropriate career path has become a laborious undertaking for both parents and students.

This study developed an enhanced career guidance tool using Machine Learning techniques to assist students in selecting the appropriate career path based on their technical and non-technical in the IT field. The developed tool necessitates input data from students, encompassing attributes such as their personalities and technical abilities. This was achieved through a combination of personality assessments and technical questions. It goes beyond traditional career guidance by offering a range of diverse career options and providing detailed explanations for the recommended paths.

1.2 Problem Statement

As students approach the conclusion of their undergraduate studies, they inevitably find themselves at a juncture where the selection of a suitable career path becomes imperative. At such a pivotal point, these students may grapple with a lack of industry knowledge, making it challenging to precisely discern the optimal trajectory for their career paths. The decision-making process extends beyond the mere selection of a career; it involves envisioning the person they aspire to become post-graduation. In recognizing the need for a comprehensive and forward-looking guidance system, this study developed an innovative career guidance tool using deep learning. The developed career guidance tool will streamline the process of determining the most fitting career trajectory. Acknowledging that students typically receive advice from various sources such as parents, lecturers, and other educational experts, the developed tool is poised to address this challenge proactively. The core of the developed tool lies in an automated career guidance that predicts the ideal career for an individual based on an objective assessment of their skills. Upon completion of the assessment, intricately embedded in the developed tool, individuals will seamlessly discover a suitable career path. This approach is not merely about making a choice; it is a strategic

move intended to significantly reduce the risk of career misdirection by aligning individuals with careers that resonate with their skills and aspirations.

1.3 Research Objective

1.3.1 Aim

The aim of this study is to develop a career guidance tool using deep learning.

1.3.2 Specific Objectives

- i). To investigate factors influencing career selection among students.
- ii). To analyze the methods currently used in career guidance.
- iii). To review the existing algorithms and models used in career guidance.
- iv). To develop a career guidance tool using deep learning.
- v). To test and validate the developed tool.

1.4 Research Questions

- i. What are the factors that influence career selection among students?
- ii. What are the existing methods currently being used for career guidance?
- iii. What are the algorithms and models being used for career guidance?
- iv. How can a career guidance tool be developed?
- v. How can the developed tool be tested and validated?

1.5 Scope of the study

This study placed its primary emphasis on forecasting the career choices of undergraduate students specifically those pursuing IT. The scope of the study is intentionally confined to the undergraduate level. Consequently, its application does not extend to encompass individuals in the earlier stages of their education, such as those in primary school, nor does it address the needs of students pursuing secondary school studies. The decision to narrow the focus to undergraduate students stems from the recognition of the unique challenges and decision-making processes that emerge during this critical juncture. The study aims to provide targeted insights and assistance

precisely when students are on the brink of making significant choices that will profoundly impact their future trajectories.

1.6 Significance of the study

This research holds notable importance as it intentionally centers its focus on predicting the career choices of undergraduate students, representing a critical phase in their lives. At its core, the significance of this study lies in its timely intervention. Graduation marks a pivotal period where students stand at the crossroads of significant decisions regarding their future paths. By homing in on this juncture, the research endeavors to offer targeted insights and interventions, aiding students in navigating the challenging landscape of career choices with informed decision-making.

This study is not merely an academic pursuit; its practical relevance extends to key stakeholders. Educational institutions, parents, and policymakers all play vital roles in shaping the educational landscape for undergraduates. The outcomes of this research are poised to provide practical insights that can inform the strategies and initiatives of these stakeholders, fostering a more supportive and effective educational environment. Moreover, the research contributes to the enhancement of existing career guidance tools. By developing a tool tailored specifically to the needs of undergraduates, this study addresses a critical gap in current tools. It sets the stage for the evolution of more sophisticated and effective guidance systems, better equipped to assist students in aligning their choices with their unique aptitudes, interests, and aspirations.

The long-term impact of this study is a focal point. Recognizing that career decisions often have far-reaching consequences, the findings and resulting tool have the potential to positively influence students' future trajectories. The ripple effects extend beyond individual students, encompassing educational institutions, families, and broader educational policies and practices. In essence, the deliberate focus on undergraduate students and their career choices positions this study as a practical and timely intervention, poised to make meaningful contributions to the field of career guidance. Its significance lies not only in academic exploration but in the potential to shape the

trajectories of young lives and contribute to the broader landscape of educational practices and policies.

Chapter 2: Literature Review

2.1 Introduction

Navigating one's career path is a pivotal undertaking for young individuals, wielding a profound impact on their future social integration and overall well-being. Among the myriad considerations in life planning, the selection of a profession stands out as a critical focal point. The intricacies entwined with career choice, necessitating the harmonization of personal attributes with the demands of a profession, render the decision-making process a complex and somewhat mysterious endeavor. To unravel this enigma, career guidance assumes a paramount role. Its primary objective is to assist individuals not only in making informed decisions regarding education, training, and occupation but also in effectively managing the trajectory of their careers.

Career guidance encompasses collective activities strategically designed to equip learners with the competencies crucial for their career development (Sharapova et al., 2023). Concurrently, career counseling zooms in on addressing the specific career-related concerns of individuals. Complementing these, career information serves as a valuable resource, shedding light on diverse courses, occupations, and career pathways. This comprehensive approach underscores the multifaceted nature of career planning, emphasizing the importance of aligning personal aspirations with practical considerations for a well-informed and fulfilling professional journey.

2.2 Empirical Literature

The field of career guidance has garnered significant attention from scholars and institutions, who are devoting substantial efforts to tackle the complex task of professional decision-making at all levels of education, ranging from primary to postgraduate. Researchers and academics are increasingly utilising machine learning to analyse large databases of student information, taking advantage of the advancements in this field. The goal is to derive significant insights on the potential career paths of students, with the purpose of identifying not just viable but ideal job matches through a thorough review of their academic experiences.

Guleria and Sood (2022) presented a framework for career counselling of students that incorporates ML and AI methodologies in their study. ML-based White and Black Box models were employed to analyse an educational dataset consisting of employability

and academic attributes that are critical for students' skill development and employment placements. White Box and Black Box models were trained within the proposed framework using an educational dataset that was collected for the research. Naive Bayes achieved the highest Recall and F-Measure scores for prediction performance at 91.2% and 90.7%, respectively, when compared to the scores of Logistic Regression, Decision Tree, SVM, KNN, and Ensemble models utilised in their research. One of the limitations of this research is the restricted sample size and scope of attributes present in the dataset utilised. It is necessary to train machine learning models using educational datasets obtained from various academic institutions. These datasets should contain a greater number of attributes that are utilised for student assessment.

A web application for a career guidance system was designed and developed by Vinesh et al. (2021). This application offers candidates appropriate department recommendations to assist them in their decision-making process. In contrast to the pre-study iteration of the current career guidance system, the proposed system yielded more precise recommendations. The classification of candidates' skill sets, and prediction of suitable departments based on candidate performance were accomplished using the K-Nearest Neighbour algorithm and the K-Means Clustering algorithm, respectively. The formation of clusters involved dividing the students' scores for each specific skill set and calculating the success rate for different departments within each cluster. For department recommendation purposes, where a higher success rate and a lower failure rate are desired, the success rate in each cluster was computed and utilised. A web application with the anticipated results was subsequently designed and developed following an exhaustive study of the career guidance system for this undertaking. However, the accuracy rate of their model is low and needs enhancements to its accuracy rate and recommend appropriate departments through the utilisation of advanced machine learning models.

The researchers (Wang et al., 2022) employed the machine learning (ML) technique eXtreme Gradient Boosting (XGBoost) to forecast the vocational orientation of college students. To do so, they utilised a real-world dataset gathered exclusively from

a single college. The dataset comprised particulars regarding the educational pursuits and professional decisions of eighteen thousand college graduates throughout their collegiate years. SHAP (Shapley Additive Explanation) was also utilised to analyse the significance of individual features and interpret the results. In accordance with the outcomes, XGBoost can accurately forecast the career path of students, as evidenced by its F1 value of 0.872, recall rate of 85.4%, and precision of 89.1%, respectively. Additionally, the interaction of characteristics among four distinct student decisions—namely, studying abroad, working, choosing to study in China, and choosing to study—was investigated. A number of academic characteristics, most notably variations in grade point average (GPA) throughout collegiate studies, are discovered to exert a comparatively greater influence on the ultimate selection of a profession. While commendable strides have been made in refining the accuracy of our models, there remains ample space for further enhancement. It's imperative to note that this study primarily focused on college students, and the geographical context of the research being China somewhat diminishes its direct applicability to the Kenyan setting.

Rai et al. (2019) proposed innovative approaches to educational data mining and machine learning that they believe will be beneficial to students. Education sectors generate vast quantities of data that are of little use to students. Students typically incur academic losses when they choose the incorrect discipline. Their research identifies students with average academic achievement and directs them to the appropriate major in which they can thrive. They obtained an approximation of the outcome by employing machine learning methods. They compiled information using data from surveys and various institutions that collected information from the actual world. Machine learning was used to recommend the most viable and desired career path based on an individual's current educational standing. Data gathered was used to construct machine learning models. Improving student performance and assisting them in achieving their objectives was their core objective.

In a study conducted by Kazeem et al. (2017), a web-based intelligent career guidance system was created to aid pre-tertiary science students in Nigeria in autonomously

selecting a career trajectory. This system operates remotely and is accessible via computer systems or mobile devices/smart phones. Applicants to Nigerian higher education institutions (universities or polytechnics) may utilise the system to explore different academic disciplines. For career recommendation, the intelligent system employs student-generated parameters including combinations of preferred science subjects, results from analyses of career interest inventories, and intelligent quotient tests. The intelligent system that operates on the web was conceptualised and executed on the basis of a rule-based expert system employing the forward chaining algorithm. The client-side interface pages, also known as the "bootstrap 3" front-end framework, were developed utilising JavaScript and HTML5. Two hundred pre-tertiary science students participated in the implementation and evaluation of the system; they were administered career choice assessments and were asked to provide feedback regarding the performance of the system. The responses indicated that the information provided by the system regarding career guidance is 80% adequate, 70% satisfactory, and 95% accurate and pertinent in recommending a career.

Table 2.1 Summary of Empirical Literature

Author	Technique	Performance Metrics	Limitations
Guleria and Sood (2022)	Naïve Bayes	F1 Score: 0.91 Recall: 0.90	-This research utilized restricted sample size and scope of attributes present in the dataset.
Vinesh et al. (2021)	KNN KMeans	Accuracy: 0.79	-The accuracy rate of their model is low and needs enhancements to its accuracy rate and recommend appropriate departments through the utilisation of advanced machine learning models.

Wang et al (2022)	XGBoost	F1:0.87 Recall:0.85 Precision:0.89	- This study primarily focuses on college students, and the geographical context of the research being China somewhat diminishes its direct applicability to the Kenyan setting. -The performance metrics of their models is still low and can be improved.
Rai et al. (2019)	Data Mining	Not Stated	-Data mining techniques can rely on outdated information and is very expensive.
Kazeem et al. (2017)	Expert System	N/A	-Expert systems requires a lot of manual work and deep knowledge of the domain. -Generating rules for expert systems is time consuming and complex endeavour.

2.3 Theoretical Frameworks

2.3.1 Trait and Factor Theory

The first career development theory was proposed by Frank Parson in the early 1900s. Parson's concept of career guidance became the foundation for what later became known as trait and factor theory (Parsons, 1909). "Trait" refers to a characteristic of an individual that can be measured through testing, while "Factor" refers to a characteristic required for successful job performance (Sharf, 2013). Thus, the terms "trait" and "factor" refer to the assessment of characteristics of the person and the job. Parson (1909) proposed that, to select an occupation, an individual should have (a) a clear understanding of him or herself--attitudes, abilities, interests, ambitions, resource

limitations, and their causes; (b) a knowledge of the requirements and conditions of success, advantages and disadvantages, compensation, opportunities, and prospects in different lines of work; and (c) true reasoning on the relations of these two groups of facts. Simply stated, trait and factor theory of career development means matching the individual's trait with requirements of a specific occupation, subsequently solving the person's career-search problems (Zunker, 2006).

This theory greatly influenced the study of job descriptions and job requirements in an attempt to predict future job success from the measurements of traits that are job related. Fundamental to this theory is the assumption that individuals have unique patterns of ability or trait that can be objectively measured and correlated with the requirements of various types of jobs. The development of assessment instruments and the refinement of occupational information are closely associated with this theory. Although there is little research supporting or refuting trait and factor theory itself as a viable theory of career development, studies that have been conducted focused on relating trait factors to one another or establishing the validity and reliability of measurements of traits and factors--aptitudes, achievements, interests, values, and personality--which have been correlated with each other by developers of tests (Sharf, 2013). As the oldest and most widely used of all career development theories, trait and factor theory focuses on the match between an individual's aptitudes, achievements, interests, values, and personality and the requirements and conditions of occupations.

2.3.2 Ginzberg's General Theory of Occupational Choice

The initial comprehensive theory of occupational behaviour was introduced by Ginzberg, Ginzburg, Axelrad, and Herma (1951). This theory was formulated through an empirical investigation that involved conducting interviews with sixty-four students attending Horace Mann Lincoln School and Columbia University, both located in New York City. Ginzberg et al. employed a cross-sectional sampling technique to select a group that they perceived to have the greatest degree of autonomy in selecting an occupation. The objective of the study was to derive broad conclusions regarding the occupational decisions that young individuals make prior to and subsequent to completing college. The researchers reached the conclusion, in accordance with their findings, that the process of occupational decision-making could be delineated along

three developmental phases, with each stage distinguished by the manner in which an individual converts their impulses and needs into an occupational selection.

They refer to the initial phase as the "fantasy period." A child develops a desire for an occupation and simulates a variety of adult situations through the use of make-believe work roles between the ages of 6 and 11. As a result of the infant making numerous decisions without being cognizant of the potential obstacles that may impede their progress, Ginzberg selected the term "fantasy" to describe the nature of decisions at this stage of development. While enjoyment is often the main factor influencing work activity selection, the fantasy period is distinguished by arbitrary decision-making that disregards reality, abilities, and potentials—the three most essential components identified by Ginzberg in the process of occupational choice (Bailey & Stadt, 1973).

