**DESIGN AND IMPLEMENTATION OF AN ADVANCED CRM SYSTEM FOR CUSTOMER SEGMENTATION AND CONVERSION PRIORITIZATION USING DATA SCIENCE TECHNIQUES**

**MARTINA NATASHA ADUODI SCT 222-0185/2021**

**A RESEARCH PROJECT SUBMITTED TO THE DEPARTMENT OF INFORMATION TECHNOLOGY IN THE SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY, IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE, BUSINESS COMPUTING, JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY.**

**APRIL, 2025**

# DECLARATION

This research project is my original work and has not been presented for a degree in any other University.

.......................... ..........................

Signature Date

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

.......................... ..........................

Signature Date

# ABSTRACT

This project aims to develop a Customer Relationship Management (CRM) system that integrates machine learning algorithms and predictive analytics for effective customer segmentation and conversion prioritization. The first objective is to review recent CRM models and data science techniques for customer segmentation and conversion forecasting. Building on this, the project will design a CRM system that uses customer interaction history, demographic data, and behavioural patterns to segment customers and forecast conversion likelihood. The system will also implement predictive analytics to automate client prioritization based on conversion potential. Additionally, a user-friendly interface will be developed to provide real-time insights, allowing businesses to dynamically adjust sales and marketing strategies. This project addresses current CRM limitations and explores the practical application of data science to enhance customer relationship management, aiming to improve conversion rates and operational efficiency.

Table of Contents

[DECLARATION 2](#_Toc197185387)

[ABSTRACT 3](#_Toc197185388)

[CHAPTER 1 7](#_Toc197185389)

[1.1 BACKGROUND INFORMATION 7](#_Toc197185390)

[1.2 INTRODUCTION 7](#_Toc197185391)

[1.3 Statement of the Problem 8](#_Toc197185392)

[1.4 Proposed Solution 9](#_Toc197185393)

[1.5 Objectives 10](#_Toc197185394)

[1.7 Justification 10](#_Toc197185395)

[1.8 Proposed Research and System Methodology 11](#_Toc197185396)

[1.9 Scope 12](#_Toc197185397)

[CHAPTER 2 13](#_Toc197185398)

[2.1 Introduction 13](#_Toc197185399)

[2.2 Theoretical Review 13](#_Toc197185400)

[2.3 Case Study Review 14](#_Toc197185401)

[2.4 Integration and Architecture 14](#_Toc197185402)

[2.5 Summary 14](#_Toc197185403)

[2.6 Research Gaps 15](#_Toc197185404)

[CHAPTER 3 16](#_Toc197185405)

[SYSTEM ANALYSIS AND DESIGN 16](#_Toc197185406)

[3.1 Introduction 16](#_Toc197185407)

[3.2 System Development Methodology 16](#_Toc197185408)

[3.3 Feasibility Study 17](#_Toc197185409)

[3.3.1 Economic Feasibility 17](#_Toc197185410)

[3.3.2 Technical Feasibility 18](#_Toc197185411)

[3.3.3 Operational Feasibility 18](#_Toc197185412)

[3.3.4 Legal and Ethical Feasibility 18](#_Toc197185413)

[3.3.5 Environmental and Time Feasibility 19](#_Toc197185414)

[3.4 Requirements Elicitation 19](#_Toc197185415)

[3.4.1 Data Collection 19](#_Toc197185416)

[3.4.2 Sampling Techniques 19](#_Toc197185417)

[3.4.3 Relevance to the Problem and Objectives 20](#_Toc197185418)

[3.5 Data Analysis 20](#_Toc197185419)

[3.5.1 Data Cleaning and Organization 20](#_Toc197185420)

[3.5.2 Statistical Analysis 20](#_Toc197185421)

[3.5.3 Visualization of Findings 21](#_Toc197185422)

[3.5.4 Key Insights from the Data 22](#_Toc197185423)

[3.6 System Specification 24](#_Toc197185424)

[3.6.1 Functional Requirements 24](#_Toc197185425)

[3.6.2 Non-Functional Requirements 25](#_Toc197185426)

[3.7 Requirements Analysis and Modeling 25](#_Toc197185427)

[3.7.1 Requirements Analysis 26](#_Toc197185428)

[3.7.2 Requirements Structuring 26](#_Toc197185429)

[3.7.3 Modeling Tools 27](#_Toc197185430)

[3.8 Logical Design 29](#_Toc197185431)

[3.8.1 System Architecture 29](#_Toc197185432)

[3.8.2 Control Flow and Process Design 31](#_Toc197185433)

[3.8.3 Design for Non-Functional Requirements 34](#_Toc197185434)

[3.9 Physical Design 35](#_Toc197185435)

[3.9.1 Database Design 35](#_Toc197185436)

[3.9.2 User Interface Design 37](#_Toc197185437)

[CHAPTER FOUR: 43](#_Toc197185438)

[SYSTEM IMPLEMENTATION AND TESTING 43](#_Toc197185439)

[Introduction 43](#_Toc197185440)

[4.1 Technologies Used 43](#_Toc197185441)

[4.1.1 Actual System Implementation 44](#_Toc197185442)

[Login Interface 44](#_Toc197185443)

[● 45](#_Toc197185444)

[Dashboard 45](#_Toc197185445)

[● Key Information Displayed: 46](#_Toc197185446)

[Student Profile Management 46](#_Toc197185447)

[● Features: 46](#_Toc197185448)

[4.1.1 Conversion Prediction Model Integration 49](#_Toc197185449)

[● Implementation Details: 49](#_Toc197185450)

[● Code (Illustrative Prediction Call): 50](#_Toc197185451)

[1. Model Loading 50](#_Toc197185452)

[Algorithm Selection Rationale: Why Random Forest? 51](#_Toc197185453)

[4.1.2Additional Interfaces 52](#_Toc197185454)

[● Analytics and Reporting: 52](#_Toc197185455)

[● Task Management: 53](#_Toc197185456)

[4.2 Testing 53](#_Toc197185457)

[4.2 Types of Tests Performed 53](#_Toc197185458)

[Summary of Test Cases and Results 54](#_Toc197185459)

[4.3 Project Appraisal 55](#_Toc197185460)

[4.2.3 Limitations 55](#_Toc197185461)

[4.3.2 Strengths 55](#_Toc197185462)

[4.3.3 Conclusions 56](#_Toc197185463)

[4.3.3 Recommendations 56](#_Toc197185464)

[References 58](#_Toc197185465)

# CHAPTER 1

INTRODUCTION

## 1.1 BACKGROUND INFORMATION

Customer Relationship Management (CRM) has become an essential strategy for businesses aiming to improve customer interactions, streamline internal processes, and drive overall profitability. Traditionally, CRM systems have been used to manage functions such as lead generation, client tracking, and sales management. By centralizing customer data, these systems provide businesses with valuable insights, helping them better understand their clients and foster long-term relationships.

However, the evolving business environment has exposed some limitations in traditional CRM systems, particularly in prioritizing customer engagement based on the likelihood of conversion. With the increasing volume of leads generated through various digital channels, distinguishing high-potential clients from those less likely to convert has become a significant challenge. Consequently, businesses often spend substantial time and resources on low-probability leads, resulting in inefficiencies and lower profitability.

To address this issue, there is a growing interest in enhancing CRM systems with data science and analytics capabilities. By incorporating customer segmentation techniques, businesses can categorize clients into distinct groups based on behavioural patterns, demographic data, and interaction history. This enables a more focused and targeted approach to customer engagement, allowing sales and marketing teams to prioritize high-conversion prospects and optimize their efforts.

The integration of data-driven customer segmentation into CRM systems represents a promising solution for improving conversion rates and resource allocation. By leveraging advanced analytics, businesses can gain deeper insights into customer behaviors and preferences, enabling them to make more informed decisions and enhance their overall sales strategies.

## 1.2 INTRODUCTION

Customer Relationship Management (CRM) systems play a vital role in helping businesses manage client interactions, streamline sales processes, and enhance customer satisfaction. Despite their widespread adoption, many traditional CRM systems struggle to effectively prioritize customer engagement based on conversion likelihood. This can result in inefficient use of resources, as sales teams may focus on leads with a low probability of conversion while overlooking those with higher potential.

In response to these challenges, this research project proposes the development of an advanced CRM system that incorporates customer segmentation using data science and analytics. The goal is to create a solution that not only manages client information but also classifies and prioritizes customers based on their likelihood of conversion. By integrating data-driven insights into the CRM workflow, the system aims to enhance resource allocation and improve overall conversion rates.

By automatically segmenting clients into distinct categories based on key indicators such as behaviour, demographics, and engagement history, the proposed CRM system will enable businesses to focus their efforts on high-potential leads. This approach aims to make the customer engagement process more efficient and effective, ultimately driving better outcomes for sales and marketing teams.

## 1.3 Statement of the Problem

Despite the widespread adoption of Customer Relationship Management (CRM) systems, many businesses continue to struggle with efficiently converting leads into paying customers. Traditional CRM systems focus primarily on storing customer data, tracking interactions, and managing sales processes. However, they often lack the capability to effectively prioritize client engagement based on the likelihood of conversion. This gap results in several critical issues:

**1. Inefficient Resource Allocation**

Businesses frequently face the challenge of distributing their sales and marketing efforts evenly across all leads, regardless of their potential to convert. According to industry reports, over 50% of generated leads are never followed up on effectively due to poor prioritization strategies. This inefficiency results in a significant waste of time and resources, reducing the overall productivity of sales teams.

**2. Reliance on Subjective and Outdated Methods**

Current CRM systems often rely on manual assessments or simplistic lead scoring models that do not capture the complexities of customer behavior. Sales representatives typically use their judgment to determine which clients to focus on, leading to inconsistent and biased decision-making. This approach fails to utilize the vast amount of customer data available, missing the opportunity to gain deeper insights into customer preferences and behaviors.

**3. Lack of Advanced Customer Segmentation**

Most CRM systems use basic filters or demographic criteria to segment clients. However, this approach does not account for the dynamic and multifaceted nature of customer interactions. Studies show that businesses using advanced segmentation techniques can increase their conversion rates by up to 30%. The absence of such advanced, data-driven segmentation in traditional CRM systems limits their ability to accurately differentiate between high-value and low-value leads.

**4. Inadequate Focus on Conversion Optimization**

While existing CRM solutions are designed to manage lead generation and client engagement, they often lack robust features for analyzing conversion potential. This leads to a scenario where a high volume of leads is generated, but only a small percentage is effectively converted into customers. For instance, in many industries, average conversion rates can be as low as 2-5%, highlighting the need for better client prioritization and engagement strategies.

**5. Impact on Business Outcomes**

The inability to effectively prioritize clients based on their likelihood of conversion has a direct impact on business performance. It results in lower conversion rates, higher customer acquisition costs, and wasted sales efforts. According to recent market research, businesses that implement data-driven prioritization strategies can see a 15-20% increase in sales productivity, emphasizing the potential benefits of a more advanced CRM system.

**Research Problem**

The central problem addressed by this research is the lack of advanced customer segmentation and prioritization capabilities in traditional CRM systems. Despite the vast amount of customer data available, many businesses fail to leverage this data to optimize their sales efforts. This project aims to develop a CRM system that integrates data science and analytics to automatically segment clients and prioritize those with the highest conversion potential. By addressing this gap, the proposed solution seeks to enhance resource allocation, increase conversion rates, and ultimately improve business profitability.

## 1.4 Proposed Solution

This research aims to develop an advanced CRM system that integrates data science, machine learning, and predictive analytics to enhance customer segmentation and lead prioritization. Unlike traditional CRMs, this solution will focus on leveraging data-driven insights to optimize resource allocation and improve conversion rates.

**Key Operations of the System**

**1. Customer Segmentation**

The system will use machine learning algorithms to classify clients based on interaction history, demographics, and behavior, identifying high-potential leads.

**2. Predictive Analytics**

It will apply predictive models to forecast conversion likelihood, helping sales teams focus on clients with the highest probability of converting.

**3. Real-Time Insights**

A user-friendly interface will provide real-time data, enabling dynamic strategy adjustments based on up-to-date customer insights.

**4. Automated Prioritization**

The system will automatically rank leads by conversion potential, optimizing sales efforts and reducing time spent on low-probability clients.

**Adopting Modern Techniques**

The solution draws on recent advancements in global and regional CRM models, focusing on enhancing customer insights through cutting-edge analytics rather than outdated methods, to deliver a highly efficient, data-driven CRM system.

## 1.5 Objectives

**General Objective**

To develop a data-driven CRM system that utilizes customer segmentation and predictive analytics to prioritize high-conversion leads, enhancing sales efficiency and improving conversion rates.

1.6 Research Questions**Specific Objectives**

1. To conduct research on recent CRM models and data science techniques for customer segmentation and conversion forecasting.
2. To design a CRM system that integrates machine learning algorithms for customer segmentation based on interaction history, demographics, and behavioural data.
3. To implement and test predictive analytics features within the CRM system to forecast the likelihood of lead conversion and automate client prioritization.
4. To develop a user-friendly interface that provides real-time insights for dynamic adjustment of sales and marketing strategies.
5. How can recent advancements in CRM models and data science techniques be utilized to enhance customer segmentation and conversion forecasting?
6. What machine learning algorithms are most effective for segmenting customers based on interaction history, demographics, and behavioural data?
7. How can predictive analytics be integrated into a CRM system to accurately forecast lead conversion potential and automate client prioritization?
8. What features and design elements are necessary to create a user-friendly CRM interface that provides real-time insights for optimizing sales and marketing strategies?

## 1.7 Justification

This research aims to address inefficiencies in traditional CRM systems by developing a data-driven solution for better customer segmentation and lead prioritization. Businesses often struggle to allocate resources effectively, resulting in wasted time and low conversion rates. By incorporating machine learning and predictive analytics, the proposed CRM system will enable businesses to focus on high-value leads, improving sales efficiency and profitability.

The solution will contribute to the CRM field by introducing advanced analytics for smarter client prioritization. Given the increasing importance of data in business operations, this research is highly relevant, offering a timely solution to optimize customer engagement and enhance competitive advantage.

## 1.8 Proposed Research and System Methodology

The research and system development for the CRM system will follow the **Agile methodology**, which allows for flexibility, iterative progress, and frequent feedback from stakeholders. This ensures the project can adapt to evolving requirements and improves the system throughout the process.

