ENHANCING EDUCATIONAL ALIGNMENT: DEVELOPING A COMPREHENSIVE ALGORITHM FOR COMPLEMENTING PERSONALIZED COURSE PLACEMENT IN KENYA.

JOY MACHUKA

AN ACTION RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE MASTERS OF SCIENCE DEGREE IN DATA SCIENCE AND LEADERSHIP

AMSTERDAM TECH

DEPARTMENT OF MASTER OF SCIENCE

(JUNE 2025)

This dissertation is available for library use on the understanding that is copyright material and that no quotation from the dissertation may be published without proper acknowledgement.

# Declaration

I Joy Machuka hereby declare that this research project is my original work and has not been presented for a degree in any other university I know of.

Signature ………JBM…………………… Date …………30/06/2025…………

Joy Machuka

This research project has been submitted for examination with my approval as the University Supervisor.

Signature ……………………………………. Date ………………………………..

Leonidas Souliotis

Signature ……………………………………. Date ………………………………..

Steve Marshall

# Dedication

I dedicate this project to my family; mum and dad for their unfailing all-round support, constant encouragement and prayers during the period of undertaking this master’s degree. With their careers in the education system, I hope this project is of impact and starts up some conversations within the system.

# Acknowledgement

This project would not have been possible without the Almighty God’s providence of time and resources. I thank my parents for pushing me to complete my project, my friends for pushing me to start the masters and they kept encouraging me to complete. The support of our lecturer Leo, Steve and Monique for their guidance. Monique and Steve really led us through the action learning research approach that was handy during the writing of this report. I acknowledge the use of ChatGPT (OpenAI, 2025) as a supportive tool during this research project. The AI tool assisted with language refinement and clarification of ideas, under my full direction and critical supervision.

**Table of Contents**

[Declaration ii](#_Toc202139066)

[Dedication iii](#_Toc202139067)

[Acknowledgement iv](#_Toc202139068)

[Project Summary viii](#_Toc202139069)

[Chapter 1: Introduction 1](#_Toc202139070)

[Chapter 2: Literature Review 3](#_Toc202139071)

[2.1 Overview 3](#_Toc202139072)

[2.2 Theoretical Framework 3](#_Toc202139073)

[2.3 Aims and Objectives 4](#_Toc202139074)

[2.4 Research Questions 5](#_Toc202139075)

[2.5 Conceptual Framework and Boundaries 5](#_Toc202139076)

[2.6 Visual Conceptual Framework 6](#_Toc202139077)

[Chapter 3: Methodology 7](#_Toc202139078)

[3.1 Overview 7](#_Toc202139079)

[3.2 Research Approach 7](#_Toc202139080)

[3.3 Research Design 7](#_Toc202139081)

[3.4 Population and Sample Size 8](#_Toc202139082)

[3.5 Course Placement Model Development 8](#_Toc202139083)

[3.5.1 Data Collection 8](#_Toc202139084)

[3.5.2 Action Learning Reflection on Data Collection 9](#_Toc202139085)

[3.5.3 Data Processing and Analysis 9](#_Toc202139086)

[3.6 System Development Methodology 10](#_Toc202139087)

[3.7 Ethical Considerations 13](#_Toc202139088)

[Chapter 4: Project Activities in Cycles 14](#_Toc202139089)

[4.1 Project Activities in Cycle 1 14](#_Toc202139090)

[4.1.1 Constructing: Understanding the Problem 14](#_Toc202139091)

[4.1.2 Planning: Strategy and System Design 14](#_Toc202139092)

[4.1.3 Taking Action: Prototyping and Data Testing 15](#_Toc202139093)

[4.1.4 Evaluating Action: Feedback, Challenges & Iteration 15](#_Toc202139094)

[4.2 Reflection on Cycle 1 15](#_Toc202139095)

[4.3 Project Activities in Cycle 2 15](#_Toc202139096)

[4.3.1 Scaling and Refining Strategies 15](#_Toc202139097)

[4.3.2 Iterating for Organizational Impact 16](#_Toc202139098)

[4.3.3 Implementing and Presenting Organizational Change 16](#_Toc202139099)

[4.4 Reflection on Cycle 2 16](#_Toc202139100)

[Chapter 5: Project Findings 17](#_Toc202139101)

[5.1 Overview 17](#_Toc202139102)

[5.2 Key Findings from Data Collection 17](#_Toc202139103)

[5.2.1 Difficulty Accessing Secondary Data 17](#_Toc202139104)

[5.2.2 Insights from Primary Data Collection 17](#_Toc202139105)

[5.2.3 Data Challenges Encountered 18](#_Toc202139106)

[5.3 Key Findings from Data Analysis 18](#_Toc202139107)

[5.3.1 Dataset Structure 18](#_Toc202139108)

[5.3.2 Skills, Interests, and Extracurricular Trends 18](#_Toc202139109)

[5.3.3 Academic Trends and Grade Progression 19](#_Toc202139110)

[5.3.4 Subject Performance 19](#_Toc202139111)

[5.3.5 Psychometric and Behavioral Attributes 19](#_Toc202139112)

[5.3.6 Cognitive and Physical Indicators 19](#_Toc202139113)

[5.3.7 Age Distribution 19](#_Toc202139114)

[5.3.8 Statistical Observations 19](#_Toc202139115)

[5.4 Key Findings from Data Modeling 19](#_Toc202139116)

[5.4.1 Performance of the Random Forest Model Evaluation 19](#_Toc202139117)

[5.4.2 Performance of the XGBoost Model 20](#_Toc202139118)

[5.4.3 Performance on the Secondary data Models 21](#_Toc202139119)

[Chapter 6: Conclusions and recommendations 23](#_Toc202139120)

[6.1 Addressing factors Influencing Career Choice 23](#_Toc202139121)

[6.2 Integrating Multiple Data Sources 23](#_Toc202139122)

[6.3 Balancing Academic Capabilities with Personal Aspirations 23](#_Toc202139123)

[6.4 Comparing my System to the Existing Placement Process 23](#_Toc202139124)

[6.5 Summary and Reflections 23](#_Toc202139125)

[6.6 Conclusions 24](#_Toc202139126)

[References 25](#_Toc202139127)

**Table of Figures**

[Figure 1: Visual Conceptual Framework 6](#_Toc202139015)

[Figure 2: Rapid Application Development Model 1(Kissflow, 2025) 11](#_Toc202139016)

[Figure 3: RAD Multi Module Model (PREP INSTA, 2024) 12](#_Toc202139017)

# Project Summary

Choosing a university course is one of the most crucial decisions one would make. This is because this shapes career trajectories and life’s fulfilment. Of course, one could change career trajectories midway. Sometimes the transition is smooth, sometimes it comes at a cost of wasted time and resources. However there’s never really a right course, but I’d like to think that there’s the best fit course that aligns with ourselves and is most likely to bring more positive results. Unfortunately, the current system, at least for Kenya is primarily based on national exam grades and a limited list of course preferences. This often results in mismatched placements. Students are most times forced to adapt to the given courses. Some adapt quiet well and take off successfully, but for some they struggle in vain and eventually are forced to change course.

To address this challenge, this project is focused on designing, developing, and testing an inclusive course placement algorithm to complement the selection process. The algorithm employs an all-round data driven approach that integrates diverse factors such as the student’s academic performance, hobbies, personal interests, socioeconomic factors and behavioural traits, to recommend courses that best align with each student's unique profile.

Ultimately, the aim of the project is to provide a solution by developing an algorithm that complements students, parents, guardians, and educational institutions in making informed course selection decisions. The project is to be implemented using the agile software development methodology to ensure iterative design, feedback incorporation, and continuous improvement. By minimizing mismatched placements, the algorithm seeks to optimize resource utilization, enhance student satisfaction, and contribute to the overall productivity and effectiveness of the education system.

