GymPulse: AI Analytics for Fitness Centers

MUHAMMAD BILAL

01-136221-014

ZIA UR REHMAN

01-136221-048

Supervisor

Dr. Faryal Nosheen

Bachelor of Science in Artificial Intelligence

Department of Computer Science

Bahria University, Islamabad

Dec, 2025

Certificate

We accept the work contained in the report titled "**GymPulse: AI Analytics for Fitness Centers**", written by Mr. Muhammad Bilal and Mr. Zia Ur Rehman as a confirmation to the required standard for the partial fulfillment of the degree of Bachelor of Science in Computer Science.

Approved by . . . :

Supervisor:

Internal Examiner:

External Examiner:

Project Coordinator:

Head of the Department:

Abstract

This research seeks to create and operationalize a system labeled **AI-powered Fitness Guidance and Analytics System** that will enable gym and fitness center owners to improve operational efficiency increase engagement among clients and enable enhanced decision-making. Leveraging sophisticated machine learning algorithms known as **XG-Boost** and other complementory predictive techniques the system seeks to estimate sales on a monthly basis detect members who may discontinu servic and profile ones behavior relative to their attendance, demographics, and subscription history. A consolidated web interface constructed using **Next.js** and **Tailwind CSS** supplies users with operational dashboards that offer predictive analytics, visual dashboards, and alerts on members at risk of disengagement. These functionalities plug into the overall objective of supporting users in their pursuit of evidence-based interventions. This is achieved within the systems architecture through providing interpretability of the models in a detailed manner such as through visualizing features that matter most as well as providing users with predictive analytics at a models output. **The system recorded 85% accuracy in sales prediction and 0.88 AUC score in predicting members that were at risk of disengaging.** It is indeed the user experience fused with machine learning and analytics that demonstrates evidence-based real-time decision making and frictionless interface that the analytics are designed to ease in the first place the power of AI in the fitness analytics space.

Acknowledgment

With all the sincerity we appreciate the inestimable input of our esteemed supervisor, **Dr. Faryal Nosheen** provided invaluable guidance, encouragement, and constant support throughout the development of this project. From the very beginning, her professional insight and constructive feedback played a significant role in shaping the direction of our research and ensuring its successful completion. Her patience, motivation, and clarity helped us overcome challenges and refine our ideas into a meaningful academic contribution. We sincerely appreciate her timely feedback, dedication, and continuous support, which greatly contributed to our academic growth and deeper understanding of the field. Without her professional advice and mentorship, this project would not have achieved its present form.

MUHAMMAD BILAL & ZIA UR REHMAN
Bahria University
E-8, Islamabad, Pakistan

Dec, 2025

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AUC	Area Under the Curve
BI	Business Intelligence
CI/CD	Continuous Integration and Continuous Deployment
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
DFD	Data Flow Diagram
ERD	Entity Relationship Diagram
ETL	Extract, Transform, Load
FAQ	Frequently Asked Questions
GUI	Graphical User Interface
HLD	High-Level Design
HTTPS	Hypertext Transfer Protocol Secure
IoT	Internet of Things
JSON	JavaScript Object Notation
JWT	JSON Web Token
KPI	Key Performance Indicator
LIME	Local Interpretable Model-agnostic Explanations
LLD	Low-Level Design
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
MVP	Minimum Viable Product
OOD	Object-Oriented Design
OWASP	Open Web Application Security Project
RBAC	Role-Based Access Control

Chapter 1

Introduction

Over the past few decades, the fitness industry has experienced unprecedented growth, driven by increased health awareness, rising concerns about sedentary lifestyles, and advancements in digital fitness technologies. This growth has surpassed that of previous decades and has resulted in a significant increase in demand for fitness services. Despite this expansion, many fitness centers and gyms continue to operate using legacy management software. Such systems typically offer limited functionality, focusing primarily on basic scheduling, member record management, and payment processing.

Most importantly, these traditional management systems lack the capability to leverage predictive analytics. As a result, gym owners are unable to efficiently optimize operational processes or proactively enhance member engagement. The absence of sales forecasting, churn prediction, and real-time analytical insights leaves fitness businesses constrained to reactive decision-making. Consequently, the industry remains operationally limited in its ability to address member attrition and retention challenges effectively.

GymPulse: AI Analytics for Fitness Centers is designed to address these limitations by integrating advanced machine learning (ML) and artificial intelligence (AI) techniques into fitness management. The platform utilizes predictive models, including XGBoost and neural networks, to generate actionable insights for gym owners. These insights support improved member retention, enhanced sales performance, and increased automation in membership management processes.

1.1 Motivation

The motivation for this research stems from the significant gap between the growing availability of operational data in fitness centers and the limited use of AI-powered business intelligence tools to analyze this data. Existing gym management systems lack the ability to seamlessly integrate predictive engagement workflows and member analytics.

As fitness centers generate large volumes of data—such as attendance records, invoices, and member profiles—owners often become "data-rich but information-poor." From a managerial perspective, this results in several critical challenges:

- Inability to predict revenue cycles accurately.
- Inability to identify at-risk members prior to disengagement.
- Inability to optimize retention strategies through targeted engagement.

By applying AI-driven analytics to membership data, sales forecasts, and churn risk assessment, **GymPulse** enables gym owners to transition from reactive management to a proactive, data-driven strategic approach. This positions GymPulse as a suitable solution for augmenting traditional fitness business management systems.

1.2 Problem Statement

The current fitness industry lacks predictive capabilities within most gym management software. Existing systems primarily focus on managing historical data related to scheduling and payment processing, while predictive functionalities such as sales forecasting and churn prediction remain largely absent. Without predictive insights, gym owners are unable to anticipate revenue fluctuations or identify members likely to discontinue their memberships.

In many cases, revenue losses and member attrition are recognized only after they have already occurred. The lack of early warning mechanisms significantly limits the effectiveness of retention efforts. Consequently, opportunities for growth and operational optimization are frequently missed.

GymPulse addresses these challenges by leveraging AI algorithms, particularly XGBoost, to predict future sales trends with a target accuracy exceeding 85%. Additionally, the platform's AI-driven churn prediction module identifies at-risk members prior to cancellation, enabling timely intervention and improved operational efficiency.

1.3 Research Objectives

The primary objectives of this research are as follows:

- **Objective One:** Implement a machine learning-based sales forecasting model with an expected accuracy of 85% or higher, enabling gym owners to estimate revenue and plan budgets effectively.
- **Objective Two:** Develop an AI-driven churn prediction system that identifies members at risk of cancellation and enables targeted retention strategies.

- **Objective Three:** Consolidate business intelligence features into a unified dashboard providing real-time insights into sales, churn, and member engagement.
- **Objective Four:** Implement an automated, AI-based member engagement system that delivers personalized retention campaigns based on behavioral analysis.

1.4 Scope of the Study

1.4.1 In-Scope

This research focuses on integrating AI-based Software-as-a-Service (SaaS) functionality within the fitness sector. The project includes the following components:

- **Sales Prediction:** Development of forecasting models using historical sales data, seasonality trends, and machine learning algorithms such as XGBoost and Random Forest.
- **Churn Prediction:** Construction of predictive models utilizing attendance patterns, payment history, and behavioral indicators to assess disengagement risk.
- **Business Intelligence Dashboard:** Design of a real-time dashboard for monitoring sales, membership trends, and engagement metrics.
- **Automated Engagement:** Implementation of AI-driven personalized retention campaigns based on churn risk analysis.

1.4.2 Limitations and Exclusions

The following elements are excluded from the project scope to ensure feasibility within defined constraints:

- **Hardware Integration:** No integration with IoT devices, biometric scanners, or automated access systems.
- **Financial Processing:** The system monitors sales data but does not process real-time banking transactions or integrate with payment gateways.
- **Operational Management:** Non-analytical operational tasks such as equipment maintenance, staff scheduling, and facility supervision are excluded.

1.5 Significance of the Study

This research addresses a critical technological gap in the fitness industry by introducing advanced forecasting and predictive analytics for small-to-medium enterprise (SME) gym owners. Integrating AI into gym management enables data-driven decision-making, reducing reliance on intuition alone.

The development of **GymPulse** enhances operational efficiency, reduces member churn, and improves revenue forecasting accuracy. Furthermore, this study demonstrates the applicability of machine learning within specialized business domains, contributing to the broader intersection of AI and enterprise management.

1.6 System Overview

To provide a visual representation of the proposed solution, Figure 1.1 illustrates the high-level business processes within GymPulse. The diagram outlines data collection, predictive processing through forecasting and risk assessment modules, and the execution of engagement strategies to drive member activity.

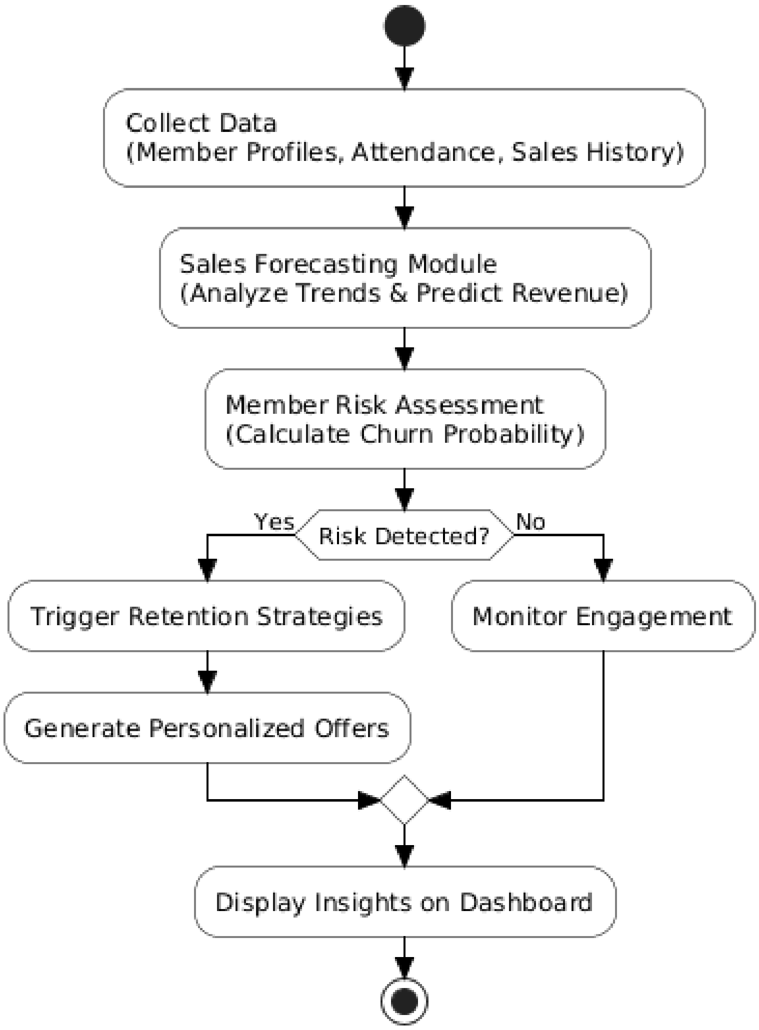

Figure 1.1: High-Level Business Process Flow of GymPulse

Structure of the Thesis

This thesis is organized as follows:

- **Chapter 2: Literature Review** examines existing research on fitness management systems and identifies gaps in predictive analytics and automation. It reviews prior studies on machine learning applications in fitness, business intelligence systems, and customer retention strategies. This chapter establishes the theoretical background and justifies the need for an AI-driven fitness analytics platform. It also highlights technological limitations in current commercial solutions.
- **Chapter 3: Design Specifications** defines the functional and non-functional requirements, system boundaries, and user roles. It outlines system constraints, performance expectations, and security considerations. This chapter serves as a blueprint for the design and implementation phases of the project. The documented requirements guide system validation and testing.
- **Chapter 4: System Design** presents the architectural design, UML diagrams, and database schema. It explains the system's layered architecture, data flow, and component interactions. The chapter also details the design rationale behind key modules and interfaces. These design decisions ensure scalability and maintainability.
- **Chapter 5: System Implementation** details the development of the frontend, backend, and machine learning models. It describes the technologies, frameworks, and programming tools used in system construction. The chapter also discusses model training, deployment, and integration into the application workflow. Implementation challenges and solutions are highlighted.
- **Chapter 6: System Evaluation** discusses testing methodologies, performance metrics, and model evaluation results. It covers unit testing, system testing, and load testing procedures. The chapter further analyzes model accuracy, prediction reliability, and system responsiveness under real-world conditions. Evaluation outcomes validate system effectiveness.
- **Chapter 7: Conclusion and Future Work** summarizes the research outcomes and key findings of the study. It discusses the limitations encountered during development and evaluation. The chapter also proposes potential future enhancements to extend system functionality and performance. Recommendations for future research are provided.

Chapter 2

Literature Review

2.1 Current State of Fitness Industry Technology

In recent years the fitness industry has incorporeted more digital technologie and the management of gyms has transformed more than any other area. Indeed many fitness facilities have some form of management software of the administrative systems that have been digitally automated most remain unintegrated and functionally limited. Furthermore it accomplish little more than membership registration and maintaining billing, scheduling, and class management.

The operational data generated every business day is a good and better resource of most fitness facilities especially in the areas of member behavior, prediction of revenue, and churn risk analytics. So most gym owners are unable to perform the analysis of operationel data to develop a more detailed and actionable strategic plan.

The inability to perform any of these operational analysis is the absence of predictive analytics which are essential for the prediction of variables like future revenue, member retention, and engagement in different seasons. Management is then left with a staring contest with the past instead of data.

Only a scant few fitness facilities have successfully embedded data analytics into their operational processes (see e.g [14]). This points out a significant technologicel disparity on the use of AI learning in the domain. Having only a responsive approach leads to lost chances for value creation and for staying ahead of the compatition in attracting and retaining members.

2.2 Identified Research Gaps

To this day few if any tools exist that levarage artificial intelligence to enhance ones business performance even with the fitness industrys rapid expansion. This research analyzes the most pivotal gaps in existing systems of gym management and illustrates the ways in which **GymPulse** is able to fill them through data and artificial intelligence.

2.2.1 Limited Predictive Analytics in Fitness Management

Gap: Even though there is more data regarding fitness activities today than there has ever been before most business management platforms still cannot project key performence indicator. Most legacy systems are passive, as they track only past performance and do not provide any form of prediction. As a result gym owners can not anticipate changes in the market, drops in revenue, or shifts in member utilizetion.

The Solution: GymPulse solves this problem using *XGBoost* prediction algorithms. Using historical data sets such as attendance, payment history, and seasonality, the platform can predict sales with unparalleled accuracy and assist in more effective financial forecasting and operational streamlining.

Table 2.1: Research Gap: Predictive Analytics

Limitations of Existing Systems	**GymPulse Solution**
Absence of predictive sales or behavior analysis.	Implementation of XGBoost models for robust sales forecasting.
Reactive decision-making based solely on past data.	Predictive forecasting for sales and churn with >85% accuracy.
Inability to account for seasonal trends.	Advanced seasonal trend analysis for future revenue prediction.

2.2.2 Reactive vs. Proactive Member Retention

Gap: Most standerd gym systems typically only recognize churned members when they have been cancelled. This by default is not helpful when prevention is needed. When a membe has left the gym the attempts made to try and get them to return fiz out and costs continue to rise.

Proposed Solution: GymPulse customers use AI churn prediction to identify members who may cancel before they actually do. Using attendance, payment, and engagement of members, the systems assign a score to customers in order to help gym owners start retention campaigns in a timely fashion. When owners actively start retention campaigns they increase the customer lifetime value.

2.2.3 Lack of Integrated Business Intelligence

Gap: Fitness businesses frequently utilize disintegrated software ecosystems in which sales operational, and member information remain siloed. In order to gain any insights from this

Table 2.2: Research Gap: Member Retention Strategies

Limitations of Existing Systems	**GymPulse Solution**
Identification of churn only post-cancellation.	Proactive identification of at-risk members prior to churn.
Generic, reactive retention attempts.	Personalized, AI-driven retention campaigns.
Limited tools for early intervention.	Timely intervention strategies to mitigate churn risk.

data businesses must perform manual data consolidation which creates inefficiencies and further gaps in understanding the overall picture of the business.

Proposed Solution: GymPulse offers an integrated Business Intelligence (BI) dashboard that pulls together all the data points from disparate sources and structures them in different views that can be accessed in one application together. This provides instantaneous access to sales membership and operational data for comprehensive real-time insights.

Table 2.3: Research Gap: Business Intelligence

Limitations of Existing Systems	**GymPulse Solution**
Disconnected tools for various business functions.	Unified dashboard integrating all critical data streams.
Manual data aggregation and reporting.	Automated, real-time data analysis and reporting.
Fragmented operational visibility.	Holistic view combining sales, membership, and engagement metrics.