**1. Research Methodology**

The research will use a **qualitative approach**, focusing on literature reviews and case studies to explore current CRM systems, customer segmentation, and predictive analytics. This will help identify gaps and inform the design of the proposed system.

**2. System Design and Development Methodology**

The CRM system will be developed using **Agile** in short sprints (2-4 weeks), enabling continuous development and frequent testing. Key features like customer segmentation, machine learning models, and the user interface will be developed using:

* **Python** for backend and machine learning,
* **JavaScript** and **HTML** for frontend,
* **Flask** for the backend framework.

Continuous testing will be performed throughout each sprint to ensure system components are integrated and functioning as expected.

**3. Justification for Agile**

**Agile** is chosen for its iterative nature, allowing the system to evolve based on stakeholder feedback. This is particularly important for the machine learning and predictive analytics components, which require constant refinement and testing. Agile ensures the system remains aligned with business goals and user needs.

**4. Research Life Cycle**

The research will follow these phases:

* **Phase 1: Research** – Literature review and requirement gathering.
* **Phase 2: Design** – System architecture and model design.
* **Phase 3: Development** – Iterative development of features.
* **Phase 4: Testing** – Continuous testing and feedback.
* **Phase 5: Evaluation** – Performance evaluation and system refinement.

By using **Agile**, this approach ensures flexibility, continuous improvement, and the timely delivery of an advanced CRM system tailored to optimize customer segmentation and lead prioritization.

## 1.9 Scope

This CRM and business analytics project focuses on developing a solution for customer segmentation and conversion prioritization for businesses, particularly in sectors like e-commerce, retail, and services. The system will classify customers based on interaction history, demographic data, and behavior to optimize conversion rates.

**Target Group/Population:**  
Businesses with high customer interaction, primarily in urban areas, will benefit from this CRM system.

**Limitations:**

* **Data Availability**: Quality and completeness of customer data may affect model accuracy.
* **Methodology**: The use of machine learning algorithms might not capture all customer behavior nuances.
* **Resources**: Limited by time, data, and computational power.
* **Integration**: Initial version may not integrate fully with all external systems.

**Study Confines:**  
The project will focus solely on customer segmentation and conversion prioritization. It will not cover customer acquisition, marketing campaigns, or full system integration in its first phase.

This scope ensures the project remains focused and manageable while delivering actionable insights for businesses.

# CHAPTER 2

**LITERATURE REVIEW**

## 2.1 Introduction

This chapter provides an overview of the literature related to Customer Relationship Management (CRM) systems and their integration with machine learning (ML) and predictive analytics for customer segmentation and conversion prioritization. The increasing application of machine learning and predictive analytics in CRM systems has transformed how businesses manage customer relationships, enhance targeting, and improve conversion rates. CRM systems, which traditionally focused on managing customer data and interactions, are now evolving with the help of advanced data science techniques to optimize engagement and sales strategies.

The purpose of this literature review is to examine existing CRM models, explore the theoretical underpinnings of machine learning in CRM, evaluate case studies of CRM systems with predictive capabilities, and discuss integration strategies for implementing such systems. The review will also identify key research gaps, which this project aims to address.

## 2.2 Theoretical Review

This section identifies and defines the key concepts and variables relevant to the integration of machine learning and predictive analytics into CRM systems. The key concepts include:

* **Customer Relationship Management (CRM)**: A system for managing interactions with current and potential customers, focusing on improving business relationships, enhancing customer satisfaction, and driving sales.
* **Machine Learning (ML)**: A subset of artificial intelligence (AI) that enables systems to learn from data and make predictions or decisions based on it, improving over time.
* **Predictive Analytics**: The use of statistical algorithms and machine learning models to forecast future outcomes, such as the likelihood of a customer converting to a sale.
* **Customer Segmentation**: The practice of dividing customers into distinct groups based on specific criteria (e.g., demographics, behaviors) to tailor marketing efforts effectively.

The theoretical divisions relevant to this research involve different machine learning techniques used for customer segmentation and conversion prioritization:

* **Demographic Segmentation**: Classifying customers based on demographic attributes like age, gender, income, etc.
* **Behavioral Segmentation**: Segmenting customers according to their behavior, such as purchase history, interactions, or website usage.
* **Predictive Segmentation**: Using machine learning models to predict future customer behaviors, such as the likelihood of purchasing or engaging with specific content.

Each segmentation technique has its advantages and limitations:

* **Demographic Segmentation**: Easy to implement but lacks personalization.
* **Behavioral Segmentation**: Provides more insights but may require large datasets for accurate analysis.
* **Predictive Segmentation**: Offers highly accurate predictions but relies on advanced machine learning models and high-quality data.

## 2.3 Case Study Review

Several CRM systems have successfully integrated machine learning and predictive analytics to enhance customer segmentation and conversion prioritization. Notable implementations include:

* **Salesforce Einstein**: Salesforce uses AI and predictive analytics to automate lead scoring, helping businesses prioritize high-conversion leads. It has demonstrated success in increasing marketing efficiency and sales outcomes.
* **HubSpot CRM**: HubSpot employs machine learning to segment customers based on behaviors and interactions, allowing for more targeted sales and marketing strategies.
* **Copper CRM**: Integrated with Google Workspace, Copper CRM uses machine learning and natural language processing (NLP) to automate sales processes and predict conversion opportunities based on customer interactions. Its integration with Google tools makes it particularly effective for businesses that already use Google’s suite.

These systems show that while predictive analytics and machine learning can significantly enhance CRM performance, challenges include data integration, system complexity, and the need for accurate real-time data.

## 2.4 Integration and Architecture

Implementing machine learning and predictive analytics within CRM systems requires careful consideration of integration and system architecture. Possible integration strategies include:

* **Cloud-Based CRMs**: Platforms like Salesforce and HubSpot are cloud-based and allow for seamless integration with machine learning models and data analytics tools, offering scalable solutions for businesses.
* **On-Premise Solutions**: Businesses that require more control over their data may choose on-premise CRM solutions, though these require more resources for integration and maintenance.
* **Data Integration**: Integrating data from diverse sources—such as customer interactions, purchase history, social media, and website analytics—is crucial for accurate predictions. Proper integration ensures a unified customer profile and enhances segmentation accuracy.

The architecture of a CRM system that incorporates machine learning might rely on **APIs** for data sharing, **microservices** for flexibility, and **cloud infrastructure** for scalability. Moreover, addressing the integration of **real-time data** and ensuring the system can make **data-driven decisions** quickly are essential for maximizing effectiveness.

## 2.5 Summary

This chapter has explored key concepts and theoretical frameworks surrounding CRM systems, machine learning, and predictive analytics. It has examined how machine learning is applied to CRM systems for customer segmentation and conversion forecasting, with case studies demonstrating the benefits and limitations of these technologies. Additionally, it discussed the integration challenges and strategies necessary for implementing these systems effectively in real business contexts.

## 2.6 Research Gaps

While machine learning and predictive analytics have been widely applied in CRM systems, there are several research gaps:

* **Real-Time Data Integration**: Most CRM systems do not fully leverage real-time data for dynamic decision-making. There is an opportunity to enhance systems with more responsive, up-to-the-minute predictive analytics.
* **Small Business Adaptability**: While larger businesses benefit from advanced CRM tools, small and medium-sized businesses (SMBs) face challenges in adopting such technologies due to complexity and cost.
* **Data Quality and Scalability**: Effective machine learning models rely on high-quality data. Many businesses struggle with data consistency and cleanliness, which can hinder the full potential of predictive analytics.

This research aims to address these gaps by developing a scalable, real-time CRM system that integrates machine learning algorithms for improved customer segmentation and conversion prioritization, making advanced CRM solutions more accessible to a wider range of businesses.

# 

# CHAPTER 3

# SYSTEM ANALYSIS AND DESIGN

## 3.1 Introduction

This chapter presents an in-depth analysis and design of the proposed CRM system. It follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, ensuring a structured and iterative approach to system development. The chapter covers the feasibility study, requirements elicitation, data analysis, system specifications, and both logical and physical design aspects.

## 3.2 System Development Methodology

This research adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, a robust and flexible framework tailored for data-centric system development. CRISP-DM is particularly well-suited for this project, which involves integrating a CRM application with a business analytics solution. The methodology supports iterative development and continuous refinement, making it ideal for aligning business goals with technical implementations.

The CRISP-DM process is divided into six interrelated phases:

1. Business Understanding  
   This phase focused on understanding the problem domain—low conversion rates and unoptimized client engagement in international study placement. Key objectives were defined, such as enhancing customer segmentation and prioritizing high-conversion prospects through data analytics.
2. Data Understanding  
   In this phase, data was collected from past client interactions, demographic profiles, and engagement logs (e.g., school visits, office inquiries). Exploratory analysis was carried out to identify patterns, anomalies, and insights relevant to user behavior and conversion.
3. Data Preparation  
   The collected data was cleaned, transformed, and formatted for analysis. This included removing duplicates, handling missing values, and engineering features like lead source, engagement frequency, and qualification status.
4. Modeling  
   Machine learning and statistical models were developed to classify clients based on conversion likelihood. Predictive models such as decision trees and logistic regression were explored to automate customer segmentation and scoring.
5. Evaluation  
   The performance of the models was evaluated against business metrics like conversion rate, prediction accuracy, and cost-benefit implications. This step ensured that the models aligned with the strategic goals of improving outreach efficiency and revenue.
6. Deployment  
   The final CRM and analytics system was integrated into the business workflow. A user-friendly interface was developed for consultants to access real-time insights, track client interactions, and prioritize follow-ups based on model outputs.

This structured approach ensures that both the analytical and operational components of the system are built with clear alignment to the business problem, resulting in a solution that is data-driven, scalable, and impactful.

## 3.3 Feasibility Study

A comprehensive feasibility study was conducted to evaluate the viability of developing a CRM and Business Analytics solution for the consultancy firm. The study assesses the project from multiple perspectives: economic, technical, operational, legal and ethical, and environmental feasibility. This analysis ensures that the system is achievable within the available resources, aligns with the company's operations, complies with relevant regulations, and provides a sustainable development pathway.

### 3.3.1 Economic Feasibility

Cost Analysis:  
The project leverages freely available and open-source technologies, including Python, Pandas, Flask, HTML, JavaScript, and Google Sheets, significantly reducing development costs. Since the design, development, and integration are being handled individually, there are no additional labor or outsourcing costs involved. The overall development budget is estimated at Kshs. 25,000, allocated as follows:

* Research and Data Collection: Kshs. 5,000 – for accessing data, conducting interviews, and preliminary analysis.
* System Design and Development: Kshs. 10,000 – for prototyping, UI/UX design, and back-end integration.
* Testing and Evaluation: Kshs. 5,000 – for system testing, user feedback, and refinements.
* Marketing and Outreach: Kshs. 5,000 – for branding and reaching initial users or clients.

Return on Investment (ROI):

The CRM is projected to help the consultancy convert 10 additional clients per month, each bringing in a commission of Kshs. 3,000.

* Monthly revenue: 10 clients × 3,000 = Kshs. 30,000
* Annual revenue: 30,000 × 12 = Kshs. 360,000
* Operational savings (e.g., automated follow-ups): Kshs. 10,000/year
* Referrals and retention benefits: Kshs. 5,000/year

Annual Net Benefits = 360,000 + 10,000 + 5,000 = Kshs. 375,000  
Total Net Benefits over 3 years = 375,000 × 3 = Kshs. 1,125,000

ROI = (1,125,000 − 25,000) / 25,000 × 100 = 4,400%

Interpretation:  
With an ROI of 4,400%, every Kshs. 1 invested returns Kshs. 44 in 3 years, making the project highly profitable and economically viable.

### 3.3.2 Technical Feasibility

The system is being developed using tools and platforms well within the developer’s expertise. Skills in web and Android development, UI/UX design, and data analysis make the project technically achievable. The development is being guided by the CRISP-DM methodology, ensuring a structured and efficient process that includes phases such as business understanding, data preparation, modeling, evaluation, and deployment.

Technologies used include:

* Backend: Python (Flask)
* Frontend: HTML, JavaScript
* Data Handling: Pandas, Google Sheets
* Hosting & Storage: Cloud-based platforms
* Security: Encryption and access control

The combination of open-source tools and in-house expertise confirms that the project is technically feasible.

### 3.3.3 Operational Feasibility

The system is designed to match how the consultancy engages students—via fairs, school visits, and office walk-ins. It will help staff track leads, segment clients, and prioritize follow-ups, which aligns with existing workflows.

A simple and intuitive user interface is being prioritized to minimize the need for extensive training. Staff adoption is expected to be smooth, with clear benefits in efficiency and decision-making.

Thus, operational integration is highly feasible and likely to enhance the consultancy’s service delivery.

### 3.3.4 Legal and Ethical Feasibility

The project is being developed in line with Kenya’s Data Protection Act and relevant international standards. The following legal and ethical considerations are being implemented:

* Informed Consent: Clients will be informed before their data is collected or used.
* Data Security: Encryption, secure authentication, and access control mechanisms will be implemented.
* Ethical Use: The system promotes transparency, fairness, and privacy, ensuring ethical data handling and client interactions.

These measures ensure the solution is legally compliant and ethically sound.

### 3.3.5 Environmental and Time Feasibility

The system is being developed incrementally based on the CRISP-DM process, enabling continuous progress, feedback, and testing. By using cloud-based tools like Google Sheets, the project minimizes its environmental footprint, avoiding physical infrastructure or paper-based processes.

Development is being conducted remotely and independently, allowing flexible scheduling and efficient resource use. Given the scope, timeline, and development approach, the project is both environmentally and temporally feasible.

## 3.4 Requirements Elicitation

A structured data collection process was designed and implemented to gather requirements directly from potential users and stakeholders within the student placement consultancy sector. This approach ensures that the final system is aligned with real-world business needs and user expectations.

### 3.4.1 Data Collection

To collect the necessary data, multiple tools were utilized:

* **Questionnaires**: A comprehensive questionnaire was developed to gather information on user needs, current workflows, pain points, and expectations for the CRM and analytics system. It included both open-ended and closed-ended questions to enable both quantitative and qualitative analysis. The questionnaire was distributed digitally via Google Forms.
* **Interviews**: Semi-structured interviews were conducted with key staff members at the consultancy, including counselors and marketing personnel. These interviews provided deeper insights into lead handling, client conversion, and follow-up processes, capturing practical challenges and expectations that may not surface through questionnaires alone.
* **Observation**: Direct observation of client interactions during school visits and walk-ins was carried out to understand how data is currently captured and how follow-ups are performed. This helped identify inefficiencies and usability gaps that the proposed system should address.