The tentative period (ages 11 to 17) signifies a growing preoccupation with the actualities of life during the second stage. This phase, which is additionally subdivided into transition, interest, capability, and value stages, is distinguished by the acknowledgment of the challenges associated with determining a future profession, the selection of activities predominately based on personal preferences and interests, the assessment of one's capability to operate effectively in areas of interest, the recognition of objective criteria that must be considered when selecting an occupation, and a transition from subjective considerations—interest, capabilities, and values—to objective criteria. As stated by Ginzberg et al., the transition stage, which coincides with the conclusion of high school, is a critical juncture in an individual's development. The ultimate phase, known as the realistic period (encompassing individuals aged 18 and older), is comprised of the following: exploration, crystallisation, and specification. During this stage, one obtains the necessary education and experiences to resolve occupational choice dilemmas; develops the capacity to synthesise various internal and external factors that influence one's decision; and specialises and plans within one's chosen field. The individual realises that in order to achieve his or her goals, a compromise must be reached with the constraints of the opportunities that are currently available in the actual world.

The significance of Ginzberg's research and resulting theory lies in the conclusion that occupational choice is a developmental process spanning a minimum of ten years, as

opposed to a singular decision. Additionally, the theory posits that the occupational selection process is predominantly irreversible and culminates in a state of compromise. Thus, in the years preceding a decision, an individual attempts to strike an equilibrium between a variety of subjective factors—interests, abilities, and values—and the realities of possibilities and constraints. Ginzberg maintained his stance on the notion of compromise on the grounds that no person ever arrives at a career decision that entirely fulfils their primary needs and desires. However, he conceded to the notion of "optimisation" as a more applicable formulation, arguing that each career choice involves a "comparison of probable benefits and costs" (Bailey & Stadt, 1973).

2.3.3 Super's Developmental Self-Concept Theory

Prominent psychologists and theorists, including Hull, Bandura, Freud, Jung, Adler, Rank, Murray, Maslow, Allport, and Rogers, among others, provided the foundational assumptions upon which Super's Life-Span theory of career development was built. Super (1957) constructed a model for this investigation by consolidating prior generalisations of life stages into the following categories: growth (birth to 14 years), exploration (ages 15 to 24), establishment (ages 25 to 44), maintenance (ages 45 to 64), and decline (above 65 years). The foundation of his theory is the premise that geographical factors (such as country of origin) and psychological factors (such as inherited predisposition) influence other facets of career advancement. The social-economic structure of the environment and the formation of psychological traits are among these elements that are assimilated into the self-concept. Personality formation and achievement are influenced by psychological attributes such as intelligence, aptitude, interests, and special aptitudes. Conversely, social-economic factors, including but not limited to the individual's community, school, family, peer groups, and the state of the economy and labour market, impact employment practises and job structure. These external conditions force individuals to interact with one another. Self-development occurs as a consequence of the interplay between psychological and socioeconomic elements.

Fundamental to Super's developmental theory is the notion of self-concept, which pertains to an individual's perception of themselves and their circumstances. Self-concept, as opposed to trait and factor theory, which emphasises objective or external

measures of the self (e.g., aptitude tests and interest inventories), pertains to individuals' subjective perceptions of themselves and society. The self-concept is formed through phenomena such as self-differentiation, role-playing, exploration, and reality testing. Vocational development, according to Super, is the formulation and application of one's self-concept. According to Super (1990, 1994), self-concept is "a composite of an individual's biological attributes, the social roles they assume, and the assessment of how others perceive them". The way in which people interact and perceive themselves is indicative of their personality, requirements, values, and interests, and these perceptions evolve throughout an individual's lifetime. In order to underscore his commitment to self-concept, Super devised inventories that centre on the assessment of significant roles and values that transpire throughout various life stages. A child's self-concept, according to Super, is formed through exploratory behaviour that results in the acquisition of occupational knowledge, imitation of influential figures, and the development of interests.

Individuals play significant roles in various aspects of life, including work, study, community service, home and family, and leisure activities. The significance of these roles can be discerned through an individual's level of engagement, dedication, or appreciation for a particular activity. The fundamental phases of career development—exploration, establishment, maintenance, and disengagement—are situated within the framework of roles (Super, 1953, 1990, 1994). The exploration stage (ages 15 to 25) comprises the substages of crystallisation (individuals specify their aspirations), specification (they specify their preferences in order to locate an employer), and implementation (individuals finalise plans to achieve their career objectives prior to labour). The exploration stage consists of efforts to gain a better understanding of occupational information, select career alternatives, decide on occupations, and begin working. Subsequent to exploration, individuals enter the establishment stage (ages 25 to 45), wherein they begin a job that is likely to signify the commencement of their professional careers. During this phase, they undertake the following responsibilities: stabilising (acclimating to a particular job and fulfilling its requirements to ensure continued employment), consolidating (the individual gains confidence in their job and strives to demonstrate their competence and dependability to others), and advancing.

Depending on company policies and other variables, the subsequent stage, maintenance (ages 45 to 65), consists of the following: holding (concerned with retaining the secured position), updating (enhancing one's expertise and knowledge in the field to preserve the status), and innovating (making advancements or novel contributions to the profession). Disengagement from work (caused by physical health issues, outdated knowledge and skills, and other factors) is the last stage (ages 65 and older). It consists of deceleration (reducing work responsibilities or time spent on work), retirement planning (directly addressing retirement arrangements in later life; includes activities such as planning and organising activities to be undertaken in retirement), and retirement living (addressing concerns such as friends, family, and friends). Numerous revisions have been made to Super's (1953) initial life-span theory of career development in order to enhance clarity and integrate recent research discoveries into its ten propositions. Super increased the number of propositions to fourteen in 1990.

According to Super (1953), career decisions are inseparable from life adjustments and constitute a continuous process (Bailey & Stadt, 1973). Super offers a critique of alternative theories, notably Ginzberg's, which neglects to consider the continuity of preference development throughout the phases of selection, entry, and adjustment. A career development theory, according to Super, should elucidate the mechanisms by which opportunities, interests, capabilities, and values are compromised. According to Super (1963), an individual endeavours to manifest his or her self-concept through the selection of a profession that allows for personal expression; an individual chooses a profession whose demands align with his or her self-perception. The stages involved in the development of self-concept are delineated as formation, translation, and implementation. Exploration of the self and the environment, self-differentiation, self-identification, role-playing, and reality testing are additional components of the formation stage. Translation can be accomplished in a multitude of methods. For instance, becoming cognizant of one's critical attributes within a particular domain might prompt an exploration of that vocation. The culmination of the process is the implementation of self-concept, which entails commencing professional training, finishing one's education, and entering the workforce. The development of an individual's sense of self commences during infancy and is fundamentally a "investigative process that continues for the duration of one's life". According to Super

(1963), the degree of career fulfilment individuals experience is contingent upon the degree to which they are capable of putting their self-concept into practise. Super's theory is founded upon two fundamental concepts: firstly, that vocational development is a "continuous, generally irreversible process"; and secondly, that it follows a structured pattern (p. 42).

2.4 Factors Influencing Career Selection Among Students

2.4.1 Interest

Interest (INT) is a fundamental requirement in the process of career selection. It can be defined as an effective response that generates curiosity, concern, and attention (Ahmed et al., 2017). In accordance with this definition, students determine their professional paths in accordance with their personal interests. Alexander et al. (2011) demonstrated that the career decisions of students were impacted by their level of interest in information technology as a subject. Ahmed et al. (2017) posit that the primary determinant influencing business students' career decisions is their "interest in the subject." Suutari (2003) hypothesised that interest and vocation selection decision-making are intricately intertwined. Furthermore, students' career decisions are significantly influenced by their individual interests.

2.4.2 Self-Capacity

According to the hypothesis of Alexander et al. (2011), self-efficacy is associated with career choice and, consequently, the capacity to envision viable career options. Additionally, it was determined that the evident self-efficacy of students influenced their career advantages, capabilities, goals, and confidence in seeking a desired profession. According to Taylor and Popma (1990), the level of general self-efficacy predicted career investigation abilities, while Bergeron and Romano (1994) found that self-efficacy beliefs influence undergraduate students' major selection and career decisions. The research results underscored the importance of the self-efficacy construct and placed particular emphasis on how students cultivated their self-efficacy in alignment with their personal attributes and professional pursuits.

The pursuit or investigation of career information is an endeavour to acquire knowledge regarding one's personal attributes in order to attain suitable professional objectives (Zhang & Huang, 2018). A significant proportion of undergraduate students are unprepared for the professional realm due to their limited capacity for career discernment and selection of tasks that fall beyond their area of expertise (Karacay, 2018). The conceptualization of self-efficacy in career decision-making will be influenced by the psychological motivation component that is essential for this process.

2.4.3 Career opportunities

Field preferences may be influenced by the way in which students perceive employment prospects. Students are more inclined to select a field that provides a broader range of employment prospects (Nguyen et al., 2023). The ability to acquire a reputable social status and income, as well as to effect positive change in society, are the most significant factors that Ahmed et al. (2017) identifies as influencing an individual's choice of occupation. The career decisions of individuals have a significant impact on society as a whole, thus these influences determine its future. In a study of business students, Ahmed et al. (2017) discovered that the influence of prospective job opportunities, financial outcomes, and ease of achieving grades on students' decision-making was negligible or unrelated.

2.5 Techniques Used For Career Guidance

2.5.1 Expert Systems

2.5.1.1 Rule-based Expert System

Rule-based systems enable the accessibility of computational mechanisms that are present in the majority of ES. These systems have been utilised in a variety of fields for a very long time (Gunwant et al., 2022). In rule-based systems, the knowledge base is represented by a collection of straightforward and readily comprehensible IF-THEN rules. By utilising these IF-THEN rules as inference techniques, novel logical conclusions can be deduced from prior knowledge. It is noteworthy that equivalent logical statements can serve as effective representations of both the rules and the facts within a rule-based system. The benefits of rule-based systems have been substantially

enhanced by the consolidation of numerous restrictions and language-based additional constructs, as well as the tight integration between language and evaluation mechanisms. The general architecture of a rule-based system is depicted in Figure 2.1.

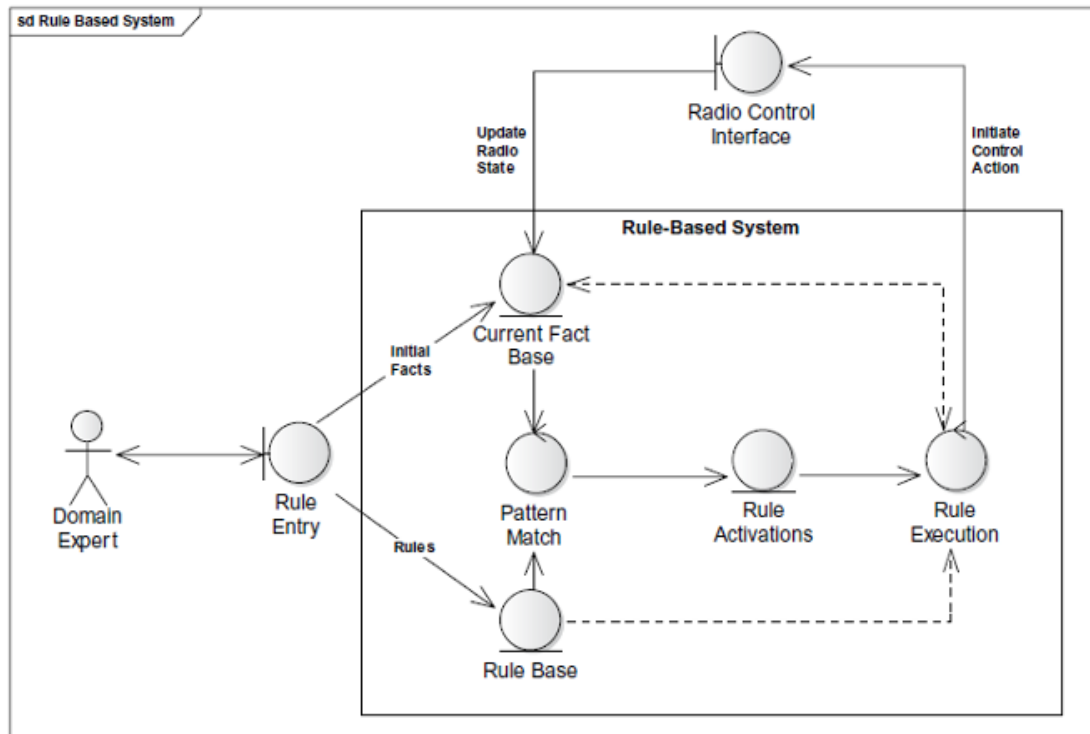


Figure 2.1 Rule Based Expert System Architecture (Gunwant et al., 2022).

2.5.1.2 Case-based Expert System

The case-based ES is an experience-based system utilised to solve new problems through the adaptation of identical problems that have been effectively solved in the past. Case-based systems enable the development of an innovative intelligent technology capable of both resolving a given issue and adjusting to evolving circumstances. It is founded upon the fundamental premise that individuals resolve a multitude of challenges through the accumulation of experience. Humans transform the patterns of cases into principles, with "the case" serving as the primary unit of knowledge.

An Excel file containing a dataset of 1,000 cases was utilised to implement a software-based system described by Ezenkwu et al. (2015) through a MATLAB-based graphical

user interface. The case-based system was designed to assist tertiary students or those who are having difficulty finishing their courses. The counsellor gathered information regarding the academic progress of students across ten different subjects. A second web-based system, devised by Daramola et al. (2014), has been effectively implemented at Covenant University. This system contains a dataset that contains records of students who have graduated. This system facilitates the course selection process for both university faculty and students. A Java ES shell was utilised in the design of the system, and Apache Tomcat was deployed as the web server. Mundra et al. (2014) have developed an electronic supplementary (ES) that operates on software and tutors secondary and high school students. A Java-based interface, the MBTI model, machine learning, and rule-based systems were utilised in the development of the system. Career recommendations were generated utilising the subject-specific essential data grades of the students and personality factors computed using the MBTI model. In order to evaluate and calculate various factors, generate reports, and expand the knowledge base for unknown cases, three modules were developed. A software-based system has been constructed utilising the nearest neighbour data mining technique (Ansari, 2017). The students assessed the eleven distinct components of the system using a five-point Likert scale in accordance with their individual preferences. The system furnishes guidance at the primary, secondary, and higher-level levels to the students. The knowledge base is consistently augmented through the updating of solutions in situations where primary-level guidance is absent. Twelve attributes pertaining to aptitude, disposition, skill, and interest were incorporated into a software-based ES developed by Waghmode and Jamsandekar (2016). The Weka processing filters were utilised to implement the data-mining technique, and the data set was obtained from the counsellor. After calculating a number of factors via a questionnaire, the system was tailored specifically for secondary school students.