# Chapter 1: Introduction

Who or what decides how our life pans out? Do we strategically plan our lives or we just go by the wind. I tend to believe that it’s a bit of both. We can’t always have life go our way. If that were the case, the world would be one happy place. In most cases, our life’s plan involves a career. For some, it's defined by talent; for others, entrepreneurship or education, leading us to our focus - careers.

This project stems from personal experiences that exposed flaws in Kenya’s education system. A friend pursued two bachelor’s degrees due to an initial misalignment. She started with law but lost interest midway and later pursued computer science, her true passion. She fit in perfectly in tech but this came after incurring losses; both financial and emotional costs. This was caused by the ineffective placements. Secondly, I helped my sister select courses. We were to only consider her grades and preferences, neglecting other critical factors. This really limited us and put us in a box. Last but not least, my experience with self-development courses further affirmed that we thrive when doing what we love. What better way to boost career productivity than pursuing a course aligned with your passion? This ultimately enhances national productivity.

This has created in me a desire to ensure that less or no student is affected by the limitations of the current system. The need for some sort of change in the system is undeniable. Throughout this study, I will use data science to complement this placement process. In my opinion increasing productivity is the leading motivating factor because it in turn increases the country’s productivity. These experiences inspired the idea of a comprehensive placement algorithm, one that factors in academic, social, and personal elements. That way, there is hope that such a system would reduce the number of mismatches, placing students on a right track.

The university placement system in Kenya, primarily based on the Kenya Certificate of Secondary Education (KCSE) exam results, has significant limitations in ensuring equitable access to higher education. The current placement is based on KCSE performance and availability of university program slots. However, this often overlooks important factors such as personal talents, cumulative academic performance over the school years, students’ skills and even possibly socio economic factors. Many students miss out on their preferred courses despite meeting the requirements due to limited university capacity. While interuniversity transfers are possible, the process is complex and favours self-sponsored students (Nation, 2021). This leads to academic dissatisfaction, disengagement, wasted resources and in some cases, failure to complete their education. Concerns over mismanagement of placement data have even been raised in parliament (Thiong’o J. - The Standard, 2024). Following the 2024 KCSE exams, panic arose due to placement uncertainties. This is because it is a common problem for many students to get courses they did not select or even miss placement. Further fuelling the anxiety is failure by the then education minister to communicate the selection process (Nyaundi L. - The Standard, 2025).

The general objective of this study: to design and develop an integrated algorithm that leverages comprehensive data, including cumulative academic records, personal interests, behavioural traits and possibly socioeconomic factors to recommend the most suitable courses for students therefore complementing the existing placement process.

The research questions include identifying key factors influencing career choices, which ultimately determine university course selection. Thereafter, the study explores how these factors can be integrated into a useful model. The goal is to develop an algorithm that balances students' aspirations, abilities, and personalities for better course recommendations to complement the existing model. Finally, I intend to compare this proposed system’s effectiveness with the current one.

Feedback from my lecturers raised questions, such as, “What is the future of good education?” and “Who decides what is good?” This led me to shift my perspective and drive this study as a complementary solution rather than a definitive one. Moreover, the study is on ensuring that the algorithm provides accurate, fair, and well-rounded recommendations. The developed integrated tool uses machine learning techniques and a data-driven approach, which enabled the system to analyse historical student performance, identify trends, and suggest courses that align with individual strengths and aspirations. The study is limited by the availability of comprehensive digital student data. This is because some institutions may not have complete digital records. Similarly, not every school in the country embraces digital records for every exam. Additionally, the algorithm might not adequately account for the unpredictably shifting career goals of students over time.

The stakeholders include primary stakeholders – the high school students and their parents or guardians. Students are my immediate audience.

The secondary stakeholders include educational institutions, universities, and government policymakers. All these can adopt the algorithm to enhance efficiency in course placements.

The tertiary stakeholders involve the corporate sector and educational technology platforms. Companies can use insights from the system to align education with workforce demands.

As the developer, I play a crucial role as a tertiary stakeholder in ensuring the algorithm remains fair, transparent, and effective. My work directly impacts students, institutions, and industries. My role involves engaging with various stakeholders to understand their needs.

At the start of this project, I am taking on a leadership role by driving the vision and development of the algorithm. So far I have had the task of justifying to my lectures and colleagues on the necessity of this study. I plan to actively interact with one high school that I am affiliated to- so as to gather insights on their challenges with course selection. I will also collaborate with educational institutions and policymakers to align the system with national education goals.

My core values so far have been fairness, transparency, and data-driven decision-making. I believe every student deserves a placement that aligns with their strengths and aspirations. Through this initiative, I hope to create a meaningful impact by reducing misplacements, improving career satisfaction, and enhancing productivity.

# Chapter 2: Literature Review

## 2.1 Overview

In subsequent sections of this chapter I delve more deeply into the topic we are investigating, our aims and objectives, research statement and research questions to accompany our specific objectives. Previous research papers on using advanced algorithms in the education sector are reviewed to identify potential areas of improvement. Lastly a conceptual framework of the proposed solution is outlined.

## 2.2 Theoretical Framework

Course placement is a stage majority of us have to go through after high school especially in Africa. Right after high school, we have to go to college and take a bachelors course or a diploma/certificate for those who did not qualify. The course we end up pursuing most often than not shapes our career for future days.

In the American education system: after completing high school (12th grade), students have two options for post-secondary education: vocational training (typically a year or two) or higher ed (typically a two-year associate’s degree or four-year bachelor’s degree in an academic program) (Shorelight Team, 2023) (India-IELTS, 2024). During the bachelor’s degree, students take core subjects in the first two years and have to select major(s) to focus on in the latter two years (Sciencespo-grenoble.fr., 2024).

Similarly, in British system, students enter further education for two years before university, focusing on specific subjects they plan to pursue at the university level (Collegepond, 2025). In contrast, in Kenya, university placement is determined by the government, using an algorithm that prioritizes the national exam scores (Nielsen, 2024). The allocation process is often criticized for lacking flexibility, as it doesn't account for students' individual preferences or talents, relying heavily on exam results to make placements.

This issue is not unique to Kenya. In South Africa, between 2008 and 2015, approximately 50% of students who met the admission requirements for a Science program failed to meet the academic standards to pass (Abed, Ajoodha, Jadhav, 2020). This highlights a broader issue within global educational systems: the need for more personalized and equitable placement processes.

There are a number of comparisons between these education systems. In the American and British systems, when students want to major, they make applications for their majors. Sometimes, in the American system, students may need to apply to different schools if their majors are not offered at their current college, or if their performance did not meet the entry requirements. So basically, applications are made directly to the universities/colleges.

In Kenya and most countries in Africa, university placement is done by the government. The placement is then done with an algorithm that typically prioritize a student's final high school exam results (like Kenya's KCSE) as the primary factor, assigning them to programs based on their achieved grades, considering available capacity at different universities against specific program choices made by the student. This system is often criticized for potential inequities based solely on exam scores and limited flexibility for individual needs. Nation (2009) states that the criteria is primarily based on the grade scored in the final national exam and the specific subject points scored. The candidate gets to check their subject performance and look at the list of universities and courses and make choices 1a, 1b, 1c, 2, 3 and 4 of courses they would love to take. These courses are picked also from specific universities. Thereafter depending on demand for the courses in various universities, the board comes down and confirming with the capacity of various universities the students gets a course and university allocated to their profile. Sometimes, the demand from the list of choices was too high and therefore, student ends up missing all courses from their choices. They are given a second chance to make choices. This time, most universities are fully occupied and therefore the options left are more of remains. Sometimes the worst happens and they miss from even their second choice; this leads to the candidate being thrown to courses not even of their choice. The allocation process has undergone several reforms, yet the core approach remains largely unchanged, with many students still facing frustration and disappointment due to the limitations of the system (Business Daily, Lynet, 2022).