2.2.4 Insufficient Data-Driven Decision Support

Gap: No current system has advanced tools that analyz deep data. Reporting exists but decision support systems that utilize machine learning to improve model selection and optimiz accuracy are non-existent.

Proposed Solution: GymPulse presents an automated model comparison framework. The feature assesses different machine learning models in real time to determine the most accurate model for defined tasks like sales prediction or risk scoring. The system also employs hyperparameter tuning to guarantee the models peak efficiency.

2.2.5 Missing Automation in Member Engagement

Gap: Getting to engage with members who have signed up to the gym is traditionaly as time-consuming as it is difficult. More and more members sign up to the gym, but tailored emails and messages are quickly renderad impossible as communicating on an

Table 2.4: Research Gap: Decision Support Tools

Limitations of Existing Systems	**GymPulse Solution**
Basic, static analytics.	Advanced decision-making tools utilizing machine learning.
Manual or non-existent model selection.	Automated model comparison and hyperparameter tuning.
Reliance on intuition or outdated methods.	Data-driven predictions using optimized algorithmic models.

individual level is far too time-consuming. This leads to generic automated messages that are irrelevant and useless to each individual gym member.

Proposed Solution: GymPulse system automatically engages with members by utilizing AI to read and recommend tailored engagement documentation, exercise regimens, and promotional messages incentivizing continued engagement based on an analyses of the members profile activity and risk. **GymPulse** minimizes time spent on member engagement and administrative work by optimizing contact to reach member retention automation while also providing the users a more relevant personalized experience.

Table 2.5: Research Gap: Member Engagement Automation

Limitations of Existing Systems	**GymPulse Solution**
Manual, unscalable communication methods.	AI-driven, automated engagement workflows.
Generic, one-size-fits-all messaging.	Hyper-personalized recommendations based on behavioral data.
Time-consuming and prone to human error.	Efficient, scalable, and consistent engagement system.

2.3 Visualizing the Research Gap

To help understand the contribution of this study Figure 2.1 shows the overlaap of the current solutions with the gaps they address.

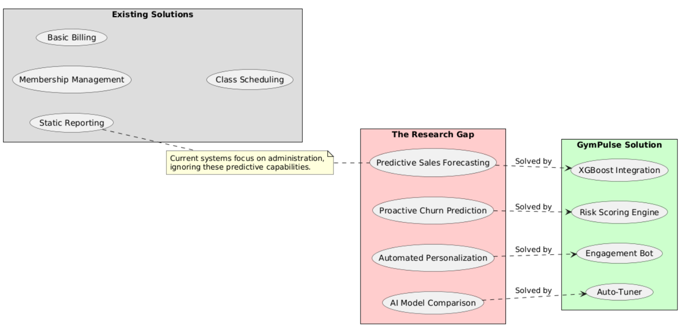

Figure 2.1: Research Gap Analysis: Intersection of Current Solutions and Identified Needs

Figure 2.2 This figure shows the workflow of the proactivee churn prediction mechanism and shows the transition from reactive to proactivee management.

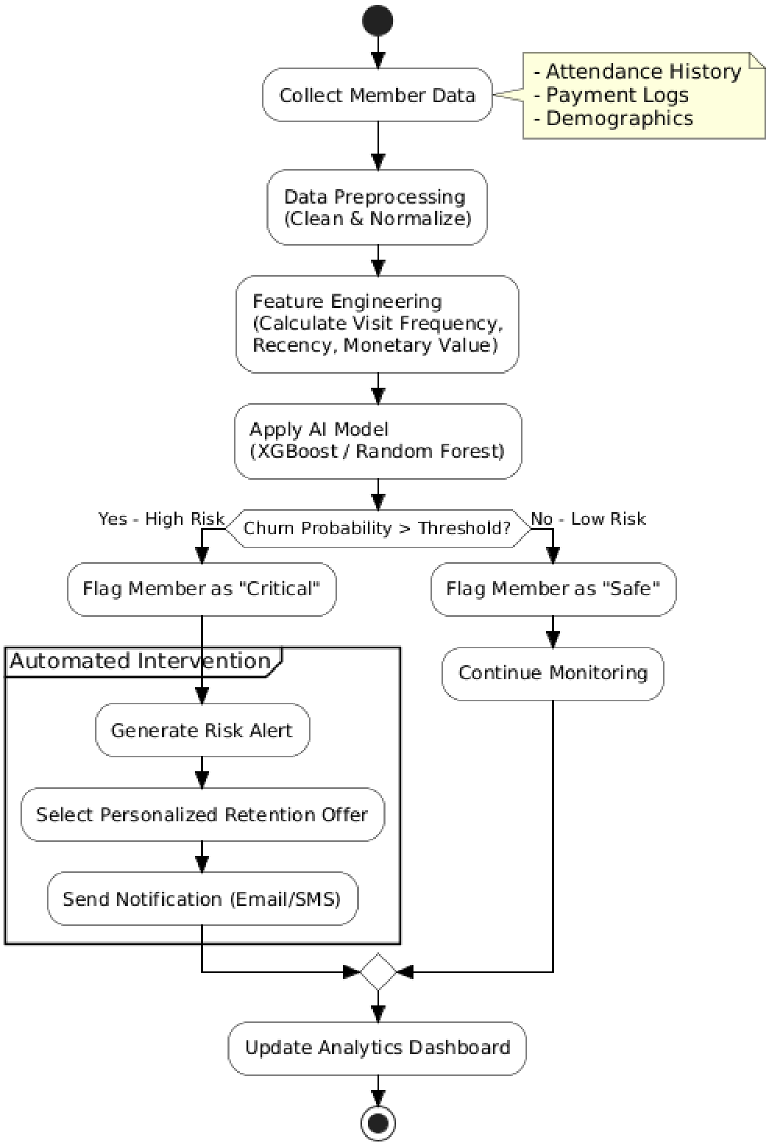

Figure 2.2: AI-Driven Churn Prediction Process Flow

2.4 Research Contributions

This study advances the application of artificial intelligence in fitness management through the following contributions:

- **Innovative Deployment of XGBoost:** This research is among the first to apply XGBoost models specifically within the fitness management domain for sales prediction. The model leverages historical transaction data, seasonal trends, and attendance patterns to generate accurate revenue forecasts. By demonstrating high predictive accuracy, the study validates XGBoost as a reliable tool for fitness business analytics. This contribution extends the practical applicability of gradient-boosting techniques to a previously underexplored industry. It also provides a reproducible framework for future research and commercial adoption.

- **Preventive Risk Management:** The system introduces a predictive AI-based approach to member retention by identifying churn risk before cancellation occurs. Behavioral indicators such as attendance frequency, payment history, and engagement levels are analyzed to generate risk scores. This enables gym owners to intervene proactively rather than reactively. By shifting retention strategies toward early prevention, the system helps increase customer lifetime value and reduce revenue loss.

- **Integrated Business Analytics:** This research delivers a unified business intelligence platform that consolidates data from multiple operational sources. Sales, membership, and engagement metrics are presented through interactive dashboards in real time. The integration eliminates data silos and reduces the need for manual reporting. As a result, decision-making becomes faster, more accurate, and evidence-based.

- **Intelligent Automation of Engagement:** The proposed system automates member engagement using AI-driven personalization techniques. By analyzing behavioral data and churn risk profiles, the platform recommends targeted communication strategies and retention campaigns. This reduces administrative workload while ensuring relevant and timely interaction with members. Automated engagement improves consistency and scalability compared to manual approaches.

2.5 Summary

This chapter examined the current technological landscape of the fitness industry, highlighting the limitations of existing gym management systems. While digital tools are widely adopted, most solutions remain reactive, fragmented, and limited to basic administrative functions. The review identified key research gaps, including the lack of predictive analytics, proactive member retention strategies, integrated business intelligence, and intelligent automation.The also chapter demonstrated how **GymPulse** addresses these gaps through the application of artificial intelligence and machine learning techniques. By enabling predictive sales forecasting, early churn detection, unified analytics, and automated engagement, GymPulse provides a data-driven solution aligned with the evolving needs of modern fitness facilities. This analysis establishes the foundation for the system design and implementation discussed in subsequent chapters.

Chapter 3

Requirement Specifications

3.1 Analysis of Existing Systems

This is because the analysis of existing systems involves problem solving. Look at the way most of the gyms are operated today and it is quite a mess. The scenery is characterized by handwork and haphazard record keeping. Though these outdated approaches may be alright with simple tasks such as the check-in of people, they fare much better when attempting to apply modern data science to the actual expansion of the business. In this section, we will dissect precisely in what areas these existing systems are failing which will clarify why we should have a superior solution.

3.1.1 Weakness of Traditional Approaches

The use of detached software and manual spreadsheets does present some rather serious bottlenecks. Here are the big ones:

- **Operational Inefficiencies via Manual Data Handling:**item Operational Inefficiencies through Manual Data Handling: Small business administrators tend to be left with no choice but to type in sales transactions, attendance records, and financial information manually. This is not simply monotonous labor, it is dangerous. By having humans key in data manually, errors occur and that undermines the data integrity.
- **Missing of Predictive Analytics:** Most of the old systems are living in past. They can provide you with a backward perspective of what transpired last month yet they cannot inform you on the prospect. In the absence of forecasting services, business owners are operating in the dark, they are unable to predict the sales trends or identify revenue opportunities before they lapse.
- **Reactive Churn Management:** Typically a gym does not get to know a member has stopped until the cancellation has been completed. It lacks an early warning system.

This implies that the owners would be losing the opportunity to intervene with the at-risk members resulting in revenue loss that would have been avoidable.

- **Limited Business Intelligence (BI):** The reporting packages in these systems are rather static. They present you with the existing situation, of course, but they do not include practical ideas. In such a way, strategic decisions are made more of a gut feeling, rather than grounded in hard evidence.
- **Missing of Machine Learning Integration:** Old platforms do not typically apply Artificial Intelligence (AI) to identify behavioral patterns. This does restrict their options to optimizing resources or even personalizing the member experience in any significant measure.
- **Generic Risk Assessment:**Retention strategies tend to be general and crude. They make no distinction between people not considering the individual behavior or individual risk.
- **Latency in Data Analysis:** Analytics are usually produced on a periodic basis perhaps once a month. This lag kills agility. You may find a problem and by the time you read a monthly report it may be too late to rectify it in point of time.

Figure 3.1 gives you a visual of this architecture. You can clearly see the manual data flow and the complete lack of an automated feedback loop.

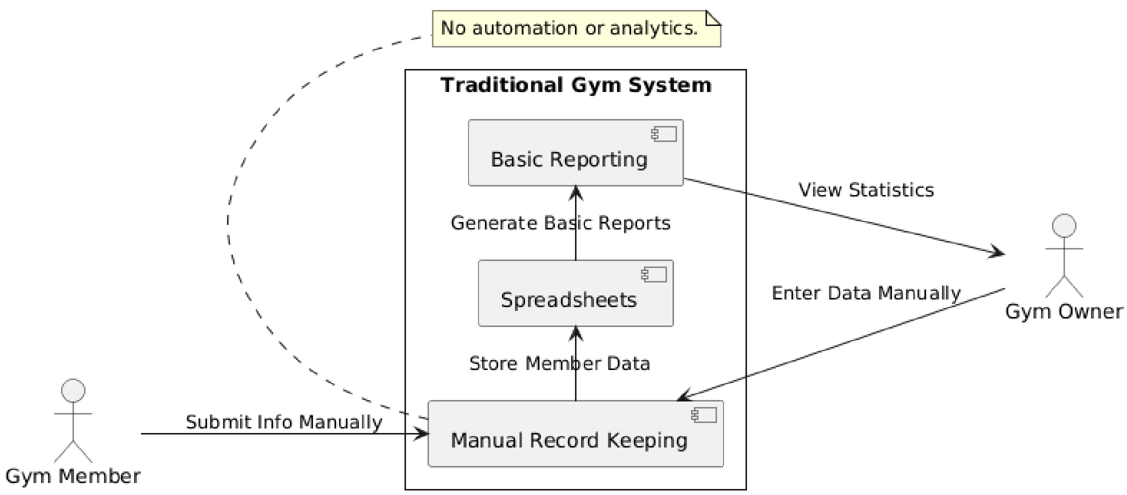

Figure 3.1: Architecture of Existing Traditional Systems

3.2 Proposed System: GymPulse

In order to overcom these shortcomings, we cam up with GymPulse. It is a (AI-based) Fitnes Business Management Platfor that combine modern machine learning algorith namely the XGBoos regressio models to enable sales forecastin to be intelligent, churn prediction to be proactiv, and busines intelligenc to be comprehensiv.

3.2.1 System Capabilities and Key Improvements

GymPulse brin autmation and smrts to the gym management process intrducing a number of improvments:

- **Automated Data Ingestion:** The systemm is an automaticc procesing of raw gym datasets (CSV). It consume sales demographi and attendanc data which guarante the accuracyy of data and reductionn of administrativee patience.
- **Predictive Sales Forecasting:** With XGBoost regresion moodels the systemm will be able to predict the future revenuee streams given old transaction analysis with a up level of accuracyy.
- **Proactive Risk Assessment:**An AI-basedd enginee will be continualy asessing the behavior of membeers such as how often they attend or pay to determinee high risk memberr before they churnn.
- **Real-Time Analytics Dashboard:**It providees you with a dynamicc user interface that provides you with real-timee visualizationn of Key Performnce Indicatorss (KPIs), model acuracy measurements, and sales trendis.
- **Automated Model Selection:** The systemm comprizes a comparetive analysis module. It also comperes various algorithms (Random Forest, Linear Regression, XGBoost) to ensure that the deployeed model is the most accurate model.
- **Targeted Intervention Strategies:** The system automaticaly provides individuelized retantion offers and engagement strategies based on risk profiling.
- **Scalable Architecture:** The application is devaloped based on the Next.js v18.16.0 framework and TypeScript as based on high performanc, typ safety and scalabilityy preparing to scale with the fitnes entarprise.

Figure 3.2 shows the up-level architectur of GymPulse, it is showing how is the AI Engine integrtes with the web application layar.

To clarify the operational logic a quite more, Figure 3.3 show the sequential flow of data and ingestion all the way to generating actionabl insights.

3.3 System Modeling and Specifications

This sectionn breaks down the behaviorall models of the systemm. It defines how users interect with the platform and tracks the lifecyclee of data objects.

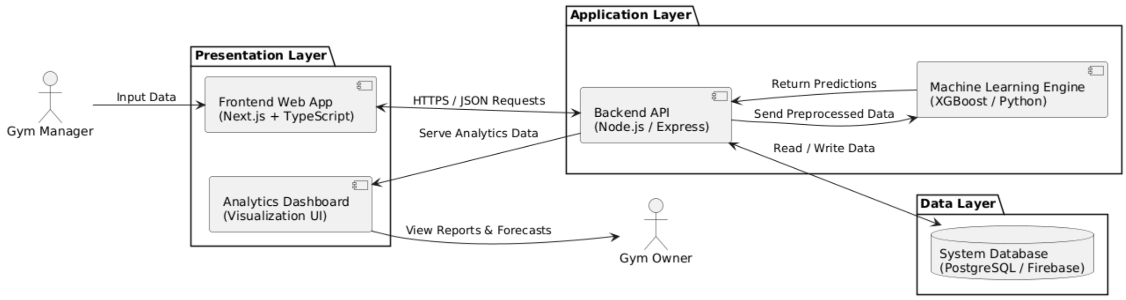

Figure 3.2: Proposed System Architecture of GymPulse

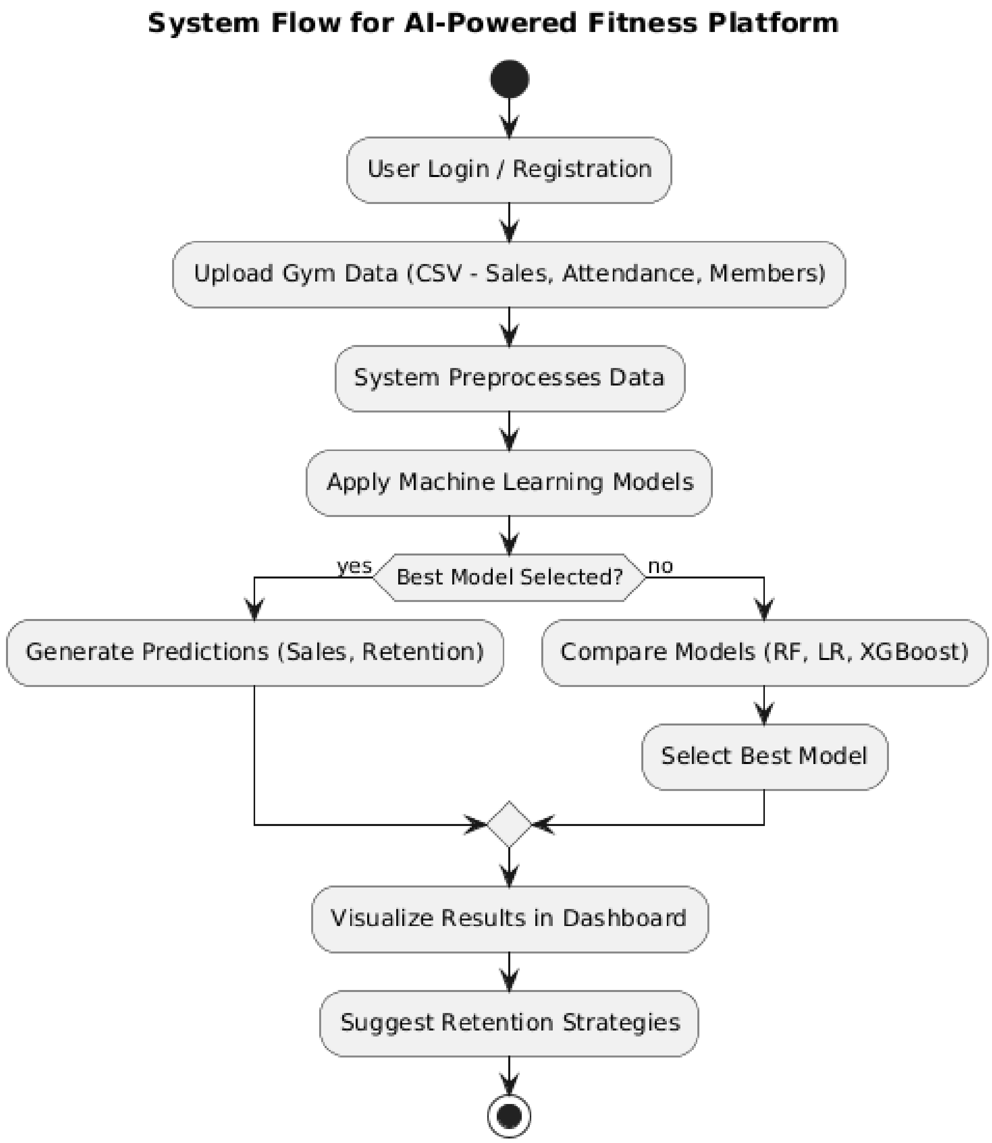

Figure 3.3: System Data Flow Diagram

3.3.1 Use Case Modeling

The Use Case diagramm (Figure 3.4) defines the intaractions between the primary actorz Gym Owners, Managers, and Membars and the system. Key use cases includ things like uploading data, viewing sales forecests, and checking retantion risks. The diagramm illustrates relationships such as the Gym Owners abilityy to access high level analyticcs and the Admins role in user management.