2.5.2 Machine Learning

Machine learning (ML) is an advanced branch of artificial intelligence that has the potential to revolutionize the current era (Nichols et al., 2018). The rapid and recent advancements in both the design and algorithms of computers, as well as the expansion of datasets, have resulted in a significant increase in computer proficiency across various sectors. The effectiveness and productivity of a machine learning solution are

contingent upon the inherent qualities and attributes of the data, as well as the proficiency of the learning algorithms. To address the vast range of data in various machine learning situations, a variety of machine learning algorithms have been created, including supervised, unsupervised, semi-supervised, and reinforced methods (Pugliese et al., 2021).

Indeed, despite the growing interest in these domains over the past decade, we have only just started to explore the full potential of machine learning algorithms in analyzing systems that enhance their performance via experience. Regrettably there are still numerous unresolved inquiries pertaining to (1) the algorithm's ability to effectively learn from specific types and quantities of data, (2) the algorithm's resilience in the face of errors in its modeling assumptions or training data, and (3) the feasibility of devising a successful algorithm for a given learning problem (Pugliese et al., 2021).

Machine learning encompasses the creation and implementation of algorithms that, instead of being explicitly programmed to produce certain outputs (i.e., actions) in response to particular inputs from the environment, assess the data and its characteristics, and make decisions using statistical methods. Typically, machine learning algorithms are adaptive and enhance their performance by acquiring new data, a process referred to as "learning" (Pugliese et al., 2021). Machine learning algorithms can be categorized into four main groups: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, as seen in Figure 2.2.

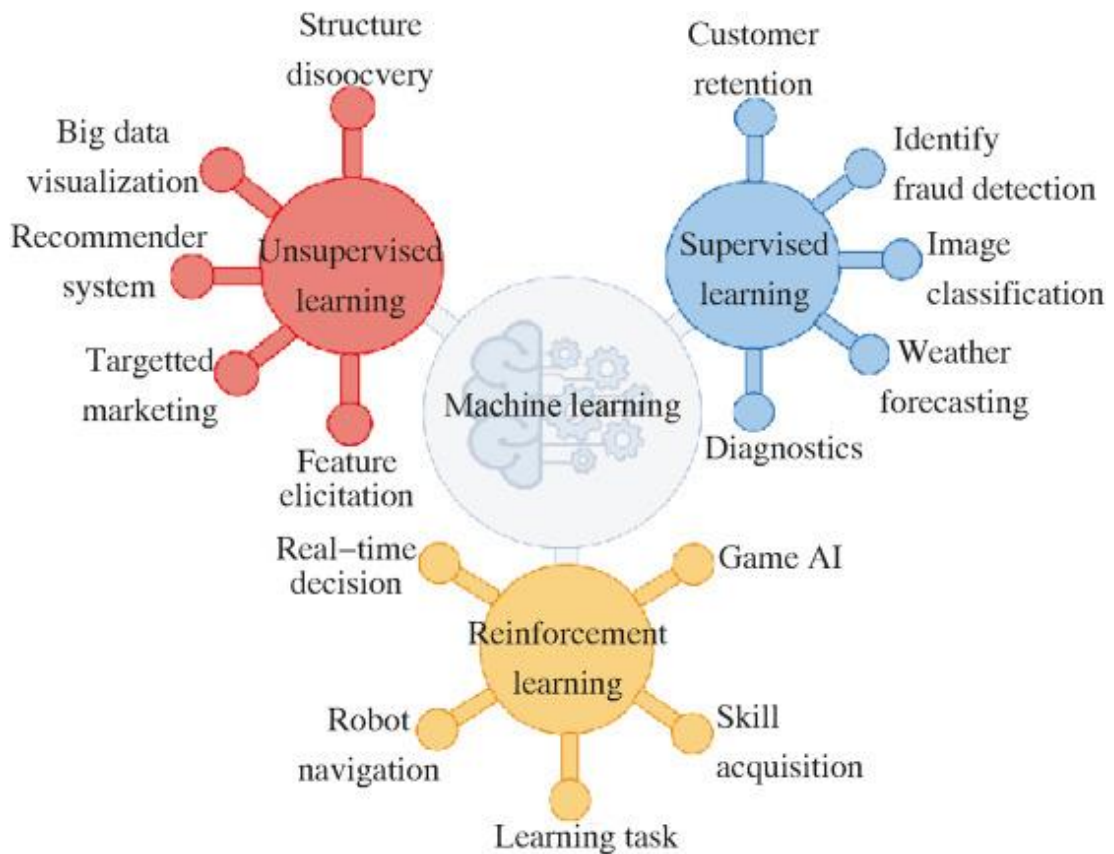


Figure 2.2 Types of Machine Learning (Pugliese et al., 2021)

2.5.2.1 Supervised Learning

Supervised learning is dependent on machine learning tasks that aim to acquire knowledge of a function that establishes a relationship between an input and an output, using a set of example input-output pairs. Therefore, this learning process involves comparing the computed output and the predicted output. Learning entails calculating the error and adjusting achieve the desired output. Some methods that fall into this category are Naïve Bayes classification, linear and logistic regression, and Support Vector Machines (SVMs). Applied supervised learning encompasses various applications such as automated response to incoming messages, particularly beneficial for large corporations, as well as facial recognition, which serves as a security measure in ATMs, surveillance areas, closed circuit cameras, the criminal justice system, and image tagging on social networking platforms like Facebook. Supervised learning can also be seen in the field of epileptic seizure detection. This involves creating detectors that are specific to each patient and can promptly identify the beginning of a seizure. By doing so, it helps to prevent physical harm and death with a high level of accuracy.

2.5.2.2 Unsupervised Learning

Unsupervised learning is a method of analyzing datasets that do not have any labels or human intervention. In unsupervised learning, the algorithm effectively partitions the samples into distinct classes only based on the properties of the training data, without any accompanying labels. The unsupervised techniques include k-means clustering, principal component analysis, and autoencoders. An example of unsupervised learning is the automated recognition of a user's friends on social networking platforms such as Facebook or Google, or the determination of the highest volume of emails sent to a specific individual and organized into distinct groups.

2.5.2.3 Semi-Supervised Learning

Semi-supervised learning is a technique that combines supervised and unsupervised methods by using both labeled and unlabeled data (Zhou and Belkin, 2014). The primary objective of a semi-supervised learning model is to enhance prediction accuracy beyond what can be achieved using only labeled data. This method is extensively employed in machine translation, fraud detection, data labeling, and text classification.

2.5.2.4 Reinforcement Learning

Reinforcement learning is based on a set of algorithms that usually work in a sequential manner to automatically determine the most effective behavior in a specific environment in order to enhance its efficiency. This technique is driven by the characteristics of the environment. During each iteration, a reinforcement algorithm, also known as a "agent", takes action and predicts the characteristics that will occur in a future iteration based on past and current characteristics. A reward or penalty is then assigned based on the accuracy of the prediction. Hence, it serves as a potent instrument for training artificial intelligence models that have the potential to enhance the operational effectiveness of complex systems, including robotics, autonomous driving, manufacturing, and supply chains. The reinforcement techniques employed include TD (λ) with function approximation, gradient temporal difference learning, and the least-squares method. A remarkable instance of reinforced learning is the algorithm capable of autonomously determining the ideal tension and optimal

direction for a certain cutting path, whether performed by a laparoscopic surgeon or an automated cutting tool.

2.6 Models and Frameworks

2.6.1 Random Forests

Random Forest (RF) is a machine learning technique that is employed for both classification and regression tasks. The method, developed by Breiman (1996), combines Breiman's bagging sampling technique (1996a) with the random selection of features, independently introduced Ho (1995), Ho (1998), and Amit and Geman (1997). This combination aims to create a set of decision trees with controlled variation. Bagging involves constructing each decision tree in the ensemble by sampling with replacement from the training data. Based on statistical analysis, it is probable that approximately 64% of the occurrences will appear at least once in the sample. The instances present in the sample are called in-bag instances, whereas the remaining instances (about 36%) are referred to as out-of-bag instances. Every individual tree inside the ensemble functions as a foundational classifier to ascertain the class designation of an unclassified instance. The classification process involves employing majority voting, in which each classifier contributes one vote based on its anticipated class label. The class label that receives the greatest number of votes is ultimately utilized to categorize the instance.

2.6.2 Artificial Neural Networks

An artificial neural network (ANN), usually referred to as a neural network, is a computational model that can process information to address problems such as classification and regression. It refers to the aspect of artificial intelligence that draws inspiration from the structure and functioning of the human brain and nervous system (Tian et al., 2021). The motivation behind ANN stems from the fascination with the human brain's ability to perform intricate computations, which differs significantly from the typical methods employed by digital computers. The human brain possesses a superior capacity compared to numerous information-processing systems. It is capable of executing intricate, nonlinear, and parallel processing by arranging its structural elements (neurons) to accomplish tasks such as precise predictions, pattern recognition, perception, motor control, and more. It is far more efficient than the

current fastest digital computer. One instance is the complex operation of the cognitive process known as human vision. This system facilitates our comprehension and acquisition of the fundamental elements of the environment, providing us with the necessary information to effectively engage with it. The brain frequently carries out perceptual recognition tasks, such as identifying voices within a complex scene, in approximately 100-200 milliseconds. In contrast, less complicated activities often require more time, even when executed on a high-performance computer. Figure 2.3 depicts the representation of a biological neuron.

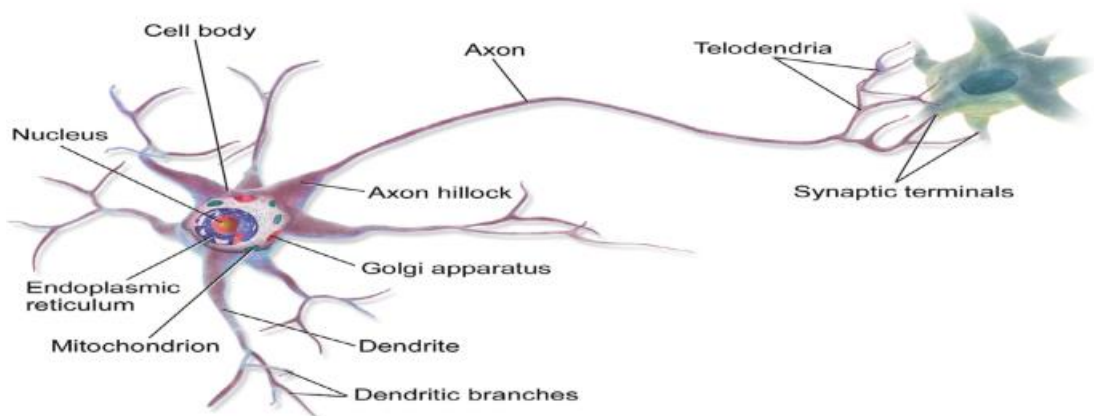


Figure 2.3 Biological Neuron (López et al., 2022).

ANN are computers that mimic the functioning of the human brain and construct a complex network of artificial neurons or processing units, numbering in the hundreds or even thousands. The artificial neural network is accomplished by creating a computational learning algorithm that does not require explicit programming of all rules. Instead, it is capable of autonomously forming its own rules of behavior based on accumulated experience. Neural networks can be practically implemented because they are highly parallel computing systems composed of numerous interconnected processing units (neurons). These neurons learn from their surroundings, and the synaptic weights store the strengths of the connections between neurons. The learning algorithm's task involves sequentially and supervisedly adjusting the synaptic weights of the network in order to achieve a given aim (López et al., 2022). There is data indicating that the collaboration of neurons can acquire intricate linear and nonlinear input-output connections through the utilization of sequential training methods. It is noteworthy that while the source of inspiration for these models differs significantly

from that of statistical models, the fundamental components of both types of models are remarkably similar.

2.6.3 Support Vector Machines

Support Vector Machines (SVMs) are a collection of interconnected techniques used for supervised learning, which may be applied to solve both classification and regression issues. The SVM classifier, introduced by Vapnik in 1995, has gained popularity over the past decade due to its strong theoretical basis. The advancement of effective implementations resulted in the use of various applications. Within the statistical learning framework, learning refers to the process of estimating a function based on a collection of samples, known as the training sets. In order to accomplish this, a learning machine must select a single function from a predetermined set of functions. This function should minimize a specific risk, known as the empirical risk, which measures the likelihood that the estimated function differs from the true function (which is currently unknown). The level of risk is contingent upon the intricacy of the selected set of functions and the training set. Therefore, a learning machine must identify the optimal collection of functions, based on their complexity, and select the most suitable function from that collection.

2.6.4 K-Means

K-means is a popular unsupervised learning technique used to solve clustering problems (Haraty et al., 2015). The methodology follows a direct and uncomplicated way to categorize a given amount of information into a predetermined number of groups (excluding k groupings). The main concept is to define k centroids, with each centroid representing one group. The centroids must be positioned strategically to create a unique region with a wide range of impacts. Therefore, the optimal choice is to position them at the greatest possible distance from each other. The next phase involves pairing each point in a given set of information with the nearest centroid until all points have been assigned to a group. After completing the initial stage and performing an unexpected aggregation automatically, it is necessary to reconfigure k new centroids as the barycentre's of each group as a result of the previous step. Once these k new centroids are generated, it is necessary to construct another connection between the existing set of centroids and the nearest new centroid. A cycle will be

generated. As a consequence of this cycle, it is evident that the k centroids will modify their assigned regions, ultimately resulting in the centroids no longer changing their positions (Haraty et al., 2015).