All data collection tools were designed to align with the project’s core objectives: increasing conversion rates, improving lead prioritization, and enhancing operational efficiency.

### 3.4.2 Sampling Techniques

Stratified sampling was employed to ensure that feedback was collected from a diverse range of participants involved in the client engagement process. The sample was deliberately structured to include counselors, data entry clerks, marketing staff, and prospective students. This approach ensured that the requirements gathered were representative of the different roles and use cases within the consultancy.

To determine an appropriate sample size, the total number of staff and stakeholders regularly involved in client onboarding and follow-up processes was taken into account. The target sample included:

* At least 5 counselors
* 3 marketing staff
* 2 administrative or data entry staff

This sampling strategy provided a balanced and manageable dataset, enabling the extraction of meaningful insights while maintaining efficiency in terms of time and resource allocation.

### 3.4.3 Relevance to the Problem and Objectives

The data collected directly informs the functional and non-functional requirements of the CRM system. For example:

* Questions related to follow-up challenges and lost leads revealed the need for automated reminders and lead scoring.
* Feedback on difficulties in tracking students from initial contact to conversion highlighted the need for a visual pipeline/dashboard.
* Interviews exposed the need for segmenting students by qualification level (e.g., KCSE, diploma, degree), influencing the system’s categorization logic.

By aligning data collection with the research objectives, it ensured that all inputs were actionable and useful in shaping a user-centered solution.

## 3.5 Data Analysis

A comprehensive analysis of the data collected during the requirements elicitation phase was conducted using Microsoft Excel. The primary objective of this analysis was to uncover patterns, user preferences, system expectations, and operational pain points. These insights played a critical role in guiding the design and development of the CRM and business analytics system, ensuring alignment with user needs and organizational goals.

### 3.5.1 Data Cleaning and Organization

Prior to analysis, the dataset was cleaned to eliminate incomplete or inconsistent responses. Data entries were also standardized to ensure compatibility with Google Sheets analytical tools. This included tasks such as formatting values uniformly and converting categorical responses into numerical codes where necessary, thereby enhancing the accuracy and efficiency of subsequent analysis.

### 3.5.2 Statistical Analysis

Google Sheets was used to apply basic statistical techniques to analyze the survey data collected from stakeholders involved in the student consultancy process. The following methods were employed:

* **Frequency Distribution**: This technique was used to identify the most commonly preferred CRM features. For instance, *Lead Tracking*, *Visual Pipeline Tracking*, and *Analytical Dashboard* received the highest frequency of top ratings (mostly 5s), indicating their strong importance to users.
* **Percentage Analysis**: Percentages were calculated to determine the proportion of users facing specific operational challenges. The results showed that a majority (over 70%) of respondents reported issues such as *manual student categorization*, *difficulty tracking conversion progress*, and *lack of performance analytics*.
* **Cross-Tabulation**: This method was applied to compare CRM feature preferences and reported challenges across different staff roles (e.g., counselors, admin staff, marketing staff). For example, counselors emphasized the need for *automated follow-up* and *conversion monitoring*, while marketing staff highlighted the need for *event tracking* and *communication integration*.

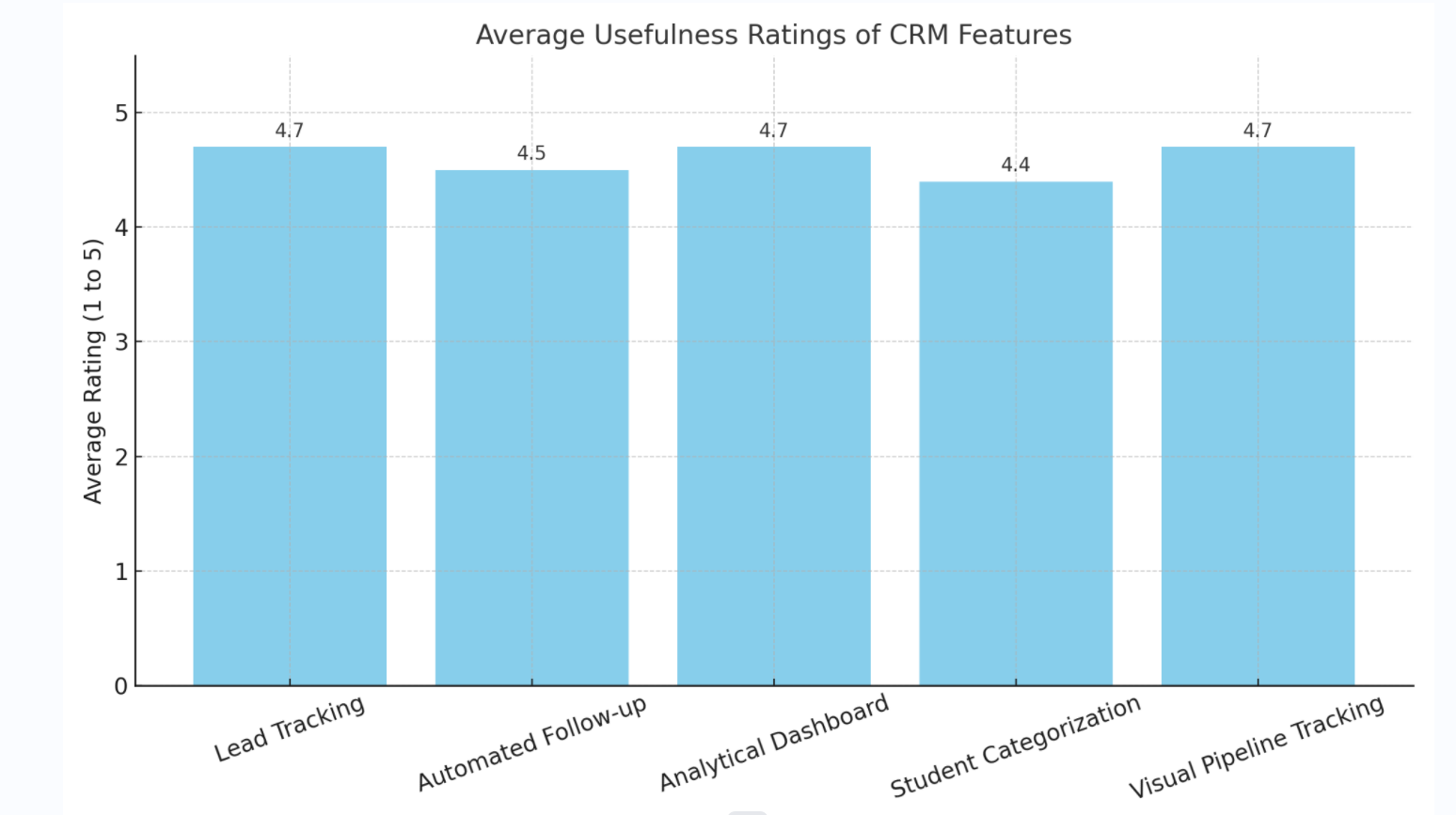
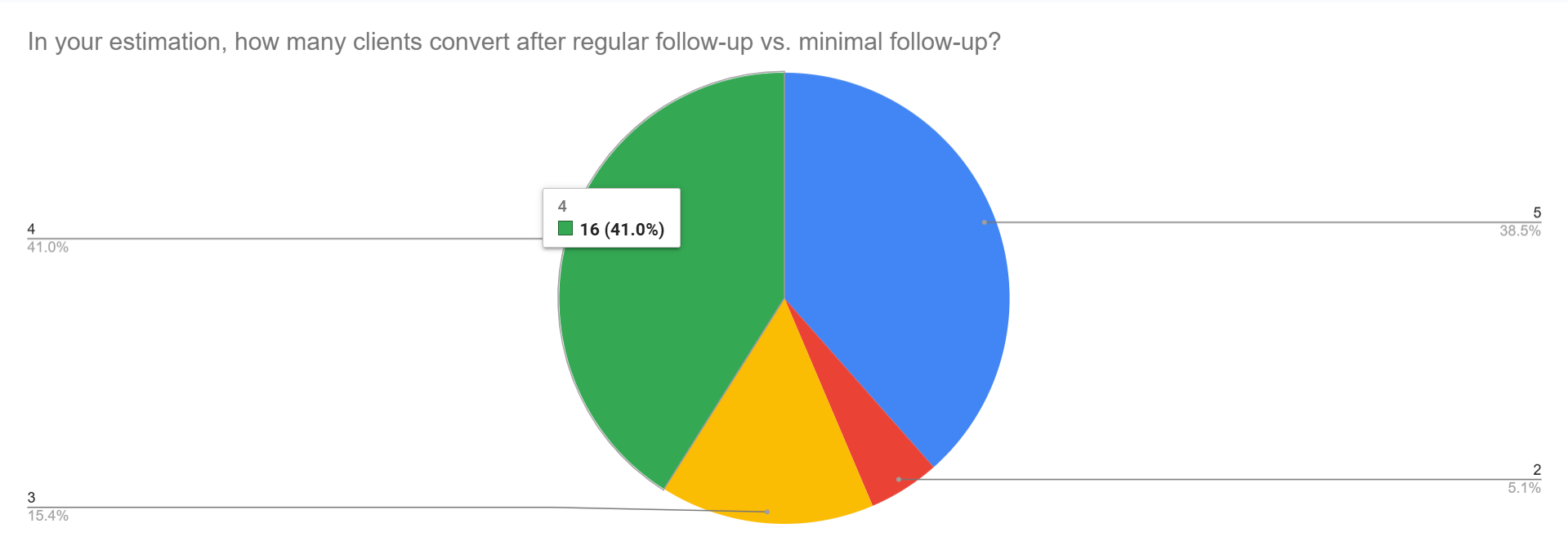
These statistical techniques provided a clear, data-driven understanding of user needs, identifying key areas for improvement in the CRM system.

### 3.5.3 Visualization of Findings

To communicate the results effectively, key insights were visualized using analytical tools:

* Bar Graph:
  + Showed the distribution of key challenges

Forms response chart. Question title: What are the main challenges you face when managing clients? (Select all that apply)
. Number of responses: 10 responses.

* Bar Graphs:
  + Compared features users expect from the CRM (e.g., lead tracking, reminder automation, analytics dashboard, student categorization).
  + 
* Pie Chart:
  + Demonstrated increase conversion rates based on follow-up frequency, as reported by counselors.  
      
    

### 3.5.4 Key Insights from the Data

**1. Most Valued CRM Features (Based on Ratings)**

* **Lead Tracking, Visual Pipeline Tracking, and Analytical Dashboard** were rated highest (average of **4.7/5**).
* **Automated Follow-up** also scored high with **4.5/5**.
* **Student Categorization** scored slightly lower but was still crucial at **4.4/5**.

**Takeaway**: Users strongly prefer tools that offer clear visualisation, data analytics, and lead monitoring.

**2. Common Challenges Faced**

* Most common issues:
  + Manual categorization of students
  + Difficulty tracking conversion progress
  + Lack of analytics to assess performance
  + Poor follow-up mechanisms
* These challenges are interlinked with the most requested CRM features, indicating strong alignment between pain points and desired tools.

**3. Client Tracking & Dashboard Preferences**

* 100% of respondents said they would prefer a **visual dashboard** to monitor student progress.
* A majority reported **frequent loss of client tracking** due to manual processes.

**Takeaway**: There's a strong need for a **centralized visual tracking system**.

**4. Follow-Up Behavior & Conversion**

* All respondents except one noticed a **positive correlation** between **follow-up frequency and client conversion**.
* Most follow up **daily or weekly** — showing that **follow-up is a top operational priority**.

**5. Client Conversion Estimates**

* Respondents estimate higher conversion with regular follow-ups:
  + Regular follow-ups: 4–5 clients convert
  + Minimal follow-ups: 2–3 clients convert

**Takeaway**: **Automated follow-ups could significantly improve conversion rates**.

**6. Suggested CRM Improvements**

Users suggested:

* Automated reminders & follow-ups
* Customizable dashboards
* WhatsApp/SMS integration
* Mobile-friendly CRM version
* Event and document management
* Bulk SMS and client feedback tools

**Takeaway**: Users need a **comprehensive, multi-functional CRM** that is mobile and communication-friendly.

**7. Demographics**

* Respondents include **Counselors, Admin Staff, Visa Advisors, and Marketing Staff**.
* Majority have **less than 3 years** of experience — CRM should be **intuitive and beginner-friendly**.

## 3.6 System Specification

As the system developer, I outline the detailed system specifications derived from the business requirements and data analysis. The system is tailored to assist consultancy staff in efficiently managing client interactions, monitoring the conversion pipeline, and generating actionable business insights.

### 3.6.1 Functional Requirements

These describe the core features and operations the system must support:

* **Lead Management**
  + Ability to add, edit, and categorize prospective students.
  + Tag clients based on education level (e.g., high school, bachelor’s, master’s, PhD).
* **Interaction Tracking**
  + Record and track student interactions across multiple touchpoints (fairs, school visits, office visits).
* **Conversion Pipeline**
  + Visualize and manage the conversion journey: Contacted → Interested → Documented → Applied.
* **Automated Reminders**
  + Notify staff for follow-up based on inactivity or deadlines.
* **User Access Control**
  + Role-based access for administrators, counselors, and marketing staff.
* **Analytics Dashboard**
  + Real-time metrics on conversion rates, interaction frequency, and counselor performance.
* **Document Upload & Notes**
  + Attach supporting documents and internal notes to each client profile.
* **Feedback & Surveys**
  + Allow collection of feedback post-application to improve services.

### 3.6.2 Non-Functional Requirements

These define system qualities and constraints:

* **Usability**
  + Clean, intuitive interface with minimal training required.
  + Mobile-friendly and responsive design for use on tablets and phones.
* **Performance**
  + Load dashboard and reports within 3 seconds.
  + Handle at least 500 concurrent users without lag.
* **Security**
  + Encrypted data storage for client information.
  + Secure login with two-factor authentication (2FA) for admins.
* **Scalability**
  + System should scale easily as more clients and users are added.
  + Designed using modular components to allow future feature integrations.
* **Maintainability**
  + Clean codebase and documentation for easy maintenance and updates.
* **Availability**
  + Target uptime of 99.5% to support operational continuity.
* **Backup & Recovery**
  + Daily automated backups with the ability to restore previous versions in case of data loss.