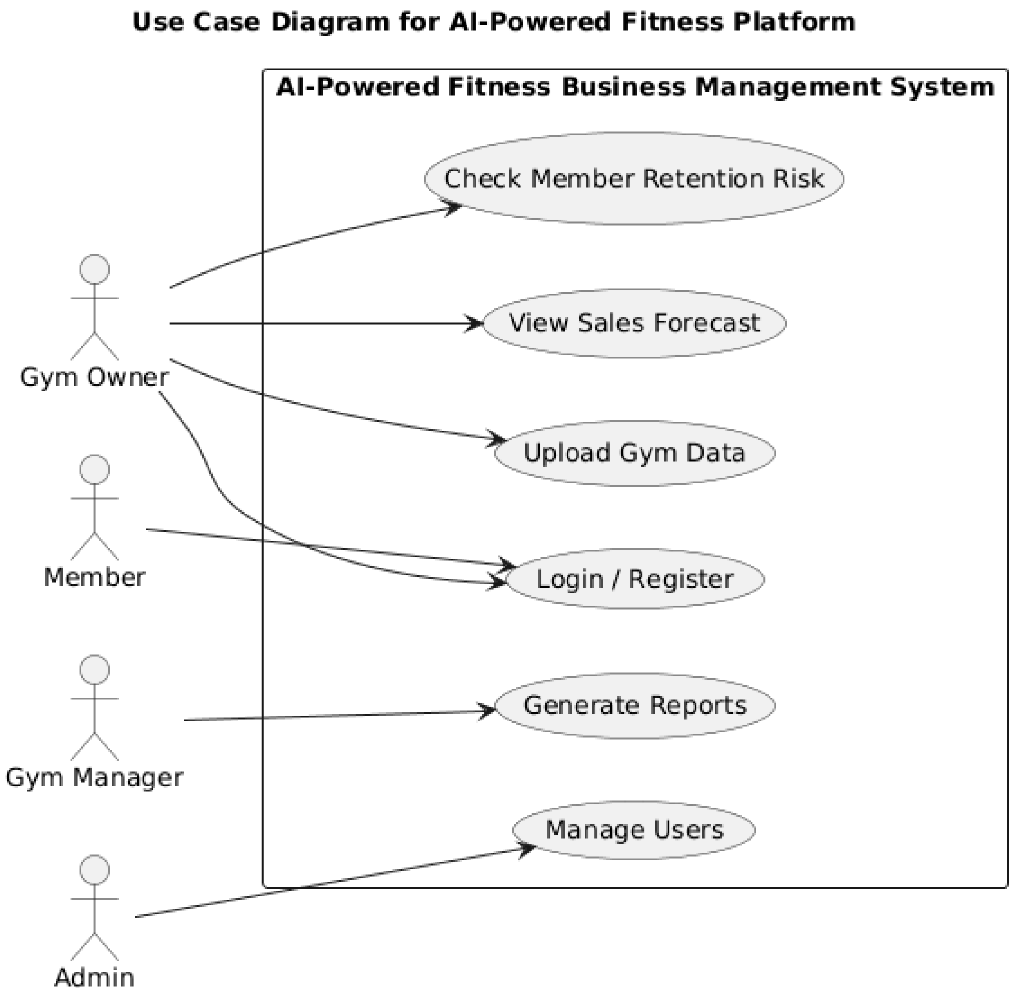

Figure 3.4: Use Case Diagram showing Actor-System Interactions

3.3.2 Member Lifecycle State Modeling

To define the business rules for member retention, the State Machine Diagram (Figure 3.5) illustrates the transitions of a member's status. It visualizes how an Active member may transition to High Risk based on AI predictions, which then trigger targeted retention offers. From there, the member may either return to an Active state or ultimately churn. The diagram clearly represents the conditions and events that govern each state change, ensuring consistency in system behavior. It also supports decision-making by aligning automated retention strategies with real-time member engagement patterns.

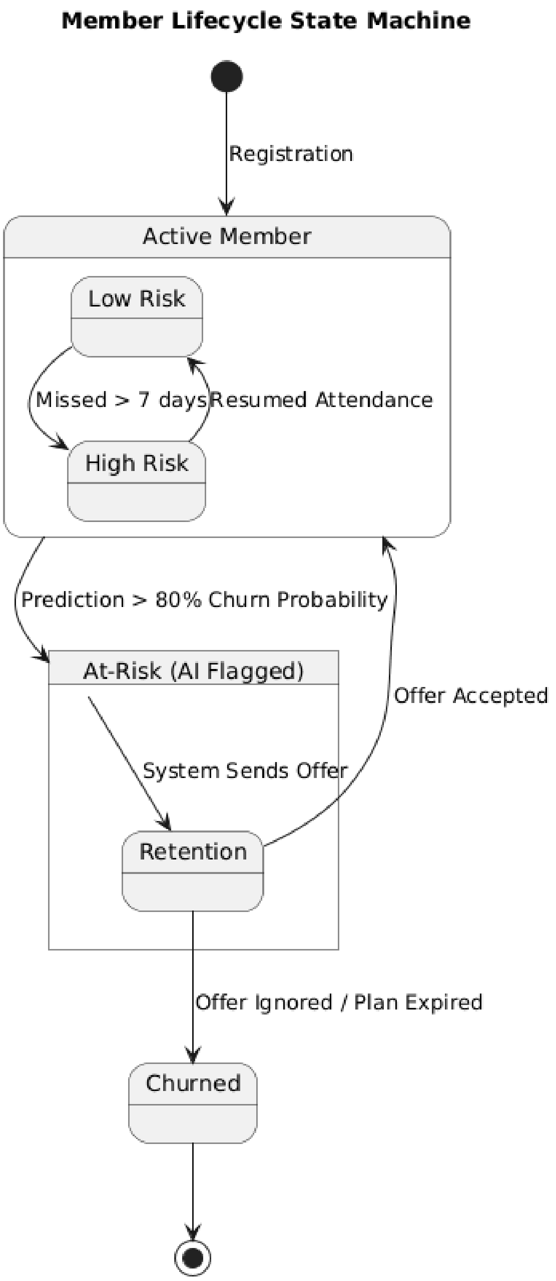

Figure 3.5: State Machine Diagram: Member Lifecycle and Risk Transitions

3.3.3 System Context

The System Context Diagram (Figure 3.6) delineates the boundaries of the GymPulse system. It identifies external entities such as the Gym Owner, Members, and Administrators, showing the flow of inputs (data uploads) and outputs (analytics and reports) between the system and its environment.

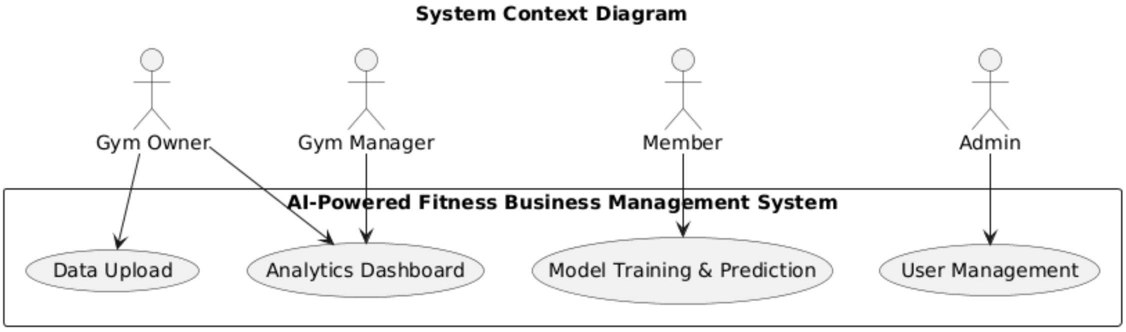

Figure 3.6: System Context Diagram

3.3.4 Data Flow Modeling

The Figure 3.7, provides a high level view of data movement within the system. It traces the flow of raw CSV data from the *Data Collection Module* through preprocessing and the *Prediction Engine*, and finally to the *Visualization Dashboard* and database storage.

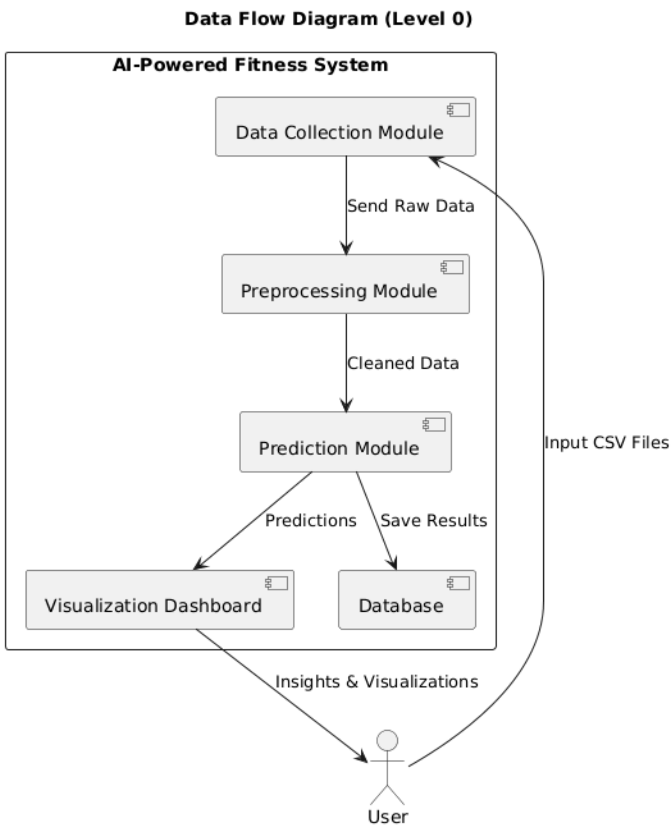

Figure 3.7: Data Flow Diagram (Level 0)

Chapter 4

Design

4.1 System Architecture

The GymPulse architecture is designed to facilitate gym management through data optimization and predictive analytics. Rather than functioning solely as a record-keeping system, the platform processes operational data to generate actionable insights. The system utilizes a microservices-based architecture to ensure modularity. This structure supports independent scaling and maintenance of distinct functional areas, including user management, sales forecasting, and retention analysis, thereby avoiding the limitations of a monolithic codebase.

The system adheres to a standard client-server model, divided into two logical layers:

1. **Presentation Layer (Frontend):** Developed using **Next.js** and **TypeScript**. This layer prioritizes type safety and responsiveness to minimize runtime errors and maintain interface consistency.

2. **Application & Data Layer (Backend):** This layer executes core business logic. A **Python-based API** manages communication between the client interface and the XGBoost inference engine. Data persistence is handled through a hybrid approach using **Firebase** and **SQL** to maintain relational integrity while allowing for flexible schema adjustments.

Figure 4.1 illustrates the high-level interaction between these components, mapping the data flow from user interaction on the dashboard through backend processing to database storage. It highlights the separation of concerns between the presentation and application layers. This architectural view clarifies how data is securely transmitted, processed, and persisted to support analytics and decision-making.

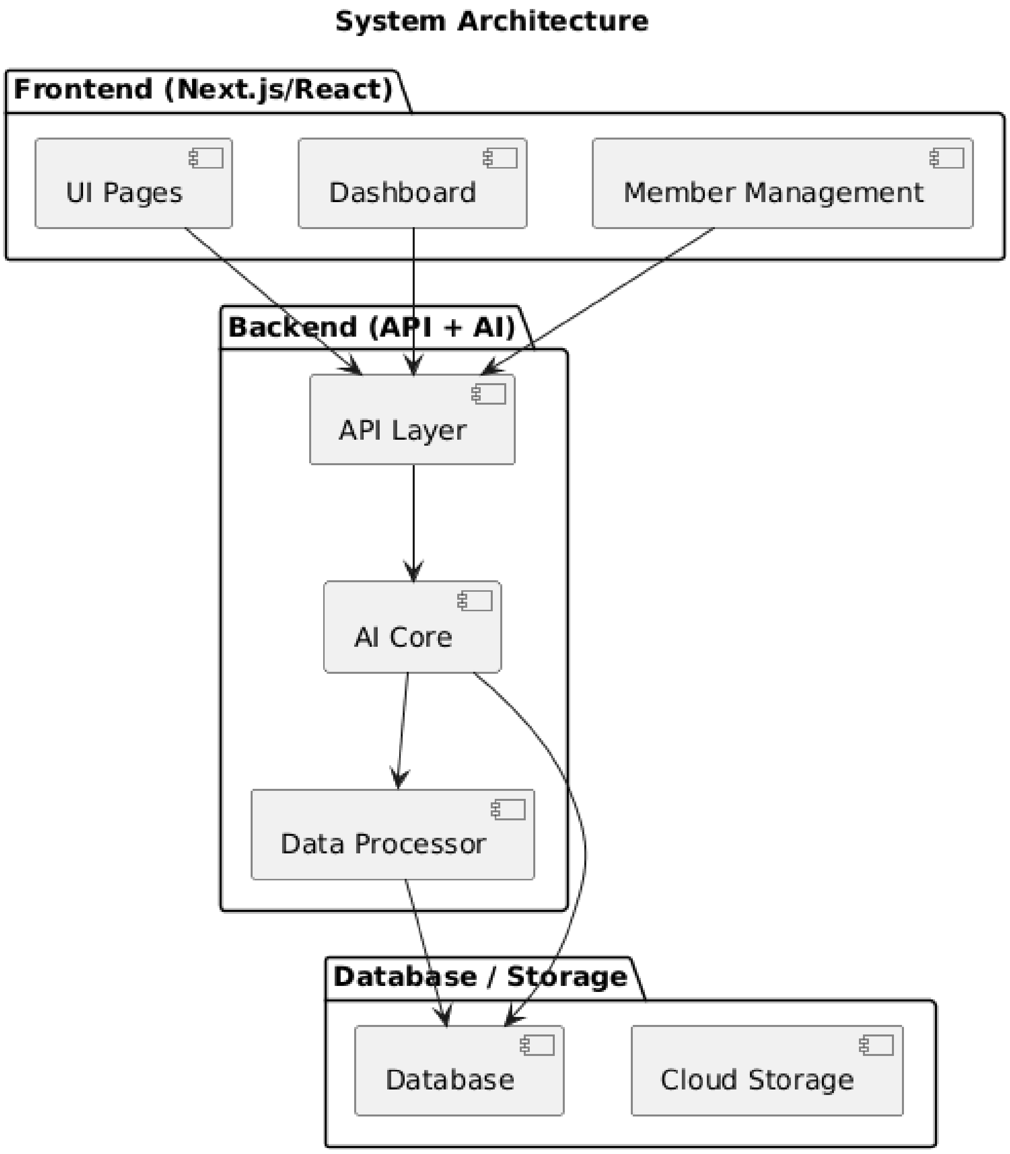

Figure 4.1: System Architecture Diagram

4.2 Design Constraints and Trade-offs

System design involves balancing conflicting technical requirements. The development process required managing specific constraints to optimize performance, security, and delivery speed.