Model	Advantages	Limitations
Random Forests	<p>-Random Forest excels in predictive accuracy by aggregating multiple decision trees. In career guidance tools, this robustness ensures reliable predictions regarding suitable career paths based on various input factors such as skills, interests, and academic achievements.</p> <p>-The ensemble nature of Random Forest minimizes overfitting, making it effective for handling diverse datasets common in career guidance scenarios.</p>	<p>-Random Forest models can be computationally expensive, particularly when dealing with a large number of decision trees. In the context of career guidance tools, this resource intensity might impact real-time responsiveness, especially if the tool is deployed on platforms with limitations in computational resources.</p>
Artificial Neural Networks	<p>-Artificial Neural Networks (ANNs) excel at processing and understanding complex patterns and data. In career guidance tools, this ability allows for the creation of highly personalized recommendations. ANNs can analyze an individual's skills, preferences, and career history</p>	<p>- The performance of ANNs heavily relies on the quality and representativeness of the data used for training. In career guidance, if the training data is biased or incomplete, the recommendations generated by the network may also exhibit biases or inaccuracies.</p>

	to provide tailored advice, ensuring that career guidance is more relevant and effective.	
Support Vector Machines	<p>-Support Vector Machines (SVMs) are particularly advantageous when data exhibits non-linear patterns. In the realm of career guidance, where the relationships between various factors (e.g., skills, preferences, market demands) can be intricate and non-linear.</p> <p>-SVMs excel in capturing and understanding these complex associations. This makes them suitable for modeling the nuanced nature of career decision-making.</p>	-Support Vector Machines involve parameters like the kernel type and regularization parameter. The performance of SVMs is sensitive to the appropriate selection of these parameters.
K-Means	-K-means clustering is computationally efficient and well-suited for organizing large datasets into distinct clusters. In the context of career guidance, this efficiency allows for the quick categorization of diverse sets of skills, qualifications, or career preferences.	- The outcome of K-means clustering is sensitive to the initial placement of centroids.

2.6 Gaps in the Existing Systems

In spite of the commendable strides made in the field of career guidance, there remain notable gaps that current research has yet to address comprehensively. The prevailing research landscape heavily leans on conventional machine learning models for predicting career paths. In the face of rapid advancements in machine learning and the formidable capabilities of deep learning, there exists a considerable opportunity to enhance the accuracy of these predictions. Deep learning, with its capacity to discern complex patterns and relationships in data, holds promise for refining career guidance methodologies.

The attributes employed by some researchers in existing studies have been limited in scope, leading to suboptimal precision and recall rates. A more nuanced and comprehensive set of attributes, made possible through deep learning techniques, could substantially improve the precision of career predictions. This advancement is crucial for providing students with more accurate and tailored guidance as they navigate critical decisions about their future. Furthermore, a noteworthy gap in the literature lies in the lack of focus on the Kenyan context. This study seeks to rectify this omission by delving into the intricacies of career guidance within the Kenyan educational landscape. Recognizing that cultural, economic, and regional factors play a significant role in shaping career trajectories, this research aims to contribute valuable insights that are directly applicable to the unique circumstances of undergraduate students. This study aims to fill these gaps in the current understanding of career guidance by introducing a novel tool. This tool, harnessing the power of deep learning, aspires to serve as a lifeline for undergraduate students on the cusp of choosing their paths in the industry. By leveraging cutting-edge technology and contextualizing insights within the Kenyan milieu, the proposed tool represents a substantial contribution to the advancement of career guidance practices.

2.7 Conceptual Model

The conceptual model precisely delineates the operational architecture of the developed career guidance tool, showcasing the seamless incorporation of state-of-the-art technologies to provide students with career guidance. The student historical data will be collected from the trustworthy Kaggle platform, which is a vital source for

varied datasets. The data will then undergo thorough preprocessing. This crucial process guarantees that the data is processed and improved to serve as the fundamental basis for training our advanced deep learning algorithms. The core of the model is based on the use of sophisticated deep learning algorithms. By employing an iterative approach, the model undergoes rigorous training, utilizing the capabilities of neural networks to convert unprocessed data into a highly useful and intelligent tool that may offer valuable career advice.

Students will actively interact with the tool using a user-friendly interface, where they enter important information that is crucial to the career coaching process. The synergy between technology and user ensures that the tool effectively accommodates the unique requirements and ambitions of every learner. The API call initiates the model's inference process, allowing it to thoroughly assess the presented data and facilitate a smooth interface. After the inference is finished, the tool will convey the results back to the user and stored to the database. Figure 2.4 depicts the conceptual model, illustrating the interaction of data collection, preprocessing, deep learning, user input, model inference, and results storage.

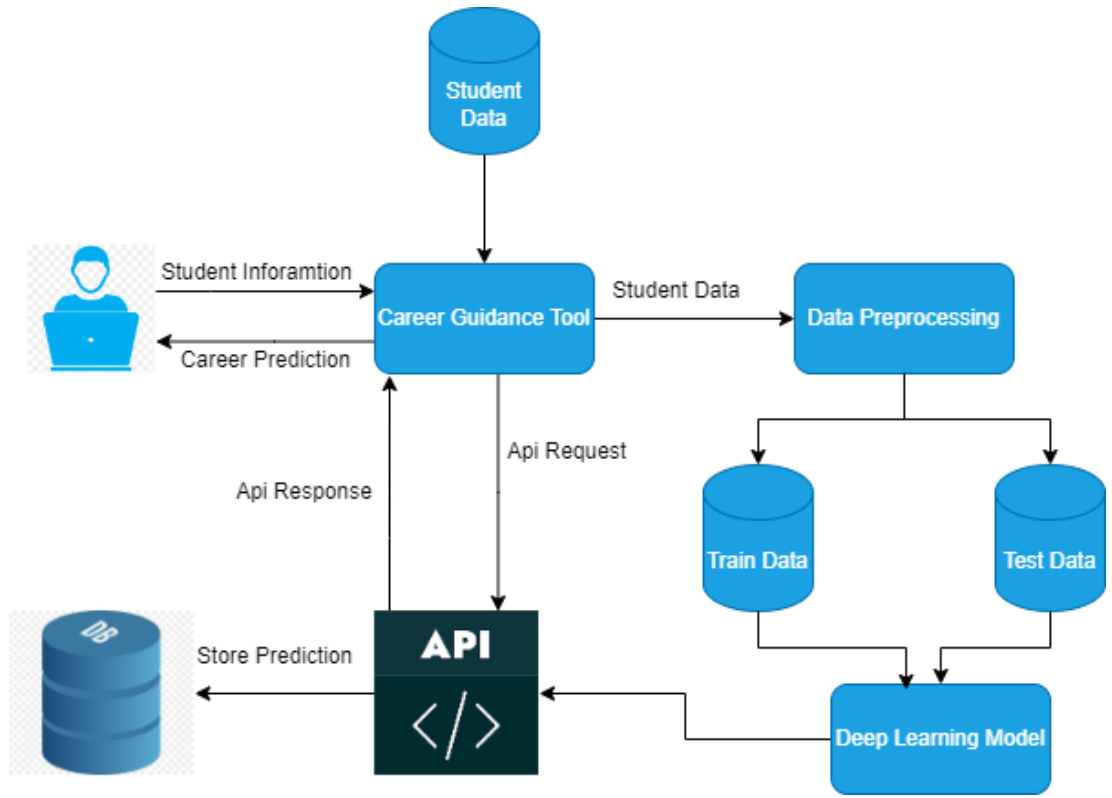


Figure 2.4 Conceptual Model.

Chapter 3: Research Methodology

3.1 Introduction

Research methodology entails the systematic study of the step-by-step methods employed in conducting a particular study (Kumar, 2005). The methodology used in this study is strategically designed to align with the specified objectives, serving as a roadmap toward resolving the identified problem statement.

3.2 Research Design

The concept of "research design" encompasses the comprehensive approach used by the researcher to integrate different elements of the study efficiently and logically, with the aim of successfully addressing the research topic (Cooper & Schindler, 2014). The research design functions as a comprehensive framework for the gathering, measurement, and analysis of data. The study design phase holds significant importance as it centers on the methodologies employed by the researcher to

accomplish their objectives. The most prevalent study designs in academic studies encompass descriptive, exploratory, and explanatory approaches. A descriptive research strategy is a scientifically rigorous approach used to elucidate the features or behaviors exhibited by the population being investigated.

Conversely, the exploratory research design pertains to a form of unstructured investigation undertaken to get preliminary knowledge regarding the study issue. This type of research typically does not encompass a sampling strategy, research objectives, or a questionnaire. The explanatory research design is a methodology employed to establish causality or investigate the extent to which one or more variables can account for the causes or effects of one or more other variables (Tobi & Kampen, 2017). This study utilized a descriptive research strategy, which is a scientific approach used to depict the features or behaviors of the population under inquiry. The utilization of the descriptive design in this study was motivated by its suitability for projecting findings to a broader population under investigation, based on a representative sample. This rationale justifies the adoption of this particular research methodology.

3.3 Population and Sampling

3.1.1 Population

This research utilized student data which contained information such as their skillset both technical and non-technical to predict suitable career path.

3.3.2 Sampling Size

A sample size refers to a selected portion of the population. According to Saunders et al. (2019), it is imperative for the sample size to be sufficiently large in order to instill the necessary confidence in both the data and the researcher. As the sample size increases, there is a greater emphasis on achieving precision and a higher level of confidence. The researcher must possess the ability to estimate the anticipated response rate, which refers to the proportion of cases within the sample that are expected to provide a response. This estimation is crucial for determining an appropriate sample size adjustment. This study used 80% of the data for training the model, 10% for testing and the remaining 10% for validating the model.

3.4 Data Collection

The study utilized secondary data from Kaggle to train the model. Kaggle is the most suitable repository for the data since the researcher cannot collect primary data from institutions due to the time and cost constraints involved undertaking such activities.

3.5 Research Quality and Reliability

In the context of qualitative research, the concept of reliability pertains to the degree of consistency observed in the responses provided by different coders when working with a given dataset. Enhancement can be achieved through the comprehensive documentation of field observations, their subsequent recording, and the transcription of digital data. The core principle behind the rigor and quality of qualitative research is in the methodical and self-reflective processes of data gathering, interpretation, and communication. The concept of validity in qualitative research differs from that in quantitative research due to the absence of organized response categories, which makes them non-mutually exclusive. The unstructured nature of qualitative inquiry is not conducive to the application of probability theory or statistical inference to a larger population (Guest et al., 2020).

Data saturation is a widely utilized qualitative research approach for ascertaining the appropriate sample sizes. Saturation is a term used in research to describe a point at which the researcher has collected enough data and has explored many qualities and topics, leaving little room for further development (Glaser & Strauss, 2017). In a general sense, saturation is employed in qualitative research as a criterion for terminating the process of data collecting and analysis. It was noted by Morse over two decades ago that satiety plays a crucial role in determining the level of severity. Saturation is a ubiquitous phenomenon observed in all rigorous qualitative investigations of high quality (Saunders et al., 2017).

3.6 System Development Methodology

Agile software development was the preferred software development methodology for this study. This is because Agile methodology recognizes that software development

is a collaborative process and needs to be treated as such (Hoda et al., 2018). The creators of the doctrine argued that these projects and software are unlike any other products or services purchased by companies. Firms do not need to go into truck factories to work with manufacturers on development. However, one cannot adopt the same approach when faced with software. All types of software need some customization, and the larger the organization, the more customization is necessary (Hoda et al., 2018). The nature of software development necessitates close cooperation between the developers and their customers.

The methodology rests on several key factors. The first is openness. The methodology requires opening effective communication channels from the beginning (Al-Saqqa et al., 2020). Secondly, the teams also have to work on common goals. This approach does not just assume that both the client and developer have the same plans. Instead, it insists on formalizing these long-term and short-term goals by a team that includes both parties. Finally, the method depends on continuous development and does not include ceremonies or ostentatious handing-over ceremonies (Al-Saqqa et al., 2020). By the time the project is over, the client should effectively control everything.

A vital advantage of this approach is that it looks at the problem holistically. Instead of considering the product or service as something that a company makes and then sells to another company, it creates a collaborative paradigm (Al-Saqqa et al., 2020). Another notable advantage is that the approach is highly flexible, and one can apply their twist to any problem. While Agile management has some complex and non-negotiable rules, it is not an all-encompassing doctrine (Al-Saqqa et al., 2020). Figure 3.1 shows the process in the agile methodology.



Figure 3.1 Agile Methodology

3.7 Utilization and Dissemination of Research Results

The research findings have substantial importance in the development of asset management tool that effectively addresses the practical requirements of organizations. The research findings will be made available on Strathmore University's digital repository, providing access to scholars who are interested in the asset management field of study. This study will provide further opportunities for academics and scholars to enhance career guidance. The research findings can be utilized institutions in Kenya such as Strathmore university to in guiding students in career selection.

3.8 Ethical Considerations / Issues

The research process shall be guided by the Code of Ethics established by the university. The researcher consistently possesses the code and will consult it as needed. For the purpose of ensuring privacy and security, the researcher intends to securely keep all data under a secured account.

Chapter 4: System Analysis and Design

4.1 Introduction

Systems analysis involves gathering requirements using both quantitative and qualitative measures. Techniques for understanding the current system and identifying new requirements (features or capabilities that must be included in the system to meet the needs of the users) include talking to current system users (if such a thing exists), observing the systems in action, and collecting and analyzing data (Conger & Mason, 2013). This chapter delves into the system architecture and the general design concerning the requirements raised by the stakeholders and the potential users of the career guidance tool. The analysis informs the requirements of the system (both functional and non-functional requirements). Unified Modelling Language (UML) diagrams have been used to explain the system's architecture and demonstrate the interaction between the target users and the system's main components.

4.2 Requirement Specifications

Requirement gathering is essential because it helps determine the new system's needs. Obtaining the functional requirements for the career guidance tool from existing literature was an essential step in ensuring the success of the tool. Reviewing existing literature, the main functional requirements necessary for developing a career guidance tool were identified. These included team structure and sources of funding, among others. This information led to the development a complete set of functional requirements for the tool, ensuring it contained all required features and capabilities.

4.2.1 Functional Requirements

The tool should allow the students to:

- i). Create an account.
- ii). Login.
- iii). View their profile.
- iv). Complete the career Evaluation Form.
- v). View the recommended career options with explanations.