In Kenya and many other African countries, the government-driven placement process often leaves little room for students to choose their own academic paths. While well-off students who can afford self-sponsored tuition may have more freedom to choose their courses, those dependent on government placements are subjected to a system that does not fully account for their unique skills, interests, or future career goals. There is an option to do interuniversity transfers after placement has been done but again the process has been made cumbersome and could only be easier for self-sponsored students (Nation, 2021).

The use of AI within the Kenyan education system remains relatively limited. The most notable application has been the utilization of automated decision-making in the process of junior secondary school allocation. However, this approach has been met with scrutiny and critique. Critics argue that the current student selection criteria for entry into secondary school is in need of a comprehensive overhaul (Diana, Arnold, Sumaiyah, 2020).

## 2.3 Aims and Objectives

This proposed system aims to be achieved by developing an algorithm to complement the selection process. The algorithm employs an integrated approach that leverages diverse data to recommend the most suitable courses for students. The general aim of the project is to design and develop an integrated algorithm. This algorithm will leverage comprehensive data, including cumulative academic records, personal interests, behavioural traits, and possibly socioeconomic factors. Its purpose is to recommend the most suitable courses for students. In doing so, it will complement the existing placement process.

More specifically, my objectives are to apply data science techniques, including machine learning, to develop an algorithm that enhances the course placement process. The goal is to bridge gaps in the current system and ensure that the algorithm accurately reflects the diverse factors that influence students' success.

My objective for us as all the stakeholders involved is to collaborate to ensure the algorithm’s design aligns with national educational goals and can be effectively integrated into the existing placement systems. My objective for others is to stir up a conversation on the possibilities of reforming the system with data science.

The research statement for this study is: "This research aims to investigate how a data science-based algorithm can improve the course placement process in Kenya by integrating academic, personal, and socio-economic data to provide more accurate and personalized recommendations."

## 2.4 Research Questions

1. What are the factors that are likely to influence the career choice to ensure effective university course recommendations for students?
2. How can different data sources such as academic records, personal interests and others be integrated into a data model?
3. How can the algorithm be developed to create a sort of balance between students’ personal aspirations, academic capabilities and personalities to make more appropriate university placements?
4. How does the effectiveness of the developed system compare to the existing system?

## 2.5 Conceptual Framework and Boundaries

The conceptual framework below links the literature review with the research problem and the research objectives. To realize the model, the following steps will be followed;

1. **Problem Immersion and Data Acquisition**  
   The initial phase involves immersion into the research context through the acquisition of relevant datasets. This stage emphasizes understanding education system and establishing early insights into the nature of the data available.
2. **Data Pre-processing as Reflective Exploration**  
   Data pre-processing will be an iterative and exploratory process. Activities such as variable identification, data cleaning, exploratory data analysis, and the treatment of missing values and outliers will be carried out in cycles. At each iteration, insights gained will inform subsequent refinements, acknowledging that data understanding evolves with each pass.
3. **Feature Selection through Experimentation and Feedback**  
   Prominent input features will be identified through experimentation with various combinations and transformations. Rather than a static procedure, feature selection will be guided by model performance, encouraging reflective cycles of inclusion, exclusion, and redefinition of variables.
4. **Model Training through Cycles of Learning**  
   Machine learning models will be trained on the cleaned and structured data, applying stratified K-fold cross-validation (with K=5) to foster robust generalization. A training–testing split of 80/20 will be maintained. This stage will emphasize iterative experimentation, with each training cycle providing insight into model behavior and informing subsequent changes.
5. **Parallel Model Development and Learning Cycles**  
   Multiple models will be developed concurrently using default hyperparameters, initiating a comparative learning phase. Each model's behavior and performance will be examined, not only for accuracy but also for insights into how they respond to the data structure and feature patterns.
6. **Evaluation, Tuning, and Relearning**  
   Evaluation metrics such as accuracy, F1-score, and top-K accuracy will serve as benchmarks for reflection. Underperforming models will be revisited through hyperparameter tuning and feature reengineering. These cycles of testing and refinement will continue until optimal configurations are reached.
7. **Recommendation Deployment and Interpretive Insights**  
   The most performant model will be selected for deployment in course recommendation.

## 2.6 Visual Conceptual Framework

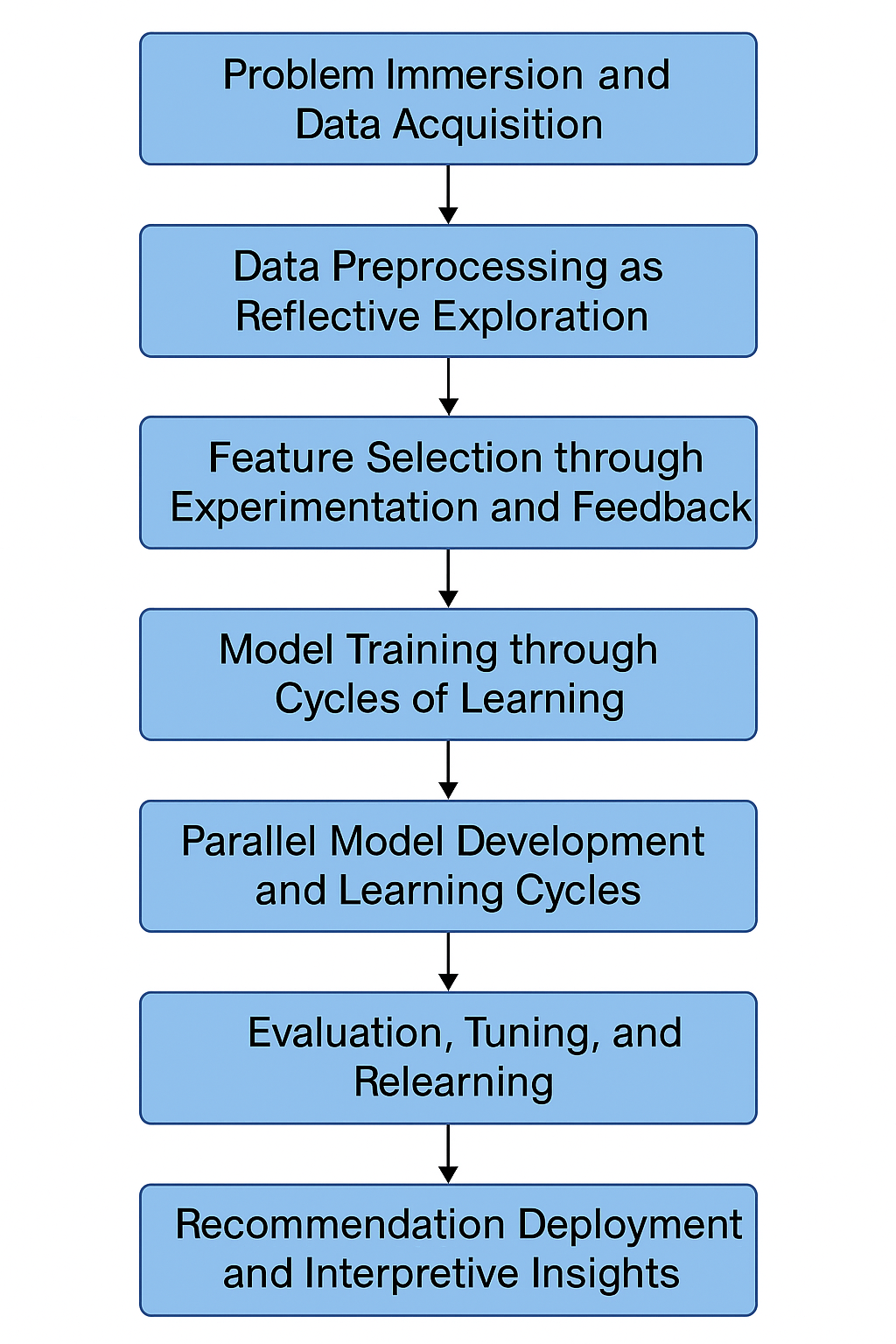


Figure 1: Visual Conceptual Framework

# Chapter 3: Methodology

## 3.1 Overview

This chapter highlights the research methodology I used and all the steps involved in the development of the model, detailing the approach, data collection techniques, data processing, and analysis. It also highlights ethical considerations, limitations, and justifications for the chosen methods.