This system specification serves as a foundational blueprint that guides the development process and ensures the final product aligns with business needs and user expectations.

## 3.7 Requirements Analysis and Modeling

The gathered requirements for the CRM and business analytics solution for the student placement consultancy were analyzed to ensure consistency, resolve conflicts, and transform business needs into structured software components. This stage also bridges the gap between user expectations and system design by modeling how the system should behave.

### 3.7.1 Requirements Analysis

After reviewing data from interviews, questionnaires, and observations:

* Dependencies Identified:
  + The dashboard feature depends on timely and accurate data entry by staff.
  + The conversion tracking system relies on consistent status updates for each client.
* Conflicts Observed:
  + Some counselors preferred a simplified dashboard, while others requested detailed analytics. This was resolved by offering two dashboard views: summary and detailed.
  + Concerns about data privacy led to the definition of strict access levels based on user roles.
* Potential Solutions:
  + Automate routine data inputs using drop-down options and validation rules.
  + Implement role-based access control to address security and data privacy concerns.
  + Design modular components for better flexibility and future enhancements.

### 3.7.2 Requirements Structuring

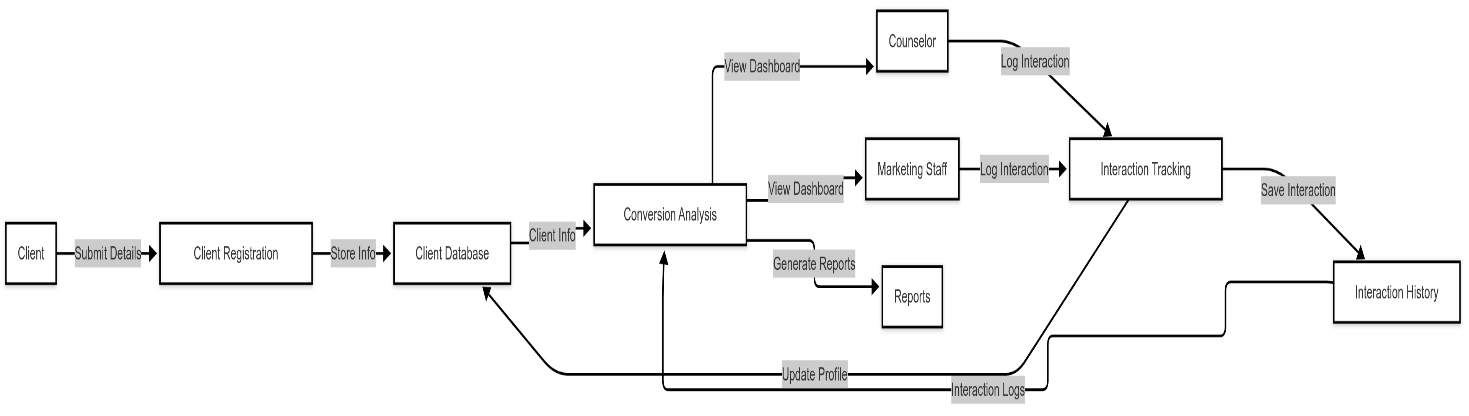
**The refined requirements are structured into software components as follows:**

| **Requirement** | **Software Component** |
| --- | --- |
| Capture leads and categorize clients | Lead Management Module |
| Track interactions and follow-ups | Client Interaction Module |
| View conversion pipeline | Conversion Funnel Tracker |
| Upload documents and notes | Client Profile Manager |
| Generate reports and analytics | Analytics Dashboard |
| Restrict access by role | User Management & Security Module |

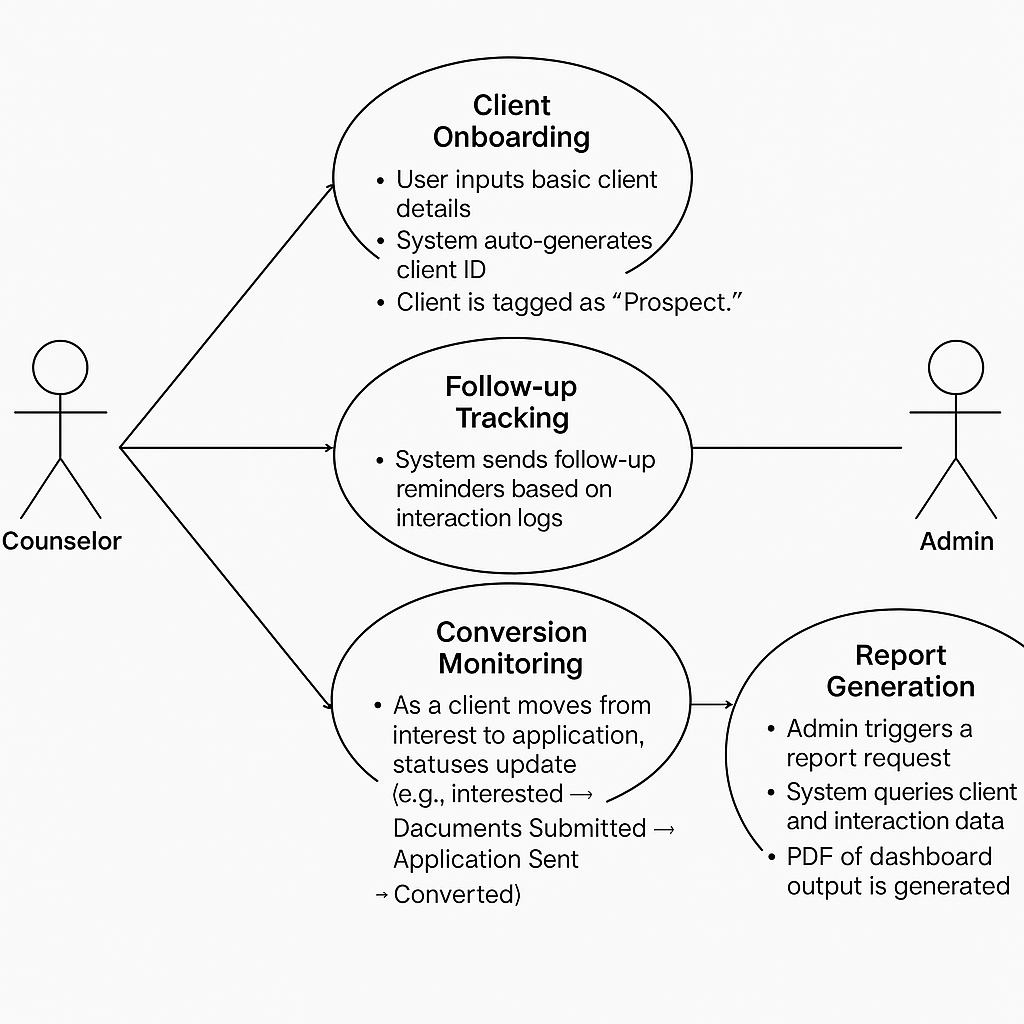
### 3.7.3 Modeling Tools

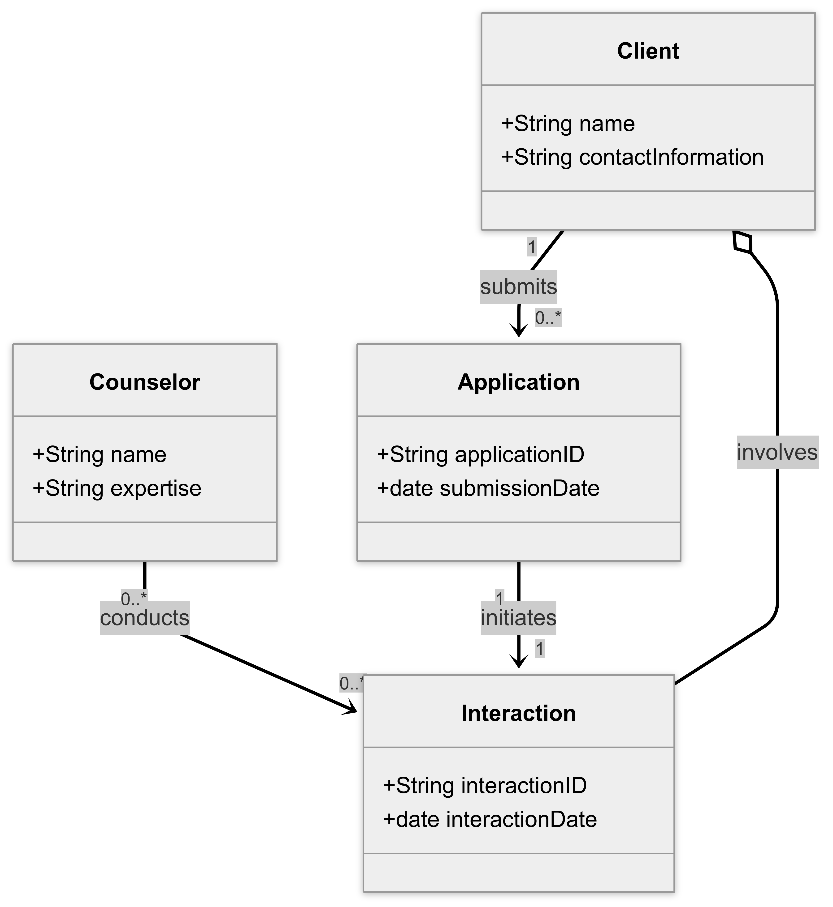
To visualize and clarify the functional requirements, the following tools and diagrams will be used:

* Low-Level Data Flow Diagrams (DFDs)  
  Shows how data moves through the system.



* Use Case Diagrams



* Conceptual and Analysis Class Diagram  
  Represent the main system entities (Client, Counselor, Application, Interaction) and their relationships.
* ****

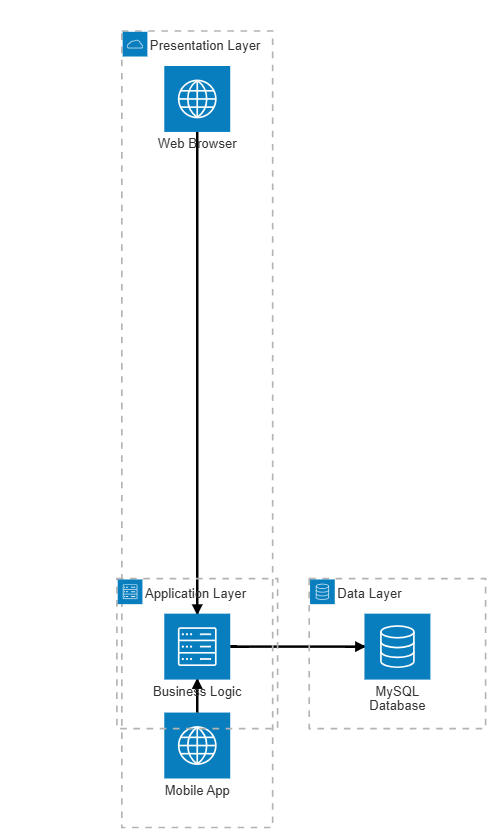
## 3.8 Logical Design

This section outlines the logical representation of the CRM and business analytics system. It captures the system’s core structure, behavior, and functionality to serve as a blueprint for the physical design and implementation phases.

### 3.8.1 System Architecture

The proposed system adopts a three-tier client-server architecture, comprising the following layers:

* Presentation Layer (Client Side)  
  This includes the user interface accessed by counselors, admins, and marketing staff via web browsers or mobile apps. It handles all user interactions.
* Application Layer (Business Logic)  
  This layer processes requests, manages business rules (e.g., client segmentation, conversion scoring), and connects the frontend with the backend. It is developed using Python/Flask and JavaScript frameworks.
* Data Layer (Database Server)  
  Manages client records, interaction logs, user accounts, and analytics data. It uses a relational database (MySQL).



**Major Components and Modules:**

| **Component** | **Description** |
| --- | --- |
| Lead Management | Captures and categorizes potential clients |
| Interaction Tracker | Logs follow-ups, meetings, and updates |
| Conversion Funnel | Visualizes client progress through the application process |
| User Management | Manages access control and user roles |
| Analytics Engine | Provides dashboards, reports, and insights |
| Document Manager | Stores application files and notes |

### 3.8.2 Control Flow and Process Design

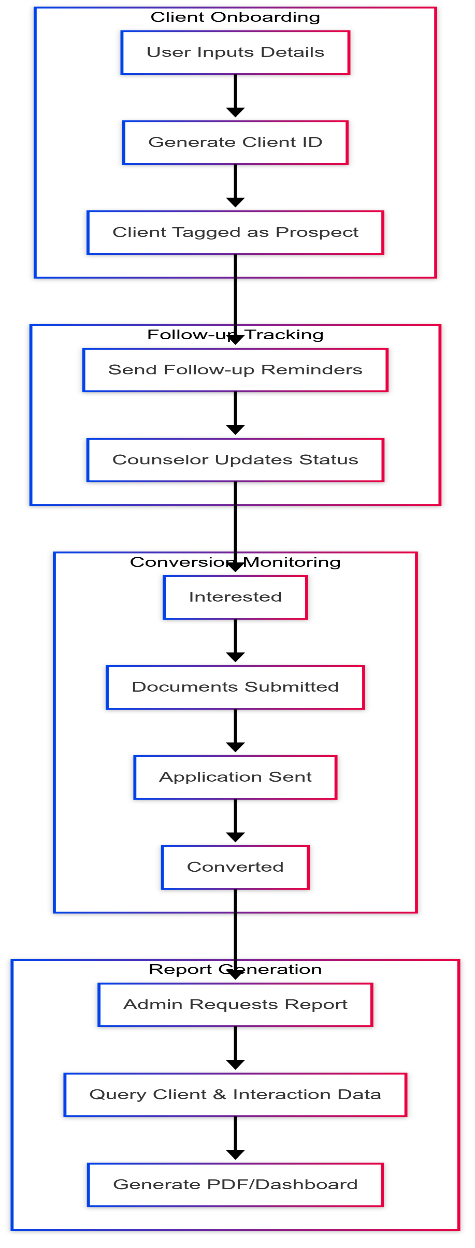
This part defines how tasks are sequenced, how decisions are made, and how data flows through the system.