4.2.2 Non-Functional Requirements

Non-Functional requirements define the tool's quality attributes. The Non-Functional requirements are as follows:

- i). The tool should be easy to use.
- ii). The tool should be secure against attacks.
- iii). The tool should be available whenever the user needs to use it.
- iv). The tool should be easy to maintain.
- v). The tool should be fault tolerant.
- vi). The tool should be compatible with various operating systems and browsers.

4.3 System Architecture

The system's architecture consists of three main components: the web interface, the machine learning model, and the database. The web interface provides a point of

interaction between the user and the model. The database stores the users' activity results for future reference whenever they need to. The system users must complete the career evaluation form to the interface as prompted. The features are then sent to the machine learning model via an API for prediction purposes. The results are stored in the database and displayed to the user, as illustrated in Figure 4.1.

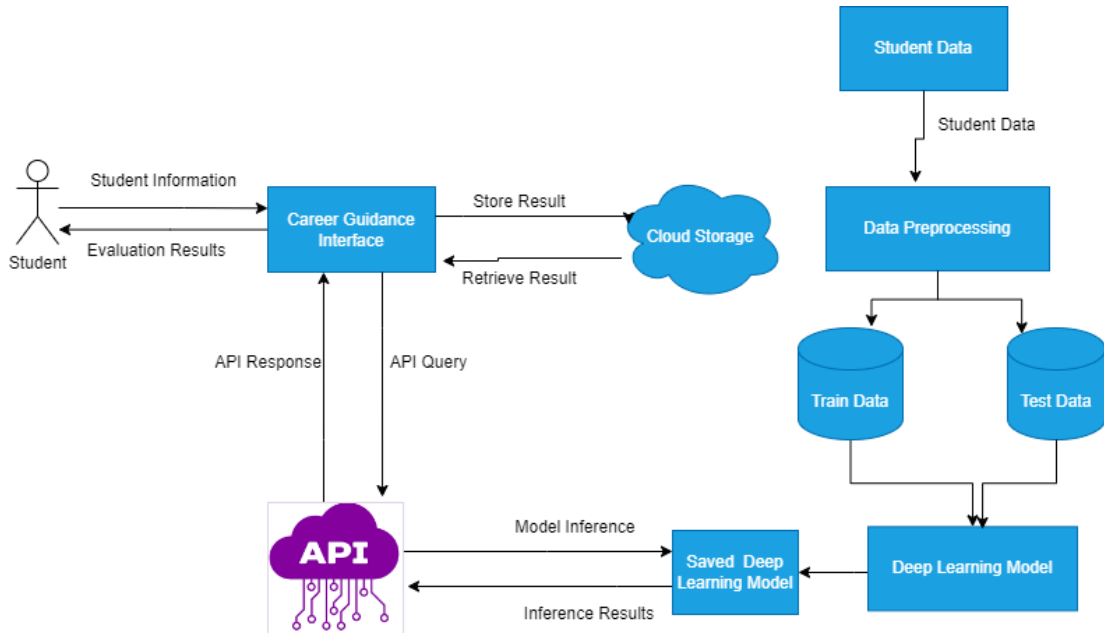


Figure 4.1: System Architecture

4.4 System Design

Systems design, a fundamental part of the development of any information system, is the procedure of outlining the system's structure, components, modules, interfaces, and data to fulfill its intended purposes (Rouhani & Lecic, 2018). Object-oriented analysis and design is the approach of choice for this study's design (OOAD). The Object-Oriented Analysis and Design (OOAD) technique is a software engineering approach to designing software systems by creating object-oriented models that abstract critical elements of the intended system and the subsequent use of the models to direct development. The model's principles and notation depict choices made during system design that will significantly affect the finished product (Carstoiu & Grigorescu, 1995).

4.4.1 Use Case Diagram

Use Case Diagrams to summarize how a user interacts with the system. Users can evaluate the system behaviour using Use Case Diagrams before writing code.

Similarly, they can be used as a blueprint throughout software development (Mule & Waykar, 2015). The Use Case diagram for the career guidance tool is shown in Figure 4.2.

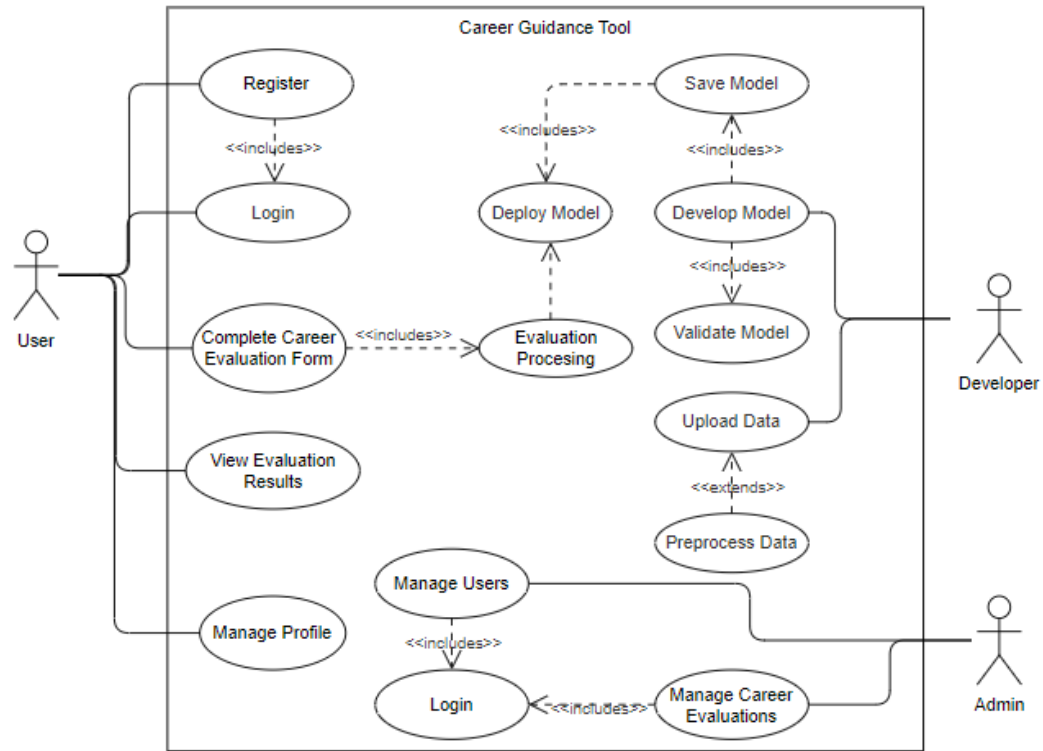


Figure 4.2: Use Case Diagram

4.4.1.1 Detailed Use Case Descriptions

Table 4.1 shows the detailed description of use cases earlier presented in Figure 4.2

Table 4.1: Description of use cases

Use Case	Pre-Conditions	Main Success Scenario	Post Conditions
Login	The user is registered on the system	-User enters login credentials -User logs in to the system and is directed to the home page.	None
Register	The user is connected to the internet.	-User enters their details on the registration form	None

		-User registration details are saved in the database.	
Complete Career Evaluation Form	The user is logged in	-User completes the career evaluation form. -Career Evaluation is initiated after the user presses evaluate button	Career evaluation result
View Profile	The user is logged in	-User clicks on the view profile button. -User profile is displayed	None
View Evaluation Results	Career Form is filled and submitted into the system	-User Clicks on View Evaluation Results -Evaluation results displayed to the user.	None
Evaluation Prediction History	User is authenticated	-Evaluation history displayed	None
View Evaluation Report	User is authenticated	-User clicks on the View Report button. - Evaluation Report is presented to the user.	None
Admin Login	None	-Admin enters login credentials into the login form. -Admin logins successfully.	None
Manage Users	Admin is logged in	-Admin selects a user. -Admin edits, updates, or removes users.	
Manage Career Evaluations	Admin is logged in	-Admin selects a career evaluation. -Admin views, edits, updates, and deletes career evaluation successfully.	

4.4.2 Class Diagram

Class diagrams are one of the most useful forms of UML diagrams because they clearly depict the structure of a system by modeling its classes, attributes, operations, and object relationships, as shown in Figure 4.3 (Bergstrom et al., 2022).

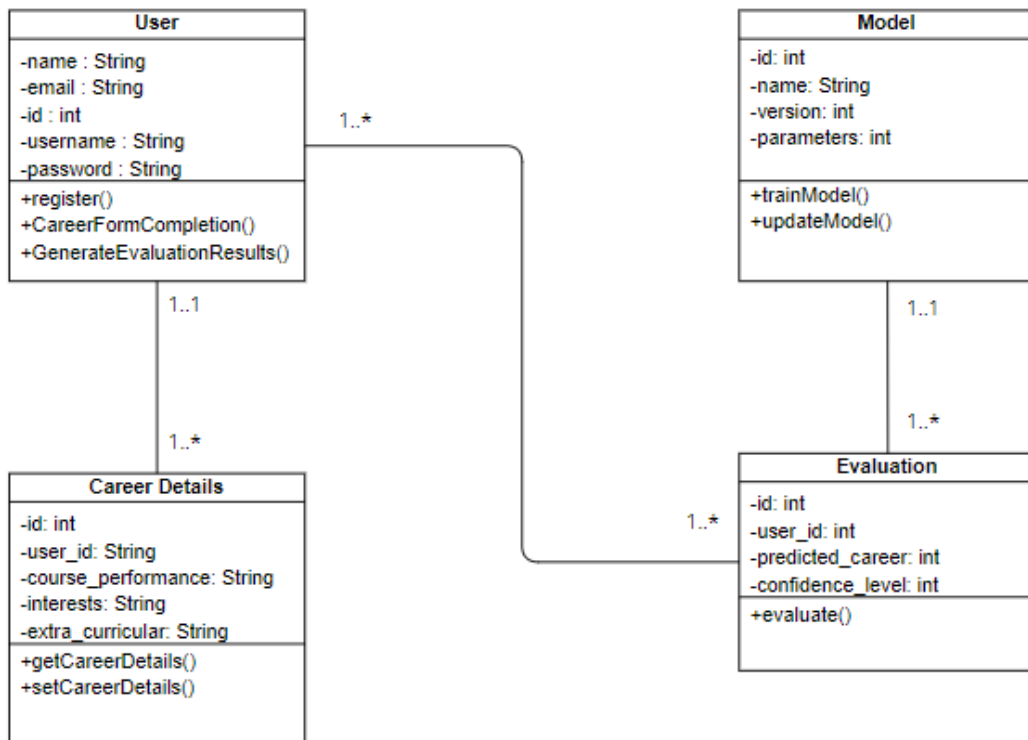


Figure 4.3 Class Diagram

4.4.3 Sequence Diagram

A sequence diagram is an interaction diagram showing the causes and effects of a chain of events (Alvin et al., 2019). The tool has five objects which interact with each other starting from the user, web interface, and Career Guidance model (comprising of the setup, training, and model). Figure 4.4 illustrates how the objects interact and the feedback mechanisms.

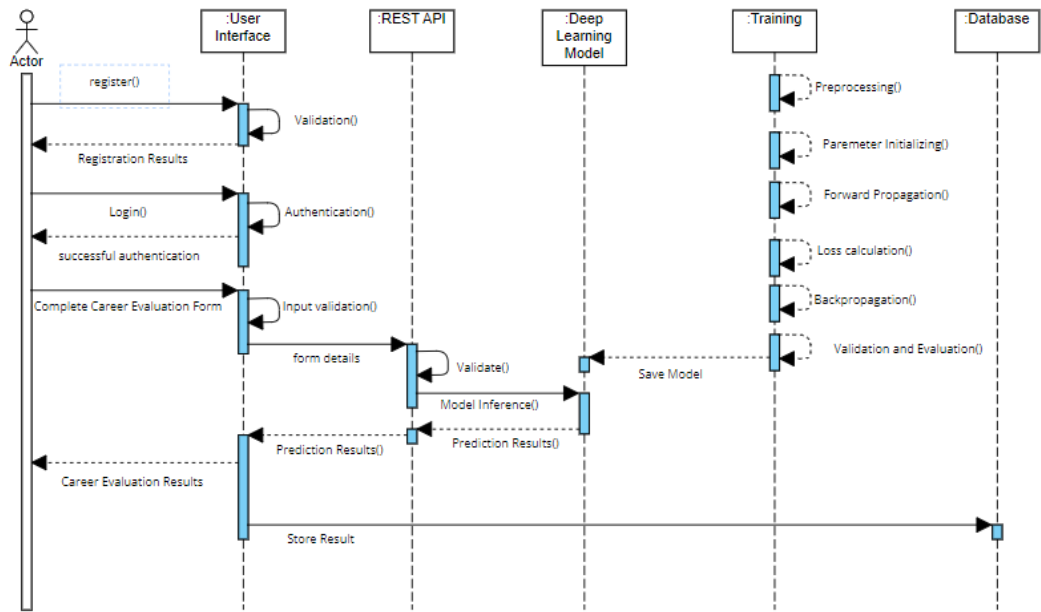


Figure 4.4: Sequence Diagram

4.4.4 Database Schema

A database schema is a blueprint for a database that outlines how different tables and models interact. The tool’s database structure is depicted in Figure 4.5.