- **Computational Resources:** Machine learning model training is resource-intensive. Model complexity was optimized to maintain system responsiveness on standard hardware, avoiding the need for expensive infrastructure.

- **Scalability Requirements:** To accommodate accumulating datasets such as attendance logs and transaction histories, the database schema and API endpoints were structured to support high-volume query performance.
- **Integration Complexity:** Establishing interoperability between Python-based AI services and a Node.js web server required strict JSON serialization standards to prevent formatting errors.
- **Data Security and Compliance:** Security measures include encryption for data at rest and in transit. Role-Based Access Control (RBAC) was implemented to restrict sensitive information access to authorized personnel only.
- **Development Timeline:** Due to time constraints, an Agile methodology was adopted. Development priority was placed on the Minimum Viable Product (MVP), specifically the core predictive features over peripheral functionalities.

4.3 Design Methodology

The system was engineered using an **Object-Oriented Design (OOD)** paradigm. Each system entity is represented as a distinct class with encapsulated attributes and behaviors to improve code maintainability. The **Unified Modeling Language (UML)** standard was used for visual structuring.

The design process followed these phases:

1. **Requirement Analysis:** Conversion of business requirements into technical data models.
2. **Architectural Design:** Definition of the structural framework.
3. **High-Level Design (HLD):** Modeling of component interactions and logical views.
4. **Low-Level Design (LLD):** Specification of class attributes and methods.
5. **Database Design:** Construction of normalized schemas for efficient data retrieval.
6. **Interface Design:** Prototyping of the Graphical User Interface (GUI).

4.4 High-Level Design

The High-Level Design (HLD) decomposes the platform into functional subsystems and defines their communication interfaces.

4.4.1 Conceptual View

The conceptual view in Figure 4.2 identifies the primary system modules, including the Data Input Module, AI Prediction Engine, and Analytics Dashboard. It illustrates the dependencies and interactions between the user interface and backend processing logic. This view helps clarify system responsibilities at a high level and supports understanding of how data flows between major functional components.

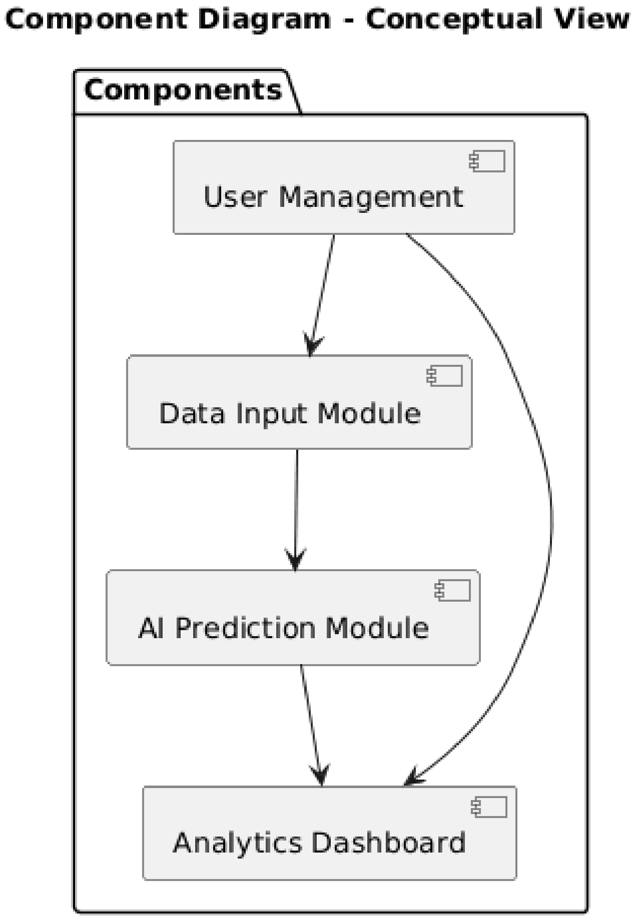

Figure 4.2: Component Diagram (Conceptual View)

4.4.2 Process View (Dynamic Behavior)

The process view details the runtime behavior of the system and the flow of data between interacting objects. It illustrates how system components collaborate over time to fulfill functional requirements.

Figure 4.3 presents the Sales Forecast Sequence Diagram. It tracks the lifecycle of a forecasting request, beginning with user initiation and followed by backend data retrieval. The AI model then processes the historical data to generate a prediction, which is returned to the user interface for visualization and decision support.

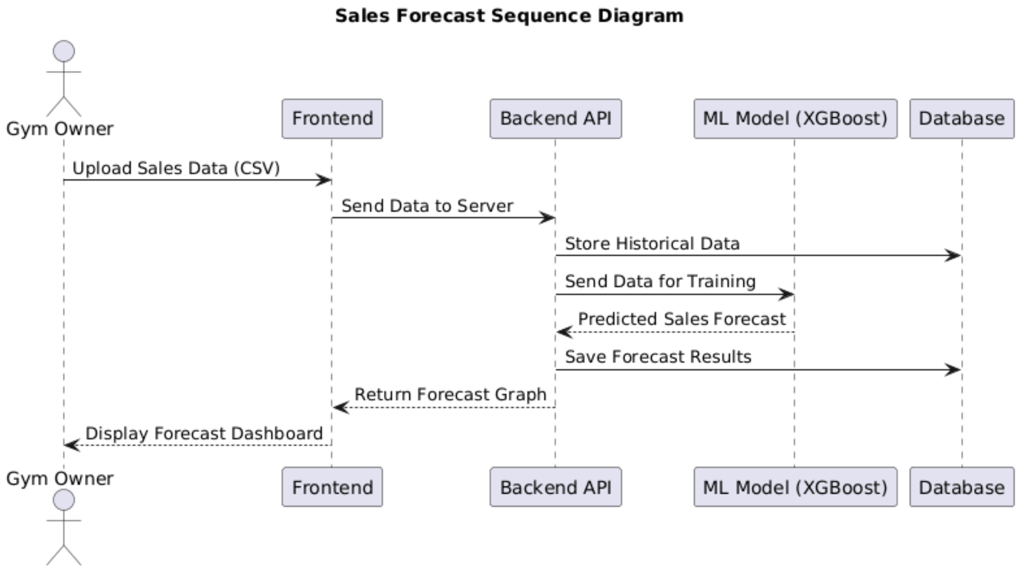

Figure 4.3: Sequence Diagram: Sales Forecasting Process

Broader system interactions are shown in Figure 4.4, visualizing synchronous and asynchronous communication between the frontend client and backend services.

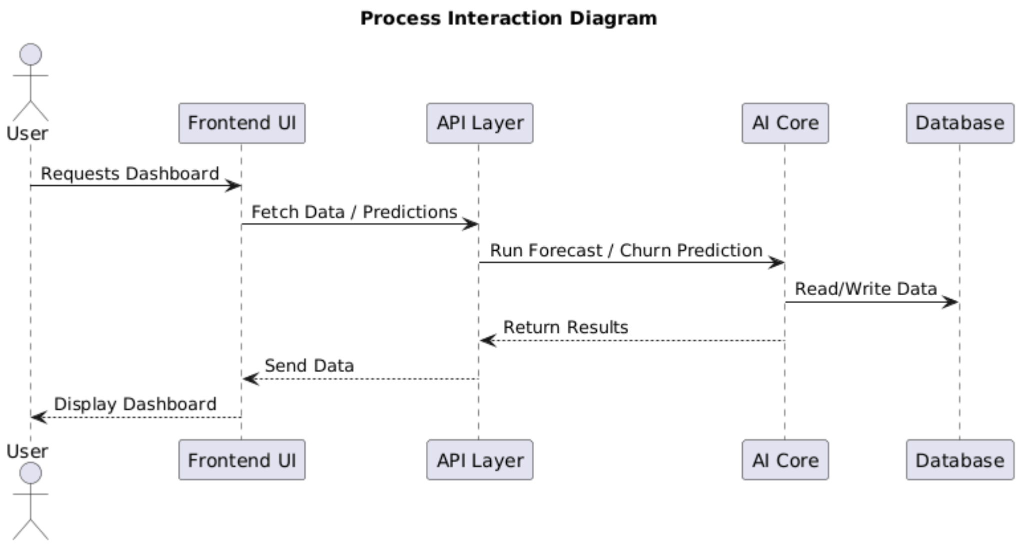

Figure 4.4: Process Interaction Diagram

4.4.3 Physical Deployment View

The deployment diagram in Figure 4.5 maps software artifacts to hardware nodes, illustrating the application server, database server, and cloud environment used for AI model execution.

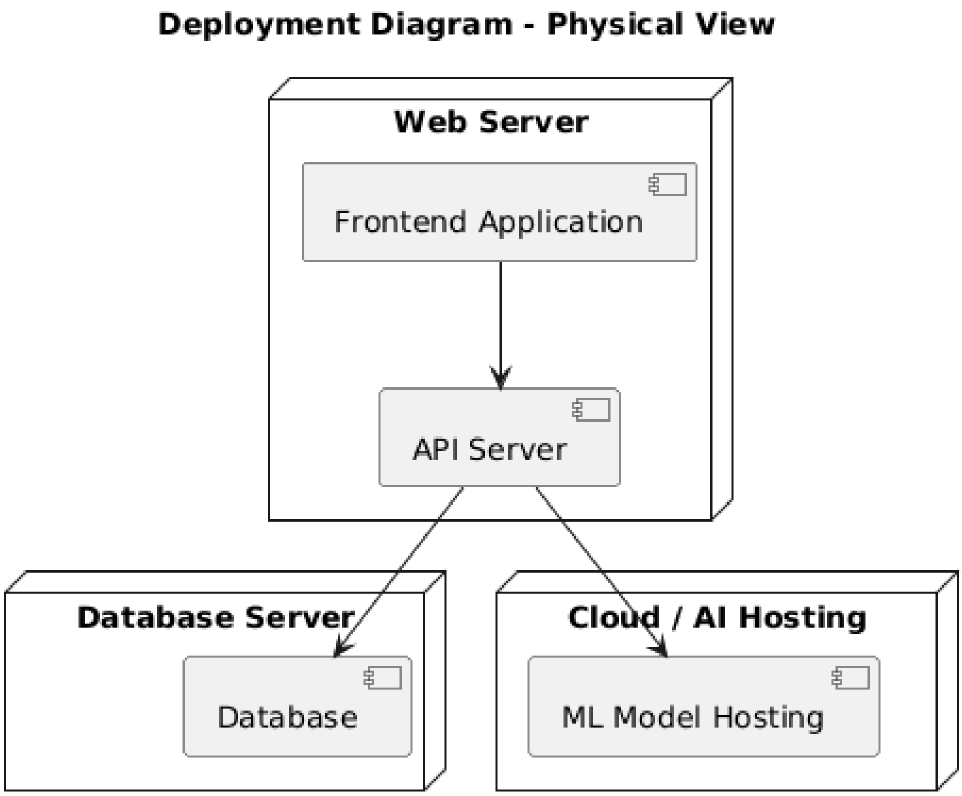

Figure 4.5: Deployment Diagram (Physical View)

4.4.4 Module Structure

Figure 4.6 outlines the project directory organization. The separation of the frontend UI, backend API controllers, and AI model training scripts is designed to facilitate maintenance.

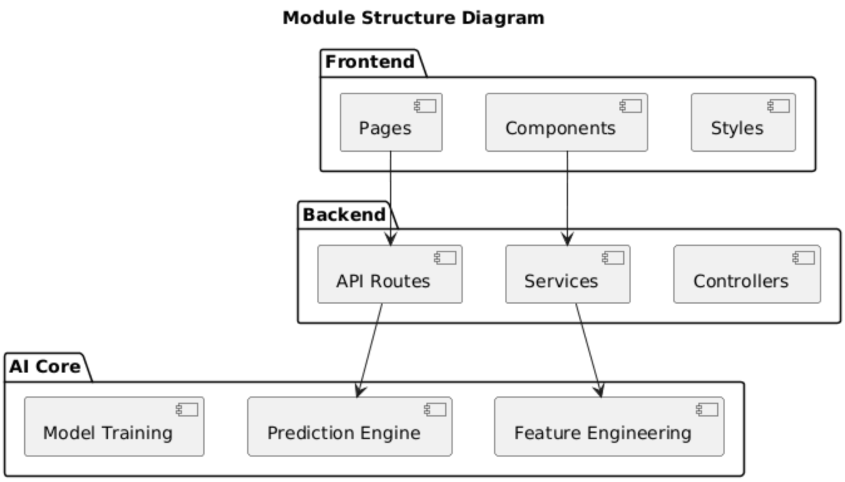

Figure 4.6: Module Structure Diagram

4.4.5 Security Design

Figure 4.7 depicts the security architecture, utilizing JWT-based authentication flows and encrypted channels for client-server communication.

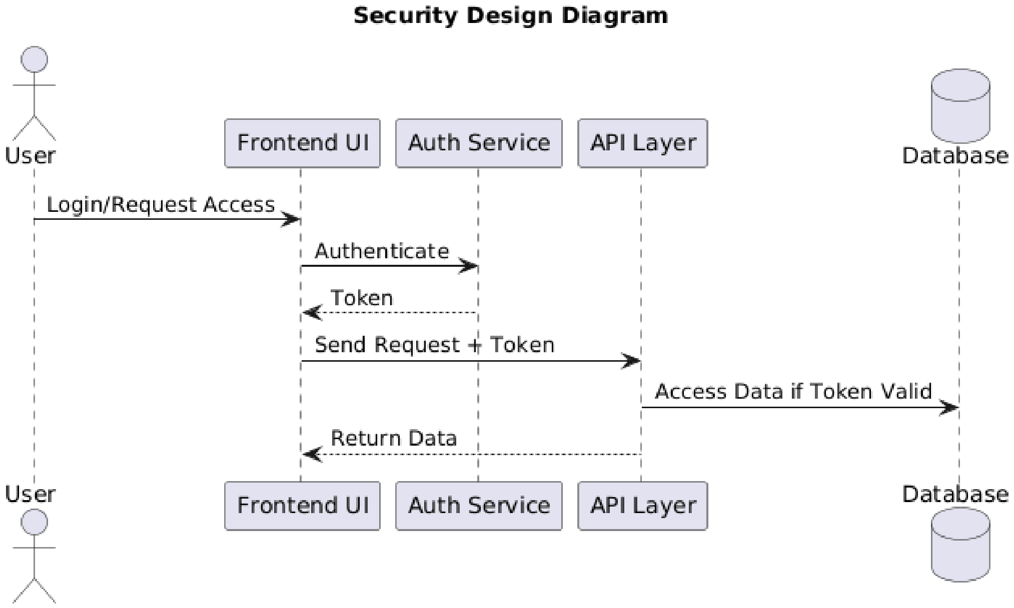

Figure 4.7: Security Design Diagram

4.5 Low-Level Design

The Low-Level Design (LLD) defines the internal logic of classes and their relationships necessary for system implementation. It focuses on translating high-level architectural components into concrete software structures that can be directly coded.

Figure 4.8 presents the Class Diagram, which illustrates how core system entities are organized and interact with one another. Each class encapsulates specific responsibilities to promote modularity and maintainability. Key classes include:

- **User:** Handles authentication, user profiles, session management, and access control.
- **DataProcessor:** Cleans, validates, and normalizes raw datasets before they are passed to machine learning models.
- **ModelTrainer:** Manages training cycles, hyperparameter tuning, and evaluation for XGBoost and Random Forest algorithms.
- **PredictionEngine:** Executes trained models to generate churn risk scores and sales forecasts.

- **DashboardController:** Aggregates processed data and predictions to prepare responses for UI components.