## 3.2 Research Approach

I used a **mixed-method approach**, integrating both **quantitative** and **qualitative** methodologies. Quantitative research design approach involves the systematic collection and analysis of numerical data to identify patterns, test relationships, or make predictions (Fischer et al., 2014). The **qualitative aspect** involves gathering insights from students, educators, and placement officers through surveys and interviews to understand the real-life challenges of university placements.

Additionally, **action research** is applied as a framework, ensuring a practical, iterative approach where findings inform refinements in the algorithm. Action research aligns with this study’s objective of improving the existing course placement process through a **data-driven solution** (McNiff, J. 2013).

## 3.3 Research Design

The research was guided by the problem statement outlined in Chapter 1, the objectives and research questions highlighted in Chapter 1, and the related works reviewed in Chapter 2 of this study. Initial plan was to get data from five schools but data accessibility was an issue for most schools. As I approached the schools, they did not have recorded historical data of the students all across their school life. Additionally, they couldn’t reach students who already completed their studies. This was an obstacle but I was led to restrategize and use one school that I was affiliated to. Therefore, Eureka High School’s dataset will serve as the primary data source for this study. Firstly, exploratory data analysis was conducted to gain a deeper understanding of the dataset. This process helped identify trends, patterns, and anomalies in the data and informed the subsequent feature selection process. Following the EDA, relevant attributes were selected to serve as inputs for the machine learning models. These models were then subjected to a rigorous training and evaluation process, culminating in the selection of the best performing model.

This research gave birth to the ultimate output i.e. a model, which has the optimal performance, that takes the data of certain number of students as input. The model feeds the data in the embedded prediction/recommendation model and gives a prediction of the recommended course. With this research design which focused on the quantitative data analysis, it was able to acquire and analyze the data of the students to come up with the right recommendation of students course. The incorporation of machine learning methodology can help to enhance the accuracy of placement, minimize the number of mismatches and assist students in making knowledge-based careers.

## 3.4 Population and Sample Size

Since the start of my data problems I had to turn to one school. In this study I will use students of Eureka high school and their academic data to contribute to the formation of the course placement model. To understand more on the individual students outside the grades, I will also survey the students in an attempt to collect information regarding their preferences, personalities and backgrounds. My data was extremely unbalanced due to the small number of data. First, I applied stratified sampling in an attempt to contain the risk of data underrepresentation particularly to the classes that employed small amounts of data and provide a fair dose of representation in my split of train and test data (Wu, Y., 2022). This did not help much hence I went a step further in using sample weighting in handling imbalance classes to guide the model when training the model to assign more weight on the underrepresented classes (Singh, K., 2020).

This was not conclusively helpful at all despite all my efforts. Models with classes having a single data sample were not running due to it. I turned to search in already available datasets even after I did it the first time but did not find any. Fortunately, this was the case when I received it in a university in Asia during the hackathon. This information was slightly varied as it had a number of rows and a number of entries per course.

## 3.5 Course Placement Model Development

### 3.5.1 Data Collection

#### 3.5.1.2 Secondary Data Sources

Primary data will be collected through:

* **Surveys & Questionnaires:** Distributed among high school students, university students, and educators to gather insights on course selection experiences, factors influencing career choices, and challenges in the current placement system.
* **Interviews:** Conducted with career counsellors and university placement officers to understand the existing system’s shortcomings and potential improvements.
* **Psychometric Assessments:** Used to analyse students’ behavioural traits, personalities, and interests to establish correlations between personality types and career suitability.

#### 3.5.1.2 Secondary Data Sources

Secondary data is gathered from:

* **Academic Records:** Historical student placement data, KCSE scores, subject preferences, and university admission records.
* **Government Reports:** Data from the Ministry of Education on university enrolments, course demand, and placement statistics.
* **Socioeconomic Datasets:** Publicly available data on income levels, access to educational resources, and regional employment trends.
* **Previous Research Studies:** Literature on factors influencing career choices, university placement algorithms, and education system inefficiencies.

### 3.5.2 Action Learning Reflection on Data Collection

Although it was expected that the collection of substantial secondary data within bodies of the government and schools was to be carried out at this stage of the research, this part of the study turned out even much more difficult and even disappointing in some cases. I expected to avail comprehensive reports of the Ministry of Education on patterns of enrolments in various universities coupled with course areas of placements. Yet, the repeated requests on this information remained without responses or with their refusal based on bureaucratic constraints and secrecy. Also, historic academic records at majority of the schools that I contacted could not be found or had no data about their alumni, hence it was hard to get post-secondary outcomes. These obstacles, in addition to shattering the initial strategy of data collection, also showed how high were gaps in availabilities of centralized education data in Kenya. These failures made me adjust the technique; I became more dependent on the primary collection of data and turned to the real-time reflection of students, educators, members of placement offices. This fieldwork has underlined the values of agility, strength, and reflective practice in carrying out field-based educational research.

### 3.5.3 Data Processing and Analysis

#### 3.5.3.1 Pre-processing and Feature Engineering

Data pre-processing carries out the cleaning and transformation of the raw data into well structured datasets that can be subjected to data mining and analytics. It enhances the quality of machine learning models through the suitability of data (Garcia & Ferrera, 2015). It includes data quality measurement, data cleansing, data transformation and data reduction. To transform the data into a modeling state, data cleaning and transformation was run as a systematic process. Before carrying out the analysis of the data, missing values and data type inconsistency were explored using the Python libraries Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn.

A univariate, bivariate and multivariate analysis was performed to get an understanding of trend and relationships involved between variables in the dataset. Among the numerical variables were academic scores and psychometric scores, whereas categorical variables were skills, interests, and extra activities. One-hot encoding of categorical variables was done to rewrite the variables in machine readable context. Their numerical characteristics were then standardized with the help of the StandardScaler in such a way that they appeared on the same scale, which is the key to the work of such models as logistic regression and neural networks.

Exploratory Data Analysis (EDA) was conducted to identify patterns, relationships and outliers in the data set. The process of cleaning was guided by a summary statistic, distribution plots, and allocation maps to facilitate the process, whereas feature selection was informed by correlation heatmaps. It is important to note that such features as Bodily Activeness, Logical Reasoning, STEM Avg\_score, and KCSE Score survived the analysis of the predicting ability.

#### 3.5.3.2 Feature Selection and Transformation

#### A comprehensive feature set was constructed by concatenating the encoded categorical features with the normalized numerical scores. Multicollinearity checks and correlation analysis were used to refine this feature set. The final dataset balanced behavioral, academic, and psychometric components to increase the model’s ability to capture complex student profiles.

#### 3.5.3.3 Machine Learning Modelling and Evaluation

The study employs the following machine learning techniques:

* **Classification Models (e.g., Decision Trees, Random Forests, Neural Networks):** Used to predict the best-fit courses for students based on historical data and feature analysis.
* **Clustering (e.g., K-Means, Hierarchical Clustering):** Groups students with similar academic and personal profiles to determine course suitability.
* **Recommendation Systems:** Implements **collaborative filtering** and **content-based filtering** to suggest courses that match students’ profiles.

To predict recommended courses, various classification models were tested including Random Forest Classifier and XGBoost. Prior to training, the dataset was split using stratified sampling to ensure equal representation of all course classes in both training and test sets.

Given the moderate class imbalance, the model training phase integrated sample weights, computed using compute\_sample\_weight(class\_weight='balanced', y=y\_train). This technique improved the model’s sensitivity to underrepresented course categories without altering the original class distribution.

The models were evaluated using classification metrics such as accuracy, precision, recall, and F1-score. Top-3 accuracy was also computed for neural network models to assess the system’s performance in providing multiple relevant course suggestions.