Key Processes:

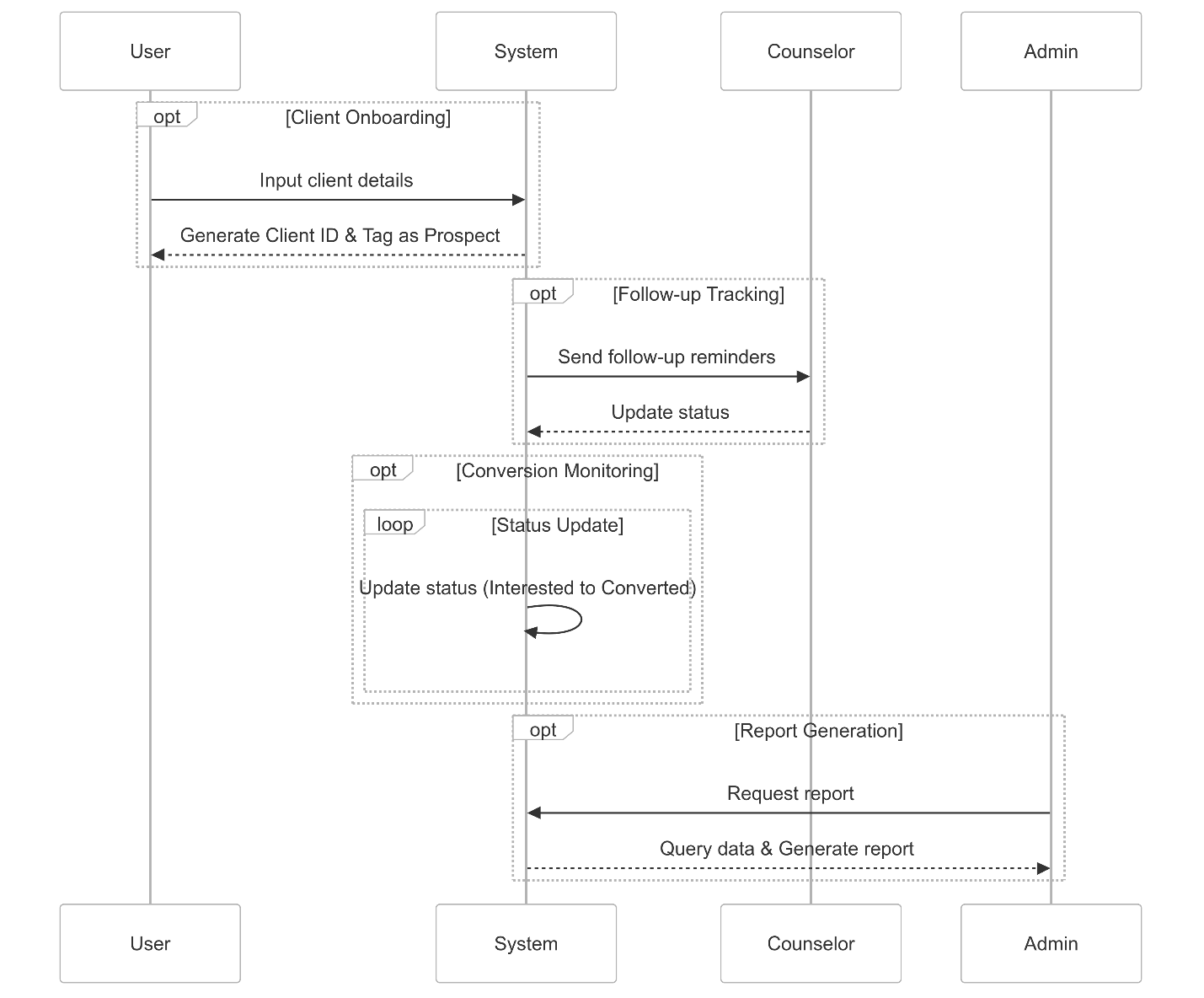
1. Client Onboarding
   * User inputs basic client details.
   * System auto-generates client ID.
   * Client is tagged as “Prospect.”
2. Follow-up Tracking
   * System sends follow-up reminders based on interaction logs.
   * Counselor updates status after communication.
3. Conversion Monitoring
   * As a client moves from interest to application, statuses update (e.g., Interested → Documents Submitted → Application Sent → Converted).
4. Report Generation
   * Admin triggers a report request.
   * System queries client and interaction data.
   * PDF or dashboard output is generated.

Modeling Tools:

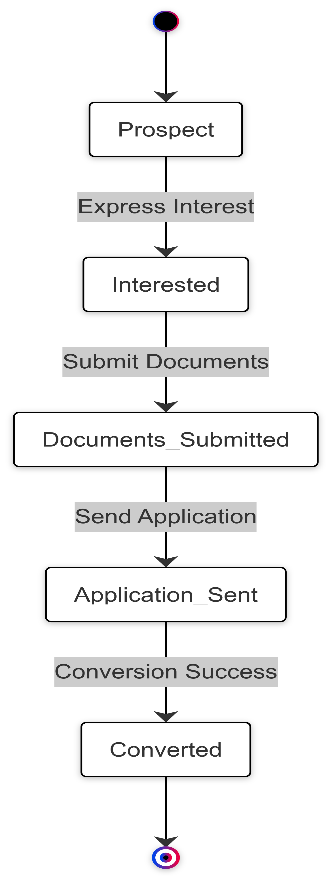
* Flowcharts: Depict data flow for onboarding, follow-ups, and reporting.



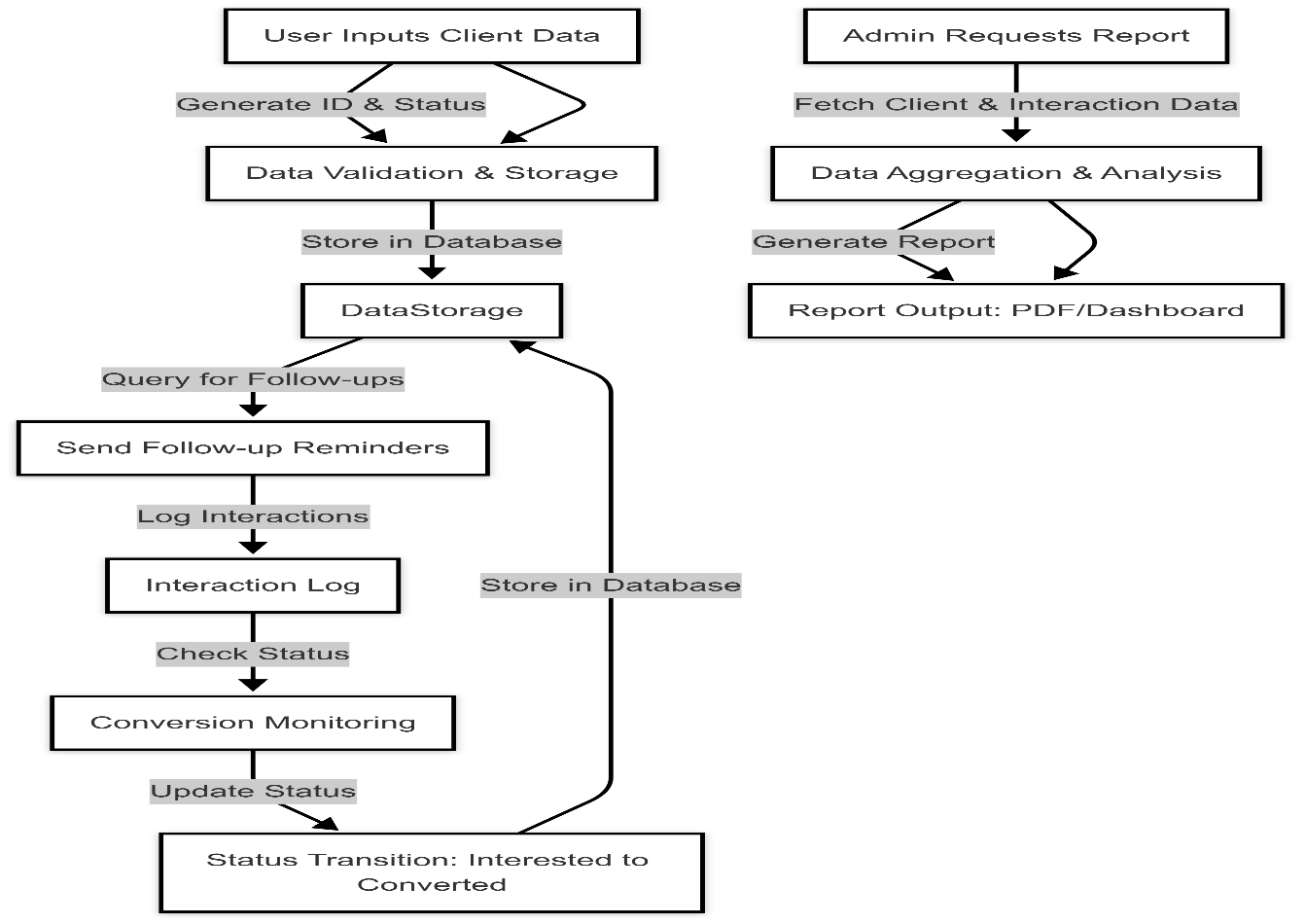
* Sequence Diagrams: Capture interactions between users, UI, and backend.



* State Chart Diagrams: Show state transitions (e.g., from "Prospect" to "Converted").



* Low-level DFDs: Detail how input transforms into meaningful reports.



### 3.8.3 Design for Non-Functional Requirements

Security Strategies:

* Role-Based Access Control (RBAC):  
  Users can only access features allowed by their roles (e.g., Admin, Counselor, Marketing Officer).
* Authentication and Session Management:  
  Secure login system with encrypted passwords and session tokens.
* Data Encryption:  
  Sensitive data (e.g., client contacts, documents) is encrypted in transit and at rest.

**Error and Exception Handling:**

* Structured exception handling ensures system stability.
* Custom error messages help users understand what went wrong without exposing technical details.
* Logs are maintained for admin review and debugging.

**Efficiency and Usability Enhancements:**

* Form Auto-complete and Validation: Reduces input errors and saves time.
* Responsive Design: Ensures accessibility on multiple devices.
* Dashboard Filters: Allows users to segment and sort client data easily.
* Loading Indicators and Feedback Messages: Improve user experience during data-heavy operations.

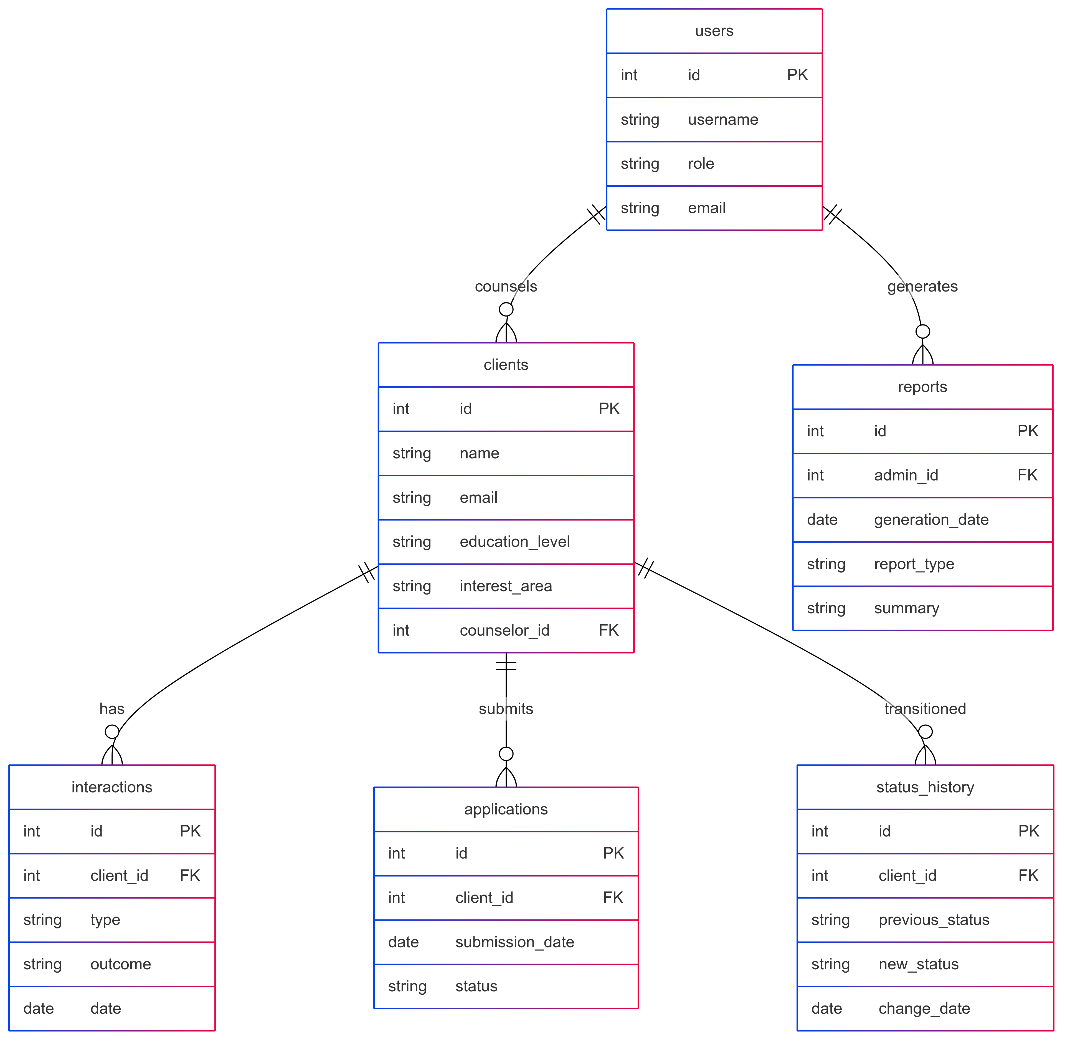
## 3.9 Physical Design

The physical design phase translates the logical design into concrete decisions about hardware, software, network, database, and UI components to support the actual implementation and deployment of the CRM and business analytics system.

### 3.9.1 Database Design

The database is designed with a focus on reliability, performance, scalability, and ease of access. MySQL is chosen as the database management system due to its robustness, support for complex queries, indexing capabilities, and security features.