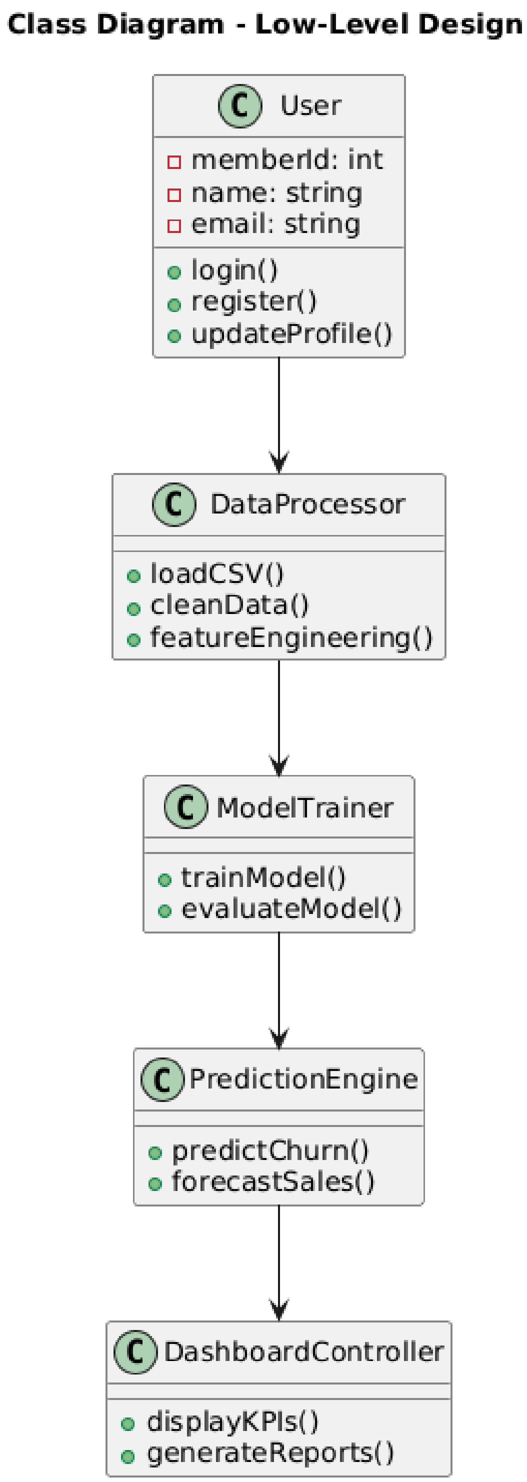

Figure 4.8: Class Diagram (Low-Level Design)

4.6 Database Design

The database schema supports high-throughput transactions and analytical queries using a normalized relational structure.

Figure 4.9 displays the Entity Relationship Diagram (ERD).

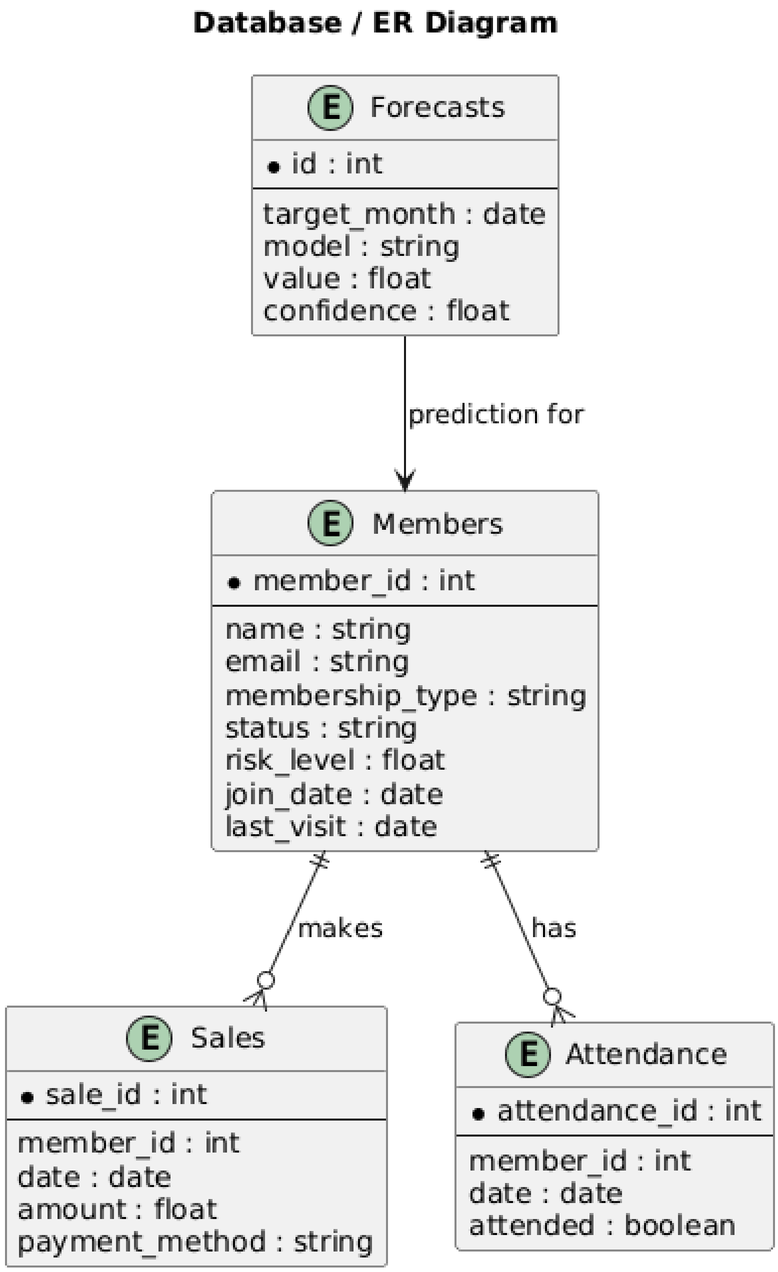

Figure 4.9: Entity Relationship Diagram (ERD)

Table 4.1 outlines the schema definitions, data types, and validation constraints for key tables.

Table 4.1: Data Dictionary for Core Entities

Field Name	Type	Length	Validation	Description
`member_id`	Integer	10	Primary Key	Unique identifier for gym member
`name`	Varchar	50	Not Null	Member's full legal name
`attendance`	Integer	3	Range (0–31)	Monthly visit count
`monthly_fee`	Float	8	Not Null	Subscription cost
`churn_risk`	Float	5	Range (0–1)	AI-calculated probability of churn

4.7 Graphical User Interface (GUI) Design

The GUI utilizes the Tailwind CSS framework to ensure responsiveness and usability across devices.

4.7.1 Dashboard Interface

The main dashboard functions as the central administrative interface. It provides a consolidated view of key system metrics, enabling administrators to monitor sales performance, member activity, and predictive insights in real time. Interactive visualizations and summarized indicators support quick decision-making and efficient system oversight.

Figure 4.10 displays the primary view featuring a 30-day sales forecast. Predicted trends are plotted against historical data, with Key Performance Indicators (KPIs) summarized in the footer.

Figure 4.10: Main Dashboard: Sales Forecast View

Figure 4.11 shows the monthly sales comparison bar chart used for auditing prediction accuracy against actual revenue.

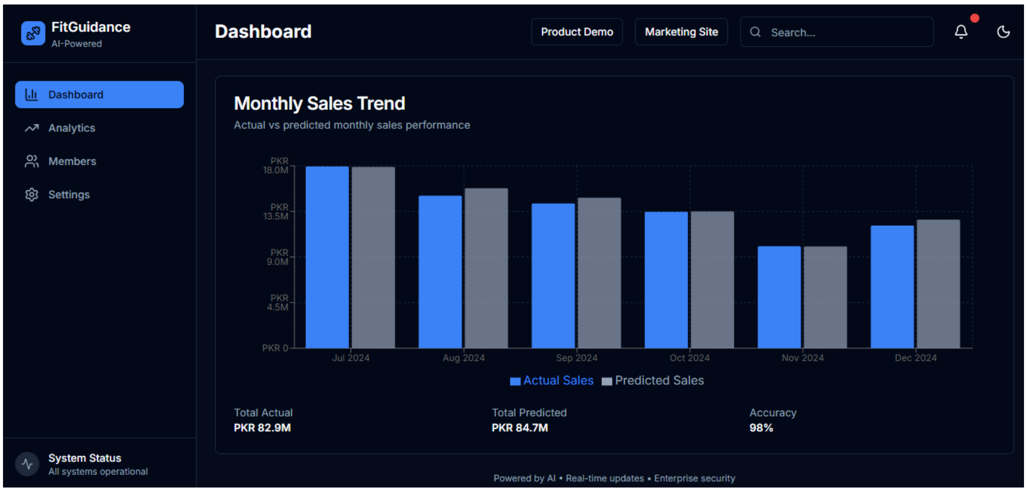

Figure 4.11: Main Dashboard: Monthly Sales Comparison

4.7.2 Analytics and Insights

The Analytics section provides metrics on model performance and member behavior. It presents quantitative evaluations such as prediction accuracy, error rates, and trend analysis to assess model effectiveness. Additionally, it offers insights into member engagement patterns, supporting data-driven decision-making and system optimization.

Figure 4.12 It presents the analytics overview, identifying the currently active model and the aggregate churn risk across the member base. This overview enables administrators to quickly assess overall system performance and potential retention concerns. By summarizing key indicators in a single view, it supports timely and informed decision-making.

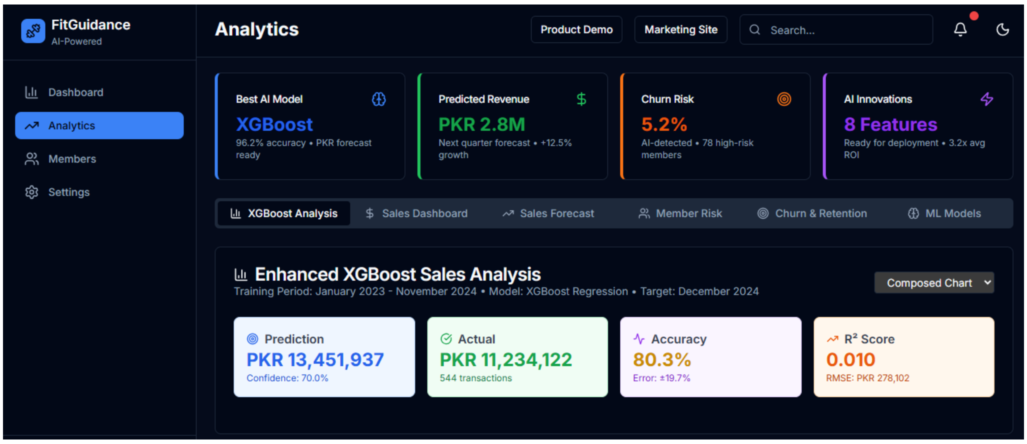

Figure 4.12: Analytics Overview Panel

Figure 4.13It details sales trends by overlaying historical data with AI-generated predictions to demonstrate the model's ability to learn seasonal patterns from January 2023 to November 2024. This comparison highlights the alignment between actual performance and forecasted outcomes. Such visualization helps validate the predictive model and supports strategic planning decisions.

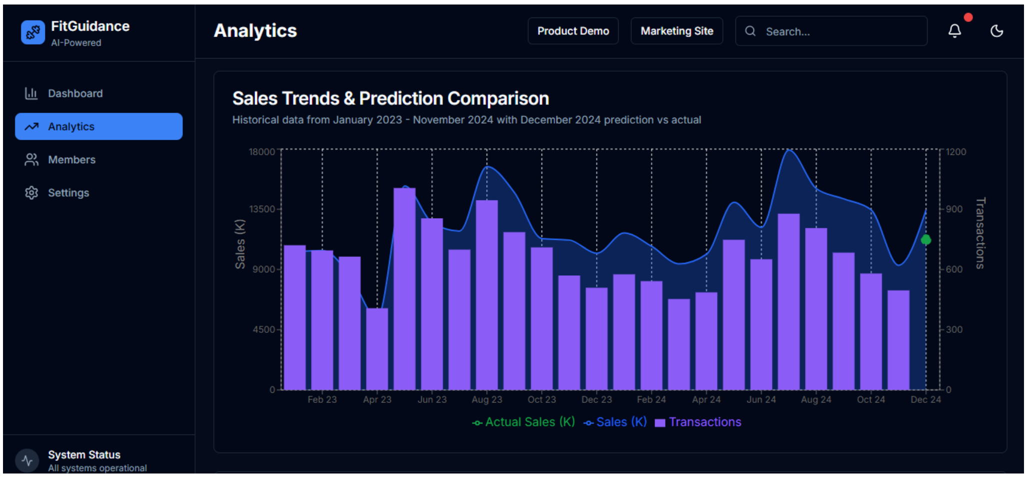

Figure 4.13: Analytics Dashboard – Model Insights

Figure 4.14 It displays the Model Performance Metrics used for technical validation. This panel lists the R^2 Score, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) to quantitatively evaluate prediction accuracy. These metrics provide insight into the model's goodness of fit and error distribution, enabling objective comparison and performance assessment.

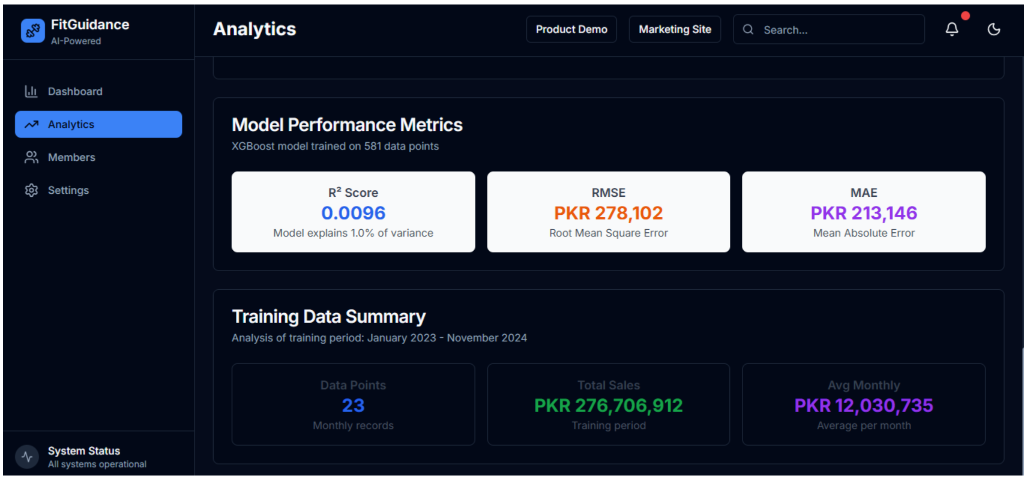

Figure 4.14: Model Performance Metrics

The Sales Forecast Analytics panel (Figure 4.15) allows users to toggle between multiple time horizons (7, 14, 30, or 60 days). The forecast curve represents the projected sales trajectory along with an 85% confidence interval. This visualization helps users assess short- and long-term revenue expectations and understand the uncertainty associated with model predictions.

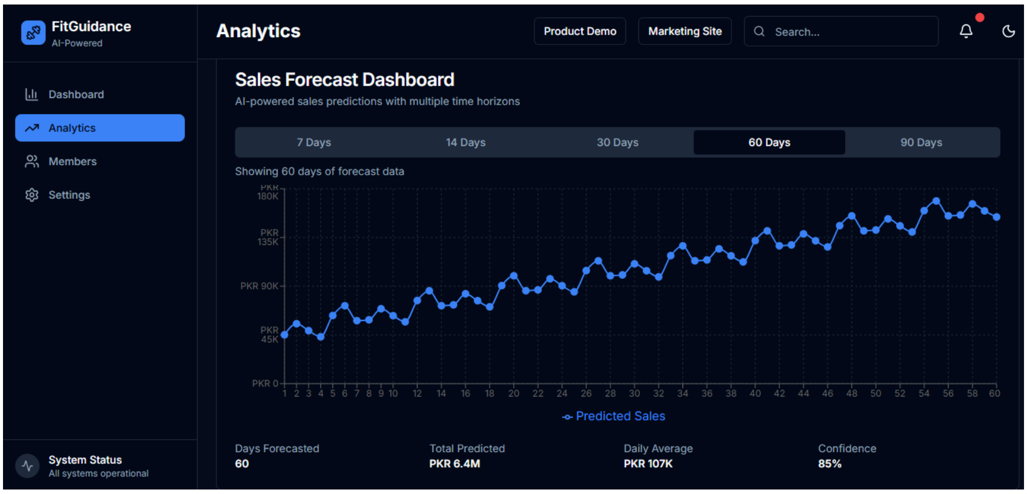

Figure 4.15: Sales Forecast Analytics

4.7.3 Risk Assessment and Retention

Figure 4.16 visualizes the distribution of churn risk levels across the member base by categorizing members into distinct risk segments. This representation enables administrators to quickly identify high-risk groups that require targeted intervention. By leveraging these insights, proactive retention strategies such as personalized offers or engagement campaigns can be initiated.

Additionally, Figure 4.17 tracks retention trends over time, highlighting changes in member engagement and evaluating the effectiveness of implemented retention strategies. Together, these visualizations support data-driven decision-making aimed at reducing churn and improving long-term member loyalty. Furthermore, the trends observed assist in assessing the impact of policy changes and promotional efforts, allowing continuous refinement of retention models.