*3.5.3.4 Feature Importance and Interpretability*

For tree-based models like Random Forest and XGBoost, feature importances were extracted to assess which factors had the most influence on the course recommendation. These insights not only validated the model logic but also guided further refinement in the feature engineering process. Features such as logical reasoning, STEM scores, and psychometric indicators ranked consistently high, confirming their critical role in accurate course placement.

## 3.6 System Development Methodology

At first, I was weighing the use of Agile approach methodology in system development because it is organized, has a cyclic process, and involves the stakeholder. As the work on the project further developed, it became apparent that the project time restraints and experimental aspect to developing the models needed a less rigid and faster-paced approach. I hence resorted to Rapid Application Development (RAD) approach. RAD made it easy to model quickly, and I could iterate on the models, improve and test them against both performance indicators and against practice (Kissflow, 2025).

In the process, models were deployed quickly, tested frequently and enhanced by using actual students data as well as by simulated inputs. This feedback cycle: that was the cornerstone of RAD as well as action research; this promoted a cycle of constant learning. Any of the versions of the models showed me some limitations or unbelievable performances, so I treated it as a lesson to more machine learning feature engineering, resolve imbalances, and test other types of architecture. Test user feedback (on the part of educators and peers) drove improvements not only in the accuracy of the model but also in the ease of use and interpretation of the end web-based system. RAD has been found particularly useful in matching the technical solution to the reality of teaching environments.

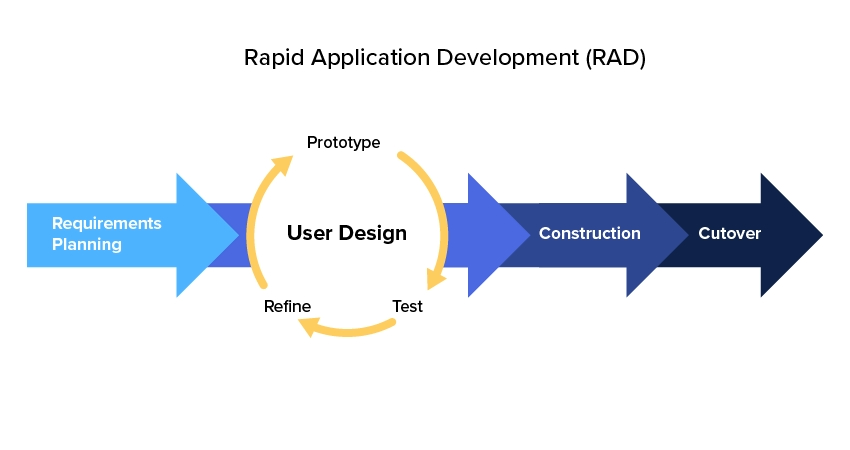


Figure 2: Rapid Application Development Model 1(Kissflow, 2025)

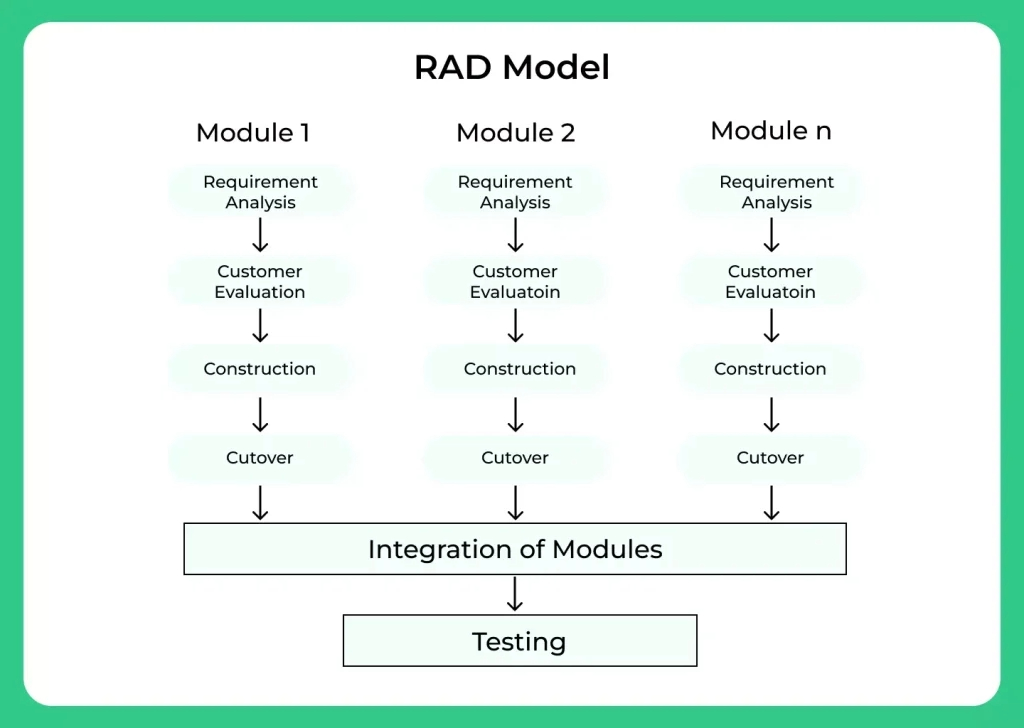


Figure 3: RAD Multi Module Model (PREP INSTA, 2024)

## 3.7 Ethical Considerations

Ethical responsibility was crucial in this project, especially due to the involvement of student data and education-related recommendations. The dataset used in this study was collected from Eureka High School with full consent from school administrations. Ethical principles of data collection were specifically aimed at education research and were accompanied with ensuring participants anonymity and confidentiality.

In case of psychometric/academic/behaviorally based data, when the information may have run against the interests of the informant, this was covered by the legal implication of the informed consent such as survey forms were issued where they could understand the purpose, extent and application of the information. Infact, no whefe was the CandidateID column used in my analysis. Additionally, algorithmic fairness was actively monitored throughout model development to avoid bias against students based on gender, background, school, or socioeconomic status. Techniques such as stratified sampling, sample weighting, and fairness-aware evaluation were integrated to uphold equity in predictions.

Academic integrity was achieved by all possible means through making references to all literature, tools, and media materials that were utilized during the development and documentation of the project.

# Chapter 4: Project Activities in Cycles

## 4.1 Project Activities in Cycle 1

### 4.1.1 Constructing: Understanding the Problem

This phase began with attending an Academic Writing workshop, that made me acquire the necessary skills to define, as well as frame the research problem. The workshop helped me understand the extent of action research better, and improved my skills of presenting the objectives of the dissertation in a practical and academic manner. Consequently, there was a marked improvement on my second submission of the introductory chapter.

At the same time, I provided literature review research to determine international best practices of placing and admitting courses and universities. I concentrated on psychometric data, historical academic records, and predictive modeling with the purpose of enhancing placement results. On the contrary, the current system in Kenya displayed a wide gap in terms of personalization, low input on career guidance and strict placement criteria. This analogy assisted me to develop a situational frame of reference about the local problem and guided the development of a solution based on both international knowledge and local facts.

### 4.1.2 Planning: Strategy and System Design

To lay a strong foundation for the system's design and methodology, I began by reviewing global models of course placement and career guidance systems. Many developed countries have successfully implemented AI-driven, data-centric guidance tools to support students in making informed academic and career decisions (Teachflow.AI, 2023). As an example, the United States, the United Kingdom and Canada use psychometric testing, student portfolios and predictive analytics to inform university placements. The tools do not only enhance the accuracy of placement but also the alignment with aspirations of students and requirements of the labor market.

Students selected using psychometric evaluation frameworks tend to exhibit higher levels of engagement, adaptability, and satisfaction in their academic paths (psico-smart.com, 2017). This observation supported the notion that the interest and personality traits of a student must be the core of any model of course placement recommendations.