**Database Schema Design**



**The database contains the following main tables:**

| **Table Name** | **Description** |
| --- | --- |
| users | Stores system user information (admins, counselors, marketing) |
| clients | Stores lead information such as name, email, education level, and interest area |
| interactions | Logs calls, emails, meetings, and their outcomes |
| applications | Tracks submitted applications per client |
| status\_history | Monitors how each client transitions through the conversion stages |
| reports | Stores summary and analytics data for visualization |

**Key Fields and Relationships**

* Each client is linked to a counselor (clients.counselor\_id → users.user\_id)
* Each application belongs to a client (applications.client\_id → clients.client\_id)
* status\_history tracks all status changes over time (status\_history.client\_id → clients.client\_id)
* interactions.client\_id → clients.client\_id

**Storage and Optimization Considerations**

* Indexes are placed on frequently queried fields (e.g., email, status, created\_at) to enhance speed.
* Foreign Key Constraints enforce referential integrity.
* Role-Based Access is applied through PostgreSQL roles to restrict data access.
* Backups and logs are maintained for data recovery and auditing purposes.

### 3.9.2 User Interface Design

The user interface is designed to be intuitive, accessible, and aligned with the actual workflows of the CRM users. The goal is to reduce complexity while maximizing productivity and visibility into the client journey.

**System Interfaces**

| **Interface** | **Purpose** |
| --- | --- |
| Login Page | Secure user access and session management |
| Dashboard | Overview of KPIs, active leads, conversion status, and alerts |
| Lead Entry Form | Input form for counselors to onboard new clients |
| Client Profile Page | View and edit client details, track status and notes |
| Interaction Log Form | Form to capture follow-ups and communication |
| Application Tracking | Timeline of each client’s application progress |
| Report Generator | Filter data and generate visual reports for management |

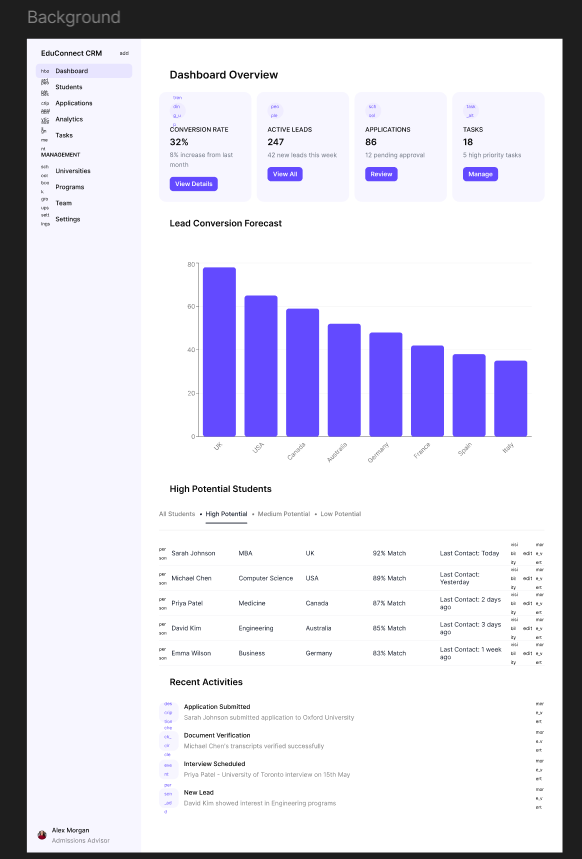
**Input/Output Form Samples**

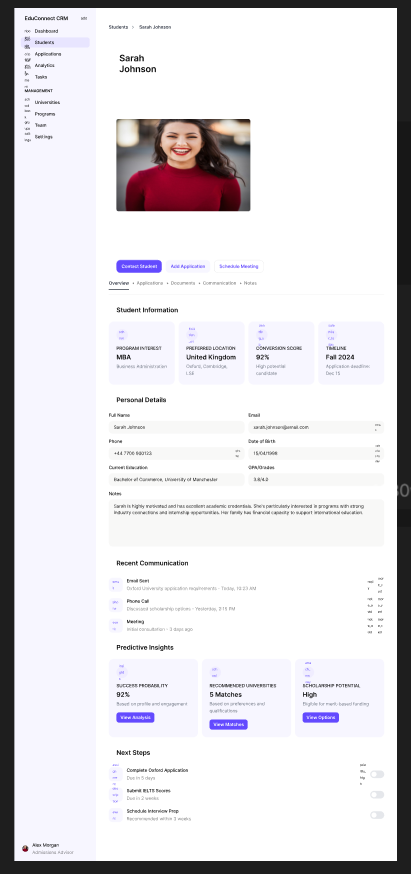
* Input: New client registration, interaction form, application update
* Output: Dashboard summary cards, downloadable client report, funnel visualization

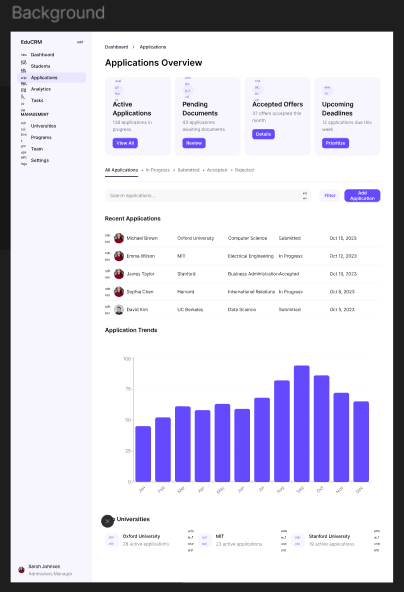
**Wireframes**

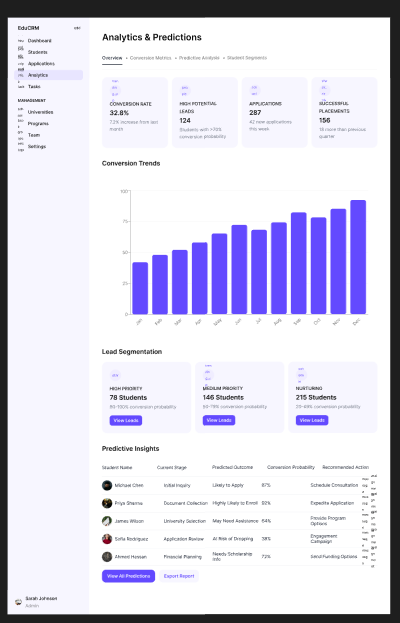
Wireframes are created using Figma and show:

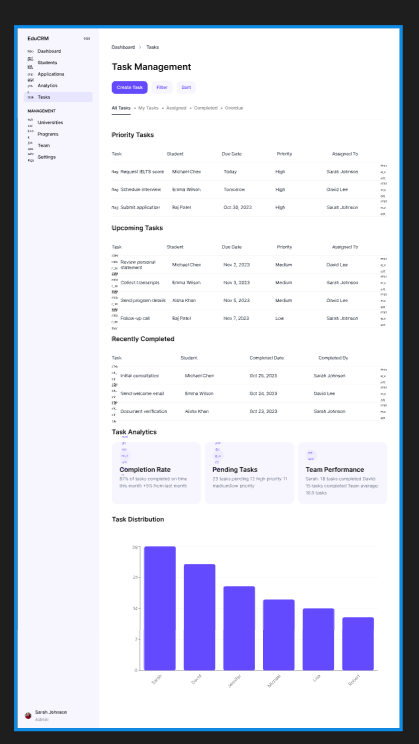
* Clear navigation sidebar with icons for the dashboard, Students, applications, Analytics and tasks
* Input forms with validation, tooltips, and progress indicators
* Analytics view with charts (bar) showing conversion trends and user performance



****

****

****

****

# 

# CHAPTER FOUR:

# SYSTEM IMPLEMENTATION AND TESTING

## Introduction

This chapter details the implementation and testing phases of EduConnect, the Customer Relationship Management (CRM) system developed specifically for educational consultants. It outlines the chosen technologies, describes the development process for key system interfaces, and elaborates on the testing strategies employed to ensure the system's functionality, reliability, and usability. The chapter concludes with a critical appraisal of the completed project, highlighting its strengths, acknowledging its limitations, and proposing recommendations for future enhancements.

## 4.1 Technologies Used

The development of EduConnect leveraged a combination of modern tools and technologies selected for their suitability in building a reliable, scalable, and maintainable web application:

1. **Backend Framework:** Python with Flask
   * *Role:* Provided the core structure for the web application, handling request routing, business logic, and integration between the frontend, database, and machine learning model. Chosen for its simplicity and flexibility.
2. **Machine Learning Library:** scikit-learn
   * *Role:* Implemented the Random Forest Classifier used for predicting student conversion likelihood based on historical data and student aflributes. Chosen for its comprehensive algorithms and ease of integration with Python.
3. **Database:** SQLite
   * *Role:* Served as the data persistence layer, storing user credentials, student information, interaction logs, and application details. Chosen for its lightweight nature and simple file-based setup, suitable for initial development and deployment.
4. **Frontend Development:** HTML, CSS, and Bootstrap
   * *Role:* Used to create the visual elements and layout of the web interface. Bootstrap was specifically employed to ensure a responsive design that

adapts well to different screen sizes and to provide pre-built UI components for faster development.

1. **Data Handling:** pandas and numpy
   * *Role:* Utilized primarily for data manipulation, cleaning, and preparation tasks, especially when feeding data into the machine learning model and processing data for analytics.
2. **Development Environment:** Visual Studio Code (VS Code)
   * *Role:* Served as the primary Integrated Development Environment (IDE) for writing, debugging, and managing the project's codebase.
3. **Version Control:** Git and GitHub
   * *Role:* Used for tracking changes to the codebase, facilitating collaboration (if applicable), and maintaining a history of revisions. GitHub served as the remote repository.

### 4.1.1 Actual System Implementation

This section describes the core modules and interfaces developed for EduConnect.

### Login Interface

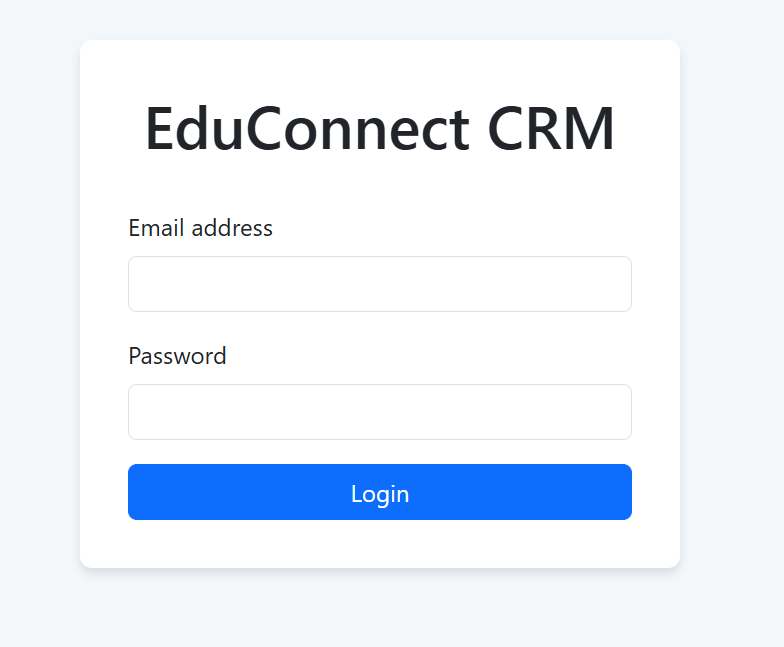
* **Description:** The entry point for registered users. This interface ensures secure access by requiring valid credentials, protecting sensitive student data and system functionalities.
* **Implementation Details:** A simple web form collects the user's email and password. Upon submission, the backend Flask route verifies these credentials against the users table in the SQLite database. Successful authentication establishes a user session; failure results in an error message. *Note: Password hashing should be implemented for enhanced security.*

Code (Illustrative Flask Route):

* @app.route('/login', methods=['GET', 'POST'])
* def login():
* if request.method == 'POST':
* email = request.form.get('email')
* password = request.form.get('password')
* user = verify\_user(email, password)
* if user:
* session['user\_id'] = user['id']
* session['user\_name'] = user['name']
* session['user\_role'] = user['role']
* return redirect(url\_for('index'))
* else:
* return render\_template('login.html', error='Invalid email or password')
* return render\_template('login.html')
* @app.route('/logout')
* def logout():
* session.clear()
* return redirect(url\_for('login'))
* # Protect all routes that require authentication
* @app.route("/")
* @login\_required
* def index():
* return render\_template("index.html", page="index")

# 

**Visual:** [Image: Screenshot of the EduConnect Login Page]



### Dashboard

* **Description:** The central hub for consultants after logging in. It provides a