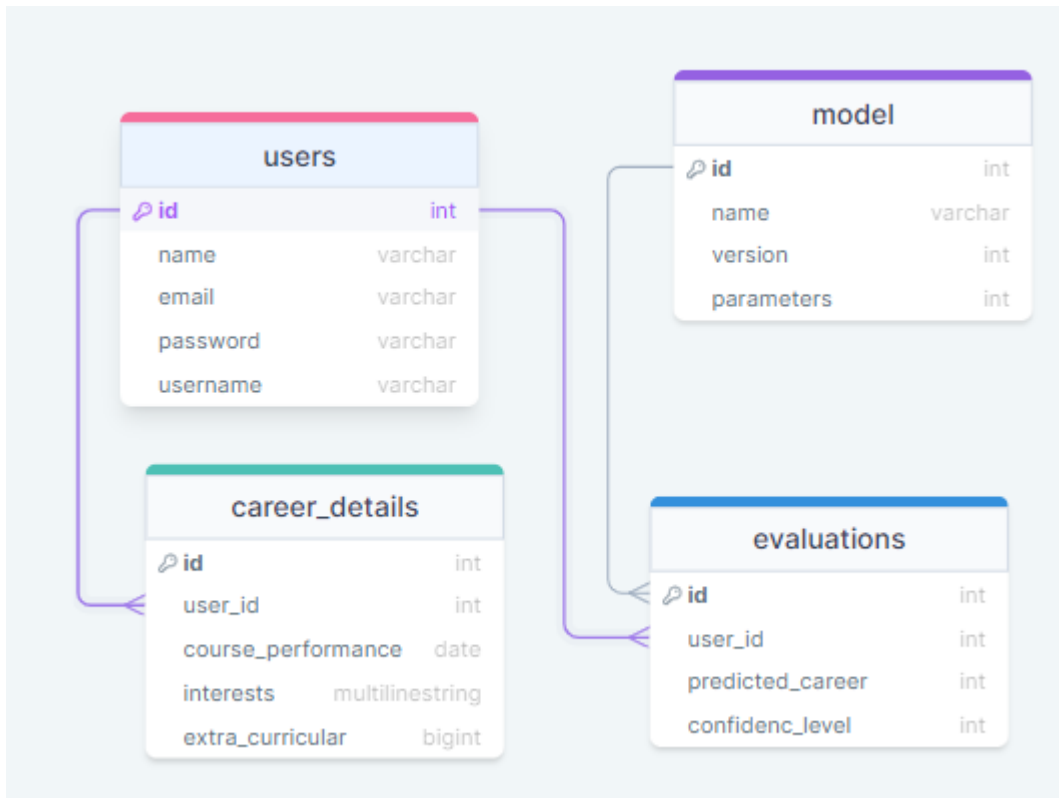


Figure 4.5: Database Schema

4.5 Wireframes

A system wireframe is a design used to inform the implementation of a system by rendering key intended features of the system based on user requirements or the developer's perspective. It is a representation of how the system functions. However, the system may provide more advanced features than the highlighted ones or features that fall short of the design (de Lange et al., 2020).

4.5.1 Home Page Wireframe

Figure 4.6 shows the landing page wireframe. This will be the first page the user will see once they visit the web application on their browser.

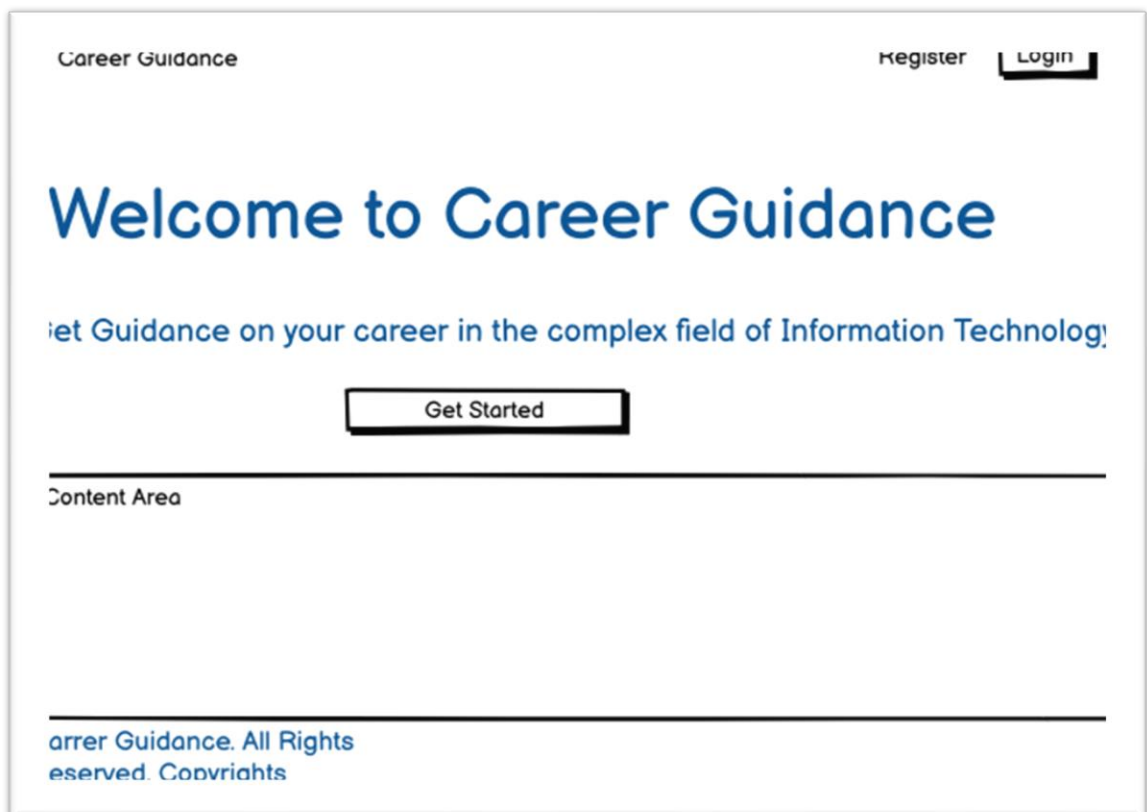
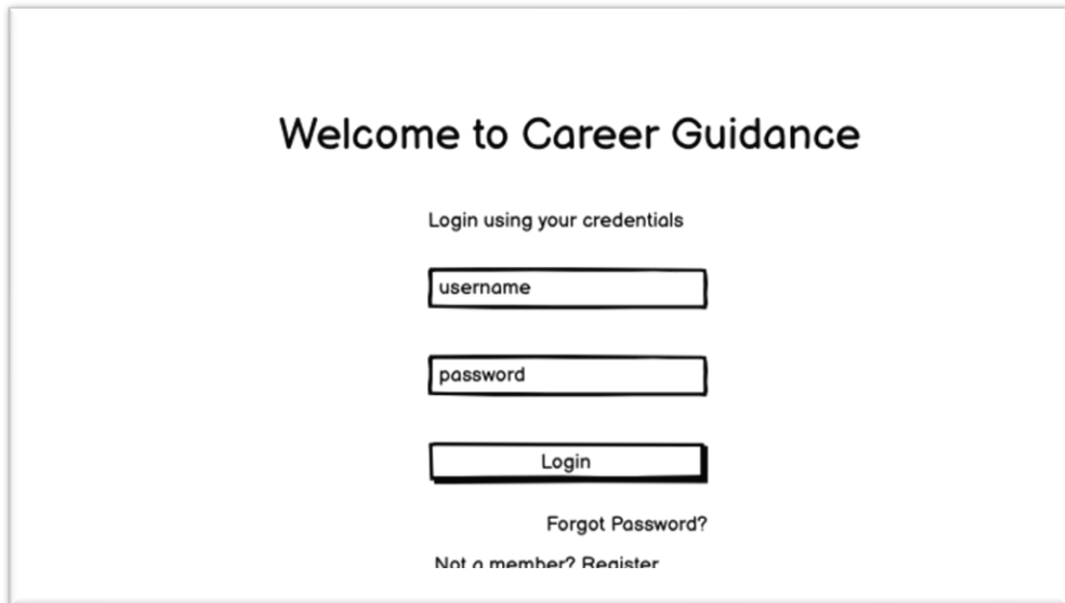


Figure 4.6: Landing Page Wireframe

4.5.2 Login Wireframe

The wireframe, as shown in Figure 4.7, contains a form that requires one to login with their correct credentials, namely Username and Password. After entering the credentials, there is an option to click the Login button.



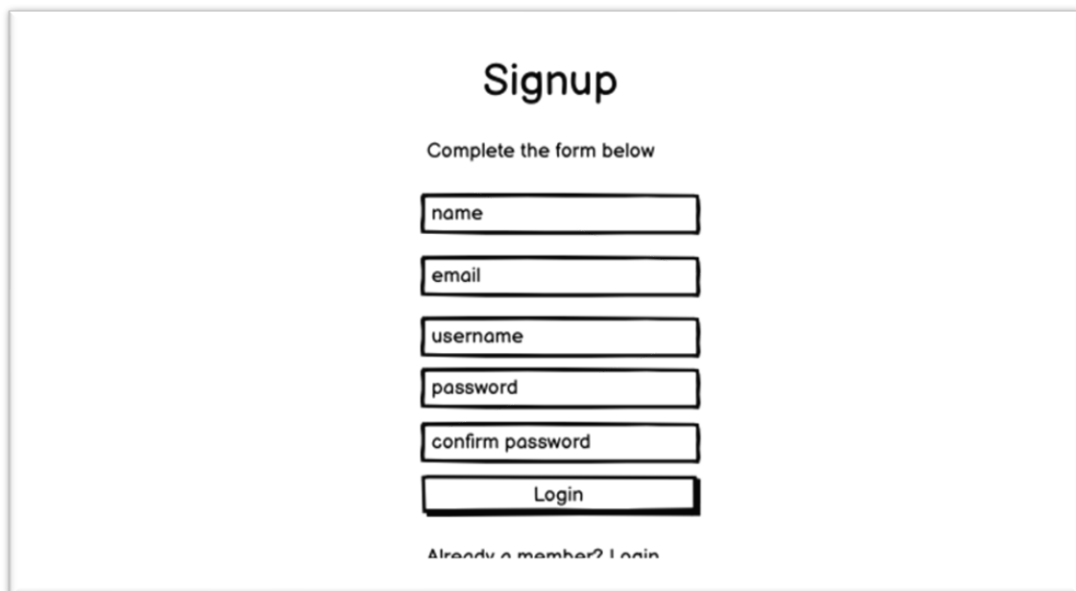
Wireframe for a login page titled "Welcome to Career Guidance". The page contains the following elements:

- Header: "Welcome to Career Guidance"
- Text: "Login using your credentials"
- Input field: "username"
- Input field: "password"
- Button: "Login"
- Text: "Forgot Password?"
- Text: "Not a member? Register"

Figure 4.7: Login Wireframe

4.5.3 Register Wireframe

The wireframe, as shown in Figure 4.8, contains a form that requires one to create an account by providing the requested information.



Wireframe for a register page titled "Signup". The page contains the following elements:

- Header: "Signup"
- Text: "Complete the form below"
- Input field: "name"
- Input field: "email"
- Input field: "username"
- Input field: "password"
- Input field: "confirm password"
- Button: "Login"
- Text: "Already a member? Login"

Figure 4.8: Register Wireframe

4.5.4 Career Evaluation Page Wireframe

The career evaluation form captures details of a student and allows the evaluation to be performed after. Figure 4.9 shows the career evaluation wireframe.

Career Guidance Welcome, Alice Logout

Career Evaluation Form

Complete the form below

Feature	Feature	Feature
Feature	Feature	Feature
Feature	Feature	Feature

Start Evaluation

Already a member? [Login](#)

Figure 4.9: Career Evaluation Form Wireframe

4.5.5 Results Page Wireframe

This wireframe illustrated in Figure 4.10, shows how the user feedback will be provided after requesting career evaluations from the system. Users will view the detailed description of the career options returned by the system.

Career Guidance Welcome, Alice Logout

Evaluation completed

Model Results

Congratulation for the assessment your career was predicted to be career x

The model was based on various x features

Figure 4.10: Evaluation Results Wireframe

4.5.6 History Wireframe

This wireframe shows the history of career evaluations on the system. A user can see the date and time of the career evaluation and the results. Figure 4.11 depicts the history wireframe.

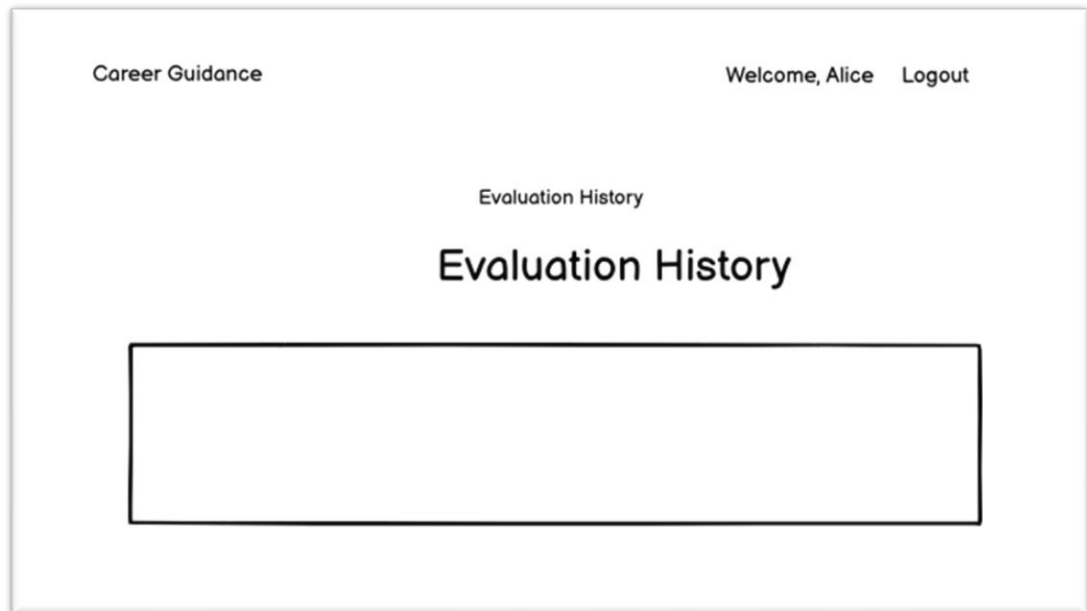


Figure 4.11: Career Evaluation History Wireframe

Chapter 5: System Implementation and Testing

5.1 Introduction

This section delved into the development, testing, and certification of the system. During the implementation phase, thorough analysis of the system's modules, development process, and functionality was conducted. The testing and validation procedures encompassed functional and usability tests aimed at assessing whether the system fulfilled its intended objectives.