Conversely, when examining the Kenyan system, I found that the majority of African countries meet some general issues in their systems, such as a shortage of access to individualized instruction, strict placement regulations, and poor incorporation of behavioral information. Compared to recommendations engines, Kenya has not yet caught up with smart course placement systems that are used to recommend courses and assist the students with their course selection process (Akala, 2021).

This is what I observed and made my planning of my model based on and I used it when deciding on the tools to use that could balance algorithmic logic with human-oriented input. I found this review justifiable to develop a hybrid system that also integrates classification models and recommendation logic; with characteristics that include academic history, interests and psychometric signals. I chose Figma to work with the wireframe design because it is familiar and convenient to iterate, and started drawing both technical stack diagram and methodological front-path to preprocess data, experiment with models and frontend implementation.

### 4.1.3 Taking Action: Prototyping and Data Testing

As part of the cycle, I submitted an assignment detailing the prototype's conceptual design, rationale for the selected methodologies, and anticipated challenges. I was familiar with figma hence opted for figma to create my wireframes for the model deployment. Feedback from this assignment from Leo provided insights into potential areas for refinement, particularly in integrating a recommendation model and comparing its performance to a classification model factors into the model.

I also encountered major data-related challenges. Many schools did not have historic records, and the Ministry of Education did not provide the data I had initially hoped to use. This pushed me to adapt my dataset strategy and rely more on synthetic data, surveys, and data augmentation techniques. Despite these setbacks, I moved forward by engineering features, building baseline models, and measuring preliminary accuracy metrics.

### 4.1.4 Evaluating Action: Feedback, Challenges & Iteration

After submitting my wireframes and model approach, I received feedback from my supervisor (Leo), which highlighted several areas for improvement, such as the need to justify model selection more clearly and to better integrate the recommendation logic into the front end. I used this feedback to refine the system’s flow and also began comparing classification models to recommendation models more critically.

Additionally, through this first cycle, I recognized the importance of having complete and representative datasets. The lack of certain socioeconomic variables and the challenges in accessing historical records forced me to rethink how I integrated user context into the models. I adjusted my preprocessing techniques, adopted new data cleaning practices, and began working on weighted class handling to address imbalances in the dataset.

### 4.2 Reflection on Cycle 1

As I move into Cycle 2, I’m more conscious of the value of co-creation. Building not just for students, but with educators, mentors, and system stakeholders. This mindset will shape my approach to interviews, testing, and iterative development moving forward.

## 4.3 Project Activities in Cycle 2

### 4.3.1 Scaling and Refining Strategies

In Cycle 2, the focus shifted from ideation to refining and scaling the model. By evaluating new combinations of features, architectures, and evaluation strategies, I enhanced the precision, equity, and usability of the course recommender using my understanding of the results of Cycle 1. One of the most important lessons was that I needed to have a more realistic view of recommendations to include top-k metrics, such as top-3 accuracy, which gave a better idea of a recommendation situation.

As I could not find a secondary set of data in the Ministry of Education and placement records, I switched to collect more primary set of data using surveys and psychometric scales that showed the strength of flexibility in action research. Lastly, I improved the wireframes by adding moving outputs and ranking of features to add more transparency and stakeholder trust, which increased the usability and adoption ability of the system.

### 4.3.2 Iterating for Organizational Impact

Iteration played an important role in this stage. I tried out more neural networks and XGBoost, were I touched up my feature engineering and introduced sample weighting to overcome class imbalance on a fair scale. The feedback provided by peers and supervisors suggested including the aspects of interests, aptitude, and behavior, which appealed to me; hence, I modified the pipeline to incorporate non-academic features and balanced the data back to cover those.

Organizationally speaking, I found that the value of the project lay beyond its technical success, as it led to discussing the need of changing career guidance and ways to empower students. I have also started creating documents such as ethical statements and guides to hand over the user. This step reaffirmed the idea that good solutions are developed by testifying, learning and adapting

### 4.3.3 Implementing and Presenting Organizational Change

This final phase of Cycle 2 involved the deployment of a working course recommendation system prototype, embedded within a user-friendly web interface. I tested the system using anonymized student data and simulated user sessions, observing how different inputs influenced recommendations. The output proved reliable, interpretable, and adaptive; an improvement from the initial static classification model.

I prepared presentations and summaries of the findings for educators and school placement officers, with emphasis on how the model could be piloted in guidance counselling workflows. Although full organizational adoption will take time, this cycle provided proof-of-concept and initiated important conversations on data-informed career guidance in secondary education.

### 4.4 Reflection on Cycle 2

As I reflect on both cycles, what stands out is how my role evolved; from a developer to a researcher, advisor, and learning facilitator. I have grown in my capacity to balance technical development with stakeholder engagement and ethical responsibility. The experience strengthened my leadership, problem-solving, and systems thinking skills, and deepened my understanding of how AI solutions can meaningfully impact educational practices when designed with empathy, inclusion, and adaptability in mind.

# Chapter 5: Project Findings

## 5.1 Overview

In this chapter; I highlight my key findings from the whole project; from data collection to the best performing model. I also highlight findings and feedback from the stakeholders. Through this chapter, I critically examine the effectiveness of the implemented model, the evidence of impact, and the new understanding generated from this inquiry.

## 5.2 Key Findings from Data Collection

The data collection phase of this project was both revealing and challenging, reflecting the complexities of working with educational data in real-world, resource-constrained settings. This section highlights the core insights, patterns, and limitations encountered during the process.

### 5.2.1 Difficulty Accessing Secondary Data

Initially, the project intended to leverage rich secondary datasets, particularly from the Ministry of Education, on student placements, KCSE results, and university admission patterns. However, these data sources proved inaccessible. Despite formal requests and follow-up communications, the Ministry did not grant access due to legal constraints and data protection concerns. This challenge significantly shifted the course of the project.

Most schools contacted lacked comprehensive historical placement records or a centralized alumni database. Many relied on physical files or partial recollections from staff, limiting the reliability and completeness of the data. This unavailability of structured secondary data forced the project to pivot toward primary data collection and redesign the model to function independently of official placement history.

Reflection: This challenge underlined the disconnect between data-driven research and current institutional data readiness in Kenya. It emphasized the need for schools to digitize and centralize academic and placement information.

### 5.2.2 Insights from Primary Data Collection

To compensate, I conducted targeted surveys, questionnaires, and psychometric assessments among students across Eureka High School. The four other schools did not have ready data thus only used one school. This generated a dataset that included:

|  |  |
| --- | --- |
| * CandidateID | * STEM school\_avg score |
| * Name | * Humanities school\_avg score |
| * Age | * Intrapersonal score |
| * Learning Disability | * Interpersonal score |
| * Physical Disability | * Grade 3 Score |
| * Skills | * Grade 6 Score |
| * Interests | * Grade 9 Score |
| * Extracurriculars | * KCSE\_Score |
| * Bodily activeness survey score | * Recommended\_Career |
| * Logical reasoning test score | * Recommended Course |

Insight: Psychometric traits provided significant explanatory power beyond grades alone—validating the mixed-method approach.

### 5.2.3 Data Challenges Encountered

Despite positive outcomes, data collection was not without its issues:

* Response rates varied across schools hence only using one school’s data.
* Some students provided inconsistent or incomplete responses, necessitating manual review and imputation techniques.
* Not all schools had updated performance data, so recent results were sometimes approximated or self-reported.
* To ensure data quality, a combination of data cleaning, normalization, and validation techniques was applied. Responses were anonymized to preserve privacy, and missing data was handled using imputation strategies.

Reflection: This process taught me how to adapt to unpredictable field realities. Rather than waiting for “perfect” data, I learned to work iteratively with what was available and make transparent decisions to ensure data integrity.

## 5.3 Key Findings from Data Analysis

Despite the limited scope of exploratory data analysis (EDA) due to the synthetic and non-original nature of the dataset, several key insights emerged from the data analysis process:

### 5.3.1 Dataset Structure

The dataset comprised 300 records and 19 features, covering a wide range of academic, behavioral, and extracurricular indicators.