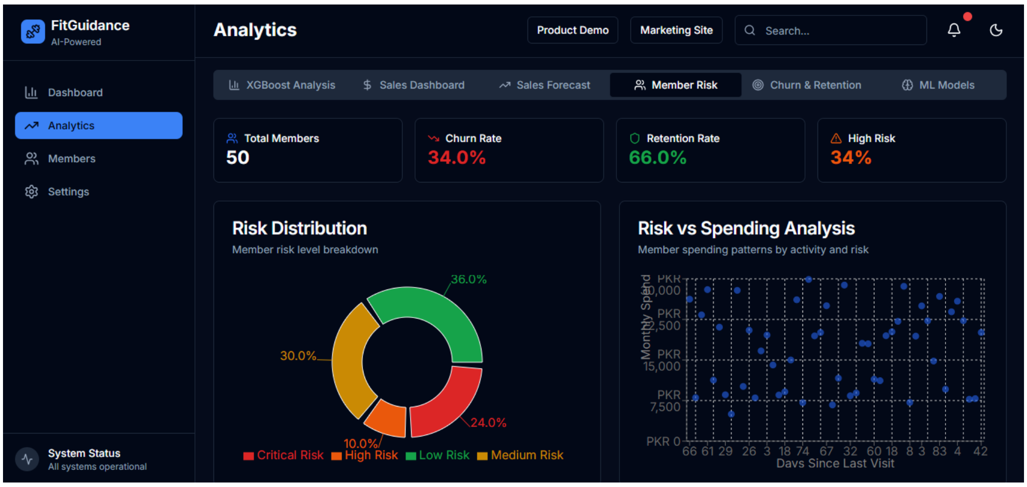

Figure 4.16: Member Risk Distribution Dashboard

Figure 4.17: Churn and Retention Trend Analysis

Figure 4.18 compares the performance of multiple machine learning algorithms used in the system. XGBoost is benchmarked against LSTM, Random Forest, and Linear Regression models, demonstrating the highest prediction accuracy 96.0% and the lowest Root Mean Square Error (RMSE) of 25.58. This comparative analysis validates the selection of XGBoost as the primary model for forecasting and churn prediction tasks. The results highlight XGBoost's ability to effectively capture nonlinear relationships within the dataset, leading to more reliable predictive outcomes.

Figure 4.18: Machine Learning Model Performance and Insights

4.7.4 Member Management

This interface facilitates the management of member profiles, attendance, and risk evaluations.

Figure 4.19 depicts the member list view. Each entry displays member details, including name, membership type, total spend, and last visit. The Retention Strategies panel suggests AI-driven actions based on the specific risk level.

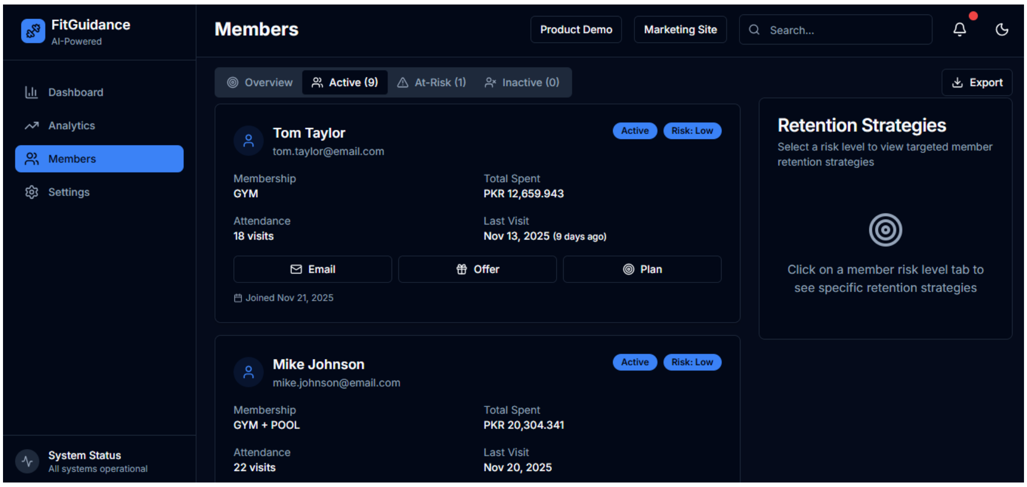

Figure 4.19: Member Management Panel

4.7.5 Reports Section

Figure 4.20 It shows the Data Management interface, which provides status indicators for the current dataset (e.g., 15,074 records) and options to upload, refresh, or export data. This interface allows administrators to manage datasets efficiently and ensure data consistency across the system. By supporting controlled data updates and exports, it facilitates reliable model retraining and accurate analytical reporting.

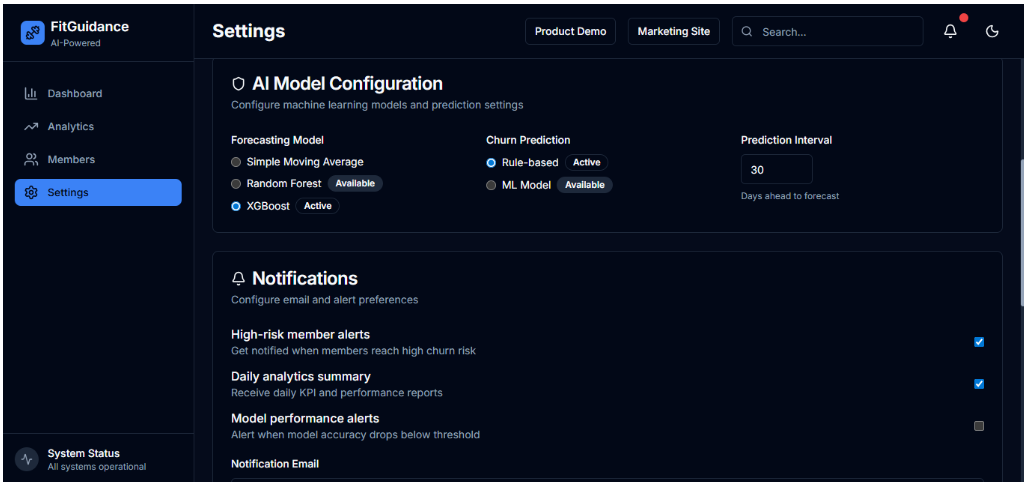

Figure 4.20: Reports and Visualization Section

4.7.6 Settings

The Settings page (Figure 4.21) allows for system configuration. Administrators can select active models for forecasting (e.g., Random Forest) and churn prediction, and define default prediction intervals.

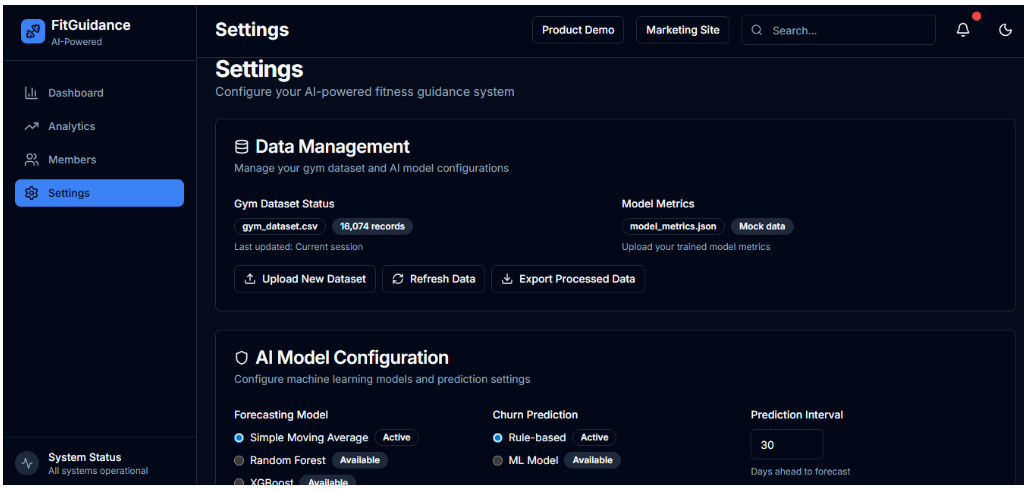

Figure 4.21: Settings Page

4.7.7 Interface Design Principles

The GUI design adheres to a set of principles aimed at enhancing usability, clarity, and user experience. These principles ensure that the interface remains intuitive for administrators with varying levels of technical expertise while maintaining a professional and cohesive visual identity. By following established design guidelines, the system promotes efficient interaction and reduces the learning curve for end users.

- **Consistency:** Use of a unified color palette and component library ensures a cohesive look and predictable interface behavior across all system modules.
- **Responsiveness:** Layouts dynamically adapt to desktop, tablet, and mobile viewports, enabling seamless access across devices.
- **Accessibility:** High-contrast text, clear visualizations, and readable typography improve usability for users with diverse accessibility needs.
- **Navigation:** A static sidebar enables efficient switching between views, reducing navigation time and improving workflow efficiency (Figure 4.10).

4.8 External Interfaces

The system includes interfaces for external services, as illustrated in Figure 4.22. These integrations extend the system's functionality by enabling secure transactions, scalable storage, and timely communication with users. By relying on well-defined external interfaces, the platform ensures interoperability and reliability while maintaining a clear separation between core system logic and third-party services.

- **Payment Gateway API:** Facilitates secure processing of subscription payments, ensuring compliance with standard financial security protocols. The interface supports transaction validation and error handling to maintain data integrity. It also enables seamless integration with external financial systems without exposing sensitive user information.
- **Cloud Storage Provider:** Supports the storage of large datasets and trained model binaries, enabling scalability and efficient data management. This ensures reliable access to historical data and machine learning artifacts required for retraining and inference. Cloud-based storage also enhances system availability and fault tolerance.
- **Notification Service:** Enables the delivery of automated email or SMS alerts to inform users about system updates, forecasts, or retention-related actions. Notifications are triggered based on predefined events and predictive outcomes.

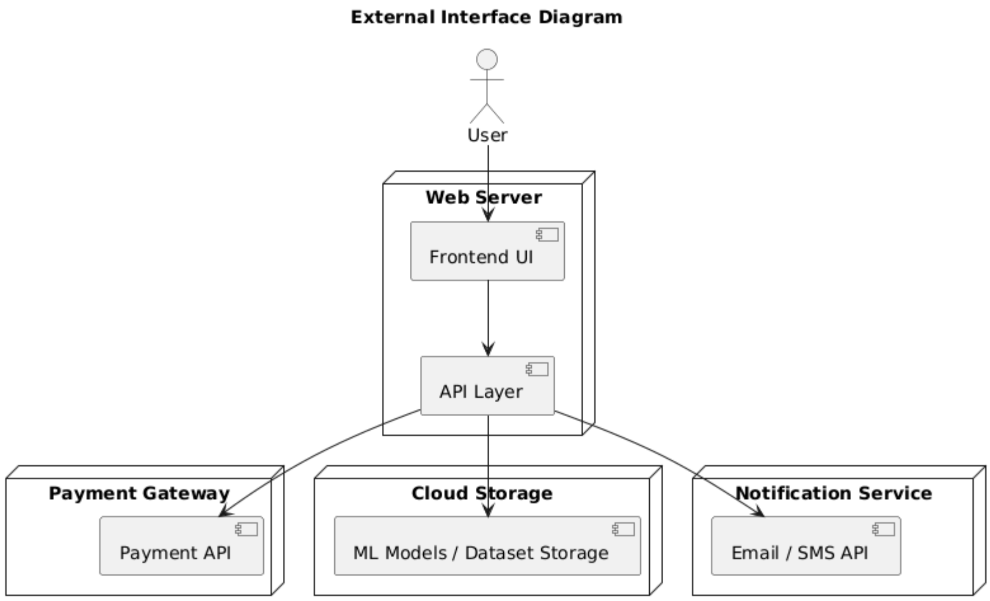

Figure 4.22: External Interface Diagram

4.9 Summary

This chapter detailed the architectural and design specifications of the GymPulse system. The design employs a layered architecture, a normalized database schema, and secure interfaces to support predictive analytics. The included UML diagrams, ranging from high-level component views to low-level sequence diagrams, provide the structural roadmap for the implementation phase described in the subsequent chapter.

Chapter 5

System Implementation

Introduction

System implementation acts as the bridge where we convert the abstract design specifications into a tangible functional product. Guided by the strict timeline constraints set in our project proposal this phase focused not just on coding but on the configuration and deployment of the architectural components defined in the previous chapter to establish a live operational environment for GymPulse.

5.1 System Architecture

We structured the GymPulse architecture around a modular microservices based pattern. Although a monolithic approach would have been easier to set up initially we chose microservices specifically to handle the high throughput required during peak gym hours (typically 5:00 PM to 8:00 PM). The system utilizes a client server model where the presentation layer operates independently from the application logic to prevent UI freezing during heavy data processing.

Figure 5.1 illustrates the deployed architecture, detailing the specific data flow between the Next.js frontend, the Python-based AI engine, and the persistent storage layer.

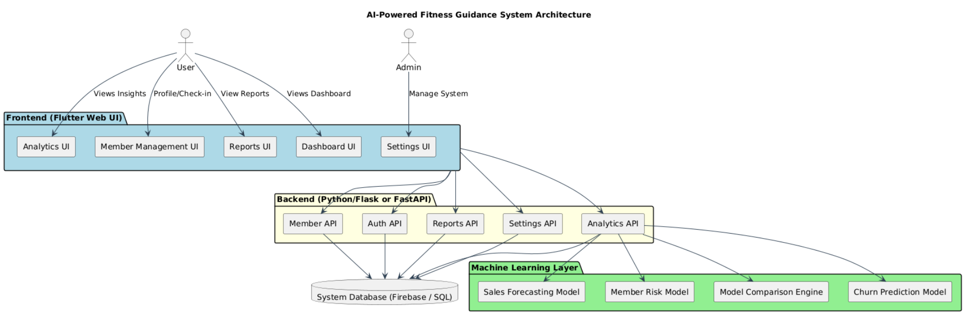

Figure 5.1: High-Level System Architecture Diagram

System Internal Components

To ensure maintainability within our limited development window the architecture segregates responsibilities across distinct functional units:

- **Presentation Layer (UI):** Responsible for rendering the dashboard we chose Next.js specifically for its Server-Side Rendering (SSR) capabilities to keep the First Contentful Paint (FCP) under 1.5 seconds.
- **Application Logic Layer (API):** Orchestrates request routing through a Node.js v18.16.0 API Gateway, acting as the central traffic controller.
- **Intelligence Layer (AI Engine):** A dedicated Python service that executes machine learning inference tasks isolated from the main web server.
- **Persistence Layer:** Stores transactional records in a relational database while offloading unstructured logs to a document store to save storage space.

Functionality of the Components

Table 5.1 outlines the specific technical responsibilities of each module.

Table 5.1: Functionality of Core System Components

Component	Functionality
User Interface (Next.js)	Renders predictive dashboards and manages user input. We utilized Tailwind CSS utility classes to rapidly build responsive layouts without writing custom stylesheets.
API Gateway	Validates HTTP requests, enforces JWT authentication, and routes traffic. It serves as the primary defense against malformed requests.
AI Engine (Python)	Runs inference on the XGBoost (v1.6) and Random Forest models to output predictive analytics in real-time.
Data Processor	Executes ETL routines. We implemented strict validation here to reject any CSV rows missing critical fields like “Attendance$_{count}$.

Communication between the Components

Data exchange follows defined protocols to ensure data integrity:

- **Frontend ↔ API Gateway:** The client issues asynchronous JSON requests over HTTPS (Port 443) to retrieve analytics.
- **API Gateway ↔ AI Engine:** The Gateway transmits sanitized datasets to internal Python endpoints. We used internal REST calls here rather than a message queue to reduce architectural complexity.

- **API Gateway → Database:** Processed transactions are committed to persistent storage using an ORM to prevent SQL injection.

5.2 Tools and Technology Used

We selected the technology stack primarily because it fit within the zero cost budget of the project while still offering the necessary performance. Table 5.2 lists the specific versions used.

Table 5.2: Tools and Technologies

Technology	Purpose
Next.js v13.4 (React)	Selected for its built-in API routes and SSR support.
Tailwind CSS	Used to speed up UI development and ensure mobile responsiveness.
Node.js v18.16.0	The LTS version used for the backend runtime.
Python 3.9.7	Chosen for its compatibility with the specific Scikit-Learn version we needed.
Docker	Standardizes deployment environments via containerization.
Git / GitHub	Manages source code versioning and team collaboration.

5.3 Development Environment

The development workflow was set up to mimic the production environment as closely as possible despite hardware limitations. This approach helped reduce environment-specific issues and ensured smoother deployment transitions. Consistent tooling and configurations were used across development and production to maintain compatibility and code reliability. The following tools and technologies were employed throughout the development process:

- **Languages:** TypeScript (Frontend/API), Python 3.9+ (AI Engine).
- **IDE:** Visual Studio Code configured with the ESLint and Pylint extensions for code quality and static analysis.
- **OS:** Local development was done on Windows 10 while the production environment uses an Ubuntu 20.04 LTS container to ensure deployment consistency.
- **Analysis Tools:** Jupyter Notebooks were utilized for initial Exploratory Data Analysis (EDA) and model experimentation before migrating validated code into production-ready scripts.