5.2 Model Components

5.2.1 Multi-Layer Perceptron (MLP)

MLP a supplement of feed forward neural network was trained and tested on the career dataset from Kaggle. MLP (Multi-Layer Perceptron) comprises three layers: the input layer, output layer, and hidden layer, as depicted in Figure 5.1. The input layer receives the input signal for processing, while the output layer performs tasks such as prediction and classification. Situated between the input and output layers, the hidden layers serve as the primary computational engine of the MLP. In a feedforward manner, data flows from the input to the output layer. The neurons within the MLP are trained using the backpropagation learning algorithm. MLPs are adept at approximating any continuous function and can address problems that are not linearly separable.

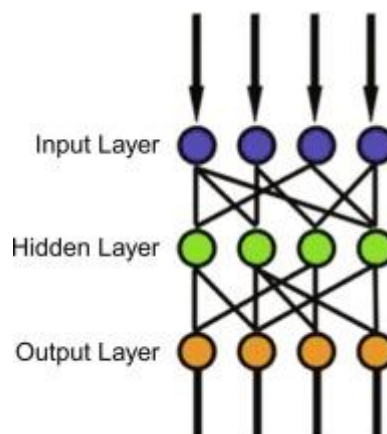


Figure 5.1: Schematic Representation of MLP

5.3 Web Application Components

The application has three primary interfaces: a homepage, a career evaluation interface, and a results interface. Using bootstrap, the front-end application's user experience was designed to be current and aesthetically pleasing. The model's application program interface was interacted with using a Laravel. Detailed descriptions of the components are as follows:

5.3.1 Home Page

The homepage provides a summary of the application and its function. It is intended to entice and urge users to utilize the system. The homepage contains a description and a search button for users to search through the career guidance tool. The home page has a button that takes users to the interface for evaluating their careers.

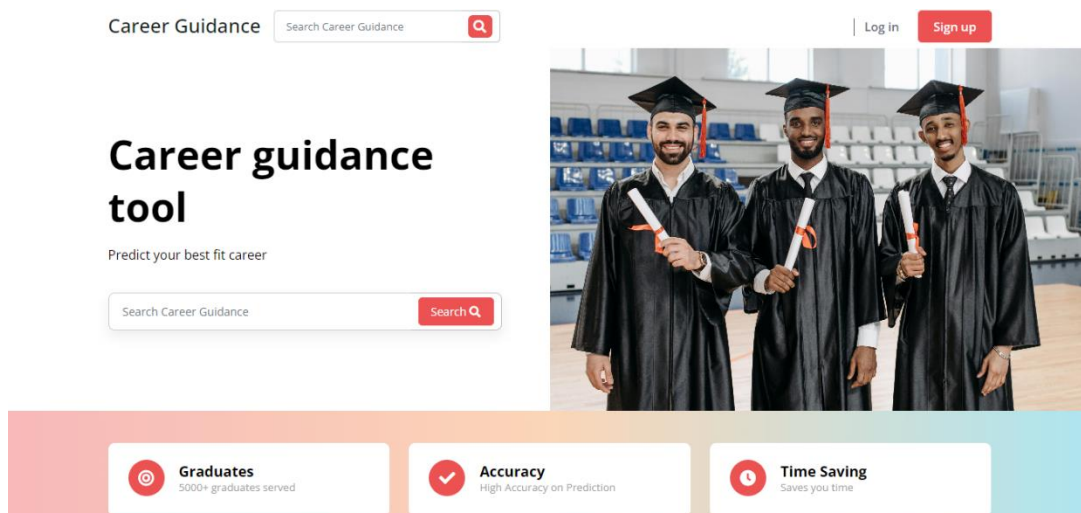


Figure 5.2: Home Page

5.3.2 Career Evaluation Interface

This interface is the main feature of the application. It allows users to evaluate their careers by completing the evaluation form. The form comprises fields prompting users to furnish information such as Academic percentage in Operating Systems, Algorithms, Programming Concepts, Software Engineering, Computer Networks, Electronics Subjects, Computer Architecture, Mathematics, Communication skills, Hours working per day, Interested Type of Books, Salary Range Expected among others. Figures 5.3 shows the career evaluation form.

Figure 5.3: Career Evaluation Form

5.3.3 Prediction Results

Once the user has completed the career evaluation form, the prediction result interface displays the predicted career. Figure 5.4 below prediction results.

Figure 5.4: Predicted Career

5.4 System Implementation

Agile software development methodology was utilized in this study and during the tool's development phase. Its capacity for continuous iteration facilitated the modification of various system versions to satisfy the study's objectives. The career

guidance tool prototype is a powerful solution that employs cutting-edge machine learning and deep learning technologies to analyze and predict the student careers.

The tool is intended to assist graduates in determining their best career before embarking on looking for jobs. The tool accurately predicts a student best fit career by examining student technical abilities, soft skills, academic performance and other pertinent data. The Flask framework integrates the machine learning model into the user interface to create a smooth user experience. Users can submit pertinent information about skills, interests, academic performance, etc., and receive a comprehensive analysis of their suitable career. In addition, the tool presents recommendations for other possible careers. The career guidance tool is an innovative solution that utilizes cutting-edge technologies to deliver essential insights into the educational industry. Its user-friendly design, and sophisticated algorithms make it an indispensable tool for students both graduates and those still pursuing their studies who wish to get the best career paths.

5.4.1 Development Environment

The system was developed in the following hardware and software environment:

- i. Python
- ii. Windows Operating System
- iii. Visual Studio Code
- iv. Google Colab
- v. Laravel (PHP Framework)
- vi. Flask (API Development)
- vii. Bootstrap CSS

5.4.2 Career Data Collection

Kaggle website provided the career dataset. Kaggle is a renowned platform that serves as a hub for data science enthusiasts, professionals, and researchers worldwide. It provides a collaborative environment where individuals can access datasets, participate in competitions, and engage in machine learning projects. Kaggle hosts a wide array of datasets across various domains, allowing users to explore, analyze, and model data for research or practical applications.

5.4.3 Data Pre-processing

The data preprocessing stage for the career guidance model began with the collection of raw data from Kaggle. This raw data included various attributes related to individuals' educational backgrounds, skills, interests, and career paths. The first step was to assess the quality of the data by checking for missing values, outliers, and inconsistencies. Any missing values were either imputed using appropriate techniques or the corresponding data entries were removed from the dataset.

Once the data was cleaned, the next step involved feature engineering, where new existing features were transformed to better represent the underlying patterns in the data. This process included encoding categorical variables, scaling numerical features, and possibly creating new composite features based on domain knowledge. After feature engineering, the data was split into training and testing sets to evaluate the performance of the model. The training set was used to train the career guidance model, while the testing set was used to assess its performance on unseen data. Additionally, techniques such as cross-validation were employed to ensure the model's generalizability and robustness. This rigorous preprocessing ensured that the model was trained on high-quality data, leading to more accurate and reliable career recommendations for individuals. Figure 5.5 shows the loading of the dataset.

```
from google.colab import drive
drive.mount('/content/drive')

path = "/content/drive/MyDrive/CareerGuidance"

df=pd.read_csv(path + "/career_guidance.csv")
data.head()
```

Figure 0.5: Load CSV file.

Figure 5. 6 shows the transformation applied through one hot encoding

```
X = OneHotEncoder().fit_transform(x)
Y = y.copy(deep=True)
```

Figure 5.6 One hot encoding

5.4.4 Training Model

The training process for the model began with the instantiation of the MLP classifier using the provided code snippet. The classifier was configured with specific parameters, including the activation function ('tanh'), the architecture of the hidden layers (consisting of three layers with 50 neurons each), and the solver ('adam'). Once initialized, the classifier was trained on the preprocessed training data (X_train) and corresponding labels (y_train) using the 'fit' method.

During the training phase, the model iteratively adjusted its internal parameters to minimize the error between the predicted outputs and the actual labels in the training data. This optimization process continued until a predefined convergence criterion was met, typically involving reaching a stable state or a specified maximum number of iterations. After the training was completed, the model's performance was evaluated using both the training and testing datasets. The accuracy score was calculated for both sets by comparing the model's predictions to the actual labels. The training accuracy score, which measures the model's performance on the training data itself, was found to be 99.93%, indicating a high level of accuracy in predicting the labels of the training samples. Similarly, the testing accuracy score, which assesses the model's performance on unseen data (X_test), was also determined to be 78.26%. Figure 5.7 shows the model training.

```
clf = MLPClassifier(activation='tanh', hidden_layer_sizes = (50,50,50), solver = 'adam') #, random_state=1)
clf.fit(X_train,y_train)
print("Training Accuracy Score: ",accuracy_score(clf.predict(X_train),y_train))
print("Testing Accuracy Score: ",accuracy_score(clf.predict(X_test),y_test))

Training Accuracy Score: 0.9992973185676541
Testing Accuracy Score: 0.7826869027543564
```

Figure 5.7: Model Training

5.4.5 Creating Model API

The API defines a function *predict_career* which takes in input features related to a student, such as technical skills, academic performance, etc., and returns a prediction of career based on the input features. The input features are pre-processed using the *MinMaxScaler* and *StandardScaler* methods from the scikit-learn library to ensure consistency in the range and distribution of the input data. The prediction is made using

a pre-trained machine learning model loaded using the *CustomUnpickler* module. The API endpoint `/predict` accepts POST requests with input data as a JSON payload. The payload contains input features such as technical skills, academic performance, etc. The endpoint uses the *predict_career* function to predict career of the student and returns the result as a JSON response, as shown in Figure 5.8.

```
def predict_career():
    if career_model:
        try:
            json_ = request.json
            print(json_)
            query = pd.get_dummies(pd.DataFrame(json_, index=[0]))
            # query.to_frame('count')
            query = query.reindex(columns=career_cols, fill_value=0)
            predict = list(career_model.predict(query))
            return jsonify({'prediction': str(predict)})
        except:
            return jsonify({'trace': traceback.format_exc()})

    else:
        return ('There is an error in Model ')

if __name__ == '__main__':
    try:
        port = int(sys.argv[1])
    except:
        port = 12345

app.run(debug=True)
```

Figure 5.8: API

5.4.6 Career Guidance Tool

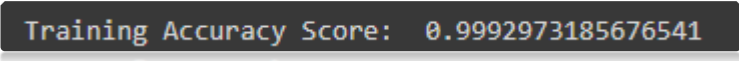
The software can only be used with an internet connection and a web browser. Data from Kaggle was used to train the model to predict career of a student using MLP.

5.5 System Testing

The web application was created utilizing the Agile software development methodology, which permits continuous iterations throughout various phases. Performance and problem issues were examined with continuous testing. The developed prototype was subjected to functional testing to determine if it met the specified functional requirements.

5.5.1 Test on model accuracy

Several metrics, including precision, recall, and F1 score, were used to evaluate the model's precision and performance. These measures would help to determine the model's capacity to correctly predict student careers. The model was tested using 20% of the dataset, achieving a maximum accuracy of 0.9993. This is represented in Figure 5.9.



Training Accuracy Score: 0.9992973185676541

Figure 0.9: Model Evaluation

Chapter 6: Discussions

6.1 Review of Research Objectives

Throughout the duration of this research thesis, four distinct objectives were identified and diligently pursued, resulting in their successful attainment. These objectives were carefully crafted to encompass various facets of career guidance research, aiming to shed light on crucial aspects of this field.

6.2 To Investigate Factors Influencing Career Selection Among Students.

The first objective of the study sought to explore the factors influencing the decision-making process of students when selecting their career paths. This involved a comprehensive investigation into the diverse determinants that shape individuals' vocational choices. The research revealed that students' career trajectories are significantly influenced by factors such as interests, self-capacity, and available career opportunities. Notably, personal interests emerged as a primary determinant shaping students' professional path, with individuals often aligning their career choices with their intrinsic passions. This notion was confirmed by the findings of Alexander et al. (2011), who highlighted the substantial impact of students' interest levels in specific subjects, such as information technology, on their career decisions.

Moreover, the study underscored the pivotal role of self-efficacy in guiding career choices and aspirations. Students' belief in their own capabilities was closely linked to their career preferences and their ability to envision viable professional trajectories. Taylor and Popma (1990) revealed that general self-efficacy levels were predictive of students' proficiency in exploring various career options, while Bergeron and Romano (1994) demonstrated the influence of self-efficacy beliefs on undergraduate students' selection of majors and career paths. Furthermore, the research elucidated that students exhibit a preference for fields offering diverse employment prospects. The allure of attaining a reputable social status, securing a stable income, and effecting positive societal change emerged as compelling factors guiding individuals' occupational choices. Ahmed et al. (2017) highlighted the significance of these factors in shaping individuals' career aspirations and decisions.

6.3 To Analyze The Methods Currently Used In Career Guidance.

The second objective aimed to scrutinize the prevailing methodologies employed in career guidance. The research uncovered that machine learning and expert systems emerged as the predominant techniques in this domain. Within machine learning, a diverse array of methodologies was explored, encompassing supervised, semi-supervised, unsupervised, and reinforcement learning. These four methodologies constituted the foundational pillars of various machine learning algorithms. Expert systems, on the other hand, manifested in two primary categories: rule-based and case-based expert systems. Rule-based systems relied on a knowledge base comprising straightforward IF-THEN rules, which served as the basis for deducing logical conclusions from existing knowledge. By leveraging these rules as inference mechanisms, novel insights could be derived from established principles.

Case-based systems, in contrast, offered an innovative approach grounded in the accumulation of experiential knowledge. This methodology operated on the premise that individuals navigate diverse challenges by drawing upon past experiences. Central to this approach was the concept of "the case," which served as the fundamental unit of knowledge. By distilling patterns from individual cases, humans derived overarching principles to address present issues and adapt to evolving circumstances.

6.4 To Review The Existing Algorithms and Models Used In Career Guidance.

The third objective of the study involved conducting a comprehensive review of the algorithms and models presently employed in the field of career guidance. This undertaking entailed delving deeply into the computational methodologies and predictive models utilized to enable well-informed career choices. Among the algorithms scrutinized were Random Forests, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-means algorithms. These algorithms have been extensively investigated by researchers for their efficacy in career prediction. Throughout the literature review process, the strengths and limitations of each algorithm were meticulously examined and discussed in detail. This critical analysis shed light on the suitability of these algorithms for addressing the complexities inherent in career guidance and prediction tasks.