### 5.3.2 Skills, Interests, and Extracurricular Trends

The most frequently occurring skills included:

* Research (88 mentions)
* Communication (79 mentions)
* Problem solving (59 mentions)
* Critical thinking (45 mentions)

Top interests were:

* Technology (81 mentions)
* Science (67 mentions)
* Engineering (44 mentions)
* Business (38 mentions)

The most common extracurricular activities involved:

* Debate club (58 participants)
* Science club (46)
* Drama club and Volunteering (31 each)

### 5.3.3 Academic Trends and Grade Progression

Academic performance showed a general decline from early to later grades:

* Grade 3 Average: 89.03
* Grade 6 Average: 72.20
* Grade 9 Average: 65.72
* KCSE Average: 67.40

This progression aligns with expectations as curriculum difficulty increases over time.

### 5.3.4 Subject Performance

STEM subjects had a slightly higher mean (72.47) than Humanities (71.16) and Languages (68.38), which was unexpected given general trends in Kenya.

All subject scores had moderate standard deviations and slightly negative kurtosis, indicating a distribution without heavy tails.

### 5.3.5 Psychometric and Behavioral Attributes

Scores were relatively high in behavioral measures:

* Intrapersonal Score: 76.14 (SD: 12.31)
* Interpersonal Score: 73.28 (SD: 13.11)

These scores suggest students may possess strong emotional and social competencies.

### 5.3.6 Cognitive and Physical Indicators

Logical reasoning scored the highest among test indicators with an average of 73.77.

Bodily activeness had an average of 67.49, with a moderate right-skewed distribution, indicating fewer students at the high end of physical activity.

### 5.3.7 Age Distribution

The average student age was 18.06 years, with minimal skewness, confirming a relatively uniform age range around Form 4 or high school graduation.

### 5.3.8 Statistical Observations

Most variables had near-normal distributions with mild skewness and negative kurtosis.

No extreme outliers or abnormalities were observed in numeric variables, making the dataset suitable for predictive modeling.

## 5.4 Key Findings from Data Modeling

Using the data as is; we started with a random forest model and the performance was bad.

### Performance of the Random Forest Model Evaluation

1. Severe Class Imbalance

* Most courses had only 1 sample in the test set.
* Several classes had 0 support, meaning they never appeared in the predictions or test set.
* Only a few courses had more than one instance (support=2), leading to extremely sparse class representation, which degrades model learning.

1. Low Overall Accuracy

* Accuracy = 38%, which is slightly better than random guessing for 60 samples across over 50+ classes.
* Macro average F1-score: 0.26 — indicates poor balanced performance across classes.
* Weighted average F1-score: 0.35 — slightly better but still weak.

1. High Precision and Recall for a Few Classes Only

Some classes had perfect scores (precision=1.0, recall=1.0), but each had only 1–2 support samples, making them unreliable or overfitted.

Example: -

* Bachelor of Arts (Design) → precision = 1.00, support = 1
* BSc in Geology → precision = 1.00, support = 1
* This suggests the model might be memorizing rather than generalizing.

1. Zero Scores for Most Classes

The majority of degree programs had:

* + Precision = 0.00
  + Recall = 0.00
  + F1-score = 0.00
* Indicates no correct predictions for those courses.

1. Inconsistent Model Performance

Classes like:

* BSc in Software Engineering and BSc in Computer Science had some predictive power but still low precision.
* Classes with higher support (support=2) had better recall in some cases, indicating potential if more training data were available.

#### Reflection

Due to the overfitting and poor performance indicators, I decided to group the courses. I realized a number of courses were labelled differently but were the same. I therefore did the cleaning because I tried an XGBoost model but because some classes only had 1 occurrence, it was impossible. After data cleaning of the course’s column, atleast there was now no course appearing only once.

### Performance of the XGBoost Model

1. **Balanced Data Distribution Improved Slightly**

* Unlike the earlier Random Forest model, I ensured no class had only one occurrence in the training set.
* This helped XGBoost at least identify a few classes correctly (some precision=1.00, recall=1.00), which indicates learning did occur for better-represented classes.

Overall, model performance is still low

#### Reflection

Due to the still low performing model and still experiencing class imbalance and insufficient data; I opted to go look for an already existing dataset from online sources. After a long hard frustrating search, I came across one dataset from github of data from a university. The dataset had 3535 rows which was better than our primary data. So again, I adapted to the new change.

### Performance on the Secondary data Models

The dataset had 35 unique courses and 3535 total rows. Although it was almost perfectly distributed with several duplicates. The analysis of the secondary data using three different models; Random Forest, XGBoost, and a TensorFlow Neural Network; demonstrated consistently strong performance across all models, indicating that the input features were highly predictive for course classification.

1. Random Forest Model:

The Random Forest classifier achieved outstanding results with an overall accuracy of 100% across 707 samples. Most courses, including Animation, Graphics and Multimedia, Bachelor of Architecture, Bachelor of Education, and various B.Sc. and B.Tech. programs, recorded perfect precision, recall, and F1-scores of 1.00, indicating no misclassifications. Only a single class (Bachelor of Business Studies) showed a slightly lower precision of 0.92, but still achieved a recall of 1.00, resulting in an excellent F1-score of 0.96. The confusion matrix confirmed that predictions were highly accurate and correctly mapped.

1. XGBoost Model:

The XGBoost model produced almost identical results to the Random Forest classifier. The overall accuracy remained at 100%, with nearly all courses attaining perfect precision, recall, and F1-scores. Similar to the Random Forest, only the Bachelor of Business Studies had a slight dip in precision to 0.92, but recall and F1-scores stayed high at 1.00 and 0.96, respectively. This consistency reinforced that XGBoost was equally effective in learning the patterns within the dataset.

1. TensorFlow Neural Network:

The Tensor Neural Network demonstrated strong learning and rapid convergence. Training logs showed that the model improved significantly from an initial accuracy of 25% to a final training accuracy of 99.5%, with a validation accuracy stabilizing at 99.65%. The final test evaluation yielded a remarkable test accuracy of 99.72%, matching the high performance observed in the tree-based models. The detailed classification report showed perfect precision, recall, and F1-scores for nearly all classes, with only the Bachelor of Business Studies again slightly lower at a precision of 0.92, consistent with the pattern seen in the other models.

#### 5.4.3.1 Reflection

Overall, the findings indicate that all three models; Random Forest, XGBoost, and the Neural Network; achieved near-perfect classification performance on the secondary data. The minor variation observed in the Bachelor of Business Studies class may suggest slight overlaps in features or similarities with other business-related programs that could benefit from further feature engineering or more granular data. These results demonstrate that the available secondary data was clean, well-structured, and sufficient for building robust classification models for course placement. The high accuracy confirms the reliability of the models in recommending or predicting courses with a high degree of confidence.

# Chapter 6: Conclusions and recommendations

## 6.1 Addressing factors Influencing Career Choice

The first research question asked: What are the factors likely to influence career choice to ensure effective university course recommendations for students?

These factors were obviously identified with the help of the rich secondary data. The columns had diverse academic subjects indicators (e.g., Physics, Mathematics, Chemistry), personal interests (Drawing, Dancing, Singing, Photography), co-curricular activities (Sports, Travelling, Yoga), language preferences (English, Hindi, French) and even a clear vocational domain (Engineering, Doctor, Pharmacist, Journalism) amongst others.

This wide range of features shows that grades cannot be used to decide about the placement of students.

## 6.2 Integrating Multiple Data Sources

The second research question asked: How can different data sources such as academic records, personal interests, and others be integrated into a data model?

The secondary dataset managed to merge several dimensions in one structured database, which makes it a perfect test example to work with in integration. They could establish deriving the complicated relationships and learning patterns by converting the specific personal interests, leisure activities, behavioural tendencies, discipline capabilities, and the perceived profession desired into matrix values.