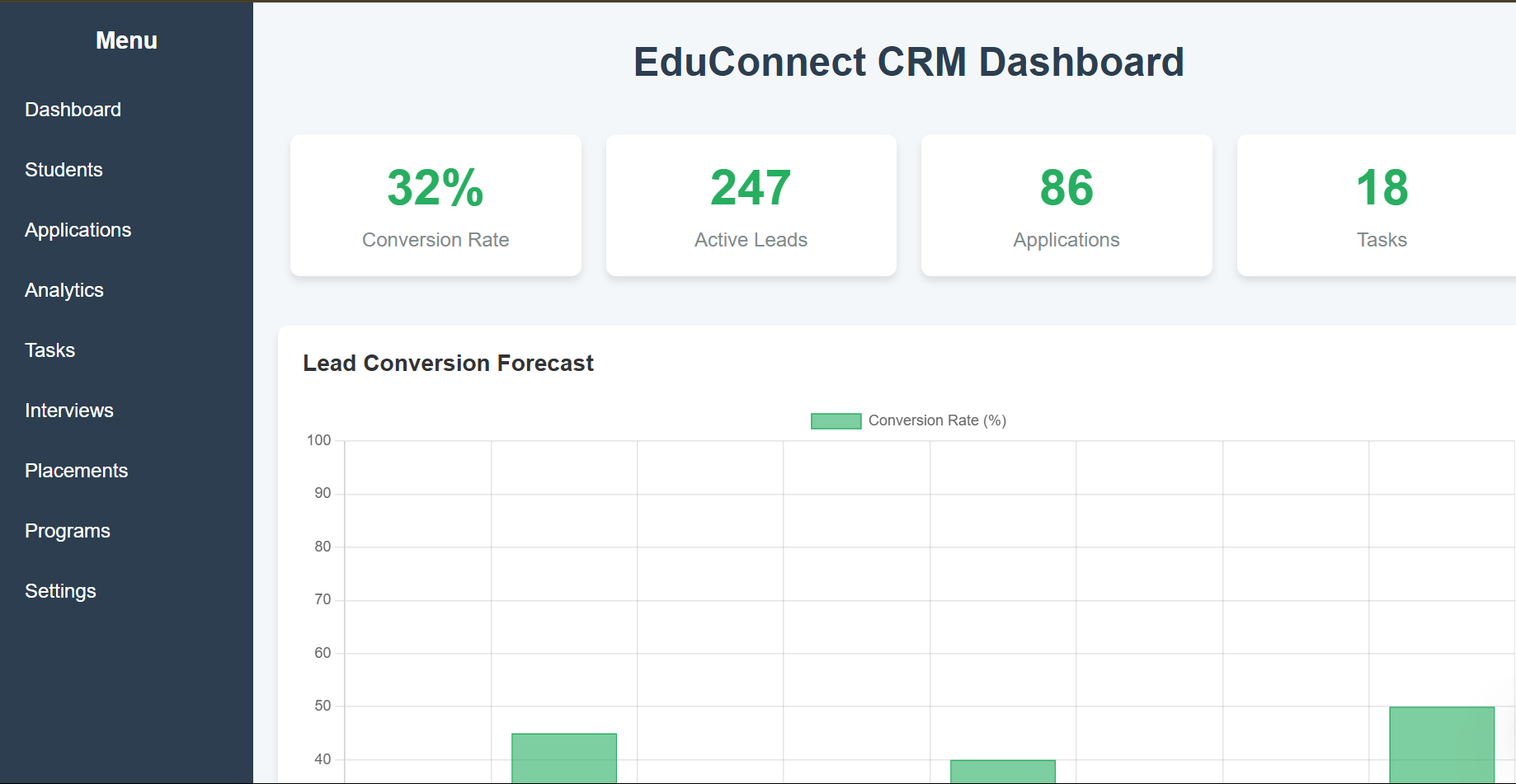
high-level overview of key activities and metrics, enabling users to quickly grasp their workload and priorities.

* **Implementation Details:** The dashboard aggregates data from various parts of the system. It displays summaries such as the number of new leads, upcoming tasks or appointments, recent student interactions, and potentially highlights students with high conversion probabilities calculated by the ML model. The frontend uses HTML, CSS, and Bootstrap components, populated with data fetched by Flask routes.

# Key Information Displayed:

* + Summary statistics (e.g., total students, conversion rate).
  + List of upcoming tasks or follow-ups.
  + Quick view of recent student activity or high-priority leads.
  + Navigation links to other system modules.

**Visual:** [Image: Screenshot of the EduConnect Dashboard Interface]



## Student Profile Management

* **Description:** Allows consultants to create, view, update, and delete (CRUD) student records. This module serves as the central repository for all information related to a specific student.

# Features:

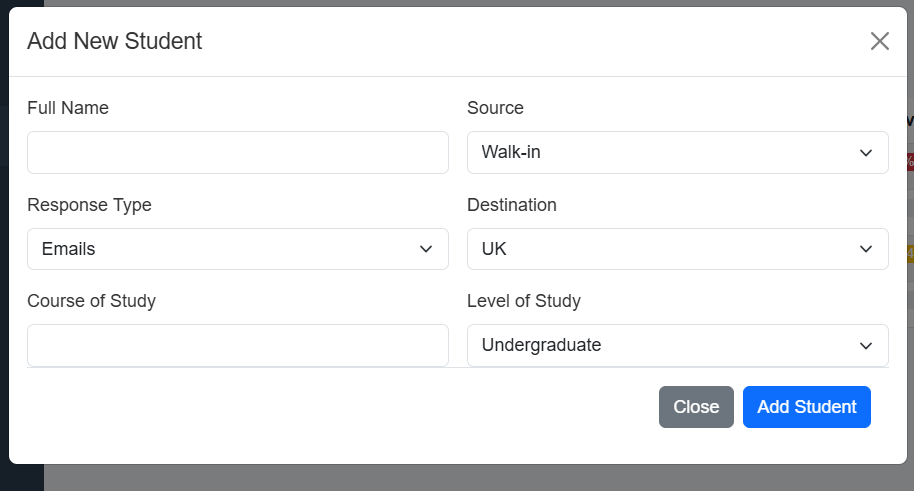
* + **Student Information:** Fields for name, contact details (email, phone), academic background (level, grades), preferred study destinations/courses, lead source, etc.
  + **Interaction Logging:** A chronological record of communications (calls, emails, meetings) with the student, including notes and dates.
  + **Document Upload (Optional):** Facility to aflach relevant documents (e.g., transcripts, passport copies).
* **Database Implementation:** Specific tables in SQLite (students, interactions)

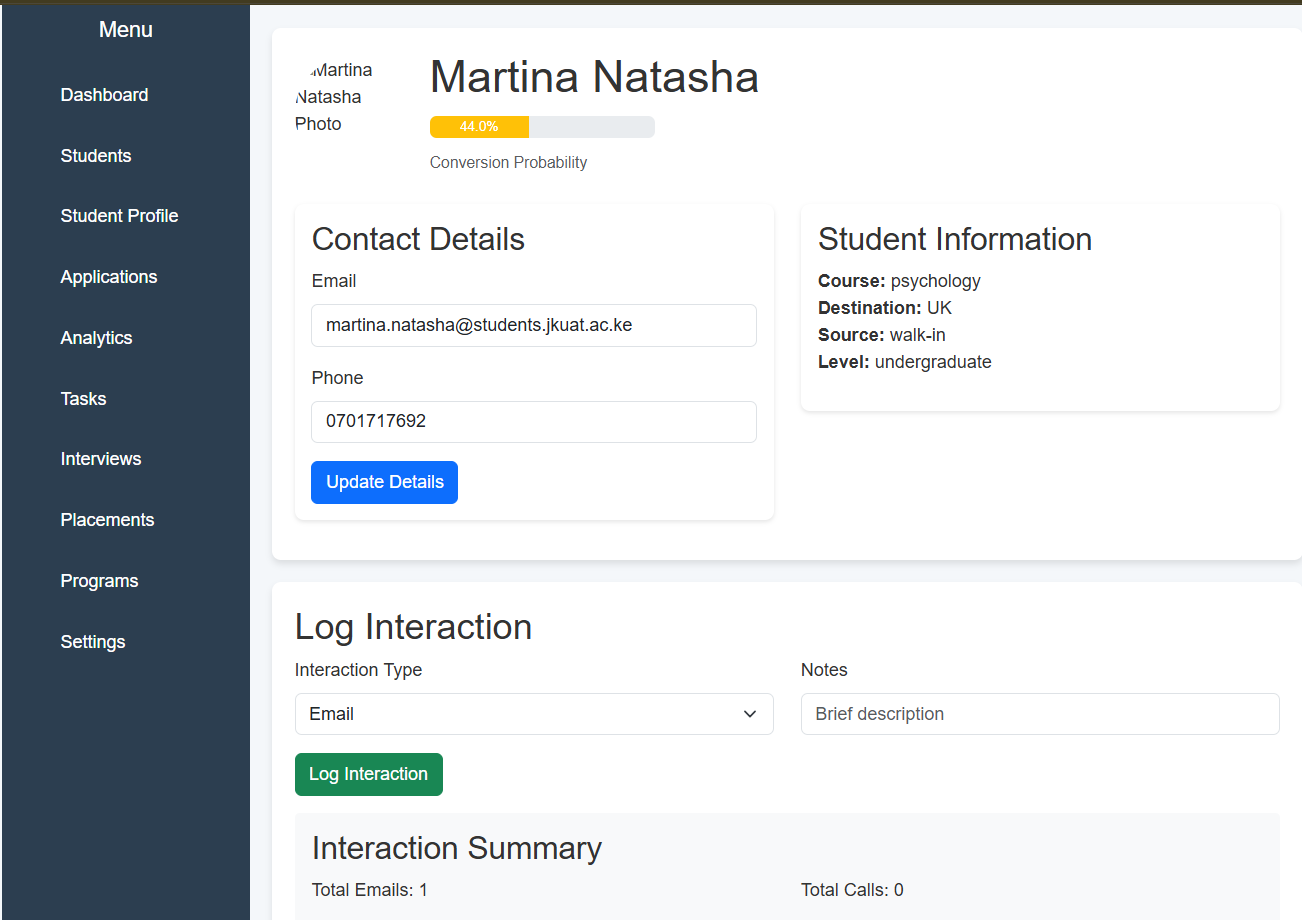
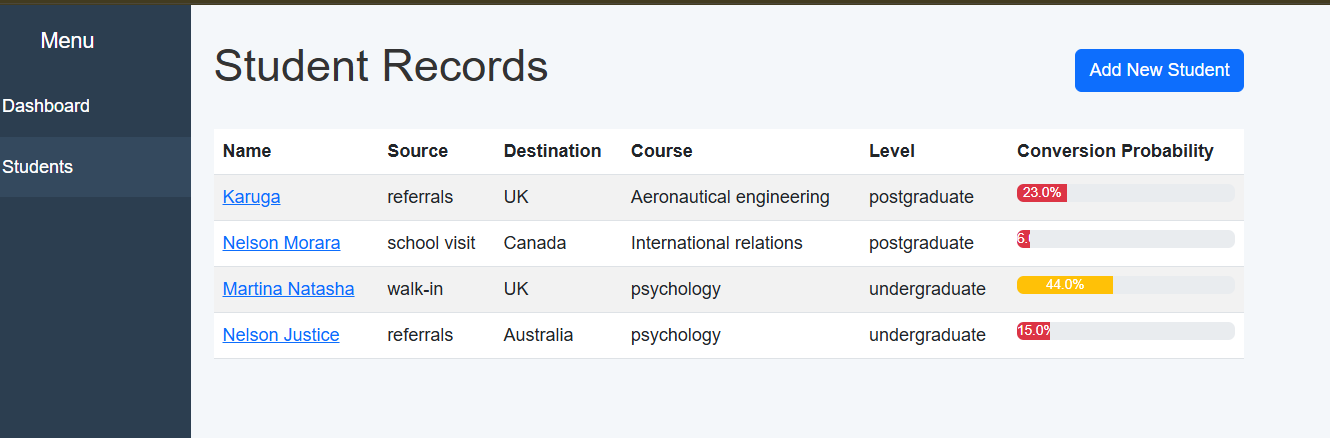
store this data. Backend routes handle form submissions to add new students or update existing ones.

Code (Illustrative Add Student Logic):

* def add\_student(student\_data):
* """Add a new student to the database"""
* conn = sqlite3.connect(DB\_PATH)
* c = conn.cursor()
* c.execute('''
* INSERT INTO students (
* name, source, response, destination,
* course\_of\_study, level\_of\_study,
* conversion\_probability, created\_at,
* email\_responses, call\_responses
* ) VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?)
* ''', (
* student\_data['name'],
* student\_data['source'],
* student\_data['response'],
* student\_data['destination'],
* student\_data['course\_of\_study'],
* student\_data['level\_of\_study'],
* student\_data.get('conversion\_probability', 0),
* datetime.now(),
* student\_data.get('email\_responses', 0),
* student\_data.get('call\_responses', 0)
* ))
* student\_id = c.lastrowid
* conn.commit()
* conn.close()
* return student\_id

[Image: Screenshot of the Student Profile View/Edit Interface]





### 4.1.1 Conversion Prediction Model Integration

* **Description:** Integrates the pre-trained machine learning model (Random Forest Classifier) to predict the likelihood of a student converting (i.e., successfully enrolling through the consultancy).

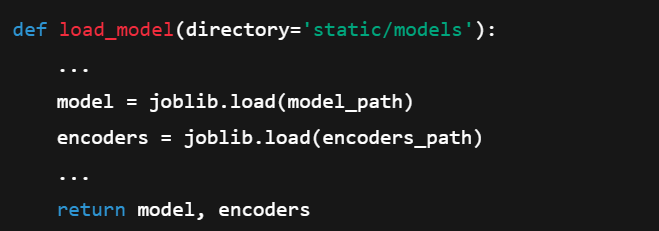
# Implementation Details:

* + **Model Loading:** The trained scikit-learn model is saved (e.g., using joblib or pickle) and loaded into the Flask application at startup or on demand.
  + **Feature Extraction:** When a prediction is needed (e.g., on the student profile page or dashboard), relevant features (like lead source, interaction frequency/type, academic level, preferred destination) are extracted from the database for that student.
  + **Prediction:** The predict\_proba() method of the loaded model is used to get the probability of conversion. This score is then displayed in the user interface, often alongside the student's details.