6.4 To Develop a Career Guidance Tool Using Deep Learning.

The fourth objective of the study focused on the creation of a career guidance tool utilizing deep learning techniques. This goal was successfully accomplished through the development of the tool, which involved the implementation of a deep learning model. Notably, the deep learning model achieved an exceptional accuracy rate of 99.93%, outperforming all other models scrutinized during the literature review process. Furthermore, the consumer tool was developed and seamlessly integrated with the model through an Application Programming Interface (API), ensuring efficient functionality and user accessibility.

6.5 To Test and Validate The Developed Tool.

The fifth objective involved the thorough testing and evaluation of the developed tool. Rigorous testing and validation procedures were conducted to gauge the functionality and dependability of the tool. This pivotal phase aimed to validate the accuracy and effectiveness of the tool in real-world settings, ensuring its practical applicability and reliability. Remarkably, the model attained an impressive accuracy rate of 99.93%, affirming its robust performance and utility.

Chapter 7: Conclusion and Recommendation

7.1 Conclusions

Career guidance plays a vital role in individuals' lifelong career development journey, as highlighted by Keshf and Khanum (2021). Their research outlines five distinct stages individuals navigate through in their career development. The initial stage begins during schooling years, where individuals start contemplating their future paths. Subsequently, around the age of 15, the exploration phase commences, characterized by the exploration of various career options and the fine-tuning of preferences. This stage involves crystallization, specification, and implementation of career choices. As individuals reach approximately 24 years of age, they transition into the establishment stage, where they strive to carve out their space and establish themselves in their chosen careers. Following this, the maintenance stage typically occurs around 40 years of age, with the decline stage beginning around 60 years of age. This research comprehensively investigated various aspects of career guidance, aiming to enhance the effectiveness and accessibility of career decision-making processes among students. The findings of the research underscored the paramount importance of factors such as interest, self-efficacy, and career opportunities in shaping students' career trajectories. Notably, personal interests emerged as pivotal determinants in career decisions, alongside self-efficacy beliefs that significantly influenced students' career aspirations and confidence in pursuing desired professions.

In examining current methodologies employed in career guidance, machine learning and expert systems emerged as prominent techniques. The exploration of various machine learning approaches, including supervised, semi-supervised, unsupervised, and reinforcement learning, highlighted the diverse computational techniques utilized to facilitate informed career decisions. Additionally, expert systems, particularly rule-based and case-based systems, were identified as instrumental in harnessing accumulated experience to address career-related challenges. Furthermore, a comprehensive review of algorithms and models utilized in career guidance revealed the widespread adoption of techniques such as Random Forests, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-means algorithms. The detailed examination of these methodologies provided valuable insights into their

respective advantages and limitations, informing the subsequent development of the career guidance tool.

The culmination of this research thesis resulted in the successful development of a sophisticated career guidance tool leveraging deep learning methodologies. Leveraging cutting-edge deep learning techniques, the developed tool exhibited remarkable accuracy, surpassing existing models reviewed during the literature review phase. Integrated with an intuitive user interface and accessible via an API, the tool represents a significant advancement in facilitating informed career decision-making processes among students. Through rigorous testing and validation processes, the developed tool demonstrated exceptional accuracy and reliability, affirming its practical utility and effectiveness in real-world scenarios. With an impressive accuracy rate of 99.93%, the tool holds immense promise in empowering students with the insights and guidance needed to navigate their career paths successfully.

This research contributes valuable insights and tools to the field of career guidance, offering a comprehensive framework for enhancing career decision-making processes and facilitating students' journey towards fulfilling and rewarding career opportunities. This study underscores the transformative potential of data-driven approaches in empowering individuals to make informed and strategic career choices by leveraging cutting-edge technologies and methodologies.

7.2 Recommendations

This research led to the development of an innovative career guidance tool to assist students in career selection decisions. The researcher recommends the following:

- i). Recognizing the diversity among individuals in terms of backgrounds, interests, and career aspirations, the development of career guidance tools that cater to the specific needs of diverse populations is imperative. This includes considering factors such as gender, ethnicity, socioeconomic status, and geographic location to ensure that guidance interventions are inclusive and equitable.

- ii). Conducting longitudinal studies to assess the long-term impact of career guidance interventions on individuals' career trajectories and overall well-being is critical. By tracking participants over an extended period, researchers can evaluate the sustained effectiveness of guidance tools, identify potential challenges or barriers to success, and inform continuous improvement efforts.
- iii). Collaborating with industry partners, employers, and professional organizations can provide valuable insights into current workforce needs, industry trends, and skill requirements. By involving stakeholders in the development and validation of career guidance tools, researchers can ensure that the tools align with real-world demands and effectively prepare individuals for successful career transitions.

7.3 Limitations of the Study

The study was confined to graduates and undergraduate students, thus restricting its generalizability to other student demographics such as postgraduates and secondary school students. Additionally, the research relied on secondary data for model training, necessitating cautious interpretation and application of the research findings.

7.3 Future work

While the current research successfully achieved its objectives several avenues for future research present themselves.

- i). First, future research could expand the scope of factors influencing career selection beyond those explored in this study. Exploring additional variables such as socio-economic background, cultural influences, and personality traits could provide a more comprehensive understanding of career decision-making processes.
- ii). Further refinement of methodologies in career guidance is warranted. Investigating novel techniques and integrating emerging technologies such as natural language processing (NLP) and reinforcement learning could enhance the efficacy and accuracy of career guidance tools.
- iii). To enhance the relevance and timeliness of career guidance, researchers should explore the incorporation of real-time data and trends from the labor market, education sector, and industry developments.

References

- Ahmed, K. A., Sharif, N., & Ahmad, N. (2017). Factors Influencing Students' Career Choices: Empirical Evidence from Business Students. *Journal of Southeast Asian Research*, 2017, 1–15. <https://doi.org/10.5171/2017.718849>
- Alexander, P. M., Holmner, M., Lotriet, H. H., Matthee, M. C., Pieterse, H. V., Naidoo, S., Twinomurinzi, H., & Jordaan, D. (2010). Factors Affecting Career Choice: Comparison Between Students from Computer and Other Disciplines. *Journal of Science Education and Technology*, 20(3), 300–315. <https://doi.org/10.1007/s10956-010-9254-3>
- Bailey, L. J., & Stadt, R. W. (1973). *Career Education: New Approaches to Human Development*. Parkwest Publications.
- Breiman, L. (1996a). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (1996b). Stacked regressions. *Machine Learning*, 24(1), 49–64. <https://doi.org/10.1007/bf00117832>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Daramola, O., Emebo, O., Afolabi, I., & Ayo, C. (2014). Implementation of an Intelligent Course Advisory Expert System. *International Journal of Advanced Research in Artificial Intelligence*, 3(5). <https://doi.org/10.14569/ijarai.2014.030502>
- Ezenkwu, C. P., Johnson, E. H., & Jerome, O. B. (2017, March 1). *Automated Career Guidance Expert System Using Case-Based Reasoning Technique*. Research Gate. <https://www.researchgate.net/profile/Chinedu-Ezenkwu->

2/publication/314286511_Automated_Career_Guidance_Expert_System_Using_Case-Based_Reasoning_Technique/links/58bff57f458515bc83906ebc/Automated-Career-Guidance-Expert-System-Using-Case-Based-Reasoning-Technique.pdf

- Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. *Systems Science & Control Engineering*, 2(1), 602–609. <https://doi.org/10.1080/21642583.2014.956265>
- Gavhane, P., Shinde, D., Lomte, A., Nattuva, N., & Munjal, M. (2020). Career Path Prediction Using Machine Learning. *Www.ijrst.com*), 5. <https://www.ijrst.com/paper/7326.pdf>
- Ginzberg, E., Ginzburg, S. W., Axelrad, S., & Herma, J. L. (1951). *Occupational choice*. Columbia University Press.
- Guleria, P., & Sood, M. (2022). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, 1–36. <https://doi.org/10.1007/s10639-022-11221-2>
- Gunwant, S., Pande, J., & Bisht, R. K. (2022). *Gunwant*. JI4d.org. <https://jl4d.org/index.php/ejl4d/article/download/648/825?inline=1>
- Haraty, R. A., Dimishkieh, M., & Masud, M. (2015). An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data. *International Journal of Distributed Sensor Networks*, 11(6), 615740. <https://doi.org/10.1155/2015/615740>

- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. <https://doi.org/10.1109/34.709601>
- Howell, F. M., Frese, W., & Sollie, C. R. (1977). Ginzberg's theory of occupational choice: A reanalysis of increasing realism. *Journal of Vocational Behavior*, 11(3), 332–346. [https://doi.org/10.1016/0001-8791\(77\)90029-x](https://doi.org/10.1016/0001-8791(77)90029-x)
- Kazeem, A., Ismaila, B., Moses, K. B., & Ibam, E. O. (2017, September). *Development of a Web-based Intelligent Career Guidance System for Pre-Tertiary Science Students in Nigeria*. Research Gate. https://www.researchgate.net/publication/319911593_Development_of_a_Web-based_Intelligent_Career_Guidance_System_for_Pre-Tertiary_Science_Students_in_Nigeria#:~:text=This%20research%20therefor e%20developed%20a,various%20fields%20of%20study%20in
- Kerka, S. (1998). *Clearinghouse on Adult, Career, and Vocational Education*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=965846c0e25e346782d6a28775f96c0048f01c77>
- Keshf, Z., & Khanum, S. (2021). Career Guidance and Counseling Needs in a Developing Country's Context: A Qualitative Study. *SAGE Open*, 11(3), 215824402110401. <https://doi.org/10.1177/21582440211040119>
- López, O. A. M., Montesinos López, A., & Crossa, J. (2022). Fundamentals of Artificial Neural Networks and Deep Learning. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, 379–425. https://doi.org/10.1007/978-3-030-89010-0_10
- Mirzanevedshahzad, S., Takdees, F., & Ahmed, M. (2018). Factors Influencing on Students' Career Choices: Multivariate Analysis. *International Journal of*

Contemporary Applied Researches, 5(6).

<https://www.ijcar.net/assets/pdf/Vol5-No6-June2018/01.pdf>

Nguyen, T. L., Nguyen, H. T., Nguyen, N. H., Nguyen, D. L., Nguyen, T. T. D., &

Le, D. L. (2023). Factors affecting students' career choice in economics majors in the COVID-19 post-pandemic period: A case study of a private university in Vietnam. *Journal of Innovation & Knowledge*, 8(2), 100338.

<https://doi.org/10.1016/j.jik.2023.100338>

Nichols, J. A., Herbert Chan, H. W., & Baker, M. A. B. (2018). Machine learning:

applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, 11(1), 111–118. <https://doi.org/10.1007/s12551-018-0449-9>

Njogu, S. W. (2019). *ANALYSIS OF FACTORS INFLUENCING CAREER CHOICE AMONG PUBLIC SECONDARY SCHOOL STUDENTS IN MERU COUNTY, KENYA*. Kemu; kemu.

<http://repository.kemu.ac.ke/bitstream/handle/123456789/802/Sarah%20Wambeti%20Njogu.pdf?sequence=1&isAllowed=y>

Parsons, F. (1909). *Choosing a Vocation*. Brousson Press.

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach:

Global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4. <https://doi.org/10.1016/j.dsm.2021.12.002>

Rai, A. (2019, April). *Career Counseling Classification using Advance Machine Learning*. Research Gate.

https://www.researchgate.net/publication/332751781_Career_Counseling_Classification_using_Advance_Machine_Learning

Reid, H. (2016). *Introduction to Career Counselling & Coaching*.

<https://doi.org/10.4135/9781473958142>

- S. Alvin Leung. (2008). The Big Five Career Theories. In *International Handbook of Career Guidance* (pp. 115–132). https://doi.org/10.1007/978-1-4020-6230-8_6
- Schutt, D. A. (2007). *A Strength-based Approach to Career Development Using Appreciative Inquiry*. National Career Development Assn.
- Sharapova, N., Zholdasbekova, S., Arzymbetova, S., & Zaimoglu, O. (2023, March). *Efficacy of school-based career guidance interventions: A review of recent research*. Research Gate.
https://www.researchgate.net/publication/369672495_Efficacy_of_school-based_career_guidance_interventions_A_review_of_recent_research
- Sharf, R. S. (2013). *Applying career development theory to counseling* (6th ed.). Brooks/Cole Cengage Learning.
- Shmilovici, A. (2005). Support Vector Machines. *Data Mining and Knowledge Discovery Handbook*, 257–276. https://doi.org/10.1007/0-387-25465-x_12
- Super, D. E., Starishevsky, R., Matlin, N., & Jean Pierre Jordaan. (1963). *Career development : self-concept theory - essays in vocational development*. College Entrance Examination Board.
- Taylor, K. M., & Popma, J. (1990). An examination of the relationships among career decision-making self-efficacy, career salience, locus of control, and vocational indecision. *Journal of Vocational Behavior*, 37(1), 17–31.
[https://doi.org/10.1016/0001-8791\(90\)90004-1](https://doi.org/10.1016/0001-8791(90)90004-1)
- Tian, Y., Shu, M., & Jia, Q. (2021). Artificial Neural Network. *Encyclopedia of Mathematical Geosciences*, 1–4. https://doi.org/10.1007/978-3-030-26050-7_44-1
- Vapnik, V. N. (2010). *The nature of statistical learning theory*. Springer, Cop.

- Vignesh, S., Shivani Priyanka, C., Shree Manju, H., & Mythili, K. (2021, March 1). *An Intelligent Career Guidance System using Machine Learning*. IEEE Xplore. <https://doi.org/10.1109/ICACCS51430.2021.9441978>
- Wang, Y., Yang, L., Wu, J., Song, Z., & Shi, L. (2022). Mining Campus Big Data: Prediction of Career Choice Using Interpretable Machine Learning Method. *Mathematics*, *10*(8), 1289. <https://doi.org/10.3390/math10081289>
- Wang, Z., Liang, G., & Chen, H. (2022). Tool for Predicting College Student Career Decisions: An Enhanced Support Vector Machine Framework. *Applied Sciences*, *12*(9), 4776. <https://doi.org/10.3390/app12094776>
- Zunker, V. G. (2006). *Career counseling : a holistic approach*. Thomson/Brooks-Cole.

Appendices

Appendix A: Similarity Report

Appendix B: Ethical Clearance Letter