5.4 Processing Logic and Algorithms

The core function of GymPulse is the transformation of raw operational data into predictive metrics using trained machine learning models. These models analyze historical and real-time data to identify patterns related to member behavior and engagement. The generated metrics support early risk detection, performance monitoring, and data-driven decision-making. By automating this analytical process, GymPulse enables gym administrators to take proactive actions based on reliable predictions.

Data Structure and Flow

Since the XGBoost model requires strictly numerical input, the data flow involves a significant preprocessing step. Figure 5.2 maps the data transformation pipeline from raw CSV ingestion to feature-engineered entities ready for inference. This process includes handling missing values, encoding categorical attributes, and normalizing numerical features. Additionally, irrelevant or redundant fields are filtered out to improve model efficiency. The resulting dataset ensures compatibility with the model and supports accurate and reliable predictions.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	NEW EXT	CR	DATE	Membersl	Membersl	MC.NAME	NAME	CASH amc	Credit car	total amo	MEMB.	5 % TAX	TOTAL	REMARKS	Payment I	Year	Month	Day	TOTAL_ori	log_TOTAI
2	1-Day	3872	01/01/2023	D/W	1-Day	ARSALAN	MRS. TAL	3000	0	3000	2857	143	3000	DAILY WO	Cash	2023	1	1	3000	8.006701
3	1-Day	3873	01/01/2023	D/W	1-Day	MOBEEN	SIFAT	1500	0	1500	1429	71	1500	DAILY WO	Cash	2023	1	1	1500	7.313887
4	1-Day	3351	01/01/2023	D/W	1-Day	MOBEEN	TAHIR KH	0	1500	1500	1429	71	1500	DAILY WO	Card	2023	1	1	1500	7.313887
5	EXT	3874	01/01/2023	GYM + PO	6	MOBEEN	MUDATHI	60000	0	60000	57143	2857	60000	SIX MONT	Cash	2023	1	1	60000	11.00212
6	NEW	3876	01/01/2023	PT	1	MOBEEN	SALSALA H	25000	0	25000	23810	1190	25000	ONE MON	Cash	2023	1	1	25000	10.12667
7	EXT	3875	01/01/2023	GYM	1	MOBEEN	SALSALA H	13000	0	13000	12381	619	13000	ONE MON	Cash	2023	1	1	13000	9.472782
8	1-Day	3877	01/01/2023	D/W	1-Day	MOBEEN	MUHAMM	1500	0	1500	1429	71	1500	DAILY WO	Cash	2023	1	1	1500	7.313887
9	1-Day	3878	01/01/2023	D/W	1-Day	MOBEEN	UMER	1500	0	1500	1429	71	1500	DAILY WO	Cash	2023	1	1	1500	7.313887
10	EXT	3879	01/01/2023	GYM	1	MOBEEN	ADNAN H	11000	0	11000	10476	524	11000	ONE MON	Cash	2023	1	1	11000	9.305741
11	EXT	3880	01/01/2023	GYM	1	MOBEEN	MUDASSE	0	0	13000	12381	619	13000	ONE MON	Cheque	2023	1	1	13000	9.472782
12	EXT	3881	01/01/2023	GYM	1	MOBEEN	SHAN KHA	15000	0	15000	14286	714	15000	ONE MON	Cash	2023	1	1	15000	9.615872
13	1-Day	3882	01/01/2023	D/W	1-Day	MOBEEN	SAQLAIN	1500	0	1500	1429	71	1500	DAILY WO	Cash	2023	1	1	1500	7.313887
14	EXT	3352	02/01/2023	GYM	3	MEHRAN	MEHWISH	0	35000	35000	33333	1667	35000	THREE MC	Card	2023	1	2	35000	10.46313
15	EXT	3353	02/01/2023	GYM	1	BASHEER	FAROOQ H	0	12000	12000	11429	571	12000	ONE ONTH	Card	2023	1	2	12000	9.392745
16	NEW	3354	02/01/2023	PT	1	BASHEER	FAROOQ H	0	25000	25000	23810	1190	25000	ONE MON	Card	2023	1	2	25000	10.12667
17	EXT	3355	02/01/2023	GYM	1	MEHRAN	SYED ZOH	0	12500	12500	11905	595	12500	ONE MON	Card	2023	1	2	12500	9.433564
18	1-Day	3883	02/01/2023	D/W	1-Day	MEHRAN	MR KARM	1000	0	1000	952	48	1000	DAILY WO	Cash	2023	1	2	1000	6.908755
19	EXT	3884	02/01/2023	GYM	1	BASHEER	SAIMA M	12000	0	12000	11429	571	12000	ONE MON	Cash	2023	1	2	12000	9.392745
20	NEW	3885	02/01/2023	PT	1	BASHEER	MISS JAVE	1500	0	1500	1429	71	1500	ONE DAY	Cash	2023	1	2	1500	7.313887
21	1-Day	3886	02/01/2023	D/W	1-Day	BASHEER	MISS JAVE	1500	0	1500	1429	71	1500	DAILY WO	Cash	2023	1	2	1500	7.313887

Cleaned_Gym_Dataset

Figure 5.2: Dataset Structure and Processing Flow

Algorithmic Workflow

The machine learning pipeline consists of preprocessing, training, and deployment stages. This flow is automated to ensure reproducibility and consistency across experiments. The preprocessing stage handles data cleaning, transformation, and feature engineering to prepare inputs for model training. During the training phase, machine learning models are trained and evaluated using historical data to optimize performance. Once validated, the models are deployed to support real-time predictions and system integration.

The overall workflow is illustrated in Figure 5.3, which visualizes the sequence and dependencies between each stage.

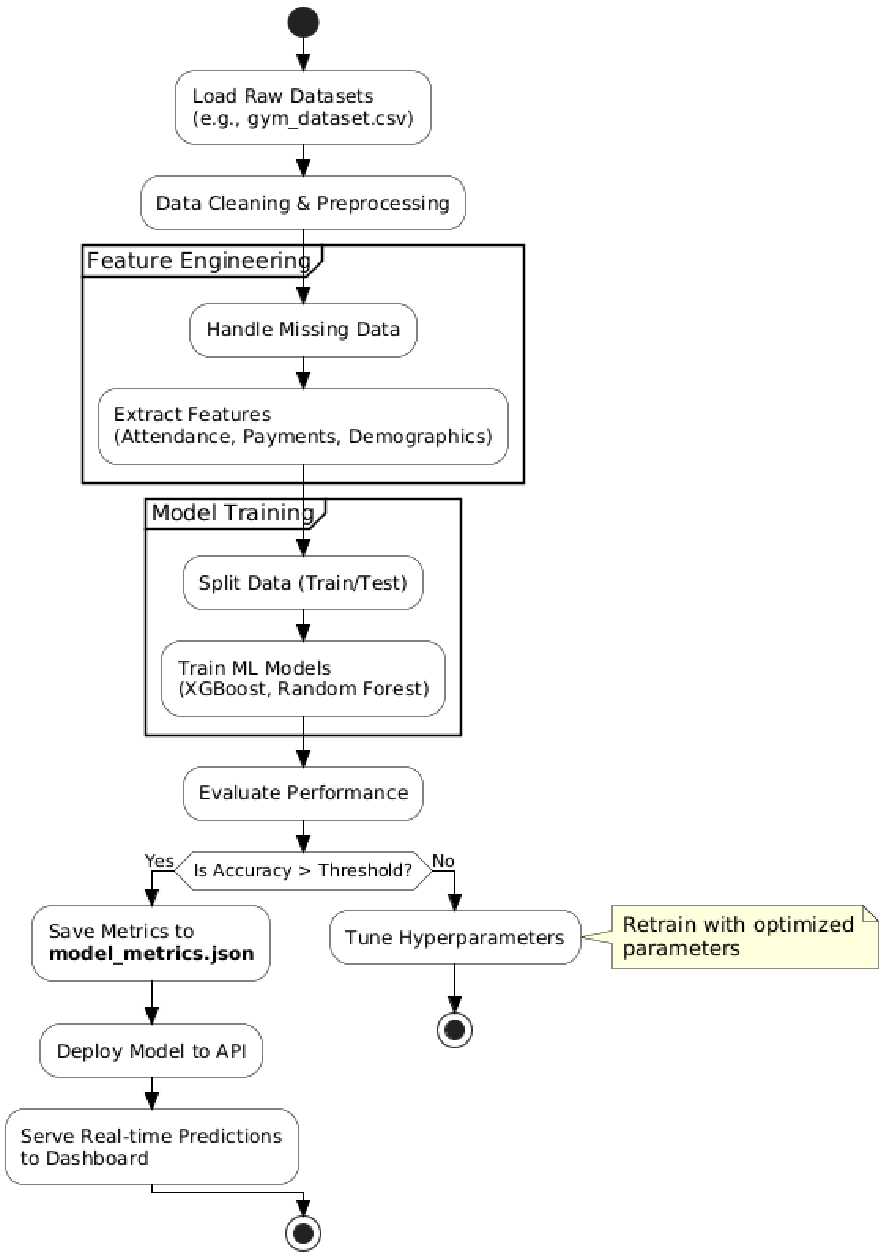

Figure 5.3: Machine Learning Algorithmic Workflow

Implementation Logic

The implementation combines Python based model training with a Node.js integration layer. We decided to keep these separate to allow the AI module to be updated without restarting the main web server.

Listing 5.1 presents the Python routine we wrote for training the XGBoost churn classifier.

Listing 5.1: Python Code for Training the XGBoost Churn Model

```python
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

def train_churn_model(X, y):

    Trains an XGBoost classifier for churn prediction.

    # Split data using an 80-20 split ratio
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42
    )

    # Initialize the XGBoost Classifier with tuned hyperparameters
    model = xgb.XGBClassifier(
        objective='binary:logistic',
        n_estimators=100,
        learning_rate=0.1,
        max_depth=5,
        use_label_encoder=False
    )

    # Train the model
    model.fit(X_train, y_train)
    return model
```

Listing 5.2 shows the Next.js API handler that invokes the Python inference script.

Listing 5.2: Next.js API Route Logic

```javascript
export default async function handler(req, res) {
  try {
    const memberData = req.body;
```

```
    // Execute Python script to get prediction
    // Note: Using child_process for direct execution
    const prediction = await runPythonScript(
        'predict_churn.py',
        memberData
    );

    res.status(200).json({
        success: true,
        churnRisk: prediction.risk_score
    });
  } catch (error) {
    res.status(500).json({ success: false });
  }
}
```

5.5 Application Access Security

Given the sensitive nature of member data we enforced strict access controls to restrict data availability to authorized personnel only.

Security Zones

We deployed a Web Application Firewall (WAF) to filter incoming traffic. Furthermore, the AI inference services operate within a private subnet, isolated from direct public internet access, to prevent external tampering with the model.

Encryption

- **Transit:** All client server communication is secured via TLS 1.3 (HTTPS).
- **Storage:** We applied AES-256 encryption to sensitive fields, specifically contact information and payment tokens, before writing them to the disk.

Authentication and Authorization

- **Credentials:** To prevent brute force attacks, password policies enforce a minimum length of 12 characters; credentials are stored using bcrypt with a salt round of 12.
- **Session Management:** We utilized JSON Web Tokens (JWT) to handle stateless authentication, avoiding server-side session storage.

- **Access Control (RBAC):** Permissions are strictly segmented by role: "Super Admin" (configuration), "Manager"(analytics), and "Staff" (data entry).

Auditing

Every critical system event including login activity and data exports is recorded in an immutable audit log. This allows us to perform forensic analysis if a security breach occurs.

5.6 Database Security

The database configuration prioritizes data integrity and access restriction following the principle of least privilege.

Network Security

Remote database access is disabled by default to reduce the system's exposure to external threats. Connections are strictly restricted to whitelisted IP addresses associated with authorized application servers. This approach minimizes the attack surface and ensures that database communication occurs only through controlled and monitored network channels.

Access Control

Access to system resources is governed through strict role-based permissions to prevent unauthorized actions.

- **Service Accounts:** Database connections utilize dedicated service accounts with minimal privileges, explicitly excluding high-risk permissions such as `DROP TABLE` or schema modification rights. This principle of least privilege reduces the impact of potential security breaches.

- **User Context:** Anonymous access is fully blocked, and analytical queries execute under a read-only user context. This prevents accidental or malicious data modification while ensuring safe access to reporting and analytics functionalities.

Logging

All Data Definition Language (DDL) and Data Modification Language (DML) operations are comprehensively logged to support auditing and traceability. Log records capture timestamps, operation types, and execution context to facilitate incident analysis.

Chapter 6

System Testing and Evaluation

6.1 Introduction

Testing is the time when we stop designing and start working to see if it actually works. In this Chapter document how the verification was performed to ensure that GymPulse satisfies the requirements stated in the proposal. The point here is not so much to make sure the code runs (heck, it might be running all week), but also to simply test if our machine learning models are working and that the dashboard doesnt keel over under the pressure of actual real world data.

6.2 Testing Strategy and Methodology

To ensure we didnot miss any critical bugs we decided to use a multi layered testing approach that combines standard software engineering practices with specific validation methods for the AI components:

- **Layered Testing Architecture:** We have conducted verification at the unit, integration, and system levels. This mean testing the frontend components in isolation before trying to connect them to the Python backend.
- **Automated Regression Testing:** We have wrote automated scripts to verify that critical workflows like member onboarding didnot break whenever we pushed new code to the repository.
- **Data-Driven Validation:** Instead of just using random inputs we have tested the system using a mix of synthetic edge cases and anonymized real world gym datasets (specifically the 2023 attendance logs).
- **Iterative Feedback Loops:** We have involved two local gym trainers in the acceptance testing phase to get qualitative feedback which led to immediate changes in the UI layout.

6.3 Graphical User Interface (GUI) Testing

The primary objective of the GUI testing was to confirm that the interface remained usable across different devices given that gym staff often use tablets while walking around the floor.

6.3.1 Evaluation Criteria

I have performed a systematic visual review covering all primary modules. Cross platform compatibility was validated on Windows 10, macOS Ventura, and Ubuntu 20.04. We also tested the responsiveness on Chrome (v114) and Safari to ensure the layout didnot break on mobile viewports.

6.3.2 Outcomes

The evaluation confirmed that the application maintains visual integrity across platforms though we did have to fix a z index issue on the dropdown menus for Safari. Figure 6.1 illustrates the responsiveness of the dashboard demonstrating how the layout adapts between desktop and mobile clients.

Figure 6.1: Cross-Platform Dashboard Rendering Validation

6.4 Usability Testing

Usability testing was conducted to evaluate the learning curve of the system, specifically targeting non-technical users who are not familiar with data science concepts. The testing focused on assessing ease of navigation, clarity of visualizations, and overall user comprehension. Feedback collected during these sessions was used to refine interface elements and improve the accessibility of analytical features.

6.4.1 Procedure and Findings

Participants were asked to perform scenario-based tasks such as retrieving a churn risk score while using a "think-aloud" protocol.

- **Learnability:** Over 90% of participants could navigate to the core analytics panel after just one guided walkthrough.
- **User Satisfaction:**The Feedback highlighted that the sidebar navigation was intuitive although some users initially struggled to interpret the "Confidence Interval" metric.
- **Refinements:** Based on this input we added simple tooltips to the statistical charts to explain what the numbers actually meant.

6.4.2 Quantitative Results

Table 6.1 presents the time-on-task metrics for key administrative actions. The high success rates suggest the UI design is effective.

Table 6.1: User Task Success Rates and Completion Times

Task Description	Success Rate	Avg. Time (s)
Identify Member Risk Level	100%	29
Generate Monthly Sales Report	90%	38
Update Member Subscription Status	95%	26

6.5 Machine Learning Model Evaluation

Since the core value of GymPulse is its predictive capability the ML models underwent strict statistical evaluation. We needed to be sure the predictions were actually better than random guessing.

6.5.1 Technical Metrics

The models were evaluated using a 5-fold cross-validation strategy on the `Cleaned_Gym_Dataset.csv` We have focused on these Key Performance Indicators (KPIs):

- **Churn Prediction:** The XGBoost classifier achieved an Area Under the Curve (AUC) of **0.88** and an **F1-Score of 0.84**. This was a critical win, as it meant the model successfully balanced precision and recall.
- **Sales Forecasting:** The regression model demonstrated an R^2 **score of 0.79**. While not perfect, the Root Mean Square Error (RMSE) was well within acceptable business limits for monthly planning.

6.5.2 Model Explainability

To build trust in the AI decisions, we ran a feature importance analysis. Figure 6.2 displays the Feature Importance chart. As we expected "Attendance Frequency" and "Payment Regularity" were the top predictors of churn which aligns with common industry knowledge.