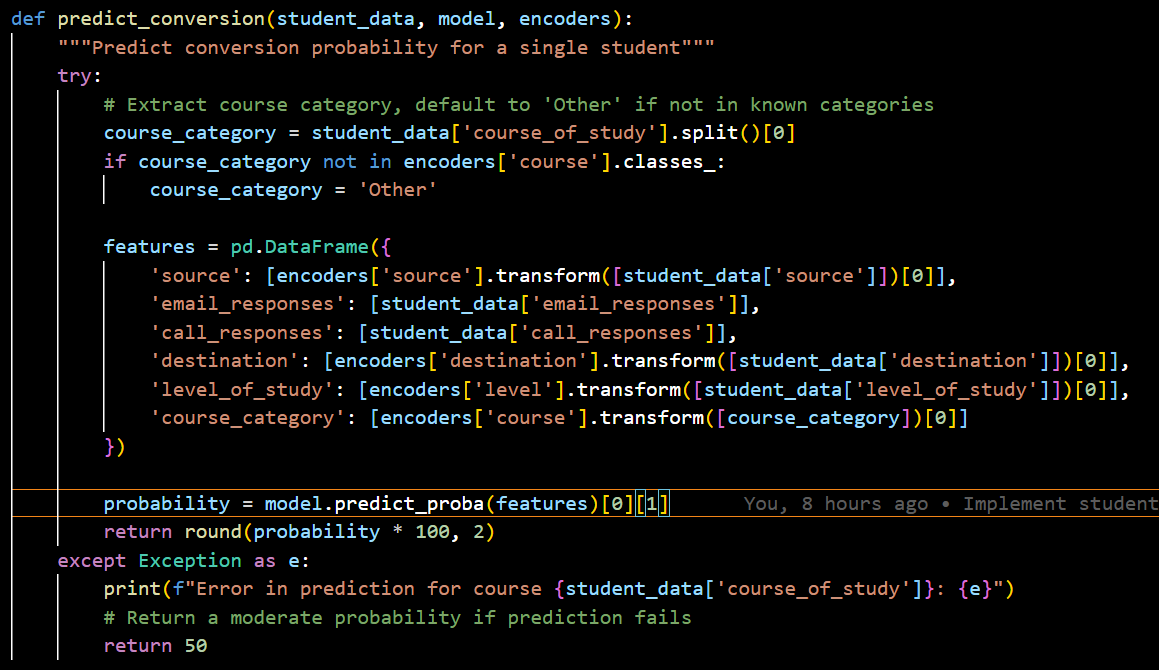
# Code (Illustrative Prediction Call):

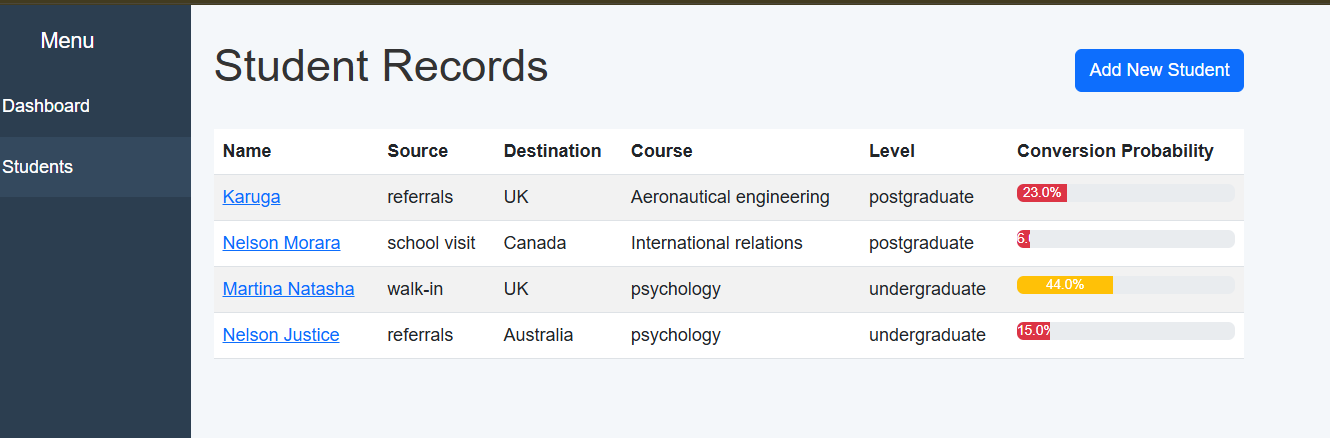
# 1. Model Loading

Model Loading: The trained scikit-learn model is saved (e.g., using joblib or pickle) and loaded into the Flask application

******

2. Feature Extraction & Prediction

****

* **User Interface:** The prediction score (e.g., "75% Likelihood to Convert") is typically displayed on the student's profile page or in summary tables.
* **Visual:** [Image: Screenshot showing Conversion Probability displayed on a Student Profile]  
  

## Algorithm Selection Rationale: Why Random Forest?

The selection of the Random Forest Classifier algorithm for predicting student conversion likelihood in EduConnect was based on several key considerations relevant to the nature of the data and the project goals:

* + - * 1. **Handling Mixed Data Types:** Student data typically includes a mix of numerical features (e.g., number of interactions, academic scores) and categorical features (e.g., lead source, preferred country, academic level). Random Forest naturally handles both types of data without requiring extensive preprocessing like

one-hot encoding for all categorical variables (though some preprocessing is still often beneficial).

* + - * 1. **Robustness to Non-linearity:** The factors influencing a student's decision to convert are unlikely to be purely linear. Complex interactions between variables (e.g., a specific combination of academic background and chosen destination might significantly increase conversion odds) are common. Random Forest, being a tree-based ensemble method, can capture these non-linear relationships and interactions effectively, which might be missed by simpler linear models like Logistic Regression.
        2. **Reduced Risk of Overfitting:** Compared to a single Decision Tree, which can easily overfit the training data by creating overly complex rules specific to individual data points, Random Forest mitigates this risk significantly. It achieves this through two primary mechanisms:

**Bagging (Bootstrap Aggregating):** Training multiple trees on different random subsets of the data.

**Feature Randomness:** Considering only a random subset of features at each split point in each tree. This ensemble approach leads to a model that generalises before to new, unseen student data.

* + - * 1. **Good Performance with Less Tuning:** Random Forest often achieves high accuracy "out-of-the-box" with default hyperparameters compared to other complex algorithms like Support Vector Machines (SVMS) or Gradient Boosting Machines (GBMS), which can require more careful and time-consuming tuning to optimise performance. This makes it a practical choice for achieving good results efficiently.
        2. **Implicit Feature Importance:** Understanding *why* the model predicts a certain conversion likelihood is crucial for consultants. Random Forest provides built-in metrics to estimate the importance of each feature (e.g., 'lead source', 'number of emails exchanged') in making predictions. This helps identify key drivers of conversion and provides actionable insights for the consultancy.
        3. **Resilience to Outliers:** The ensemble nature of Random Forest makes it relatively less sensitive to outliers in the data compared to some other algorithms.

While other algorithms like Logistic Regression (simpler, more interpretable but assumes linearity), Gradient Boosting (potentially higher accuracy but more sensitive to tuning), or SVMs could have been considered, Random Forest offered the best combination of predictive performance, robustness, ability to handle the specific data characteristics, ease of implementation, and provision of feature insights for the EduConnect CRM context.

### 4.1.2Additional Interfaces

* **Applications Tracking:**
  + *Description:* A module to monitor the status of university/college applications submifled for students (e.g., Submifled, Awaiting Decision, Offer Received, Accepted).
  + *Visual:* [Image: Screenshot of the Application Tracking Interface]

# Analytics and Reporting:

* + *Description:* Provides visualizations and reports on key performance indicators (KPIs), such as conversion rates by source, consultant performance, application success rates, and lead generation trends over time. Uses libraries like Matplotlib or Plotly integrated with Flask.
  + *Visual:* [Image: Screenshot of the Analytics/Reports Dashboard]

# Task Management:

* + *Description:* Allows consultants to create, assign, and track tasks related to student follow-ups, application deadlines, or internal processes. Often includes reminders or notifications.
  + *Visual:* [Image: Screenshot of the Task Management Interface]

## 4.2 Testing

Rigorous testing was conducted to ensure EduConnect meets functional requirements and performs reliably.

## 4.2 Types of Tests Performed

* **Unit Testing:** Focused on verifying individual functions and components in isolation. Examples include testing database interaction functions (add, retrieve, update), input validation logic, and specific calculations within the backend. Python's unittest or pytest frameworks were potentially used.
* **Integration Testing:** Tested the interaction between different modules. Examples include verifying that submitting the login form correctly authenticates and redirects, ensuring student data entered in the profile is correctly saved and retrieved, and checking that the ML model receives correct features and returns a prediction to the UI.
* **User Acceptance Testing (UAT):** Involved potential end-users (educational consultants) interacting with the system to perform realistic tasks. Feedback was collected on usability, workflow efficiency, and whether the system met their operational needs.
* **Performance Testing:** Assessed system responsiveness and stability under simulated load conditions (e.g., multiple users accessing the system concurrently) to identify potential bottlenecks, particularly concerning database access and ML model prediction times.
* **Security Testing (Basic):** Checked for common vulnerabilities like improper access control (ensuring users can only see their assigned data, if applicable) and potentially basic checks for injection flaws (though deeper security testing might be a future recommendation).

### Summary of Test Cases and Results

The following table provides a representative sample of test cases executed:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test Case ID** | **Module** | **Descripti on** | **Input Data** | **Expected Output** | **Actual Output** | **Status** |
| TC-LOG-0 1 | Login | Valid Login Credential s | Existing user email & correct password | User redirected to Dashboar d, success message | User redirected to Dashboar d, success message | Passed |
| TC-LOG-0 2 | Login | Invalid Login Credential s | Non-exist ent email or incorrect password | Remain on Login page, error message displayed | Remain on Login page, error message displayed | Passed |
| TC-STU-0 1 | Student Managem ent | Add New Student Record | Valid student details via form | Student record saved in DB,  success message | Record saved correctly, success message shown | Passed |
| TC-STU-0 2 | Student Managem ent | View Existing Student | Select a student from list | Student profile page displays correct details | Correct details displayed | Passed |
| TC-PRED- 01 | Prediction Model | Get Conversio n Prediction for a Student | Features of a sample student | A  probability score between 0% and  100%  displayed | Probability score (e.g., 68.5%)  displayed | Passed |
| TC-TASK- | Task | Create | Task | Task | Task | Passed |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 01 | Managem ent | New Task | details (title, due date, assignee) | saved and appears in task list | saved and listed correctly |  |
| TC-UAT-0 1 | Overall Workflow | Consultan t registers lead, logs interaction  , checks prediction | Simulated typical user workflow | Consultan t completes tasks efficiently | Minor usability feedback noted, workflow functional | Passed |

## 4.3 Project Appraisal

An objective evaluation of the EduConnect system reveals the following strengths and limitations:

### 4.2.3 Limitations

1. **Prediction Accuracy Dependency:** The effectiveness of the conversion prediction model is heavily reliant on the quality, completeness, and relevance of the historical data used for training and the input features provided for new predictions. Inconsistent data logging can reduce accuracy.
2. **Database Scalability:** Using SQLite, while simple for development, may pose performance limitations as the volume of data (students, interactions) grows significantly. It's less suitable for concurrent write operations compared to server-based databases (e.g., PostgreSQL, MySQL).
3. **Security Considerations:** The example implementation details (e.g., plain text password check in the snippet) might indicate areas needing security hardening, such as implementing robust password hashing and access control measures.
4. **Feature Scope:** The current version may lack certain advanced CRM features found in commercial products (e.g., automated email sequences, detailed marketing campaign tracking).

### 4.3.2 Strengths

1. **Tailored Solution:** EduConnect is specifically designed for the workflow of educational consultants, unlike generic CRMs.
2. **Centralized Information:** Provides a single platform for managing student data, interactions, applications, and tasks, reducing reliance on spreadsheets or disparate systems.
3. **Data-Driven Insights:** The integration of the conversion prediction model allows consultants to prioritize efforts based on data, potentially improving conversion rates. The feature importance provided by Random Forest enhances this.
4. **Improved Efficiency:** Streamlines common tasks like accessing student history, tracking application progress, and managing follow-ups.
5. **User-Friendly Interface:** Leverages Bootstrap for a clean, responsive, and relatively intuitive user interface.

### 4.3.3 Conclusions

The EduConnect CRM system has been successfully implemented, meeting its primary objectives of providing educational consultants with a dedicated tool for managing student relationships and leveraging predictive analytics using a Random Forest model. The key functionalities, including student profile management, interaction logging, application tracking, and conversion prediction, were developed and integrated. Testing confirmed that the system functions as intended, although areas for improvement and potential limitations were identified. The project serves as a valuable proof-of-concept and a functional tool for its target users.

### 4.3.3 Recommendations

Based on the implementation experience and project appraisal, the following recommendations are suggested for future development:

1. **Database Migration:** Plan for migration from SQLite to a more scalable, server-based database system (e.g., PostgreSQL or MySQL) to handle future growth in data volume and user concurrency.
2. **Enhance Automation:** Introduce features like automated email/SMS reminders for follow-ups, template-based communication, and automated workflows to further improve consultant efficiency.
3. **Refine ML Model:** Continuously evaluate and retrain the conversion prediction model with new data. Explore incorporating additional features or experimenting with different algorithms (e.g., Gradient Boosting, Neural Networks) to potentially improve accuracy. Implement MLOps practices for befler model management.
4. **Improve Data Quality Measures:** Implement stricter input validation and potentially tools or processes to encourage consistent and complete data entry,

which benefits both CRM usability and prediction accuracy.

1. **Security Hardening:** Conduct a thorough security audit and implement best practices, including robust password hashing (e.g., using Werkzeug security helpers in Flask), proper session management, input sanitization, and role-based access control.
2. **Expand Reporting:** Develop more sophisticated and customizable analytics

# References

1. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.  
   <https://www.the-modeling-agency.com/crisp-dm.pdf>
2. Sommerville, I. (2016). *Software Engineering* (10th ed.). Pearson Education.
   * A comprehensive guide on software development methodologies, requirements analysis, and system modeling.
3. Laudon, K. C., & Laudon, J. P. (2020). *Management Information Systems: Managing the Digital Firm* (16th ed.). Pearson.
   * Covers business needs for systems like CRMs and decision support systems.
4. Turban, E., Sharda, R., & Delen, D. (2018). *Decision Support and Business Intelligence Systems* (10th ed.). Pearson.
   * Explains how data-driven tools like predictive analytics are used in CRMs.
5. Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About Face: The Essentials of Interaction Design* (4th ed.). Wiley.
   * A foundational book on UI/UX design that supports your Figma design practices.
6. Clements, P., Bachmann, F., Bass, L., Garlan, D., Ivers, J., Little, R., ... & Stafford, J. (2003). *Documenting Software Architectures: Views and Beyond* (2nd ed.). Addison-Wesley.
   * Helps justify your logical and physical architecture decisions.
7. Hoffer, J. A., Ramesh, V., & Topi, H. (2016). *Modern Systems Analysis and Design* (8th ed.). Pearson.
   * Guides system modeling using use cases, DFDs, and ER diagrams.
8. Salesforce. (2021). *The State of CRM: How Customer Relationship Management is Evolving*. Salesforce.com.  
   <https://www.salesforce.com/resources/articles/what-is-crm/>
9. McKinsey & Company. (2020). *How advanced analytics boosts CRM performance*.  
   <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/>
10. ISO/IEC/IEEE 29148:2018. *Systems and software engineering — Life cycle processes — Requirements engineering*.

* International standard for writing and analyzing system requirements.