## 6.3 Balancing Academic Capabilities with Personal Aspirations

The third research question asked: How can the algorithm create a balance between students’ personal aspirations, academic capabilities, and personalities to make more appropriate placements?

These balances were implicitly learned in the models since hard skills (subject scores) and soft indicators (hobbies, preferences and activities) were included. As an example, a student that has good grades in Mathematics and Physics and at the same time has an interest in Drawing and Designing could be having better fit in areas like Architecture or Industrial Design rather than just settling in the default area of Engineering.

## 6.4 Comparing my System to the Existing Placement Process

The fourth research question asked: How does the effectiveness of the developed system compare to the existing system?

Although a direct, live comparison with actual placement results was beyond this study’s scope, the model outputs strongly indicate that such a system could reduce mismatches, improve student satisfaction, and better align talents with future careers. This tool can be enhanced further to be used as a complimentary system to the existing system

## 6.5 Summary and Reflections

Based on the findings, this study makes the following key recommendations:

1. Adopt Data-Driven Placement Approaches:

Placement authorities such as KUCCPS and other relevant agencies should integrate data science methods into the placement process. Doing so will help bridge the gap between students’ unique profiles and institutional placements, leading to more satisfying career paths and better student outcomes.

1. Increase Data Capture:

Schools and education stakeholders should broaden the types of data collected from students to include not only grades but also hobbies, subject preferences, behavioural traits, and self-declared aspirations. Such data can be gathered through structured surveys or career counselling sessions during secondary school.

1. Test the System using Actual Data of Students:

A pilot test to simulate the real use of the student should be undertaken to ensure the validity of the algorithmic effectiveness in a placement practice. This pilot should also gather some feedback of students, educators and institutions in order to perfect the system further.

1. Involve the Stakeholders in the Policy Formulating:

The implementation will be synergistic because different actors will be involved, including educators, policymakers, parents, students and technical team(data scientists). Guidelines, data privacy protection, and training must be established to promote ethical, transparent and responsible uses of student data.

1. Continuous Improvement:

The placement algorithm should be treated as a dynamic system. As more data becomes available and labour market trends evolve, the model should be updated and retrained regularly to remain relevant and accurate.

## 6.6 Conclusions

To sum up, these studies offer compelling evidence that course placement involving the data science has the potential of changing the way that students are matched to courses in the university. When Kenya adopts a holistic, learner-centred data, it will be able to develop a more responsive, fair and effective system of placement; and eventually enable students to follow career paths in which they will be most successful as well as being able to contribute to the growth and development of the country.

# References

Abed, T., Ajoodha, R. & Jadhav, A. (2020). A prediction model to improve student placement at a South African higher education institution. pp. 1-6. [Online] Available at: [10.1109/SAUPEC/RobMech/PRASA48453.2020.9041147](https://ieeexplore.ieee.org/document/9041147) [Accessed 24 Jan. 2025 1317hrs]

Akala, B.M. (2021). Revisiting education reform in Kenya: A case of Competency Based Curriculum (CBC). Social Sciences & Humanities Open, 3(1), p.100107. <https://doi.org/10.1016/j.ssaho.2021.100107> .

Business Daily, Lynet (May 23, 2023) How new varsity placement system works. [Online] Available at: <https://www.businessdailyafrica.com/bd/economy/the-a-z-of-new-varsity-placement-system--4244712> [Accessed 24 Jan, 2025 1023hrs]

Collegepond. (2025). UK Education System: Understanding The Structure And Academic Standards. [Online] Available at: <https://collegepond.com/uk-education-system-understanding-the-structure-and-academic-standards/> [Accessed 1 Mar. 2025].

Diana, N., Arnold, O. and Sumaiya, O. (n.d.). The Role of Artificial Intelligence in African Education Systems Case Study: The Use of Automated Decision-Making in Form One Placements in Kenya IDEATE TECH POLICY AFRICA. [Online] Available at: <https://cipit.org/wp-content/uploads/2024/09/The-Role-of-Artificial-Intelligence-in-African-Education-Systems-The-Use-of-Automated-Decision-Making-in-Form-One.pdf> [Accessed 24 Jan. 2025 14:15].

Fischer, H.E., Boone, W.J. & Neumann, K. (2014) ‘Quantitative research designs and approaches’, in Lederman, N.G. & Abell, S.K. (eds) Handbook of Research on Science Education, Volume II. Routledge, pp. 18-37.

García, S., Luengo, J. & Herrera, F. (2015) Data preprocessing in data mining, vol. 72, pp. 59-139.

India, I.I(2024). Understanding the US Education System: A Comprehensive Guide. [Online] Available at: <https://ieltsidpindia.com/blog/the-us-education-system> [Accessed 01 March. 2025 1553hrs]

Kissflow (March, 2025). RAD (Rapid Application Development): Definition, Meaning and Benefits. [Online] Available at: <https://kissflow.com/application-development/rad/rapid-application-development-rad-meaning/> [Accessed 13 March. 2025 1531hrs].

McNiff. J. (2016) Writing up your action research project

Sciencespo-grenoble.fr. (2022). Understanding the American university system 1/2: The Bachelor’s degree | Université Grenoble Alpes. [online] Available at: <https://en.sciencespo-grenoble.fr/getting-bachelors-degree-us>. [Accessed 01 March. 2025 1559hrs]

Nation (March 14, 2009) — updated on July 03, 2020 How universities select candidates. [Online] Available at: <https://nation.africa/kenya/news/how-universities-select-candidates--584634> [Accessed 23 Jan, 2025 1742hrs]

Nation (2021). How is university placement done that many students miss their course choices? [Online] Available at: <https://nation.africa/kenya/news/how-is-university-placement-done-that-many-students-miss-their-course-choices--3443462> [Accessed on 24 Jan. 2025 1244 hrs]

Nielsen, H. (2024). An Overview of the Education System in Kenya. [Online] Available at: <https://www.fawco.org/global-issues/education/education-articles/4949-education-round-up-january>. [Accessed 01 March. 2025 1633hrs]

Nyaundi, L. - The Standard (2025). Anxious wait: KCSE graduates in limbo over university placement. [Online] Available at: <https://www.standardmedia.co.ke/health/index.php/national/article/2001509753/anxious-wait-kcse-graduates-in-limbo-over-university-placement/> [Accessed 27 Jan. 2025 1324hrs].

PREP INSTA. (2024). RAD Model in SDLC. [Online] Available at: <https://prepinsta.com/software-engineering/rad-model/> [Accessed 23 Mar. 2025 1603hrs].

Psico-smart.com (2017). How are psychometric tests influencing student selection in higher education? [Online] Available at: <https://psico-smart.com/en/blogs/blog-how-are-psychometric-tests-influencing-student-selection-in-higher-education-101102> [Accessed 23 Mar. 2025].

Shorelight Team (August, 7, 2023). American Education and Higher Ed System: The US Education System Explained. [Online] Available at: <https://shorelight.com/student-stories/the-us-higher-education-system-explained/> [Accessed 01 March. 2025 1516hrs].

Singh, K. (2020). How to Improve Class Imbalance using Class Weights in Machine Learning. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/> [Accessed 20 May. 2025 2214hrs].

Teachflow.AI. (2023). The Impact of AI on School Career Guidance and Counseling - Teachflow.AI. [Online] Available at: <https://teachflow.ai/the-impact-of-ai-on-school-career-guidance-and-counseling/> [Accessed 13 March. 2025 1716hrs].

Thiong’o, J. - The Standard (2024). Agency fails to explain varsity placement rule. [Online] Available at: <https://www.standardmedia.co.ke/national/article/2001497001/agency-fails-to-explain-varsity-placement-rule> [Accessed 27 Jan. 2025 1306hrs].

Wu, Y. (2022). 7 Techniques to Handle Imbalanced Data. [online] KDnuggets. Available at: <https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html> [Accessed 23 June. 2025 0821hrs].