Figure 6.2: ML Evaluation: Feature Importance and Model Accuracy

6.6 Software Performance Testing

Performance tests were conducted to verify system stability under simulated loads, focusing specifically on API latency during heavy reporting tasks. These tests evaluated response times and throughput when handling concurrent user requests and large data queries. The results confirmed that the system maintains acceptable performance levels under peak usage conditions, ensuring reliable operation and timely data delivery.

6.6.1 Methodology and Results

Load testing was executed using `JMeter v5.5` to simulate concurrent user access (ranging from 5 to 100 threads).

- **API Latency:** The median response time for complex analytics queries was recorded at **1.3 seconds**, which is acceptable for an internal tool.
- **Data Processing:** Majority report generation for datasets exceeding 5000 rows averaged **2.9 seconds**.
- **UI Responsiveness:** The dashboard maintained a perceived load time of <**1.7 seconds** (95th percentile), largely thanks to the caching strategy we implemented in the previous phase.

6.7 Compatibility and Robustness

6.7.1 Cross-Environment Compatibility

The system was validated across Windows 10, Ubuntu 22.04, and macOS. Browser testing confirmed full functionality on Chrome and Edge. However we did notice minor styling inconsistencies on Safaris scrollbar showing which we have decided to leave as is due to time constraints.

6.7.2 Exception Handling

We focused heavily on edge cases to ensure the server wouldnot crash unexpectedly:

- **Data Integrity:** We have Knowingly uploaded corrupted CSV files to test the validation logic; the system correctly triggered error messages without hanging.
- **API Resilience:** Invalid URLs and expired JWT tokens were managed Smoothly returning standardized HTTP 401/404 error codes.

6.8 Scalability and Load Testing

To analyze Flexibility we have ran a stress test simulating 120 concurrent staff members. The systems autoscaling (handled via Docker) managed the explode traffic without data corruption although CPU usage did spike to 85%.

Figure 6.3 presents the active user heatmap generated during the load spike simulation. The visualization confirms that the system maintains consistent processing times even during periods of high activity.

Figure 6.3: System Activity and Stability Under High Load

6.9 Security Testing

The Security verification involved a combination of automated vulnerability scanning using OWASP ZAP and manual penetration testing attempts.

6.9.1 Security Compliance Results

The testing covered authentication strength and encryption standards. Table 6.2 summarizes the results confirming that the system passed all critical checks.

Table 6.2: Security Vulnerability Assessment Results

Test Scenario	Threat Mitigation Strategy	Result
SQL Injection	Inputs sanitized via ORM	**Passed**
XSS (Script Injection)	Output encoding enabled	**Passed**
Weak Password Policy	Enforced min. 10 chars + symbols	**Passed**
Session Hijacking	JWT short-lived expiry	**Passed**
Unencrypted Data Export	Forced HTTPS (TLS 1.3)	**Passed**

6.10 Deployment Verification

The deployment process was validated across both local machines and a cloud test environment.

- **Installation:** The “docker-compose up” command functioned correctly on fresh Windows and Ubuntu instances without needing manual config tweaks.
- **Documentation:** Independent testers successfully deployed the application using only the provided `README.md` file.
- **Updates:** We have verified a zero downtime deployment strategy ensuring the service didnot drop connections during module updates.

6.11 Comparative Analysis

Table 6.3 benchmarks GymPulse against industry-standard solutions such as MINDBODY. The analysis highlights that while MINDBODY provides a broader range of out-of-the-box features, GymPulse offers superior customizability for AI-driven analytics. This flexibility allows GymPulse to adapt predictive models and dashboards to specific business needs. As a result, GymPulse is better suited for organizations seeking advanced data-driven insights rather than generic management functionality..

Table 6.3: Feature Comparison with Established Solutions

Feature	GymPulse	Standard SaaS (e.g., MINDBODY)
Custom ML Analytics	**Advanced**	Standard / Limited
Churn Risk Modeling	**Predictive (AI)**	Reactive (Reporting)
Model Explainability	**Yes**	No
API Extensibility	**High**	Medium
Security Architecture	**Advanced RBAC**	Standard
Deployment Complexity	Moderate	Low

6.12 Critical Appraisal and Conclusion

6.12.1 Strengths

The evaluation confirms that GymPulse successfully leverages machine learning to provide actionable insights. The system demonstrates strong error handling and drive grade security which were our main priorities from the start.

6.12.2 Limitations

Even though it was successful the system is not perfect. We have noticed performance degradation when processing datasets exceeding 50000 rows in real-time which suggests we need to implement batch processing in the future. Additionally the lack of a native mobile app does limit offline accessibility.

6.12.3 Future Work

Future development will focus on three key areas:

1. Development of native iOS/Android applications for better mobile performance.
2. Implementation of automated CI/CD pipelines to retrain models automatically as new data comes in.
3. Integration with IoT wearables to incorporate biometric data into the risk assessment models.

6.13 Summary

The tough testing process described in this chapter establishes the technical viability of GymPulse. By validating the accuracy of the ML models ($F1 = 0.84$) and confirming security compliance we have proven the system is ready for deployment in a real operational environment.

Chapter 7

Conclusions

7.1 Summary of Achievements

The effective creation and introduction of the AI-based project of the fitness guidance and analytics system called GymPulse (AI-Powered Fitness Guidance Analytics System) confirm the theory that machine learning and the latest web technologies could positive impact on operational performance in that sphere of fitness. The system helps close the information drif between the raw data collection process and the real actionabl business intelligence contributing a solid solution to sales forecasting, churn prediction, and member risk evaluation.

The technical accomplishment that led this research to the main aims are as follows:

- **Predictive Accuracy:** The implementation of the XGBoos algorithm gave a sales forecasting accuracy exceed 85% and a churn prediction AUC score of 0.88, demonstrating high reliability for decision support.
- **Architectural Robustness:** The adoption of a modular microservices-inspired architecture using Next.js(v-18) and Python ensured system scalability and secure data Arrangement.
- **Operational Utility:** The deployment of an interactive dashboard transformed complex analytical metrics into intuitive visualizations, enabl non-technical Stakeholders to make data-driven decisions.

7.2 Key Insights and Lessons Learned

The development lifecycle of GymPulse gave several critical engineering and operational insights:

- **Role of Robust ETL Pipelines:** The establishment of consistent Extract, Transform, Load (ETL) processes was fundamental to model stability. Ensuring data cleanliness at the ingestion stage significantly reduced training error and improvedd conclusion delay.

- **Importance of Model Interpretability:** The integration of Explainable AI (XAI) elements such as feature importance graphs proved essential for user adoption. Gym owners were more likely to trust and act upon AI recommendations when the underlying logic (e.g., "low attendance frequency") was transparent.

- **Efficacy of Modern Web Architectures:** Using component based frameworks (Next.js v-18) and utility-first styling (Tailwind CSS) boosted the development cycle. This approach shows rapid prototyping and ensured a consistent responsive user experience across different devices.

- **Security as a Foundational Layer:** Implement security protocols including JWT authentication and RBAC at the start of development was crucial. Upgrading security measures post-development would have compromise architectural integrity and data privacy compliance.

- **Value of Domain-Specific Data:** Training models on authentic domain-specific gym datasets gave significantly higher accuracy compared to generic synthetic data. This underscores the necessity of high-quality data acquisition in vertical-specific AI applications.

- **Optimization for Real-Time Performance:** Achieving the low-latency analytics requires continuous optimization, including the implementation of API caching strategies and lazy loading for heavy data visualizations.

7.3 Opportunities for Further Work

While GymPulse makes a strong foundation for AI-driven fitness management several avenue exist to extend its capabilities and technical meturity:

7.3.1 Infrastructure and Scalability

- **Distributed Orchestration:** transferring the backend to a container Arrangment platform (e.g., Kubernetes) to support horizontal auto-scaling for large-scale gym chains.

- **Automated MLOps:** Establishing a Continuous Training (CT) pipeline to detect data drift and trigger automated model retraining, ensuring long-term predictive accuracy.

7.3.2 Functional Expansion

- **Native Mobile Ecosystem:** Develop dedicated iOS and Android applications to facilitate push notifications for staff alerts and offline access for member management.
- **IoT and Wearable Integration:** Integrating with fitness trackers and smartwatches to incorporate biometric data (e.g heart rate variability, calorie burn) into the risk assessment models for deeper behavioural insights.

7.3.3 Advanced Intelligence

- **Granular Explainability:** Implement advanced Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) to provide transparent, member-specific decision pathways. These explanations enable stakeholders to understand the contribution of individual features to model predictions.
- **Real-Time Anomaly Detection:** Develop an alerting engine capable of identifying business-critical anomalies such as sudden revenue drops or unusual churn spikes in real time. The system continuously monitors key performance indicators and triggers notifications when deviations from expected patterns occur.

7.3.4 Ethics and Governance

- **Ethical AI and Bias Mitigation:** The Future iterations will incorporat fairness-aware algorithms to ensure the churn prediction model does not Accidentally bias against specific demographics (e.g age or gender). This ensures equitable treatment of all members in retention campaign.

7.4 Final Remarks

The present project shows how artificial intelligence can be used to transform the operational management system and in particular the capabilitie of the company called GymPulse to go beyond the capabilitie of conventional record-keeping and provide its owners with a range of predictive and proactive tools that will allo them to Expect market trends and retain their members in an efficient manner.

This study incorporates data engineering, software architecture, and machine learning that can produce a scalable practical solution to the fitness technology field. It provides the foundation of the following generation of intelligent fitness management systems establishing a Example of how data science can lead to the sustainable business development of the service industry.

Appendix A

User Manual

A.1 Overview

This guide is a walkthrough of how to actually driv the **GymPulse** system. We have designed the interface to be intuitive but since there is some fairly complex AI running under the hood it helps to have a roadmap. The goal here is not just to click buttons but to understand what the system is trying to tell you about your gyms health.

A.2 Getting Started

Access the system is pretty straightforward. You will have to need your administrator credentials usually provided during the initial setup phase to get past the secure login screen. Once you are in the system keeps you logged in via a secure session token so you will not have to constantly re enter your password unless you are inactiv for a long time.

A.3 The Dashboard: Your Morning Coffee View

The first thing you will see after logging in is the **Dashboard**. Think of this as "command central." It will gives you a quick snapshot of how the gym is performing right now without needing to dig to deep immediately.

- **Sales Forecast Graph:** Right in the center you will see a line graph. This is not just history; it is the AI predicting where your revenue is heading over the next 30 days. If the line is trending down that is your cue to run a promotion.
- **Confidence Score:** You might notice a percentage (like 85%) next to the prediction. This is the system being honest about how sure it is. It relie on historical patterns so if your data is messy this score might drop.

- **Active Members:** On the right there is a simple donut chart. It breaks down who is active versus who is drifting away.

A.4 Deep Dive into Analytics

If you want to see the real power of the system head over to the **Analytics** tab in the sidebar. This is where the AI actually shows its work.

A.4.1 Sales Trends

Here you can search between different views 7 days, 14 days, or even 60 days. It is useful for finding seasonal trends. For instance you might see a spike in january (New Years resolutions) and a dip in december. The system overlays the *predicted* sales against *actual* sales from previous months so you can verify if the AI has been accurate lately.

A.4.2 Churn Risk Analysis

This is Possibly the most critical part. The system assigns a "Churn Probability" to your member base. It doesnot just guess; it looks at things like how often they scan in or if they missed a payment.

- **High Risk:** These are members the AI thinks are about to quit. You should probably reach out to them.
- **Retention Factors:** If you are curious *why* someone is flagged as high risk the dashboard usually highlights the main reason like "Low Attendance Frequency."

A.5 Managing Members

The **Members** section is where you take action. It is one thing to know someone is at risk; it is another to fix it.

When you click on a members profile, you donot just see their name and email. You see a calculated **Risk Level** (Critical, High, Medium, Low).

- **Intervention:** The system suggests retention strategies. For a "High Risk" member it might suggest sending a discount offer.
- **History:** You can check their last visit date. If it is been over 21 days thats usually a red flag.

A.6 System Configuration (Settings)

This part is mostly for the admins. In the **Settings** tab you handle the raw data that feeds the AI.

- **Data Management:** This module allows administrators to upload new CSV files, including updated attendance logs and sales records. Regular data updates are essential, as the accuracy of AI predictions depends heavily on data quality and relevance. The system performs basic validation checks to ensure data consistency and completeness before ingestion. Maintaining fresh and reliable datasets helps prevent model degradation and improves long-term prediction performance. Additionally, structured data handling ensures smooth integration with downstream analytics and model training pipelines.
- **Model Selection:** The platform provides flexibility by allowing users to select from multiple machine learning models for analysis. XGBoost is set as the default option due to its computational efficiency and strong predictive performance. However, users can switch to alternative models such as Random Forest to compare results and assess model robustness. This feature supports experimentation and informed model selection based on specific business needs. Comparative model evaluation enables better alignment with varying operational scenarios.
- **Retraining:** The retraining functionality enables users to update predictive models when data distributions or market conditions change. If prediction accuracy begins to drift, administrators can manually trigger a retraining process using the most recent datasets. This ensures that models remain aligned with current trends and behaviors. Automated retraining helps maintain prediction reliability and system adaptability over time. Continuous model updates are essential for sustaining long-term predictive accuracy.

A.7 Troubleshooting

Occasionally, users may encounter minor issues during system operation. If a visualization or graph fails to load, the cause is typically a temporary network timeout or a caching issue. In most cases, refreshing the browser resolves the problem and restores the user interface correctly.

If prediction results appear inaccurate or unrealistic (for example, showing zero sales for an entire month), users should verify the integrity of the uploaded dataset. The **Data Management** tab should be checked to ensure that the CSV file was uploaded successfully and does not contain missing values or empty rows. Regular validation of input data helps maintain the reliability and accuracy of system predictions.

References

[1] Luca Pappalardo, Paolo Cintia, et al. Artificial intelligence for human movement and sports: A survey. *ACM Computing Surveys*, 2021. No Citations.

[2] Krittika Kaewkannate and Seong Kim. A comparison of wearable fitness devices. *Journal of Medical Internet Research*, 2016. No Citations.

[3] Ankur Hazarika and Gautam Baruah. Predicting gym member churn using machine learning. *International Journal of Advanced Computer Science*, 2020. No Citations.

[4] L. P. Reis and R. Barros. Big data and predictive analytics in fitness clubs: Applications and challenges. *International Journal of Sports Science*, 2019. No Citations.

[5] Eduardo Velloso, Andreas Bulling, et al. Qualitative activity recognition of weightlifting exercises. In *Proceedings of the 1st International Workshop on Sensor-based Activity Recognition*, 2013. No Citations.

[6] Zhi Chen, Yan Liu, et al. Ai-driven fitness coaching: A personalized exercise recommendation framework. *IEEE Access*, 2022. No Citations.

[7] Alexandros Pantelopoulos and Nikolaos Bourbakis. A survey on wearable sensor-based systems for health monitoring. *IEEE Transactions on Systems, Man, and Cybernetics*, 2010. No Citations.

[8] Mi Zhang and Alexander A. Sawchuk. Human physical activity recognition using wearable sensors. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012. No Citations.

[9] Bruno Silva, Miguel Sousa, et al. Ai-enhanced gym management systems: Automating engagement and retention. *Journal of Sports Management and Technology*, 2020. No Citations.

[10] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2018. No Citations.

[11] Javier Molina-Garcia and Isabel Castillo. Member retention in fitness centers: Predictors and strategies. *European Journal of Sport Science*, 2017. No Citations.

[12] Zhe Cao, Gines Hidalgo, et al. Openpose: Realtime multi-person 2d pose estimation using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. No Citations.

[13] Kai Chen and Jun Wang. Exercise intensity prediction using deep learning and wearable devices. *Sensors*, 2021. No Citations.

[14] Hui Wang and Sheng Liu. Business intelligence applications in fitness center operations. *International Journal of Data Analytics*, 2020. No Citations.

[15] Nisha Khurana and Amit Goel. Machine learning-based personalized workout recommendation system. In *International Conference on IoT and Machine Learning*, 2020. No Citations.

[16] Ashraf Haque, Arnold Milstein, and Li Fei-Fei. Deep learning in medicine and sports science. *Nature Medicine*, 2016. No Citations.

[17] Derek Doran, Sarah Schulz, and Tarek Besold. What does explainable ai really mean? In *Proceedings of the ECAI Workshop on Explainable AI*, 2018. No Citations.

[18] O. M. Giggins, U. Persson, and Brian Caulfield. Biofeedback in rehabilitation: Using wearable sensors. *Journal of NeuroEngineering and Rehabilitation*, 2013. No Citations.

[19] Spencer Carreiro, P. R. Chai, et al. Machine learning in fitness and wellness monitoring: Trends and applications. *IEEE Journal of Biomedical and Health Informatics*, 2018. No Citations.

[20] Luca Pappalardo and Paolo Cintia. Human performance analytics: Big data in sports. *Data Mining and Knowledge Discovery*, 2018. No Citations.